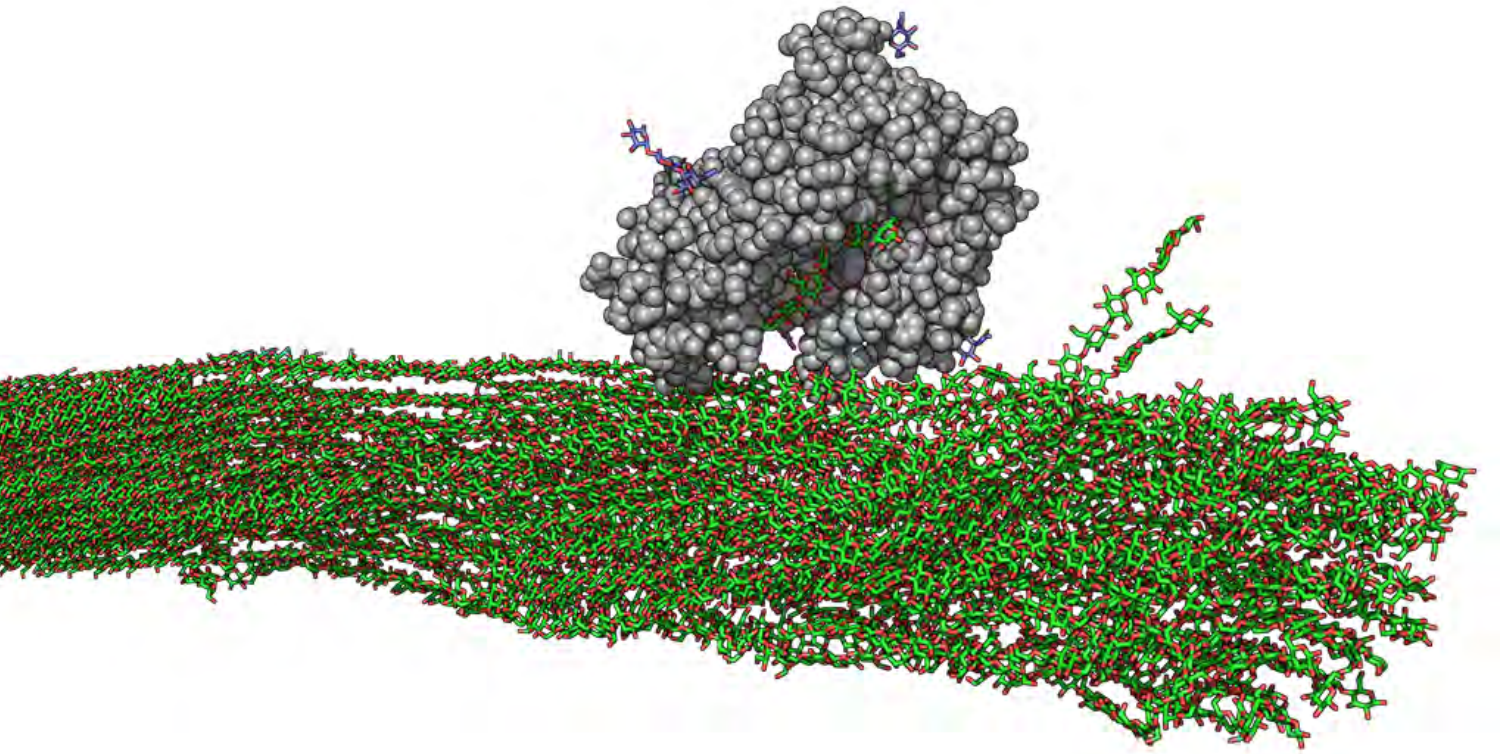
Cellulose Hydrolysis by Fungal Enzymes

Structure-function Relationships in Glycoside Hydrolases Family 7



Thesis for the degree of Doctor of Philosophy

Submitted to the Doctoral School of Science and Environment

Author

Corinna Schiano di Cola

Supervisors

Peter Westh

Kim Borch

Roskilde University, Denmark

December 2019

Cover picture: *artistic representation of the endoglucanase ReCel7B in Michaelis complex with a cellulose strand. The enzyme (gray spheres) is binding onto the surface of amorphous cellulose (green sticks) and threading a cellulose chain into the active site cleft. The N-glycans on the surface of ReCel7B are shown as purple sticks. The crystal structure of ReCel7B (PDB 6SU8) is described in chapter IV and article III appended to this thesis. The enzyme-substrate complex was created via superimposition onto TrCel7A (PDB 4C4C). Figure created using Pymol (The Pmyol Molecular Graphics System, Version 2.3.2, Schrödinger, LLC).*

Summary

Cellulose is considered the most abundant biopolymer on Earth, and it is found primarily in the cell wall of plants. It is energy rich because it consists of simple glucose molecules, but it is highly recalcitrant to biodegradation due to its crystalline structural arrangement. Despite the challenges, species from all domains of life -Eukariota, Bacteria and Archaea- have evolved powerful biocatalyst to accelerate this reaction: a complex group of enzymes, collectively known as cellulases.

On terrestrial land, very efficient cellulose degraders are filamentous fungi, which secrete synergistic mixtures of cellulases to obtain energy. Most of what is known about fungal cellulases derives from studies performed on the model ascomycete *Trichoderma reesei*, which is heavily researched both because of its relevance in the global carbon cycle and for its industrial applicability. In view of this, the same enzymatic systems available in nature are exploited in the biorefinery industry to produce sustainable fuels and bio-based products, using lignocellulose as primary feedstock. However, the necessity to use cellulases in high amount for these processes increases the overall costs and limits the expansion of biorefineries worldwide.

In this framework, the current thesis is part of a research strive aimed at improving cellulases activity for industrial bioconversions, and render them more attractive in a sustainable bioeconomy perspective. The focus of this thesis is kinetic and structural descriptions of fungal cellulases belonging to family 7 glycoside hydrolases (GH7), and particular emphasis is given on the *T. reesei* cellulase system.

Under cellulolytic conditions, *T. reesei* primarily secretes a mixture of cellobiohydrolases (CBHs) and endoglucanases (EGs). The major enzymes produced from these two classes are the CBH *Tr*Cel7A and the EG *Tr*Cel7B, both belonging to GH7. We directly compare the kinetic and binding properties of *Tr*Cel7A and *Tr*Cel7B on cellulosic substrates using steady-state kinetic models suitable for interfacial enzyme catalysis. We found that the two enzymes showed complementary differences in terms of catalytic efficiency and ability to attack different areas on cellulose, and we could ascribe these to major differences in the enzymes structure. Both *Tr*Cel7A and *Tr*Cel7B consist of two domains, a catalytic domain connected to a carbohydrate binding domain (CBM) via a flexible linker. However, in the catalytic domain the presence of eight flexible loops form a tight substrate-binding tunnel in *Tr*Cel7A, while the absence of four of these loops creates a more open cleft in *Tr*Cel7B. The functional role of the missing loops of *Tr*Cel7B was investigated by creating a number of deletions variants in *Tr*Cel7A and by comparing changes in kinetics and

binding between the variants and the wild type enzymes. Deletions in the region corresponding to loop B2 of *Tr*Cel7A consistently shifted the enzymatic properties toward the endoglucanase *Tr*Cel7B, unlike deletions in other regions, thus indicating that this loop is the main discriminant between the CBH and EG subtypes.

Further structure and sequence analyses of the loops in GH7 revealed that the tunnel/cleft shape of the catalytic domain of CBHs and EGs could be more finely described by measuring the degree of substrate coverage upon interaction with the enzyme.

Industrial enzymatic degradation of cellulose requires the use of high temperatures to achieve suitable glucose yields in a short time. Thus, thermostable cellulases are the preferred enzymes for biomass saccharification processes. We hypothesize that a temperature-induced desorption from the substrate surface is a major hurdle for these enzymes in such conditions. To tackle this problem, we studied the thermostable CBH from *Rasamsonia emersonii*, *Re*Cel7A, and designed a high affinity variant where adsorption onto cellulose was improved. The variant showed increased thermoactivation compared to the wild type, with a consequent activity improvement at industrially relevant temperatures.

Structural information is fundamental to understand enzymes' temperature adaptations and to formulate hypotheses on how to improve industrial cellulases using rational design strategies. We expand our understanding of the structure-function relationships in GH7 enzymes by solving the crystal structure of *Re*Cel7B, a highly thermostable EG from the ascomycete *Rasamsonia emersonii*. We also kinetically characterized *Re*Cel7B in comparison with the mesophilic *Tr*Cel7B. The structure of *Re*Cel7B consists of a single catalytic domain and displays a number of unique stabilizing interactions compared to *Tr*Cel7B, particularly in the amount and positions of *N*-glycosylations as well as differences in loop length and rigidity. Along with the wild type enzymes, we kinetically characterized variants of *Re*Cel7B and *Tr*Cel7B where the CBM was either added or removed to study functional adaptations in the EGs. Overall, these results highlighted the differences between mesophilic and thermophilic EGs in terms of turnover, cellulose targeting and synergy with the CBHs.

When cellulases from different classes are mixed, they display complex interactions on the cellulose surface, which reflect a combination of both promoting and unfavourable contributions. On one hand, GH7 CBHs and EGs positively cooperate to degrade cellulose, following the classical *endo-exo* synergy model. Consequently, the kinetic properties and substrate accessibility of *Tr*Cel7A and *Tr*Cel7B in synergy mixtures is quite different from the monocomponents. We explored this aspect in a kinetic perspective using a wide range

of molar combinations of the two enzymes and two different conditions of either substrate or enzyme excess. Optimal synergy revealed to be strongly dependent on the number of attack sites available on the substrate surface and conditions that favour high mole fractions of CBH. On the other hand, cellulases can experience inhibitory effects on the substrate surface due to molecules that bind and occupy the attack area of a cellulase. This negative interaction results in a reduction in activity, and we propose that this effect can be kinetically described and quantified using surface-site inhibition mechanisms. We illustrate the potentiality of this new type of inhibition kinetics using *Tr*Cel7A as a model enzyme, in combination with catalytically inactive cellulases acting as surface-site inhibitor molecules.

In conclusion, this monography is the result of a three-year study aimed at deciphering the structure-function diversity in fungal GH7 cellulases, and illustrates the numerous strategies and challenges carried out to improve cellulases' activity for industrial bioconversions.

Preface

This thesis has been submitted for the PhD degree from Roskilde University, Denmark. The work presented here has been performed during the period of October 2016 - November 2019 under the academic supervision of Prof. Peter Westh and the industrial supervision of Science Director Kim Borch. The majority of the work has been performed at the Department of Science and Environment at Roskilde University (Roskilde, Denmark) and at Novozymes (Bagsværd, Denmark). The last 14 months of work have been spent at the department of DTU Bioengineeering, Technical University of Denmark and Novozymes (Lyngby, Denmark).
The thesis describes part of the research output coming from the TEMPEN project, a collaboration between Roskilde University and Novozymes A/S, funded by Innovation Fund Denmark (Grant number 5150-00020B).

Guide to the reader

The main focus of this thesis is to improve our understanding of the structure-function relationships in cellulases from Family 7 Glycoside Hydrolase (GH7), with particular emphasis on kinetic and structural comparisons between fungal cellobiohydrolases and endoglucanases. A short introduction on the composition of the plant cell wall material, lignocellulose-degrading fungi, cellulases and their industrial applications is presented in the first chapter (**Chapter I**). Then, the kinetic and binding models used in this thesis to characterize cellulases activity are described in the second chapter (**Chapter II**), followed by a direct kinetic comparison between the major cellobiohydrolase and endoglucanase from *Trichoderma reesei*, *Tr*Cel7A and *Tr*Cel7B. In the next chapter (**Chapter III**), the major structural differences between GH7 cellulases will be illustrated, as well as the experimental protein engineering strategies we used to elucidate the functional role of a number of loops present in *Tr*Cel7A and other GH7 enzymes. The subsequent chapter is more industrially applied (**Chapter IV**), and describes the applications of thermostable cellulases, their limitations in industrial conditions and possible strategies for their improvement. In this chapter, the crystal structure of a new thermostable cellulase will be presented, the endoglucanase *Re*Cel7B, along with a thorough kinetic and structural comparison with its close orthologs. Finally, the last chapter (**Chapter V**) focuses on the type of interactions (either positive or negative) that can be established when different cellulases are mixed. Firstly, the synergistic interaction between *Tr*Cel7A and *Tr*Cel7B is investigated. Secondly, we study how the activity of cellulases can be reduced by

the presence of catalytically inactive cellulases, which bind to cellulose and inhibit enzyme activity following specific kinetics.

Disclaimer

The figures and tables presented in this thesis have been either created *de novo* or are adaptations from the appended papers. Some sections of this thesis are based on unpublished results, corresponding to Section 3.7, 3.8, 4.12 and 5.2. Materials and methods concerning these parts can be found in the referenced articles or in the referenced Appendix.

Acknowledgments

These three years of PhD work have been an exciting and challenging period of my life. I have gained a lot of experience both in research and in "normal" life. For this, I have a number of people to thank.

Firstly, I would like to thank my supervisors, Professor Peter Westh at Technical University of Denmark (former Roskilde University) and Science Director Kim Borch at Novozymes. Thank you Peter for your all your advice, motivation, inspiration and scientific feedback, and thank you Kim for your industrial insight, for teaching me how to be pragmatic in my research and for always encouraging me to achieve the best results.

Thank you to all my current and former colleagues, for great scientific collaborations and discussions, and for making the long days in the lab more fun. Particularly, I would like to thank the postdoctoral researchers Jeppe Kari, Trine Holst Sørensen and Silke Flindt Badino who were always willing to help and answer my (many) questions. Thank you to all my colleagues and friends Bartłomiej Kołaczkowski, Stefan Jarl Christensen, Nanna Røjel, Jenny Bååth, Kay Shaller, Gustavo Molina, Michael Skovbo Windhal, Radina Tokin, Nanna Rolsted Sørensen, Stine Fredslund Hansen, Malene Keller, Cynthia Segura Vesterager, Jingwen Qiu, Christine Dengsøe, Camilla Graversen and Kristina Mielec.

Thank you to all the colleagues and researchers at Novozymes: Ulla Thyssen Rosenberg, Ana Mafalda de Almeida Cavaleiro, Johan Pelck Olsen, Camilla Hindborg Kristensen, Ria Jacobsson and in particular Science Manager Kenneth Jensen for supporting some of the more ambitious projects and for always providing useful feedback.

Thank you to all my friends and colleagues at Roskilde University (RUC), Lorenzo Costigliola, Athina Andrea, Eva Marie Karlsen and Søren Hvidt. Particular thanks to Raimundo José Elicer for his feedback on statistical analysis. I would also like to thank all the RUC students I had the opportunity to meet during my teaching duties. I wish you all the best in your future studies.

Thank you to my friends, colleagues and collaborators at Technical University of Denmark (DTU). In particular Prof. Jens Preben Morth for his great insights into the

structural biology world and for making possible to solve the structure of *Re*Cel7B. Also thanks to Department Leader Bjarke Bak Christensen and Section Leader Anne Meyer to ease our work transition from RUC to DTU Bioengineering.

I also give my sincerest thanks to Associate Prof. Jerry Ståhlberg from SLU Uppsala for his great and inspiring teachings during the PhD courses, and for sharing his passion for the cellulases world.

On a more personal level, I want to thank all my family and friends, particularly Livio, Nicoló, Serena, Marta, Karina, Carsten and Sophia, Topi, Palmira, Rosa, Donatella. However my biggest thank you goes to my wonderful partner in life Philip Holck Folman for his endless love and support, as well as for great discussions on evolutionary biology and analysis of big data.

Finally, I want to deeply thank my parents Egeria Natilli and Salvatore Schiano di Cola, who always believed in me and thought me that with hard work and dedication you can achieve all your goals. *Grazie al vostro continuo supporto e all' educazione che mi avete elargito é stato possibile per me fare della ricerca scientifica la mia passione ed il mio lavoro.* I dedicate my PhD thesis to you.

List of publications

The scientific articles described in this thesis are listed below and referred throughout the text by Roman numerals. The full text version of each article is appended to the thesis.

I. **Systematic deletions in the cellobiohydrolase (CBH) Cel7A from the fungus *Trichoderma reesei* reveal flexible loops critical for CBH activity**
Corinna Schiano di Cola*, Nanna Røjel*, Kenneth Jensen, Jeppe Kari, Trine Holst Sørensen, Kim Borch and Peter Westh (2019). *Journal of Biological Chemistry,* 294(6), 1807–1815.

II. **Thermoactivation of a cellobiohydrolase**
Peter Westh, Kim Borch, Trine Sørensen, Radina Tokin, Jeppe Kari, Silke Badino, Mafalda A. Cavaleiro, Nanna Røjel, Stefan Christensen, Cynthia S. Vesterager and Corinna Schiano di Cola (2018). *Biotechnology and Bioengineering*, 115, 831–838.

III. **Structural and biochemical characterization of a family 7 highly thermostable endoglucanase from the fungus *Rasamsonia emersonii***
Corinna Schiano di Cola, Bartłomiej Kołaczkowski, Trine Holst Sørensen,Stefan Jarl Christensen, Ana Mafalda Cavaleiro, Kenneth Jensen, Kim Borch, Jens Preben Morth and Peter Westh (*The FEBS Journal, published online 21 Nov 2019, in press. DOI: 10.1111/FEBS.15151)*

IV. **A Steady-State Approach for inhibition of Heterogeneous Enzyme reactions**
Jeppe Kari*, Corinna Schiano di Cola*, Stine Fredslund Hansen, Silke Flindt Badino, Trine Holst Sørensen, Anna Mafalda Cavaleiro, Kim Borch and Peter Westh (Manuscript)

*First authorship shared

Other publications:

Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments

Andrea Strazzulli, Beatrice Cobucci-Ponzano, Roberta Iacono, Rosa Giglio, Luisa Maurelli, Nicola Curci, Corinna Schiano di Cola, Annalisa Santangelo, Patrizia Contursi, Vincent Lombard, Bernard Henrissat, Federico M. Lauro, Carlos M.G.A. Fontes and Marco Moracci (2019). *The FEBS Journal*, in press. DOI:10.1111/febs.15080

Abbreviations list

Abbreviations

ASA	Accessible surface area
BC	Bacterial cellulose
BMCC	Bacterial microcrystalline cellulose
BSA	Buried surface area
CBH	Cellobiohydrolase
CBM	Carbohydrate binding module
CD	Catalytic domain
^{conv}MM	Conventional Michaelis-Menten
DSC	Differential scanning calorimetry
DSF	Differential scanning fluorimetry
EC	Enzyme classification
EG	Endoglucanase
Endo-H	Endoglycosidase H
ESA	Exposed surface area
G1	Glucose
G2	Cellobiose
G3	Cellotriose
GH	Glycoside Hydrolase
HPAEC-PAD	High-Performance Anion-Exchange Chromatography Coupled with Pulsed Electrochemical Detection
^{inv}MM	Inverse Michaelis-Menten
LPMO	lytic polysaccharide monooxygenase
MD	Molecular dynamics
MS	Mass spectrometry

NAG	*N*-acetylglucosamine
PAHBAH	*para*-hydroxybenzoic acid hydrazide
PNGase F	Peptide-N-Glycosidase F
*p*NP-Lac	*para*-nitrophenyl β-D-Lactopyranoside
RAC	Regenerated amorphous cellulose
SHF	separate hydrolysis and fermentation
SSF	simultaneous saccharification and fermentation
T_{opt}	Temperature optimum
T_m	Thermal unfolding transition midpoint
2G	Second-generation

Enzymes abbreviations

ΔB2-1	*Trichoderma reesei* Cel7A variant (deletion in the B2 loop)
ΔB2-2	*Trichoderma reesei* Cel7A variant (deletion in the B2 loop)
ΔB2-3	*Trichoderma reesei* Cel7A variant (deletion in the B2 loop)
ΔB3	*Trichoderma reesei* Cel7A variant (deletion in the B3 loop)
ΔB4	*Trichoderma reesei* Cel7A variant (deletion in the B4 loop)
*Fo*Cel7B	*Fusarium oxysporium* Cel7B
*Hi*Cel7B	*Humicola insolens* Cel7B
*Hir*Cel7A	*Heterobasidion irregulare* Cel7A
*Ma*Cel7B	*Melanocarpus albomyces* Cel7B
*Pc*Cel7D	*Phanerochaete chrysosporium* Cel7D
*Pf*Cel7A	*Penicillium funiculosum* Cel7A
*Th*Cel7B	*Trichoderma harzianum* Cel7B
*Tr*Cel6A	*Trichoderma reesei* Cel6A
TrCel6A$_{D221N}$	*Trichoderma reesei* Cel6A variant (catalytically inactive mutant)
*Tr*Cel7A	*Trichoderma reesei* Cel7A
TrCel7A$_{E212Q}$	*Trichoderma reesei* Cel7A variant (catalytically inactive mutant)
*Tr*Cel7B	*Trichoderma reesei* Cel7B
*Tr*Cel7BΔCBM	*Trichoderma reesei* Cel7B variant (deletion of the linker-CBM)
*Re*Cel7A	*Rasamsonia emersonii* Cel7A
*Re*Cel7AinsCBM	*Rasamsonia emersonii* Cel7A variant (insertion of a linker-CBM)
*Re*Cel7B	*Rasamsonia emersonii* Cel7B
*Re*Cel7BinsCBM	*Rasamsonia emersonii* Cel7B variant (insertion of a linker-CBM)

Table of Contents

Chapter I – Introduction

This chapter will give a brief overview of the composition of the plant cell wall material and the enzymes capable of deconstructing it. Particular emphasis will be given on the structure of cellulose and fungal cellulases from glycoside hydrolase family 7 (GH7).
Two main GH7 enzymes are secreted by the model cellulolytic fungus *Trichoderma reesei,* the cellobiohydrolase *Tr*Cel7A and the endoglucanase *Tr*Cel7B, and play a central role in this thesis. The general structure and catalytic mechanism of *Tr*Cel7A and *Tr*Cel7B will also be described at the end of the chapter.

1.1 - Lignocellulose biomass composition and related enzymes

Lignocellulose is the main constituent of the plant secondary cell wall, and it can be divided into two main components: a saccharidic part, mainly composed of cellulose and hemicellulose, and a non-saccharidic part, mainly consisting of lignin. The percentage composition of these elements varies depending on the source of the lignocellulosic material (Marriott et al. 2016). Other components, scarcely present in lignocellulose, are pectins (complex polysaccharides of D-glacturonic acid) and structural proteins.

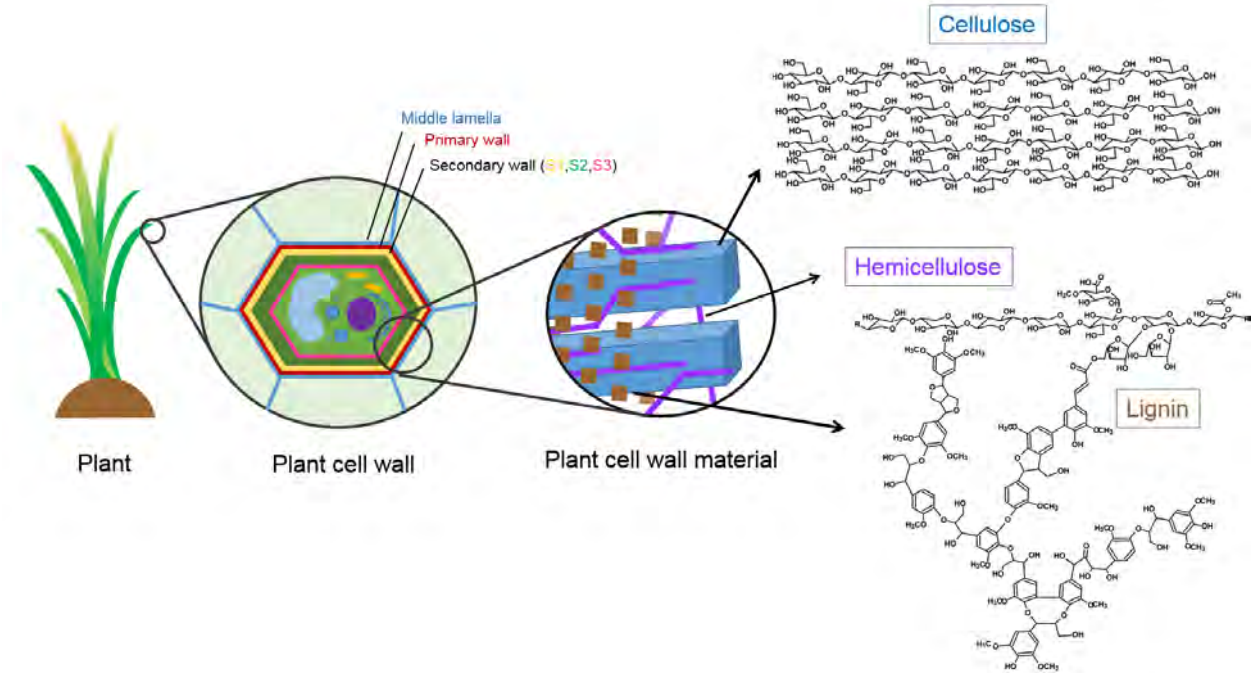


Figure 1.1. *Lignocellulose composition in plants. Lignocellulosic biomass is part of the plant cell wall and mainly consists of cellulose, hemicellulose and lignin. Molecular structures created with ChemSketch (ACD Labs, Canada).*

Fig. 1.1 shows the general structure of a plant cell wall: cellulose is arranged in highly ordered layers of microfibrils, which confer overall strenght. The cellulose material is then surrounded by less ordered polysaccharides forming the hemicellulose. This complex network is further soaked by the aromatic polymers of lignin, which is crosslinked to hemicellulose. These three main components are further described below.

- Cellulose is the most abundant (50-20%) saccharidic part of the lignocellulosic material (Pauly et al. 2008) and it consists of linear chains of glucose polymers. It is assembled in the plant cell wall in highly crystalline fibers which exclude water. Because of its high stability, it is extremely recalcitrant to microbial degradation. The structure of cellulose will be described in more detail in the next section. Enzymes that hydrolyze cellulose have been assigned to 12 different GH families: 5-9, 12, 26, 44, 45, 48, 61, and 74 (Sandgren et al. 2005).
- Hemicellulose is a less abundant (15-35%) but more complex and branched polysaccharide (Pauly and Keegstra 2008). It consists of different sugar monomers such as pentoses (eg. xylose and arabinose), hexoses (glucose, mannose, galactose), and sugar acids (glucuronic acids and galacturonic acids), which in turn can have further modifications, for instance methylation and acetylation. Hemicellulose

composition varies enormously whithin species, cell type and stage of plant delvelopment, but in general it consists of a linear backbone of sugar units connected via β-(1,4)-glycosidic bonds, with several branches that prevent the structure from crystallizing. It interacts with cellulose via hydrogen bonds and can form covalent bonds with lignin. Common hemicelluloses are xyloglucan, arabinoxylan and galactomannan (see (Pauly and Keegstra 2008) for a structural representation). Hemicellulolytic enzymes can be found in many different GH families such as GH10, 11, 16, 51, 62, 74 (Bissaro et al. 2018). Hemicellulases are also found in the carbohydrate esterase and polysaccharide lyases enzyme classes since they aid the GH action on hemicellulose by removing methyl, acetyl and phenolic esters bonds (van den Brink et al. 2011).

- Lignin is the non-saccharidic part of lignocellulose (10-30%)(Pauly and Keegstra 2008) and it consists of a complex aromatic polymer whose main purpose is to confer rigidity to the cell wall structure and increase its hydrophobicity. It lacks a repeating unit and contains hundreds of monomers of p-hydroxyphenyl, guaiacyl and syringyl, connected with a variety of linkages, the main one being β-O-4. It can also connect to hemicellulose via ferrulic acid. Oxidative enzymes, such as laccases and peroxidases, are capable of attacking lignin, while lignin results in a formidable barrier for cellulases, on which they can adsorb irreversibly (Bissaro et al. 2018). Glucuronoyl esterases from carbohydrate esterases family 15 are receiving considerable attention for their ability to cleave ester-linkages between lignin and carbohydrates (Agger et al. 2017).

1.2 - The structure of cellulose

Cellulose can be considered the most abundant biopolymer on Earth (Himmel et al. 2007). Is has been estimated that plants produce about 180 billion tons of cellulose per year globally (Puranen et al. 2014). Cellulose is simply composed of D-glucose polymers which form linear chains connected with β-(1,4)-glycosidic bonds. The repeating unit in the cellulose chain is cellobiose, a glucose disaccharide where the second unit is rotated 180° compared to the first one (Fig. 1.2). Cellulose's extreme recalcitrance to biodegradation is due its high structural stability conferred by numerous non-covalent interactions, namely hydrogen bonds and Van der Waals forces, which occur *intra*- and *inter*-chain and exclude water molecules.

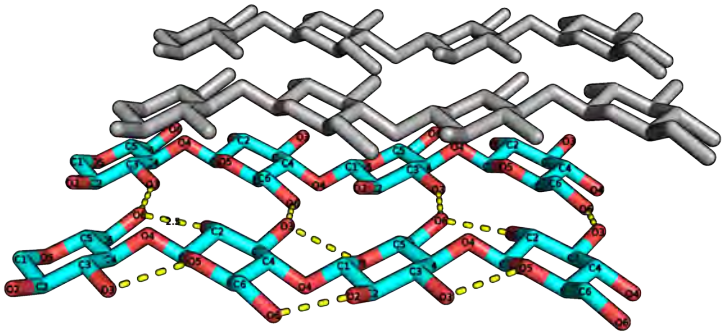


Figure 1.2. *General structure of cellulose Iα. Two sheets (gray and cyan) consisting of linear glucose chains are shown. The parallel sheets are stabilized by intra- and inter- molecular hydrogen bonds (yellow dashed lines) between the different glucose chains. Figure created using Pymol (The PyMOL Molecular Graphics System, Version 2.3.2, Schrödinger, LLC).*

Cellulose can have different degree of crystallinity depending on the stacking and the hydrogen bond arrangement of the layers. A natural occurring polymorph of cellulose is cellulose type I (Fig. 1.2), further subdivided in I_α and I_β, depending on the crystal cell symmetry. Although they can interchange, I_α is prevalent in bacterial and algal cellulose while I_β is prevalent in higher plants (Zhang et al. 2004). This type of celluloses contains *intra*-layer hydrogen bonds. Other non-natural cellulose types do exist, which have relevance in the industrial bioconversion of cellulose (chapter IV), and can be obtained from cellulose I via chemical pretreatment methods (Payne et al. 2015).

The exact mechanism of cellulose synthesis and modification during plant growth is still under research (Somerville 2006). Nonetheless, cellulose is known to be synthetized in higher plants by a complex system of proteins and enzymes bound to the plasma membrane, collectively called the *cellulose synthase* system. In general, this system relies on an enzyme complex organized in rosettes of six globules, containing glycosyl transferases which synthetizes glucose chains starting from UDP-glucose. Roughly 6-10 chains can be formed in parallel by the rosettes, after which the cellulose chains most likely self-assemble via hydrogen bonds in cellulose elementary fibrils, of a diameter of circa 2 nm. When six of these fibrils coaelesce, a 10-30 nm cellulose microfibril can be formed (Somerville 2006).

1.3 - Model cellulosic substrates

In order to study cellulases activity *in vitro*, in this thesis and appended papers we used a number of model cellulosic substrates, listed below. Such substrates allowed high experimental reproducibility, and differ in crystallinity, solubility and structure to allow kinetic characterization of a diversity of cellulases sub-types.

- Microcrystalline cellulose (Avicel PH-101) was the primary substrate used for cellulase characterization. Prepared from coniferous wood pulp, Avicel contains a mixture of crystalline and amorphous cellulose, suitable to test the activity of different cellulases subtypes (Zhang and Lynd 2004). Usually considered pure cellulose, some studies showed it might contain residual hemicellulose (<2%) (Várnai et al. 2010).
- Bacterial microcrystalline cellulose (BMCC) is not commercially available and it is prepared experimentally from bacterial cellulose via partial acid hydrolysis to remove amorphous cellulose (Cruys-Bagger et al. 2013) (Väljamäe et al. 1999). It can be used as highly crystalline model substrate for cellulase characterization.
- Regenerated amorphous cellulose (RAC) is a highly amorphous substrate. It is prepared "in house" from microcrystalline cellulose (Avicel) through cellulose dissolution with phosphoric acid followed by water precipitation, which dramatically reduces the crystallinity (Zhang et al. 2006).
- Chromogenic substrates are non-natural substrates widely used for cellulases characterization because of their simplicity to use and rapid product detection. Two examples are 4-Nitrophenyl β-D-lactopyranoside (*p*NP-Lac) (Fig. 1.3A), a soluble substrate used in Article I to measure product inhibition, and Azurine cross-linked hydroxyethyl-cellulose (AZCL-HE-Cellulose, Fig. 1.3B) an insoluble chromogenic substrate suitable for detecting endoglucanase activity (Badino et al. 2017)(Article I).

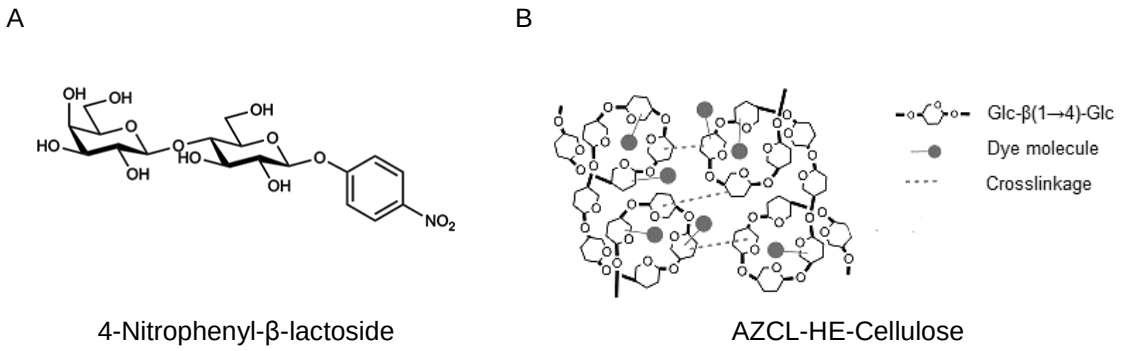


Figure 1.3. *The structure of two synthetic substrates used to measure cellulases activity. (A) para-Nitrophenyl β-D-lactopyranoside, (B) Azurine cross-linked hydroxyethyl-cellulose.*

1.4 - Lignocellulose degradation in fungi

Since lignocellulosic biomass is such an abundant energy source, organisms from all domains of life have evolved enzymatic strategies to degrade it (Cragg et al. 2015). The type and amount of enzymes produced greatly depends on the evolutionary pressure and ecological niche in which the organism reside. Fungi prevail in terrestrial, aerobic environments, while bacteria predominates in aquatic, anaerobic environments (Leschine 1995). Interestingly, enzymatic cellulose degradation has been found also in symbiotic protists of cows and termites, in Archaea, for example *Saccharolobus* and *Pyrococcus* sp. (Kim et al. 2011, Girfoglio et al. 2012) and in Metazoans like crustaceans (Kern et al. 2013). Since the major degraders of land plant biomass are fungi, they have high ecological relevance for the global recycling of carbon and nutrients (Krah et al. 2018) and industrial relevance because of their ability to secrete efficient and stable cellulolytic enzymes, as will be discussed in the next sections (Makela et al. 2014).

Fungi capable of degrading plants woody tissues can be grossly divided into three groups, depending on the type of decay they perform: white rot, brown rot and soft rot. White rot fungi mainly attack lignin, thereby creating a "bleaching effect" on the woody tissues. Most basidiomycetes belong to this group, for example *Phanerochaete chrysosporium*. Even though they secrete a diversity of laccase and peroxidase enzymes, they also possess GH7 cellulases. Most basydiomycetes are saprotrophic, but some species (such as *Heterobasidion irregulare/Heterobasidion annosum*) can also be necrotrophic (Momeni et al. 2013). Brown rot fungi degrade polysaccharides and lignin non-enzymatically, with a mechanism based on Fenton chemistry and the generation of very powerful oxidizing radical species. The non-complete degradation of lignin is probably responsible of the

brown decay. Examples of such fungi are *Fomitopsis pinicola* and *Gleophyllum trabeum* (Rytioja et al. 2014). Soft rot fungi obtain a more limited lignocellulose degradation, because of their inability to attack lignin, but they possess and excellent set of cellulases. Examples from this group are the ascomycetes *Trichoderma reesei*, *Humicola insolens* and moulds such as *Aspergillus* and *Penicillium* species (Hamed 2013, Krah et al. 2018).

Plant-degrading fungi are not only saprobes but can also be serious pathogens such as *Magnaporthe oryzae*, producing a very potent enzymatic machinery to degrade and modify the plant cell wall (Dean et al. 2012).

Among saprobes, thermophilic fungi are of particular interest for industrial applications. They belong to numerous genera and are the major organisms found in warm, aerobic and humid piles of any sort of accumulated plant material (Maheshwari et al. 2000) but also in arid environments with extreme diurnal shifts in temperatures (Powell et al. 2012). The mechanism of thermophilic fungi colonization was already proposed in the early 20th century as follows. Mesophilic organisms would start the initial decay of the plant material. By producing exotermic metabolic reactions, the overall temperature of the compost would increase to temperatures close to 40 °C, allowing the sporulation of thermophilic fungi. The latter will eventually outgrow the mesophiles and raise the temperature of the environment even further (> 60 °C) (Maheshwari et al. 2000). That is why thermophlic fungi are an excellent source of thermostable enzymes which can be exploited in biorefineries. The ascomycete *Rasamsonia emersonii* (previously known as *Talaromyces emersonii, Penicillum emersonii* and *Geosmithia emersonii*) is an example of thermophilic saprobe that can be found in wood decaying matter, soil and compost piles. It is equipped with a thermostable cellulolytic system apt to degrade a number of lignocellulosic substrates (Moloney et al. 1983). Two of its GH7 cellulases will be further described in chapter IV.

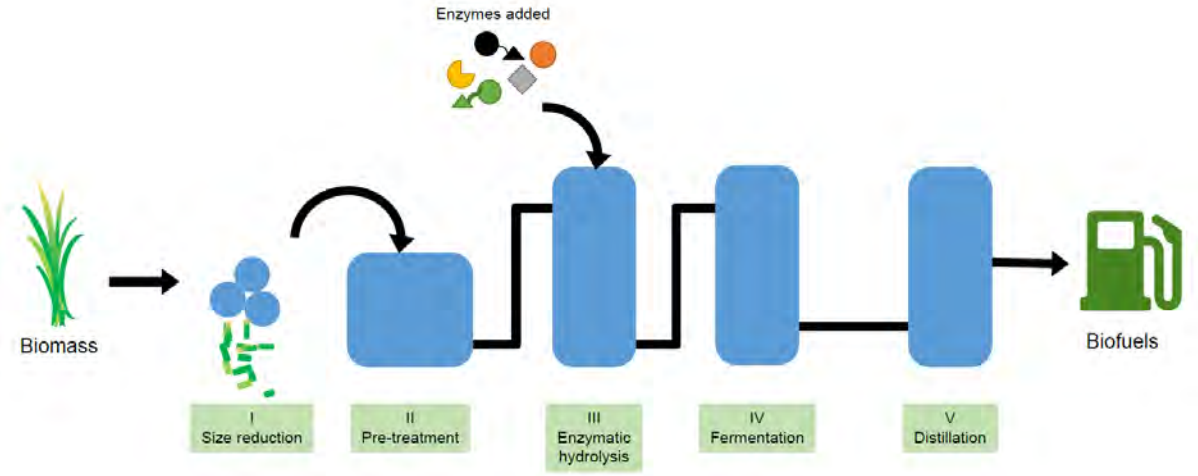
Cellulolytic fungi have developed the so called *free enzyme system* to degrade cellulose, which will be described in section 1.6. In this system, cellulases are secreted in the extracellular medium and attack the extremely recalcitrant plant cell wall material in a relatively mild physico-chemical environment (Cragg et al. 2015). However, this is not the only enzymatic strategy possible. Organism adapted to anaerobic environments, such as bacteria isolated from animal rumen and sewage sludge, rely on the use of multi-enzyme complexes called *cellulosomes*. These complexes usually remain attached to the bacterial cell surface via an anchoring protein and the interaction of a scaffoldin-dockerin protein complex. The dockerin itself is then part of a primary scaffoldin complex containing several cohesins, each anchoring a type of hydrolytic enzyme to perform cellulose biodegradation.

Enzymes used in these complexes belong to several GH families and will not be mentioned here (see (Artzi et al. 2017) for a review).

1.5 - Industrial applications of cellulases

Cellulases have been available on the market for more than 30 years and account for a significant portion (about 20%) of the world enzyme market (Harman et al. 1998). Fungal cellulases, in particular from *Trichoderma* and *Aspergillus sp.*, have been applied in different sectors (Uhlig 1998). In the pulp and paper industry, cellulases are used to significantly decrease the energy costs required in the pulping process. In particular, endoglucanases are used to reduce pulp viscosity (Oksanen et al. 1997). In the textile industry, cellulases have been efficiently implemented in the biostoning and biopolishing of clothings, and for smoothening and removing fuzz from textiles (Harman and Kubicek 1998). As in the case of pulp and paper industry, a preparation rich in endoglucanase is used. Other applications exist in other sectors such as food and feed industry, reviewed in (Kuhad et al. 2011).

In the recent decades, an intensely researched application for cellulases is within the biorefinery sector, where a more sustainable production of fuels, power and chemicals is carried out. Considering the environmental threats posed by global climate change, governments worldwide are implementing and researching different strategies to reduce human dependence from fossil fuels as primary source of energy for transportation (Li et al. 2017). Biofuels and various products produced from lignocellulosic biomass have been proposed as a promising alternative to fossil fuels and fossil-based chemicals. Among biofuels, bioethanol is of particular interest and is already being used as a blend in combustion engines for transportation in different countries (Ragauskas et al. 2006). Lignocellulose has a number of advantages compared to other feedstocks: it is low-cost, abundant, energy rich and can support the large-scale fermentations necessary for the process. The starting material can be different, however it usually derives from sources that do not compete with food industry, such as waste material from agricultural crops (sugar cane bagasse, wheat straw and corn stover) or it is harvested from plants growing on marginal lands, like the giant reed *Arundo donax* or the grass *Panicum virgatum*. (Forte et al. 2018). To understand the role of cellulases in the bioethanol industry, it is necessary to briefly describe how second-generation (2G) biofuels production facilities convert lignocellulosic biomass into ethanol. A common process is schematized in Fig. 1.4.



,

Figure 1.4 *Simplified scheme of a biorefinery process, where separate hydrolysis and fermentation occurs. The starting material (biomass) undergoes a size reduction (I) and pre-treatment (II) step, followed by enzymatic hydrolysis, where a cocktail of enzymes including cellulases is added. The subsequent product (hydrolysate) is then separated and subjected to yeast fermentation (IV). Further downstream processes such as distillation (V) give the final product (biofuels).*

After a size-reduction step (I), lignocellulosic biomass undergoes a thermochemical pre-treatment procedure (II) to render the complex material more easily accessible to enzymatic degradation. This step requires the use of high pressure and temperature, as in the case of steam explosion (Kim et al. 2007), or chemicals, as in the case of acid hydrolysis (Mosier et al. 2005). Afterwards, an enzyme hydrolysis step (also called saccharification) is initiated (III): a cocktail of different enzymes act in synergy to degrade the lignocellulosic material into simple fermentable sugars. In this step, cellulases are the major enzymatic workforce, since cellulose is the main component of plant cell wall (section 1.2). The resulting hydrolysate is finally used as energy source for a fermenting yeast like *Saccharmoyces cerevisiae* (IV). Further purifications and distillation processes (V) give the final product bioethanol or other biobased products (Payne et al. 2015). This process is known as separate hydrolysis and fermentation (SHF). Other strategies are possible, such as simultaneous saccharification and fermentation (SSF) where an engineered yeast, capable of secreting lignocellulose degrading enzymes, is used to perform the saccharification and the fermentation step at the same time (Sun et al. 2002).

Despite the clear environmental benefits, the industrial production of 2G bioethanol is still economically challenging, because the product has a low-added value and the productions costs are still high. Several solutions are possible to reduce process costs, for example

improving the pre-treatment step to reduce temperature and chemicals utilization costs. However, one of the major costs is related to the enzymes, which are needed at high dosage for the saccharification (Klein-Marcuschamer et al. 2012).
A number of enzymatic cocktails are available for this purpose. Some examples are Cellulast®, mainly composed of cellulases from *Trichoderma reesei* (Hu et al. 2015), or the more recent Cellic® Ctec2, where introduction of auxiliary activities has been claimed (Cannella et al. 2014). These mixtures are usually supplied with the addition of other preparations such as Novozym 188®, containing a β-glucosidase from *Aspergillus niger* (Rosgaard et al. 2007). The cost related to the use of these preparations can be greatly reduced by developing cellulases more efficient at degrading their substrate.

1.6 - The cellulolytic system of *Trichoderma reesei*

The most well studied organism for the production and characterization of cellulases is the mesophilic fungus *Trichoderma reesei*, commonly known as the anamorph of *Hypocrea jecorina* (Merino et al. 2007). It is a soft rot ascomycete isolated from the Solomon Islands during World War II, where it was causing damage to tents and other cotton-based materials. It was described the first time by Mary Mandels and Elwyn T. Reese, hence the name (Bischof et al. 2016). From the wild-type strain QM6a (Mandels et al. 1969) several mutants have been developed. One of the major hypercellulolytic strain is Rut-C30, with a catabolite repression resistance for cellulases production (Eveleigh et al. 1979, Peterson et al. 2012).
An overview of the cellulose degrading enzymes secreted by *T. reesei* is illustrated in Fig. 1.5 and the characteristics of these enzymes are listed in Table 1.1.

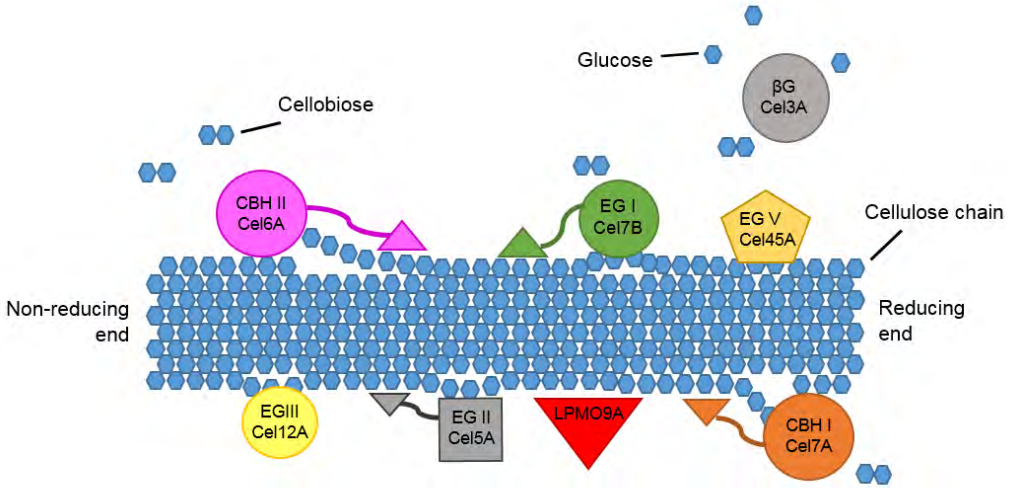


Figure. 1.5. *The cellulose-degrading enzyme system in the model fungus Trichoderma reesei. TrCel7A (CBHI, orange) and TrCel7B (CBHII, pink) are the major processive exoglucanases. TrCel7B (green), TrCel12A (yellow), TrCel5A (gray), TrCel45A (light orange) are endoglucanases with preference for amorphous cellulose. The β-glucosidase TrCel3A (gray) hydrolyzes the products of the main cellulases and relieves them from product inhibition. The non-hydrolytic enzyme LPMO9A (red) is also shown.*

Table 1.1. *General annotation of the major cellulases secreted by Trichoderma reesei. Data extracted from CAZY database (Lombard et al. 2013).*

Enzyme	Short name	Reaction mechanism	EC number	CAZy family	Amount secreted* (%)	Uniprot entry
Cellobiohydrolase I (CBH I)	*Tr*Cel7A	*retaining*, reducing-end	3.2.1.176	GH7	40-60	P62694
Cellobiohydrolase II (CBH II)	*Tr*Cel6A	*inverting*, non-reducing end	3.2.1.191	GH6	12-20	P07987
Endoglucanase I (EG I)	*Tr*Cel7B	*retaining*	3.2.1.4	GH7	5-10	P07981
Endoglucanase II (EG II)	*Tr*Cel5A	*retaining*	3.2.1.4	GH5	1-10	P07982
Endoglucanase III (EG III)	*Tr*Cel12A	*retaining*	3.2.1.4	GH12	<1-5	O00095
Endoglucanase V (EG V)	*Tr*Cel45A	*inverting*	3.2.1.4	GH45	<5	P43317
β-glucosidase	*Tr*Cel3A	*retaining*	3.2.1.21	GH3	1-2	Q12715
Lytic polysaccharide monooxygenase (formerly endoglucanase IV)	LPMO9A /Cel61A	oxidative reaction	1.14.99.54 1.14.99.56	AA9	<1	O14405

**relative percentage of the total protein produced by T. reesei RUT C-30, approximate values taken from (Rosgaard et al. 2007).*

The major cellulases can be divided in exoglucanases, endoglucanases, β-glucosidases and auxiliary enzymes.

Exoglucanases or cellobiohydrolases (CBHs) mainly hydrolyse from the cellulose chain ends and produce cellobiose as major product (Fig. 1.5). The two major secreted CBHs are the cellobiohydrolase I (CBH I) *Tr*Cel7A, acting from the reducing-end (EC 3.2.1.176) and the cellobiohydrolase II (CBH II) *Tr*Cel6A, acting from the non-reducing end (EC 3.2.1.191). They belong to GH7 and GH6, respectively. The endoglucanases (EC 3.2.1.4) act on the cellulose surface more randomly than CBHs, and are defined as catalysing endohydrolysis of (1→4)-β-glucosidic linkages to release cellodextrins. The endoglucanase I (EG I) *Tr*Cel7B is the most secreted EG by *T. reesei* and belongs to GH7. Other endoglucanases are secreted in different proportions and belong to different families (Table 1.1). Notably, this dicothomy of strictly *exo* and *endo* CBHs and EGs is an oversimplification (Kurašin et al. 2011), and will be discussed further in the next chapters.

The β-glucosidases cleave small soluble sugars released by the main cellulases (Fig. 1.5), and they are important for the regulation of cellulases expression *in vivo* (Mach et al. 1995). In addition to hydrolytic enzymes, recently discovered oxidative enzymes also act on cellulose. Lytic polysaccharide monooxygenases (LPMOs) such as LPMO9A (Fig. 1.5), have been shown to target cellulose and catalyse oxidation of the glucose units in different positions, hence breaking the glyosidic bond (Bissaro et al. 2018). However, unlike cellulases, they require a metal co-factor (Cu), an electron-donor and either O_2 or H_2O_2 for their activity (Hangasky et al. 2018). In addition to these enzymes, a number of non-enzymatic proteins are also produced, for example expansins/swollenins. They are thought to contribute to cellulose degradation by loosening up the cell wall structure with a mechanism not yet sufficiently understood (Nikolaidis et al. 2014).

Overall, these enzymes described here contribute to produce two main products: glucose and cellobiose, which can then be transported into the fungus cytoplasm to obtain energy via catabolism (Zhang et al. 2013).

1.7 - *Tr*Cel7A and *Tr*Cel7B: hydrolytic mechanism of retaining enzymes

The cellulases *Tr*Cel7A and *Tr*Cel7B are both glycoside hydrolase enzymes from family 7 (GH7) according to the CAZy classification (Table 1.1). They both catalyse the hydrolysis of β-(1→4)-glycosidic linkages and follow a *retaining* type reaction mechanism,

in which the configuration of the anomeric carbon is maintained after hydrolysis (Koshland Jr 1953).

A simplified version of the hydrolysis mechanism in *Tr*Cel7A is illustrated in Fig. 1.6, and it is similar for *Tr*Cel7B. For simplicity, the reaction mechanism shown below only describes the actual glycosidic bond breaking (for a more detailed description see (Knott et al. 2013)). *Tr*Cel7A, as other *retaining* enzymes, employs a two-step reaction mechanism. The first one, called glycosylation, involves the transfer of a proton from an acid residue (E217 in *Tr*Cel7A and E201 in *Tr*Cel7B) to the glycosidic O4 of the leaving group in subsite +1, while the nucleophile residue (E212 in *Tr*Cel7A and E196 in *Tr*Cel7B) attacks the anomeric carbon C1 in subsite -1. The stable glycosyl-enzyme intermediate is formed. In the second step, called deglycosylation, the acid residue now acts as a base and withdraws a proton from a water molecule. A nucleophilic hydroxide is formed, and attacks the glycosidic C1. Following this, the bond between the glycoside and the enzyme is broken, both the nucleophile and the acid-base are restored to their initial state, and the leaving group exits the active site.

Notably, in the active site of GH7 there is a highly conserved sequence motif EXDXXE, where a third residue, an aspartic acid, is located very close to the two catalytic glutamic acids (D214 for *Tr*Cel7A and D198 in *Tr*Cel7B). Such residue is believed to be important for the catalysis because it positions and stabilizes the nucleophile for the first hydrolysis step (Divne et al. 1994). For both *Tr*Cel7A and *Tr*Cel7B it is believed that the glycosylation step is rate-limiting in the *retaining* mechanism (Knott et al. 2014) (Zhang et al. 2013).

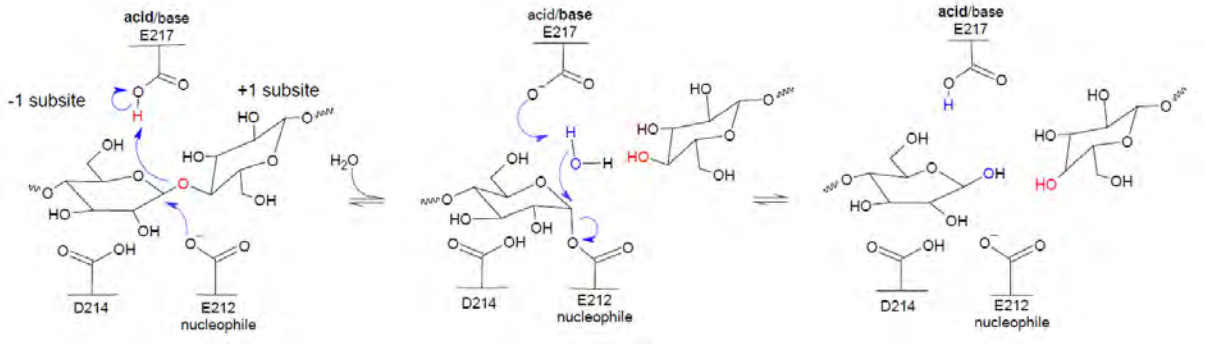


Figure 1.6. *The retaining reaction mechanism of TrCel7A on a cellulose chain. The side chains of the catalytic amino acids E212 E217 and D214 are shown, as well as the binding subsites +1 and -1. The movement of the electrons is indicated by blue arrows. Adapted from (Knott et al. 2013). The figure was created with ACD/ChemBasic 2.1 (Advanced Chemistry Development, Inc., Toronto, Canada, 2015).*

1.8 - General structure of *Tr*Cel7A and *Tr*Cel7B

Fungal GH7 enzymes contain at the N-terminal sequence a signal peptide for post-translational modifications and extracellular secretion (Saloheimo et al. 2012). The first residue of the mature peptide, a glutamine, is cyclized into pyroglutamate by a glutaminyl cyclase, with a mechanism not very well characterized in filamentous fungi (Dana et al. 2014). Such modification is also highly important for the thermal stability of the enzymes (Dana et al. 2014). The structure of GH7 enzymes consists of either a single or two domains, as illustrated in Fig. 1.7 for a number of characterized GH7. The catalytic domain (CD) is the site where the actual catalysis occurs, whereas the second domain is a carbohydrate-binding-module (CBM) which is non-catalytic and belongs to Family 1 CBM (Várnai et al. 2013). The CD and CBM are connected with an unstructured linker (Prates et al. 2018), however, most of the known and putative GH7 do not contain a CBM, which is probably related to different ecological adaptations of the organisms (Várnai et al. 2013). The CD is usually *N*-glycosylated, while linker and CBM contain O-glycosylations. Such modifications play a strong role on thermostability, folding and resistance to proteolytic attack in GH7 enzymes (Amore et al. 2017). Both *Tr*Cel7A and *Tr*Cel7B consist of CD-linker-CBM (Fig. 1.8).

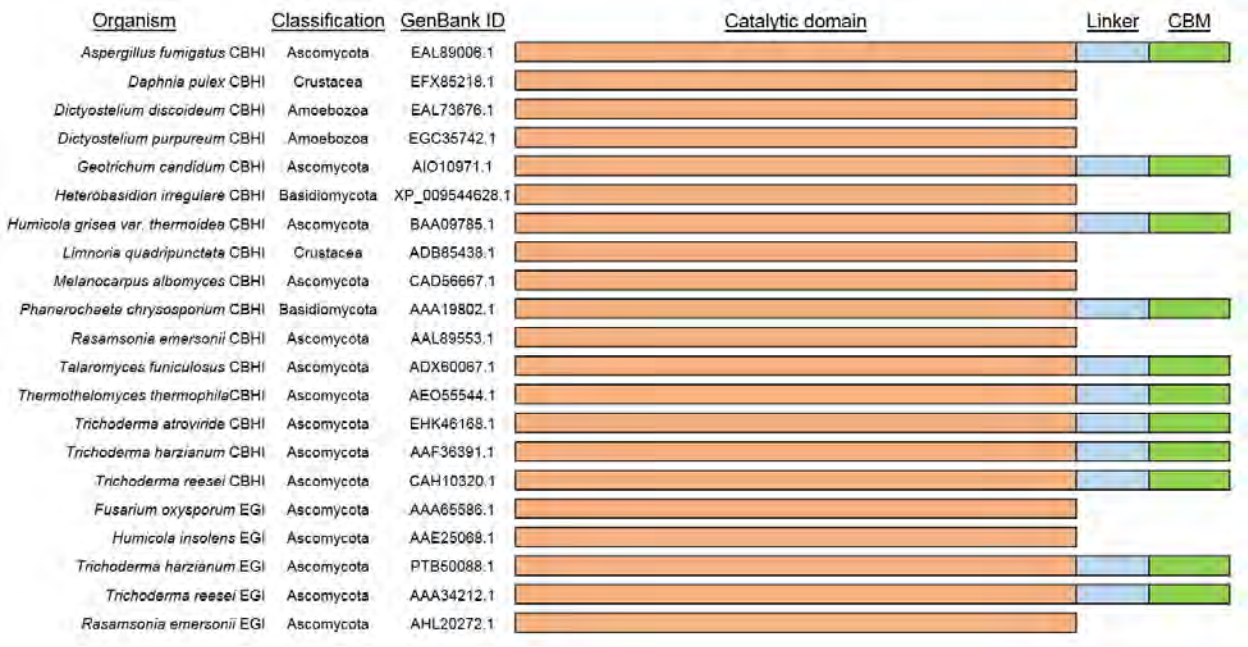


Figure 1.7. *Domain architecture of 21 GH7 with known crystal structures. GH7 cellulases are either composed of a single catalytic domain (orange) of ~450 amino acids, or the catalytic domain can be connected to a carbohydrate-binding module from family 1 (CBM, green) of ~35 amino acids, via a linker region (light blue) of ~30 amino acids. For each enzyme, the organism of origin, the enzyme subtype (CBHI or EGI), and the Genbank identifier is indicated. Figure created with Inkscape version 0.92.*

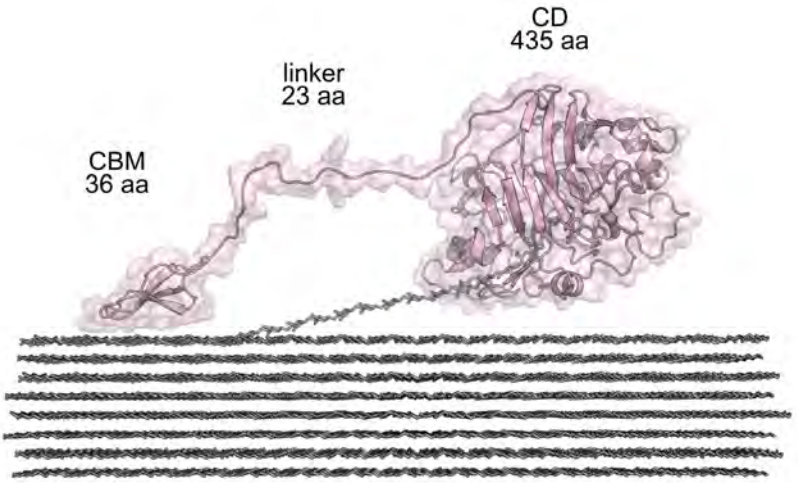


Figure 1.8. *Conceptual representation of the three-dimensional structure of TrCel7A full length. The enzyme (pink) is shown associated with cellulose (gray sticks). TrCel7A is composed of a catalytic domain (CD, PDB 4C4C) and a carbohydrate-binding module (CBM, PDB 2MWK) connected via a linker region. The number of amino acids (aa) for each region in TrCel7A is shown. The figure was created using Pymol (The PyMOL Molecular Graphics System, Version 2.3.2, Schrödinger, LLC).*

The CD of GH7 enzymes is about 50 Å wide (Fig. 1.9), and the active site is located in a tunnel or cleft formed in the CD, created by several flexible loops, which will be further described in chapter III. The tunnel/cleft is the site where a cellulose strand is threaded, after being displaced from the cellulose crystal (Knott et al. 2013). For *Tr*Cel7A and other GH7 CBHs, the CD contains sugar binding sites for up to 10 glucose units. They are named +1, +2 towards the reducing end, also called product sub-sites, and -1 to -7 towards the non-reducing end (Fig. 1.9). The glycosidic bond cleavage occurs between the -1 and +1 subsites, in accordance with the common nomenclature (Divne et al. 1994) (Kleywegt et al. 1997).

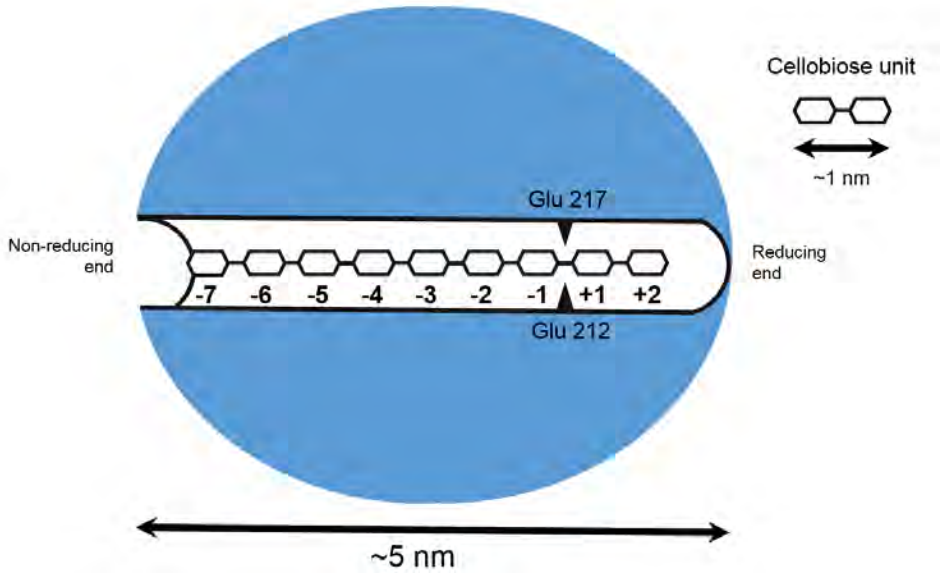


***Figure 1.9**. Schematic subsite nomenclature for TrCel7A threaded with cellononase. The substrate is stabilized by a number of residues covering subsites -7 to +2. TrCel7A cleaves the substrate from the reducing end between sub-sites -1 and +1. The catalytic residues are indicated as black triangles (Glu 212 and Glu 217 for the nucleophile and the acid/base, respectively). The size of the molecule and cellobiose have been approximated from (Divne et al. 1998) and the crystal structure (PDB 4C4C).*

The CBM from family 1 associated with GH7 enzymes consists of a polypeptide chain of about 36 residues and has a fundamental role of increasing the amount of enzyme adsorbed to the crystalline surface of cellulose (Lehtiö et al. 2003). The structure of many different CBM1s have been solved by NMR, including the CBM of *Tr*Cel7A (PDB 1CBH, Fig. 1.10A, (Kraulis et al. 1989) and *Tr*Cel7B (PDB 4BMF, (Mattinen et al. 1998)). The structure of the CBM1 forms a wedge shape stabilized by two or three disulphide bridges

and O-glycosylations (Payne et al. 2015). A number of residues, particularly the aromatics tryptophans and tyrosines, are involved in forming a planar surface which interacts with the hydrophobic crystal plane (Lehtiö et al. 2003). To illustrate the sequence conservation of the CBM from family 1, 42 CBM1 sequences from the GH7 CAZy database were selected to perform an alignment using Jalview (Waterhouse et al. 2009). The resulting alignment was analysed for sequence logos, which are graphical representations of patterns in a multiple sequence alignment (Crooks et al. 2004), shown in Fig. 1.10B. Highly conserved residues are the cysteines forming disulphide bridges (in position 8, 19, 25 and 35 in Fig. 1.10B), and a number of aromatics (position 5, 31 and 32 Fig. 1.10B), which has been previously indicated to be important for cellulose binding (Beckham et al. 2010). Notably, the positions 1 and 3 in Fig. 1.10B correspond to a threonine and a serine respectively in *Tr*Cel7A, and are O- glycosylated, which affect the thermal stability and affinity for crystalline cellulose (Chen et al. 2014).

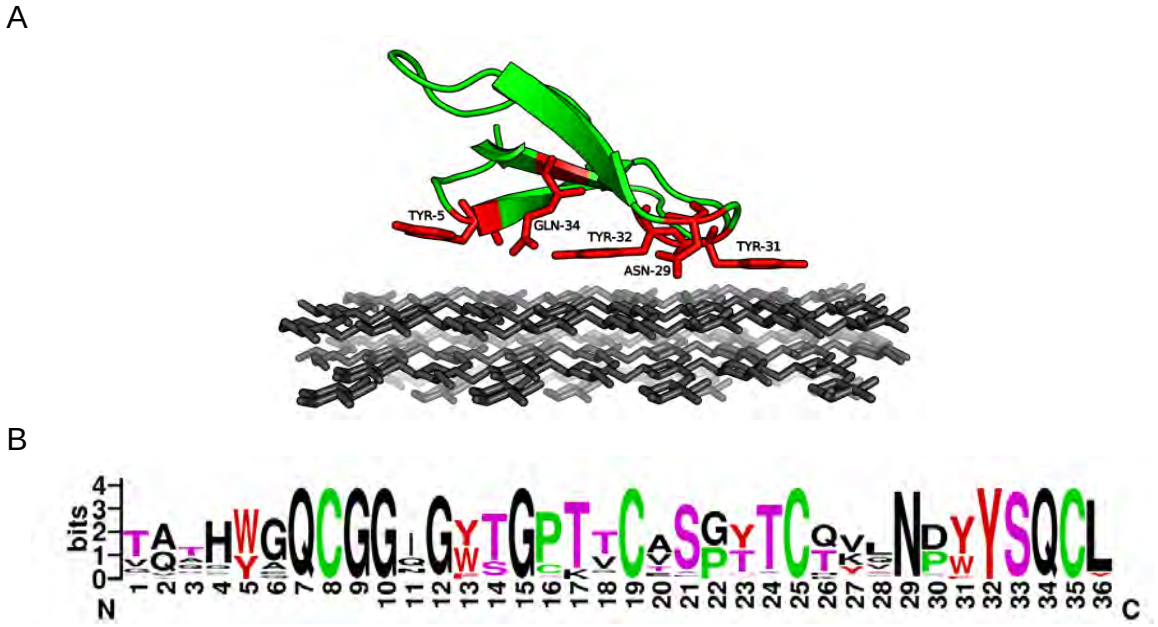


Figure 1.10. *Structure and sequence of the carbohydrate-binding module in GH7 enzymes. (A) Cartoon representation of TrCel7A CBM (PDB: 1CBH) (Kraulis et al. 1989). The CBM (green) is shown interacting with the hydrophobic face of the cellulose surface (grey). Residues forming the planar face of the CBM and close to the substrate surface are shown as red sticks. (B) LOGO representation of a multiple sequence alignment of 42 CBM1 from CAZy database (Lombard et al. 2013). The size of the 1-code amino acid letters is proportional to the sequence conservation. Color code: serine and threonine, purple; proline and cysteines, green; aromatic amino acids, red; other amino acids are colored in black. Panel A was created with Pymol and panel B with Weblogo (Crooks et al. 2004).*

Despite the high sequence similarity, the ability of the CBM to target cellulose depends on the amino acidic differences, as shown by Linder *et al.*, who reported that the binding strength of *Tr*Cel7A and *Tr*Cel7B CBMs is different (Linder et al. 1995). They investigated the binding on cellulose of these CBMs as isolated polypeptides and found out that *Tr*Cel7B CBM has stronger binding than *Tr*Cel7A CBM, and propose that this difference is mainly related to the presence of a W in *Tr*Cel7B CBM, which is substituted to a Y in *Tr*Cel7A CBM (position 5 in Fig. 1.10B). We will see in chapter III however that the overall binding of *Tr*Cel7B on microcrystalline cellulose is actually lower than *Tr*Cel7A. This suggest that a fine balance exists between enzyme activity and affinity on cellulose and that possibly the CBM might have a compensatory role for the lower binding of the CD in *Tr*Cel7B (some consideration in this regard are also made in Article III).

Chapter II - Cellulase kinetics

Many textbooks consider enzyme catalysis a biological example of homogeneous catalysis. This is true for many traditional bulk enzyme-substrate systems, where both elements are free to interact in solution. However, the kinetics of cellulases differ in this regard and can be considered an example of biological heterogeneous catalysis. This is due to the physical characteristics of the natural substrate cellulose, which can be simplistically described as an insoluble crystal (Chapter I). Once secreted in the extracellular medium, cellulases need to diffuse from the aqueous phase and adsorb onto the solvent-cellulose interface in order to perform the catalysis. This property makes cellulase kinetics different form conventional enzyme kinetics. In this chapter, a steady-state approach for interfacial enzyme kinetics will be presented, along with two steady-state kinetic models and related equations, which have been used throughout the thesis and appended articles to describe and understand the activity of different cellulases. The mathematical derivation of such equations lies outside the scope of this thesis and the reader is referred to the referenced articles.

2.1 – Steady-state models for cellulases

Several mechanistic models have been proposed to describe the kinetics of GH7 enzymes, recently reviewed in (Jeoh et al. 2017). The enzyme *Tr*Cel7A, as well as other CBHs like *Tr*Cel6A, performs hydrolysis in a processive manner. This means that the enzyme remains associated with the substrate cellulose after the first hydrolytic event and performs multiple others before dissociating from it (Jalak et al. 2010).
The enzymatic reaction on cellulose can be roughly divided into three main steps (Fig. 2.1) and described by a microkinetic scheme (Scheme 1).

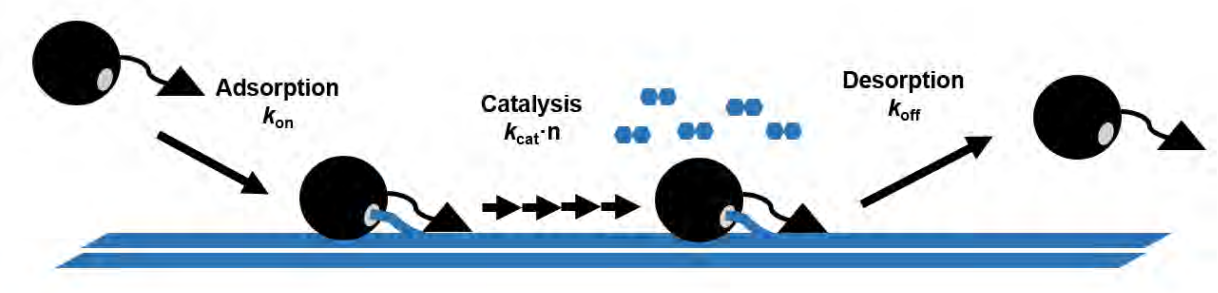


Figure 2.1. *Illustration of the action of a processive cellobiohydrolase on the cellulose surface.*

$$E + C_m \xrightarrow{k_{on}} \left(EC_m \right)_n \xrightarrow{k_{hyd}} EC_{m-n} + C$$

$$\downarrow k_{off} \qquad \downarrow k_{off}$$

$$E + C_m \qquad E + C_{m-n}$$

Scheme 1. *Simplified reaction scheme for processive cellulases, adapted from (Cruys-Bagger et al. 2013, Christensen et al. 2018)*

1. *Adsorption and threading of the cellulose strand.* The enzyme adsorbs on the cellulose surface, threads a cellulose chain into the active site and forms the ES complex. In the microkinetic scheme (Scheme 1) this process is described with the enzyme (E) and the substrate being a cellulose chain consisting of m cellobiose units (C_m). They combine together to form the Michaelis complex (EC_m). The formation of the complex is governed by the k_{on} rate constant.

2. *Catalysis.* If the association with the substrate is productive, the enzyme can perform catalysis in a processive manner. In the microkinetic scheme, the EC_m complex is formed and the enzyme can perform hydrolysis (k_{hyd}) to produce cellobiose (C) and a new complex, which now contains a reduced amount of cellobiose units (C_{m-1}). This event is governed by the rate constant k_{hyd}. n is the total number of catalytic events that the enzyme can perform without dissociation, called the processivity number. In other words, following association, the enzyme can perform n catalytic steps. In the case of processive enzymes this value is >1 and for pure non-processive enzymes, this number is theoretically 1. For *Tr*Cel7A, processivity values >10 have been reported (Cruys-Bagger et al. 2012), although the theoretical value (number of catalytic steps performed before dissociating on an ideal polymer) is much higher and never reached in practice, due to steric hindrances and obstacles on the cellulose surface (Kurašin and Väljamäe 2011).

3. *Unbinding and desorption.* In case the ES complex is bound non-productively, it simply dissociates. In the microkinetic scheme this is described by EC_m complex dissociating. The ES complex also dissociates after performing n catalytic steps. The rate constant governing the dissociation event is k_{off}.

Endoglucanases like *Tr*Cel7B are traditionally categorized as a non-processive, indicating that all the catalytic events are independent from each other. However, this simplistic distinction should be addressed with caution, since it has been shown that *Tr*Cel7A and similar CBHs can also perform endolytic attack (Ståhlberg et al. 1993, Kurašin and Väljamäe 2011, Badino et al. 2017) and EGs show some degree of processivity and produce soluble products (Murphy et al. 2012)(see also section 3.8).

Pre-steady state measurements of *Tr*Cel7A, *Tr*Cel6A done in a very short time scale (<5 s) showed that the kinetics of cellulases is characterized by an early burst phase followed by a dramatic slowdown in the reaction rate over time. These analyses also allowed to derive values for the individual rate constants k_{on}, k_{cat} and k_{off} and the processivity number n for two CBHs, *Tr*Cel7A and *Tr*Cel6A, using *ad hoc* instrumentation (Cruys-Bagger et al. 2012, Christensen et al. 2018). A distinctive burst phase has been also shown for *Tr*Cel7B (Murphy et al. 2012). Pre-steady state measurements present a number of experimental challenges related to handling insoluble material and time resolution of the instruments, therefore they are often impractical for comparisons between many enzymes. Nonetheless, these works elucidate that the rate-limiting step in cellulose hydrolysis lie outside the hydrolysis of the glycosidic bond of the substrate. In general enzymology, k_{cat} is, by definition, the slowest first-order rate constant in the forward direction of the process (Nelson et al. 2008), however the aforementioned studies show that for cellulases k_{off} is much slower than k_{hyd}, thus making k_{off} rate-limiting at steady-state. This means that for cellulase kinetics the "rate limiting k_{cat}" is actually k_{off}. To avoid confusion, in some works the rate constant governing the glycosydic bond cleavage is called k_{hyd} (Christensen et al. 2018). Thus, the specific maximal rates measured in this thesis and appended articles are usually a reflection of k_{off}.

Normally, calculating endoglucanase activity by only measuring the amount of soluble product released underestimates the total activity of the enzyme. Indeed, this could also be the case for many endoglucanases acting on cellulose, since some of the activity involves breaking the glyosidic bonds that remain associated with the insoluble part of the cellulose lattice. However, the endoglucanase *Tr*Cel7B has been showed to produce almost exclusively (97%) soluble reducing ends on RAC (Murphy et al. 2012), unlike other *T. reesei* EGs from GH5 and GH12, producing higher proportion of insoluble reducing ends. Moreover, regarding the soluble products, the same study showed that *Tr*Cel7B does not

release celloligomers as major products like a "true" endoglucanase, but rather glucose and cellobiose (Murphy et al. 2012), as we also observe on Avicel in section 3.8. Thus, the *Tr*Cel7B activity reported in this thesis reflects the release of soluble products, and we assume that this propensity is maintained for other GH7 EGs. This somewhat crude assumption was necessary to achieve a suitable balance between good throughput and reproducibility, quantitative kinetic data, and ability to compare many different GH7 enzymes.

2.2 - The conventional and inverse Michaelis-Menten approaches

In this thesis and appended papers, kinetic descriptions of the enzymes are done using two steady-state models extensively used in our research group to describe interfacial enzyme kinetics. The two approaches are somewhat complementary to each other and are illustrated in Fig. 2.2. For both approaches, initial rates are calculated using end-point measurements of 1 h, which has been recently shown to be a suitable approximation of initial rates measured with other more laborious techniques (Kari et al. 2019).

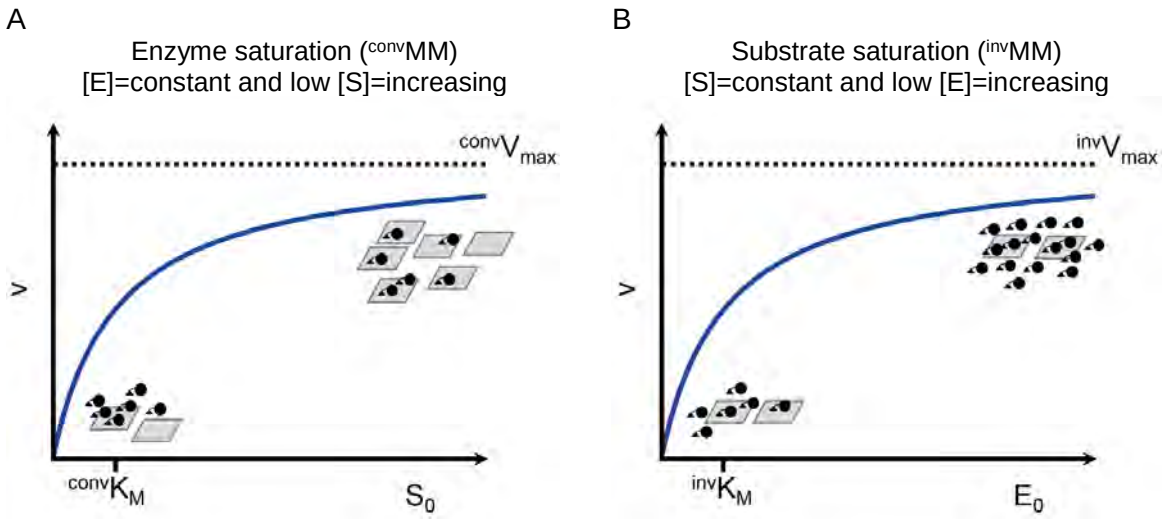


Figure 2.2. *Schematic illustration of the conventional (^{conv}MM) and inverse Michaelis-Menten (^{inv}MM) approaches. The cellulose substrate is depicted as a gray surface, while the enzyme is shown in black. (A) In the ^{conv}MM approach, steady-state rates are plotted as a function of the substrate load (S_0). A constant, low enzyme concentration (E_0) is used, and rates level off with increasing substrate loads. (B) In the ^{inv}MM approach, steady-state rates are plotted as a function of the initial enzyme concentration (E_0). A fixed, low substrate load is used, while the enzyme is in excess. Rates level off with increasing enzyme concentrations. Blue lines are saturation curves from which the kinetic parameters can be*

estimated (see main text).

The first approach used is the conventional Michaelis-Menten (^{conv}MM) model adapted for cellulase kinetics (Cruys-Bagger et al. 2013). The model is derived under the usual assumption that the substrate concentration is so high compared to the enzyme that is considered approximately constant and equal to the initial substrate concentration ($S \approx S_0$). Combining the quasi-steady state assumption of the ES complex and the mass conservation of the enzyme leads to the classical Michaelis-Menten equation:

$$^{conv}v / E_0 = \frac{^{conv}V_{max} S_0}{^{conv}K_M + S_0} \quad \text{(Eq. 1)}$$

This relationship is also illustrated in Fig. 2.2A. S is the substrate load expressed in g/L and the subscript 0 is the initial or total value. $^{conv}V_{max}/E_0$ is the maximal specific rate at saturation, and can be approximated to $^{conv}V_{max} = k_{cat}E_0\beta$, with k_{cat} being the turnover frequency of the ES complex in s^{-1}, E_0 the initial enzyme concentration in µM, and β is the probability ($0 < \beta \leq 1$) the enzyme dissociates from the cellulose strand before performing n catalytic cycles. This means that if the mean free path (n^{exp}) is smaller than the intrinsic processivity of the enzyme (n^{theo}), or $n^{exp} < n^{theo}$, then β will be low, since the enzyme will frequently perform a full processive run (n^{exp}) and stall unproductively (Cruys-Bagger et al. 2013). Although $^{conv}V_{max}/E_0$ gives a good indication on how quickly the enzyme converts the substrate, it does not reflect the overall efficacy of the enzyme on the surface (and we will argue why below). K_M is the substrate load when the reaction rate is at half saturation and has the units of g/L. For cellulases, K_M can also be considered a dissociation constant ($K_M = k_{off}/k_{on}$), since the enzyme can only escape the ES complex by dissociation as hydrolysis give rise to a new ES_{n-1} complex, and dissociation is only governed by one rate constant, k_{off} (Cruys-Bagger et al. 2013). In some instances, for example in Article III and for other cellulases (Christensen et al. 2019), it is not possible to reach full saturation under MM conditions, because the insoluble nature of the substrate sets some experimental limitations regarding handling high substrate loads (>100 g/L) accurately. We report instead the value of the specificity constant η:

$$\eta = \frac{^{conv}V_{max} / E_0}{^{conv}K_M} \quad \text{(Eq. 2)}$$

which can be calculated as the slope of the initial part of the MM curve. The specificity

constant can be considered a second order rate constant that governs the reaction rate at low substrate loads (Nelson et al. 2008) and has been used in this thesis in a comparative way as a measure of the overall efficacy of the enzymes (Sorensen et al. 2015).

The second kinetic approach used is called inverse Michaelis-Menten (^{inv}MM), and the prerequisite is a large excess of enzyme: $S_0 << K_M + E_0$ (Kari et al. 2017). Unlike enzymatic hydrolysis in bulk conditions, where the substrate would be immediately converted under enzyme excess, the insoluble nature of cellulose allows the establishment of a new steady-state.

The rationale behind the ^{inv}MM approach is as follows. When acting on cellulose, cellulases have only a certain amount of *attack sites* available on the substrate surface, that is, loci where the enzyme can bind and initiate hydrolysis. The amount of *attack sites* is most often less than the total available sites the enzyme can bind to (Christensen et al. 2018). During enzymatic hydrolysis, the total cellulose particle size remains fairly constant at short (1 h) contact times. Thus, the *attack site* concentration can also be considered constant for the same timeframe, since every time the enzyme binds and cleaves parts of the cellulose crystal, new attacks sites will be exposed. If the enzyme concentration is very high compared to the total amount of *attack sites*, it is possible to reach a new stationary condition where the *attack sites* become saturated with the enzyme (Kari et al. 2017). In this kinetic model, the enzyme concentration is assumed to be constant and equal to the total enzyme concentration, while the mass conservation is maintained for the substrate. The resulting equation is a hyperbolic function (Fig. 2.2B) very similar to the ^{conv}MM:

$$^{inv}v / S_0 = \frac{^{inv}V_{max} E_0}{^{inv}K_M + E_0} \quad \text{(Eq. 3)}$$

The inverse maximal rate $^{inv}V_{max}$, describes a saturation condition where all attack sites are complexed with enzyme. Since $^{inv}V_{max} = k_{cat} S_0 \Gamma_{attack}$, it corresponds to the product of the turnover frequency k_{cat} and the molar density of attack sites Γ_{attack} (expressed in moles per gram of substrate). As such, $^{inv}V_{max}$ takes into consideration both the catalytic speed and the ability to find sites on the cellulose surface. The substrate load (S_0) is expressed in g/L and $^{inv}K_M$ is the enzyme concentration at half saturation.

By combining the saturation parameters of Eq. 1 and Eq. 3, the attack sites factor, Γ_{attack} can be estimated:

$$\Gamma_{attack} = \frac{{}^{inv}V_{max} / S_0}{{}^{conv}V_{max} / E_0} \quad \text{(Eq. 4)}$$

Γ_{attack} is expressed in µmol/g and reflects the number of sites available for the enzyme to attack per g of substrate. Γ_{attack} will change with time, as the properties of cellulose also change with time, but it is assumed constant in the kinetic analyses used in this thesis and appended papers, where initial rates are measured after 1 h enzyme-substrate contact time. This new parameter introduced by the ^{inv}MM approach also aids in quantifying the molar substrate concentration, which is usually not known for cellulosic substrates (Kari et al. 2017).

2.3 - Cellulases adsorption

Measuring the total adsorption of the enzymes onto the cellulose surface gives indication regarding the total amount of sites available for the enzyme to bind, both as attack sites and non-productive binding sites. One simplified way to analyse this is with the Langmuir isotherm model, where the substrate coverage, Γ, is measured as a function of the free enzyme, E_{free}:

$$\Gamma = \frac{\Gamma_{max} E_{free}}{K_d + E_{free}} \quad \text{(Eq. 5)}$$

The amount of absorbed enzyme per g of cellulose, the substrate coverage, is defined as $\Gamma=(E_0-E_{free})/S_0$ and expressed in µmol/g, while K_d is the dissociation constant in µM. Γ_{max} is the saturation coverage, that is, the coverage of enzyme on the substrate at saturation and it is measured as µmol of enzyme bound per g of cellulose. This value indicates the total adsorption sites and does not specify whether the sites are available for catalysis, *i.e.* the enzyme could be bound but does not perform catalysis. In some cases, the partitioning coefficient K_p can be used

$$K_P = \frac{\Gamma_{max}}{K_d} \quad \text{(Eq. 6)}$$

especially when full coverage is not achieved, and it is measured as the slope of the initial part of the hyperbolic curve at low enzyme concentrations. Simple Langmuir isotherms can be considered an oversimplification of the binding on a heterogeneous surface like cellulose (Jalak et al. 2014) nonetheless they are valuable tools for describing the general affinity of the enzyme to the substrate.

Notably, the parameters Γ_{attack} and Γ_{max} can be compared. The former can only be calculated from kinetic data and the latter comes from adsorption measurements. Differences between Γ_{attack} and Γ_{max} give useful insights on the characteristics of cellulases and their ability to attack the substrate with different promiscuity. If $\Gamma_{max} \approx \Gamma_{attack}$, then all the adsorption sites are also attack sites for the enzyme. In this case, the cellulase can be considered an enzyme with high surface promiscuity, because it can hydrolyse many sites it binds to. If $\Gamma_{max} >> \Gamma_{attack}$, then a limited fraction of sites are attack sites for the enzyme. In this case the enzyme is considered surface specific (or with low promiscuity). Example of both types of enzyme will be described in the next section.

2.4 - Kinetic comparisons between *Tr*Cel7A and *Tr*Cel7B

The equations described in the previous section have been applied in the following, to study the kinetics of *Tr*Cel7A and *Tr*Cel7B at 25 °C. The results are shown in Fig 3.5 and Table 3.2, Chapter III.

For the ^{conv}MM approach, we used a very low enzyme concentration (50 nM) and increasing Avicel load. The maximal specific rate $^{conv}V_{max}/E_0$ is about an order of magnitude higher for *Tr*Cel7B than *Tr*Cel7A, ~2 s^{-1} compared to ~0.17 s^{-1}, Table 3.2. This higher turnover frequency for *Tr*Cel7B indicates that it is a much faster enzyme than *Tr*Cel7A.

The situation is diametrically opposed at ^{inv}MM conditions, where the substrate load is kept constant and low (8 g/L). In this case *Tr*Cel7A shows five times higher $^{inv}V_{max}$ than *Tr*Cel7B (0.025 vs 0.005 $\mu molg^{-1}s^{-1}$, Table 3.2). This suggest that despite *Tr*Cel7B being catalytically superior, it has fewer attack sites on cellulose.

The overall efficacy on the cellulose surface was quantified with the kinetic substrate accessibility Γ_{attack}, which is 0.147 $\mu molg^{-1}$ for *Tr*Cel7A compared to 0.003 $\mu molg^{-1}$ for *Tr*Cel7B (Table 3.2). As mentioned in the previous section, Γ_{attack} reflects the different site promiscuity on the cellulose surface for the two enzymes. On Avicel, *Tr*Cel7A has ~50 times more attack sites than *Tr*Cel7B, meaning *Tr*Cel7A, despite being slow, has a wide range of sites to attack on the cellulose surface. *Tr*Cel7B is more specific and attacks a small subset of sites on the cellulose surface.

The reason for this promiscuity of *Tr*Cel7A is not fully understood, but it might play a role in the synergistic degradation of cellulose (Chapter V). Since CBHs prefer more crystalline regions, while EGs more amorphous ones (Zhang and Lynd 2004), this site promiscuity could be related to increased specificity for cellulose regions that allows the enzymes to have some degree of independent (non-overlapping) sites on cellulose.

We can compare the adsorption saturation values Γ_{max} to the density of attack sites Γ_{attack} to obtain information on how many of the total binding sites are actually attack sites (Table 3.2). For *Tr*Cel7A Γ_{max} is 0.40 µmol/g and Γ_{attack} is 0.15 µmol/g, while for *Tr*Cel7B is 0.87 µmol/g and 0.003 µmol/g respectively (Table 3.2). This means that for *Tr*Cel7A about half (38%) of the sites the enzyme adsorbs to can be hydrolysed. In *Tr*Cel7B instead, only ca. 0.3% of the total sites are attack sites.

If we multiply Γ_{attack} with the substrate load used (8 g/L) we can estimate the molar concentration of attack sites. In case of *Tr*Cel7B this number is 24 nM, indicating that the number of moles of attack sites in 1 L is extremely low (24 nmoles) so the enzyme has very little substrate available for catalysis.

The site selectivity of *Tr*Ce7B is a common feature in the EGs, and it is one of the main hurdles to improve activity of the enzymes (see also chapter IV). It is difficult to rationalize

a way to increase the proportion of attack sites without losing the catalytic superiority. However, one possibility could be designing enzyme variants that could increase the total adsorption, since the binding is very low, as indicated by the high values of K_M and K_d (chapter IV).

Chapter III - Substrate accessibility in family 7 glycoside hydrolases

This chapter will describe the structure/function differences between GH7 cellobiohydrolases and endoglucanases, and identify the role of a number of flexible loops in affecting enzyme catalysis. The first part of the chapter will describe results from article I and unpublished results, where a direct kinetic comparison between *Tr*Cel7A, *Tr*Cel7B and deletion variants of *Tr*Cel7A was made, which elucidated the functional role of the loop called B2. The second part of the chapter describes other structural differences found in GH7 enzymes and some general trends.

3.1 - The GH7 enzymes in CAZy database

There are more than 5000 sequences annotated in the Carbohydrate-Active enZYmes database (CAZy, www.cazy.org) as family 7 glycoside hydrolases (GH7) (Lombard et al. 2013), although the majority of these are fragments. To date, only 84 sequences have been annotated as characterized, meaning having biochemically confirmed activity. Twenty-one crystal structures are available for GH7, one of which was solved during the course of this PhD (chapter IV). Regarding their biological distribution, GH7 encoding genes have been mostly identified in fungal genomes, however they are also present in other eukaryotes evolutionarily distant from fungi, such as Amoebozoa (Hobdey et al. 2016) and Metazoans (Kern et al. 2013). GH7s are yet to be found in the Bacteria and Archaea domains of life, although they possess cellulases from other GH families, suggesting a different evolutionary strategy for deconstructing cellulose.

A phylogenetic tree has been created with the biochemically characterized GH7 sequences in Article III (not shown), where fungal GH7 were divided into three main subgroups: ascomycotal CBHs, basidiomycotal CBHs and ascomycotal EGs. Other phylogenetic studies distinguished in more detail between fungal classes (Mertz et al. 2009). Interestingly, no basidiomycotal GH7 EGs have been identified so far (Nakamura et al. 2014).

3.2 - Overall structure of GH7 enzymes

GH7 CBHs share very high structural similarity, even between evolutionarily distant proteins (Hobdey et al. 2016). The overall fold is conserved also amongst EGs, although the few EGs structures available are less phylogenetically diverse than the CBHs (Article III). As for the secondary structure, GH7 are classified as mainly β, with a distorted β-sandwich topology (Pearl et al. 2003). The two major β sheets form the inner and outer part of the sandwich and are connected by numerous loops (blue in Fig 3.1).
In the tertiary structure, the loops close to the active site appear to be the main determinants for the overall substrate binding shape of CBHs and EGs (Payne et al. 2015). Eight peripheral loops form a tunnel-shaped active site in the CBHs, while the EGs have an open cleft shape, due to the shortening or absence of four of these loops (Fig. 3.2 and Fig. 3.3). The loop disappearance in the latter sub-group suggests an evolutionary origin of the EGs from the CBHs by events of duplications and deletions (Hobdey et al. 2016).

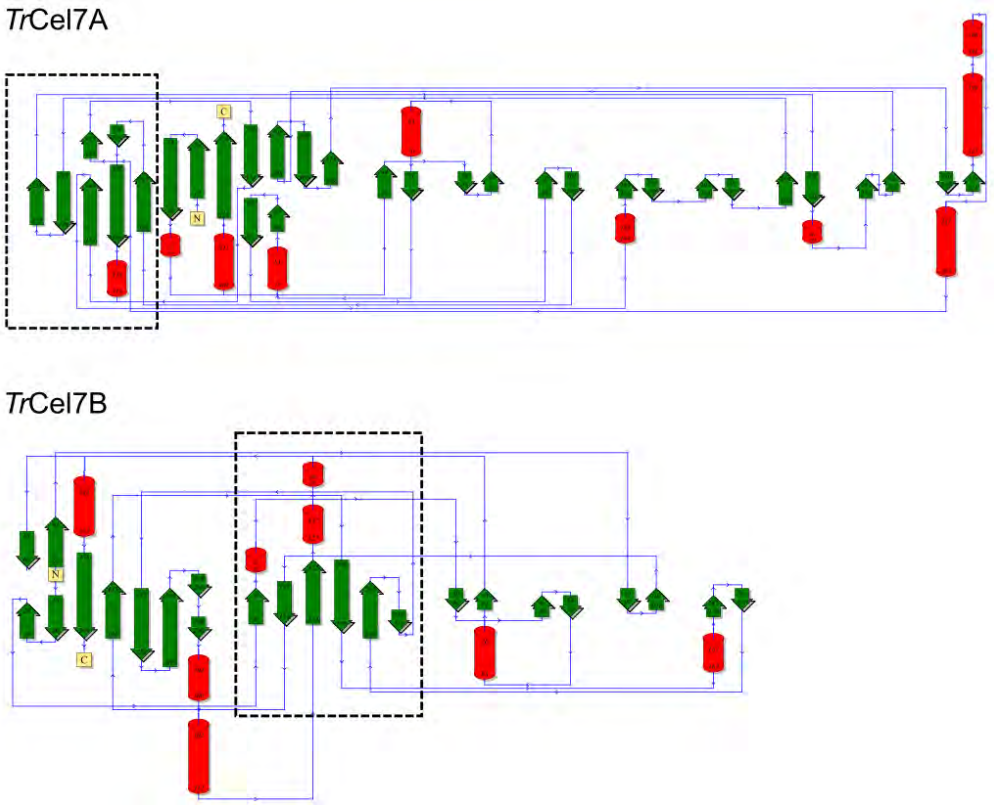


Figure 3.1*. Topology diagrams for the catalytic domain of TrCel7A (top) and TrCel7B (bottom). β-strands and α-helices are indicated by green arrows and red cylinders, respectively. The blue lines represent the connecting loops. The dashed squares indicate the inner (concave) β-sheet* (Kleywegt et al. 1997)*, where the TrCel7A tunnel or the TrCel7B cleft is formed. Figure adapted from PDB sum* (Laskowski et al. 1997).

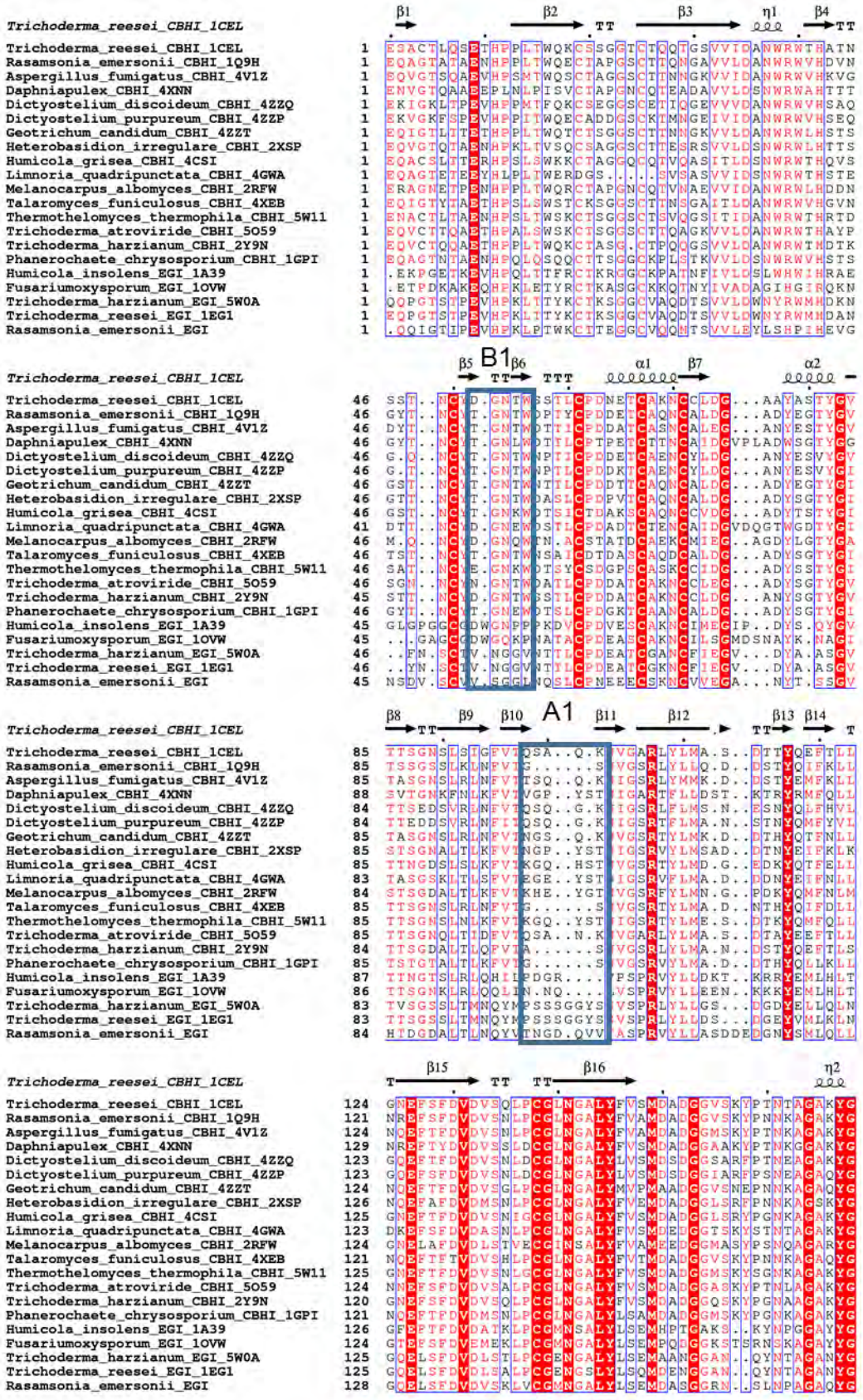


Figure 3.2. *Multiple amino acid sequence alignment of 21 GH7 cellulases with experimentally-determined structure. Secondary structure elements of TrCel7A structure (PDB 1CEL) are shown on top of the alignment with the following figures and symbols: alpha helices (helices,* α *and* η *for a* 3_{10}-

helix), beta strands (arrows, β). turns (T, strict β-turns as TT and strict α-turns as TTT). Strictly identical residues are marked in white characters on a red background, while chemically similar residues are red characters. Regions of conserved, highly similar residues are framed in blue boxes. Alignment created with Clustal Omega (Madeira et al. 2019) and rendered with ESPript 3.0 web server with default parameters (Robert et al. 2014). Blue frames indicate loop regions of interest, with loop nomenclature above (A1-A4, B1-B4). Blue stars indicate residues involved in the catalysis: catalytic nucleophile, assisting residue and proton donor/base (E212, D214 and E217 following TrCel7A numbering). Protein sequence identifiers are shown on the left with the name of the organism followed by the type of enzyme (CBHI or EGI) and the PDB accession number. For Rasamsonia emersonii EGI the PDB accession number is 6SU8 (Article III). The figure was prepared in Inkscape.

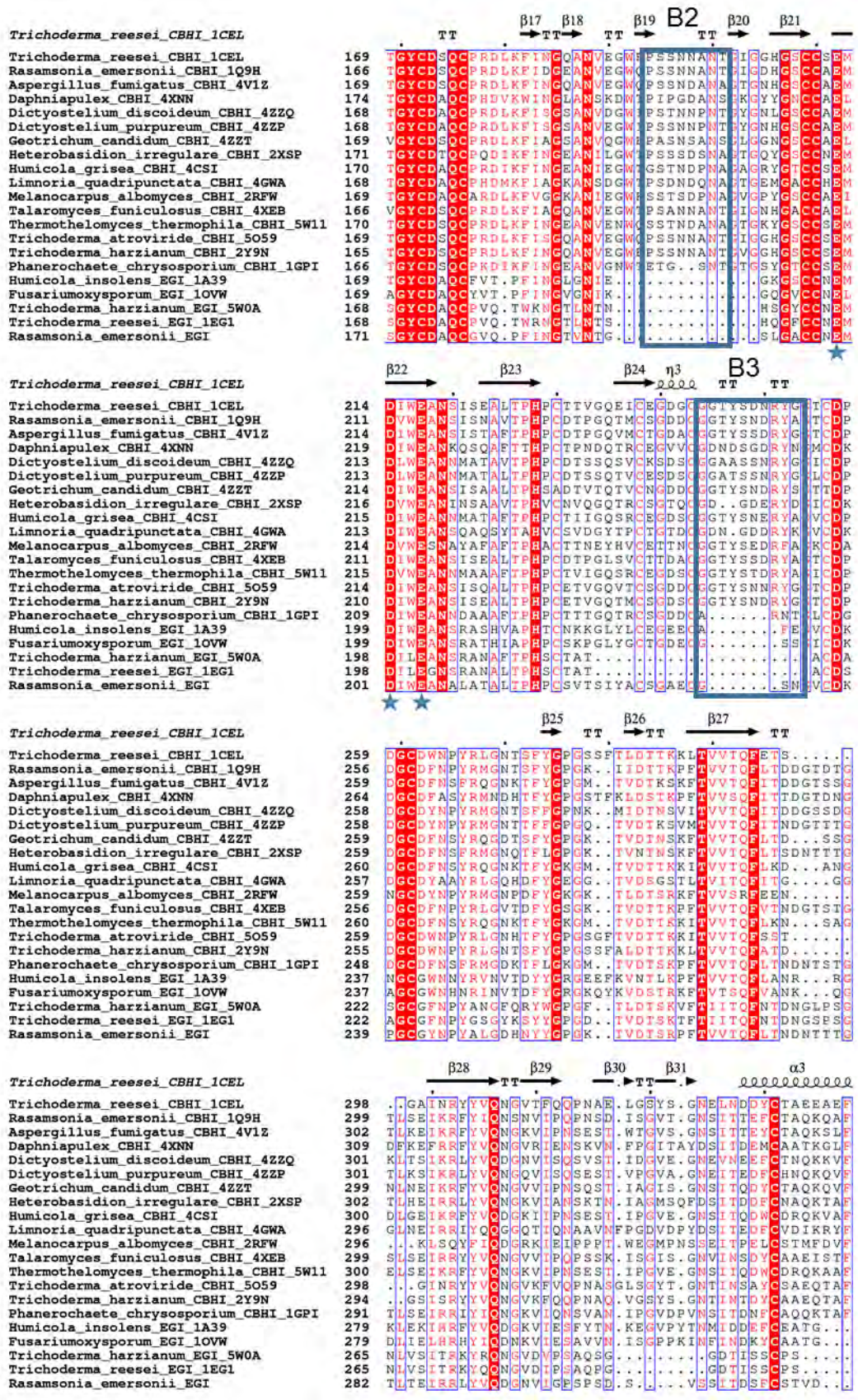


Figure. 3.2. Continued

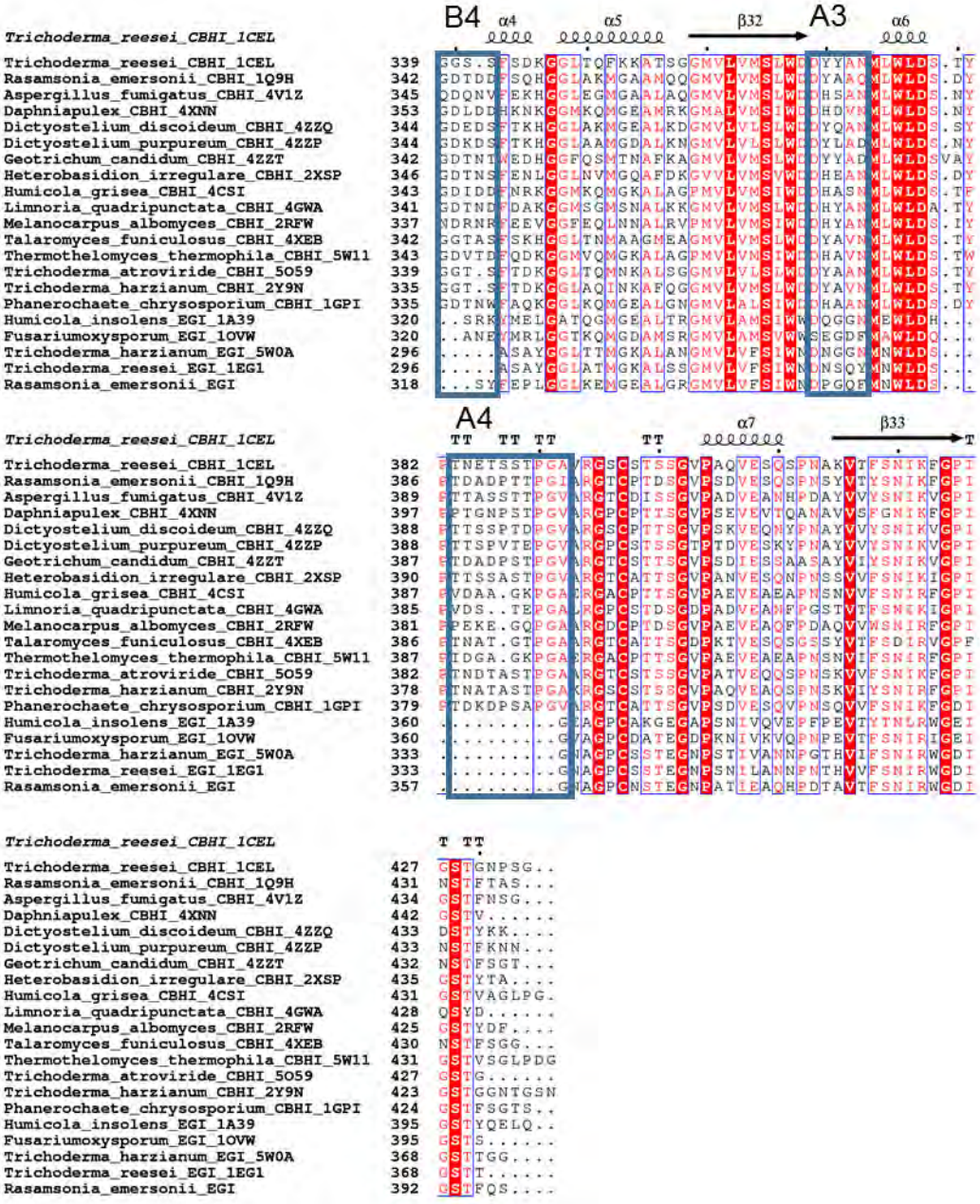


Figure 3.2. Continued.

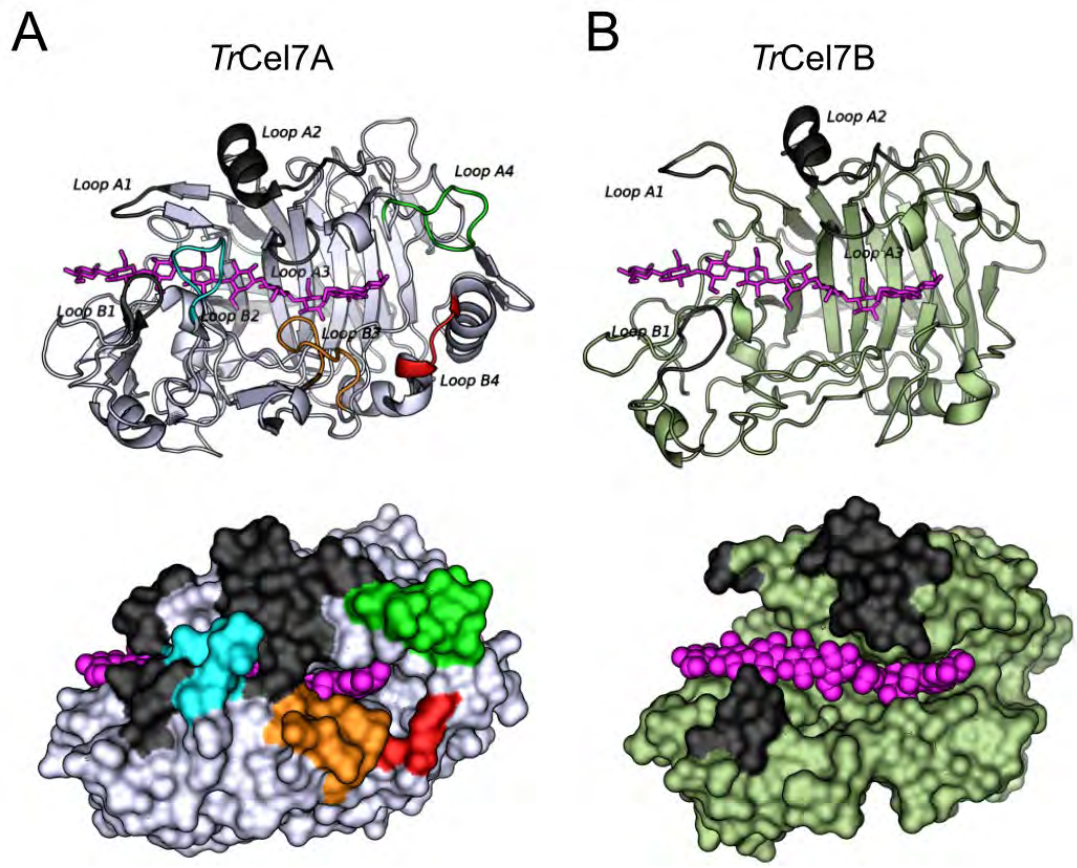


***Figure 3.3**. Structural differences between TrCel7A and TrCel7B. Cartoon (top) and surface (bottom) representation of the structures of (A) TrCel7A (gray, PDB 4C4C) and TrCel7B (green, PDB 1EG1) complexed with cellononaose (magenta sticks and spheres for top and bottom panels respectively). Substrate-binding loops that are common between the two structures are highlighted in black, while loops unique to TrCel7A are highlighted in different colours: B2 cyan, B3 orange, B4 red and A4 green. Figure created in Pymol (The PyMOL Molecular Graphics System, Version 2.3.2, Schrödinger, LLC).*

3.3 - Loop differences between GH7 CBHs and EGs

The multiple sequence alignment shown in Fig. 3.2 and the loop structures of *Tr*Cel7A and *Tr*Cel7B (Fig. 3.3) illustrate the amino acid composition and position of the major loops involved in substrate binding and catalysis. Using a previously established loop nomenclature (Momeni et al. 2014) eight main substrate binding loops have been identified in the CBHs, B1-B4 and A1-A4. The last 5 sequences in the multiple sequence alignment in Fig. 3.2 correspond to EGs, and it is possible to identify the four missing regions

corresponding to loop B2, B3, B4 and A4. In Article I, we focused on the structure-function differences between *Tr*Cel7A and *Tr*Cel7B as representative members of the GH7 CBH and EG sub-groups, respectively, with the aim of elucidating the functional role of these different loops. As described in section 2.4, *Tr*Cel7A and *Tr*Cel7B show distinct kinetics on Avicel. *Tr*Cel7B has a high turnover number compared to *Tr*Cel7A but low conformational promiscuity, which signifies a poor ability to locate many attack sites on cellulose. On the other hand, *Tr*Cel7A binds more strongly, is less selective and more processive. In light of the aforementioned characteristics, *Tr*Cel7A and *Tr*Cel7B differ in their ability to hydrolyse cellulose with different crystallinity. *Tr*Cel7A is efficient in degrading highly crystalline cellulose like BMCC, while *Tr*Cel7B effectively hydrolyses amorphous cellulose (Zhang and Lynd 2004). Moreover, product inhibition by cellobiose is different, high in *Tr*Cel7A and low in *Tr*Cel7B (Murphy et al. 2013). Some of these key differences are listed in Table 3.1.

Table 3.1. *List of main differences between TrCel7A and TrCel7B regarding kinetics on microcrystalline cellulose, cellulose adsorption, endolytic activity and processivity.*

	*Tr*Cel7B	*Tr*Cel7A	References
maximal specific rate at enzyme saturation $^{conv}V_{max}/E_0$ (s^{-1})[a]	~ 2	~ 0.2	Article I, (Badino et al. 2017)
maximal specific rate at substrate saturation $^{inv}V_{max}/S_0$ ($\mu molg^{-1}s^{-1}$)[a]	~ 0.005	~ 0.03	
Attack site density Γ_{attack} ($\mu mol\ g^{1}$)[a]	~ 0.003	~ 0.15	
experimental processivity[a]	n.d.	~ 20	(Cruys-Bagger et al. 2013), (Taylor et al. 2013), (Kurašin and Väljamäe 2011)
inhibition by cellobiose	low	high	(Murphy et al. 2013), (Gruno et al. 2004), Article I
endo-attack propensity	high	low	(Ståhlberg et al. 1993), Article I
substrate coverage	low	high	Article I, (Pellegrini et al. 2014)

[a]using Avicel as substrate. n.d.: not determined

A number of works proposed a relationship between these kinetic characteristics and the loop regions (reviewed in (Payne et al. 2015)). For example, it has been argued that the high product inhibition of *Tr*Cel7A is due to stronger interactions in the product site (Atreya et al. 2016) and that high processivity and binding of *Tr*Cel7A is related to the presence of the long tunnel loops closing together and impeding dissociation from the cellulose strand (Knott et al. 2014). Computational studies done by Silveira *et al.* (Silveira et al. 2014) linked the fluctuations of the loops to the ability of *Tr*Cel7A to perform endo-

initiations (Kurašin and Väljamäe 2011). Still, the role of each individual loops in *Tr*Cel7A absent in *Tr*Cel7B has not been not object of a systematic study.

We investigated the individual role of the loops B2 B3 B4 and A4 absent in *Tr*Cel7B by creating deletion variants in *Tr*Cel7A where each of these loop was truncated in order to resemble the structure of *Tr*Cel7B (Fig 3.4 and Article I). The deletions were designed in a way that the length of each truncation was extensive enough to allow some measurable activity difference, but not too extensive to avoid altering the folding and stability of the enzyme. This evaluation was done *in silico* by creating *Tr*Cel7A homology models with different deletions using publicly available servers such as Phyre (Kelley et al. 2015) and Swissmodel (Waterhouse et al. 2018). Based on such analyses and the fact that loop B2 is quite long (15 residues are missing in *Tr*Cel7B) we designed and expressed three deletion variants in this region: ΔB2-1, ΔB2-2 and ΔB2-3, where ΔB2-3 represents the variant with the shortest deletion (6 residues). The other deletion variants were ΔB3, ΔB4 and ΔA4, with nine, four and five residue deletions, respectively (Fig. 3.4). The deletion variant in loop B3 was inspired by an insightful work by Ossowski *et al.*, which indicated that loop B3, close to the catalytic residues, has a role in facilitating processivity of *Tr*Cel7A on cellulose (von Ossowski et al. 2003). In this latter study, a deletion of such loop in a variant of *Tr*Cel7A resulted in a decrease in activity on crystalline cellulose, product inhibition and processivity. The variant also showed increased activity on amorphous cellulose. ΔB3 has a very similar deletion and was expected to show comparable kinetics to the one previously reported.

We also designed four variants of *Tr*Cel7B where the missing loops were inserted, named iA4, iA4B3, iB4 and iB1B2A3. These variants were produced and characterized along with the *Tr*Cel7A variants and wild types. The iA4 and iB4 variants were designed in such a way that only the A4 and B4 loop, respectively, would be introduced. The former variant included an extensive insertion of 12 residues and a modification of 4 residues in *Tr*Cel7B, while the latter was a modification of one residue and an insertion of 5 residues, meant to re-introduce one α-helix missing in *Tr*Cel7B. The variant iA4B3 was designed to introduce simultaneously the A4 and B3 loop in *Tr*Cel7B, while iB1B2A3 was planned to reconstitute the B2 loop. However, in order to do so, the B1 and A3 loops were also modified to aid stabilization of the long B2 loop insertion. Unfortunately, all the *Tr*Cel7B variants showed a significant decrease in both melting temperature and in all the activity and kinetic parameters investigated, clearly indicating that the gain of function attempt was unsuccessful. Therefore, these variants will not be described further.

```
TrCel7A   177 PRDLKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEA 218
ΔB2-1     177 PRDLKFINGQANVEG-------------HGSCCSEMDIWEA 204
ΔB2-2     177 PRDLKFINGQANVEGW------------HGSCCSEMDIWEA 205
ΔB2-3     177 PRDLKFINGQANVEGWEPS------GIGGHGSCCSEMDIWEA 212
TrCel7B   176 PVQTWRNGTLNTSHQ--------------GFCCNEMDILEG 203

TrCel7A   226 TPHPCTTVGQEICEGDGCGGTYSDNRYGGTCDPDGCDWNPYR 267
ΔB3       226 TPHPCTTVGQEICEGDGCG--------GTCDPDGCDWNPYR 258
TrCel7B   210 TPHSCTATA--------------------CDSAGCGFNPYG 230

TrCel7A   319 GSYSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGM 360
ΔB4       319 GSYSGNELNDDYCTAE----GGSSFSDKGGLTQFKKATSGGM 356
TrCel7B   288 DTISSCPSASAYG---------------GLATMGKALSSGM 313

TrCel7A   368 DDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVES 409
ΔA4       368 DDYYANMLWLDSTYPT-----TPGAVRGSCSTSSGVPAQVES 404
TrCel7B   321 NDNSQYMNWLDSG-----------NAGPCSSTEGNPSNILA 350
```

Figure. 3.4. *Amino acid sequence alignment of the different loop regions of TrCel7A (Uniprot: P62694), TrCel7A loop deletion (Δ) variants and TrCel7B (Uniprot: P07981). The deletion mutants lack the residues W192–G205 for ΔB2-1, E193–G205 for ΔB2-2, S196–T201 for ΔB2-3, G245–G253 for ΔB3, E335–F338 forΔB4 and N384–S388 for ΔA4. The dashed lines indicate the regions missing compared with TrCel7A. Loop nomenclature and colour coding are the same as Fig. 3.3.*

Table 3.1. *Results from basic characterization of TrCel7A, TrCel7B and the six loop truncation variants. The thermal stability, T_m, was determined as the transition mid-point in DSC (Conditions: 0.5 mg/mL enzyme, 20-95 °C, scan rate 3.3 °C), and the three kinetic parameters k_{cat}, K_M and the cellobiose inhibition constant K_i, were derived from kinetic experiments using pNP-Lac as substrate (Conditions: 0.16-5 mM pNP-Lac, 0.5 µM enzyme, 25 °C, 30 min. K_i was estimated from similar experiments in presence of 100 and 200 µM cellobiose or from 2 and 4 mM cellobiose for TrCel7B).*

Enzyme	T_m (°C)	k_{cat} (s^{-1})	K_M (mM)	K_i (cellobiose) (µM)	Type of inhibition
*Tr*Cel7A	66.9	0.089 ± 0.001	0.78 ± 0.03	21 ± 0.8	Competitive
ΔA4	57.7	n.d	n.d	n.d	n.d.
ΔB2-1	61.9	0.055 ± 0.001	0.89 ± 0.07	32 ± 2.3	Competitive
ΔB2-2	62.9	0.108 ± 0.002	0.80 ± 0.04	24 ± 1.0	Competitive
ΔB2-3	65.3	0.079 ± 0.001	0.76 ± 0.03	22 ± 0.6	Competitive
ΔB3	65.9	0.283 ± 0.007	3.40 ± 0.14	133 ± 8	Mixed*
ΔB4	59.0	0.080 ± 0.001	0.70 ± 0.03	46 ± 2	Competitive
*Tr*Cel7B	65.8	4.8 ± 0.7	11.12 ± 2.14	5300 ± 500	Competitive

*the ratio of the two inhibition constants in the mixed mechanism (α) was 5.0 ± 2.0.
n.d: not determined.

We initially tested the activity of the *Tr*Cel7A variants using the soluble substrate analog *p*-Nitrophenyl-β-D-lactopyranoside (*p*NP-Lac) and compared it with *Tr*Cel7A and *Tr*Cel7B. We measured the kinetic parameters on *p*NP-Lac and the cellobiose inhibition constant (Table 3.1), and also estimated the stability of the variants by measuring the thermal unfolding transition midpoint (T_m) using differential scanning calorimetry (DSC). This basic characterization (Table 3.1) indicated that all the variants, except ΔA4, had a low to moderate decrease in T_m and showed measurable activity on *p*NP-Lac, suggesting that the enzymes were correctly folded. However, ΔA4 showed a significant loss in T_m and no detectable activity on *p*NP-Lac, and it was purified despite low expression yields. The A4 loop in *Tr*Cel7A contains a glycosylation site at position N384 (Divne et al. 1994), which plays an important stability role (Amore et al. 2017). We thus hypothesize that the deletion of this site in ΔA4 created strong structural rearrangements which resulted in an inactive protein. The ΔA4 variant was therefore not further studied.

As for the kinetic parameters, all variants had similar k_{cat} on *p*NP-Lac as *Tr*Cel7A except ΔB3, which increased 3-fold in activity and 4-fold in K_M, and showed reduced cellobiose inhibition (Table 3.1). These results were in agreement with what reported by Ossowski *et al.*, and suggested that deletion in B3 loop weakens the interactions in the product site (von Ossowski et al. 2003).

3.4 - Kinetic properties of *Tr*Cel7A, *Tr*Cel7B and loop deletion variants

After ensuring activity and stability of both wild type and variants, we proceeded to kinetically characterize the enzymes on Avicel by measuring the ^{conv}MM and ^{inv}MM kinetic parameters and by measuring binding as described in chapter II. The results are shown in Fig. 3.5 and Table 3.2.

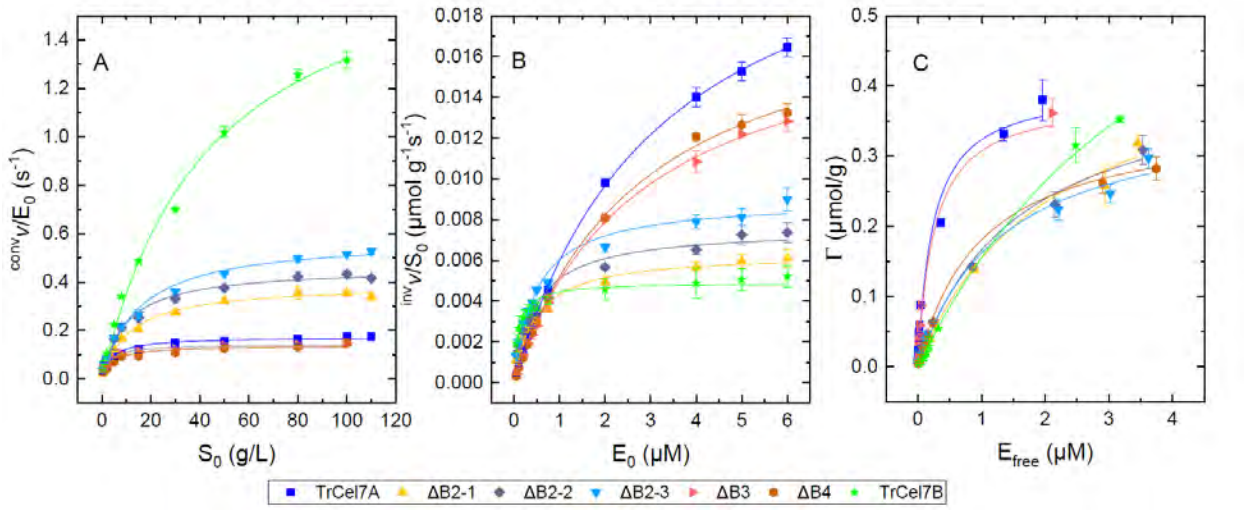


Figure 3.5. *Conventional Michaelis-Menten, Inverse Michaelis-Menten plots and binding isotherms for TrCel7A (blue squares), TrCel7B (green stars) and deletion variants ΔB2-1 (yellow up-pointing triangles), ΔB2-2 (purple rhombi), ΔB2-3 (light blue down-pointing triangles), ΔB3 (pink right-pointing triangles) and ΔB4 (brown circles). (A) Conventional Michaelis-Menten. Conditions: 0.5-110 g/L Avicel, 50 nM enzyme, pH 5, 60 min, 25 °C (B) Inverse Michaelis-Menten. Conditions: 8 g/L Avicel, 0.05-6 μM enzyme, pH 5, 60 min, 25 °C. (C) Binding isotherms. Conditions are the same as in B. Symbols represents experimental data with standard deviation, lines are the best fit for the different equations. Figure adapted from Article I.*

Table 3.2. *Steady-state kinetic, binding parameters and attack site density of TrCel7A, TrCel7B and the variants ΔB2-1, ΔB2-2, ΔB2-3, ΔB3 and ΔB4 using Avicel as substrate. Parameters derived from experiments shown in Fig 3.5.*

	^{conv}MM			^{inv}MM		Kinetic substrate accessibility	Adsorption isotherms	
Enzyme	$^{conv}V_{max}/E_0$	$^{conv}K_M$	η	$^{inv}V_{max}/S_0$	$^{inv}K_M$	Γ_{attack}	Γ_{max}	K_d
	(s^{-1}) ×10^{-3}	(g/L)	($Lg^{-1}s^{-1}$) ×10^{-3}	(μmol $g^{-1}s^{-1}$) ×10^{-3}	(μM)	(μmol/g)	(μmol/g)	(μM)
*Tr*Cel7A	172 ± 6	4 ± 0.7	43 ± 8	25 ± 0.4	3.2 ± 0.10	0.147	0.40 ± 0.03	0.24 ± 0.06
ΔB2-1	391 ± 8	12 ± 1.0	33 ± 3	6 ± 0.2	0.5 ± 0.06	0.016	0.48 ± 0.04	2.13 ± 0.41
ΔB2-2	458 ± 12	10 ± 1.0	46 ± 5	8 ± 0.3	0.5 ± 0.07	0.016	0.43 ± 0.04	1.56 ± 0.33
ΔB2-3	583 ± 19	15 ± 1.8	39 ± 5	9 ± 0.3	0.5 ± 0.06	0.015	0.38 ± 0.03	1.32 ± 0.29
ΔB3	142 ± 5	3 ± 0.5	47 ± 8	19 ± 0.3	2.7 ± 0.11	0.131	0.39 ± 0.03	0.26 ± 0.07
ΔB4	138 ± 7	4 ± 1.0	35 ± 9	20 ± 0.5	2.8 ± 0.15	0.144	0.36 ± 0.01	0.99 ± 0.09
*Tr*Cel7B	1908 ± 97	44 ± 5.1	43 ± 5	5 ± 0.1	0.1 ± 0.02	0.003	0.87 ± 0.10	4.60 ± 0.78

Deletions in loop B3 and B4 did not significantly alter the kinetics of *Tr*Cel7A, although both conventional and inverse maximal rates were slightly decreased. Deletions in B2 loop on the other hand, significantly increased $^{conv}V_{max}$, $^{conv}K_M$ and decreased $^{inv}V_{max}$, making the variants more similar to *Tr*Cel7B (Table 3.2). The variant with the shortest deletion, ΔB2-3, had the highest increase in $^{conv}V_{max}$, of almost 3-fold compared to the wild type. Increasing the extent of the deletion was not beneficial and the $^{conv}V_{max}$ decreased with ΔB2-2 and ΔB2-1. Deletion in B2 loop clearly shifted the values towards *Tr*Cel7B also in the ^{inv}MM parameters. In particular, the $^{inv}V_{max}$ in ΔB2-3 decreased almost three-fold compared to the wild type, and increasing the size of the deletion determined a decrease in $^{inv}V_{max}$, with ΔB2-1 approaching a value close to *Tr*Cel7B. As for the attack site density, the deletion in loop B2 consistently lowered Γ_{attack} by an order of magnitude (Table 3.2), indicating that the presence of the B2 loop is important for the ability of *Tr*Cel7A to locate and make productive complex on cellulose. It was interesting to see that this change in kinetic behaviour for the ΔB2 variants was not reflected in strong differences in the binding isotherms (Table 3.2 and Fig. 3.5). The overall binding did not change dramatically between the variants investigated, and Γ_{max} remained in a similar range of about 0.4 μmol/g, although the K_d value of the ΔB2 variants consistently shifted to values closer to *Tr*Cel7B.

3.5 - Activity on different substrates for *Tr*Cel7A, *Tr*Cel7B and deletion variants

Since GH7 CBHs and EGs also differ in their preference towards substrates with different crystallinity, we investigated the activity of the loop deletions on RAC and BMCC, which are mainly amorphous and crystalline cellulose, respectively (see section 1.3 and (Zhang and Lynd 2004)). Again, we observed that the ΔB2 variants display a more EG-like behaviour with a 4 to 5-fold increase in the activity on RAC and a decreased in activity on BMCC (Fig. 3.6).

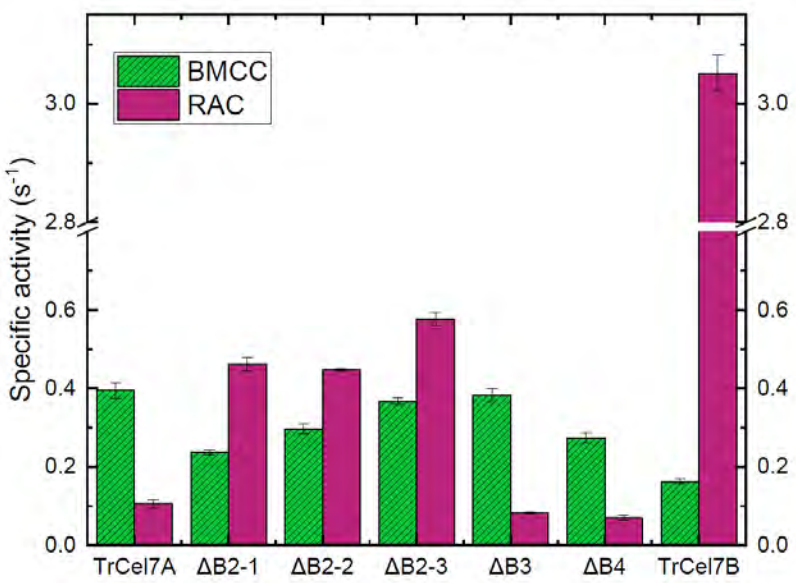


***Figure 3.6**. Specific activity on two different cellulosic substrates, regenerated amorphous cellulose (RAC, purple) and bacterial microcrystalline cellulose (BMCC, green). Conditions: 4 g/L RAC, 50 nM enzyme or 4 g/L BMCC, 100 nM enzyme, 25 °C, 60 min. Error bars are standard deviations from triplicates. Adapted from Article I.*

There was a proportionality between the activity on the different substrates and size of the loop B2 deletion: the smaller is the section of the B2 loop removed, the more the activity increased towards either RAC or BMCC, which is also reflected in the Avicel data (Table 3.2).

Overall, the results allowed us to conclude that the B2 loop is a key determinant for the difference in CBH- or EG-like activity. Among the variants investigated, ΔB2-3 was the one with the most pronounced intermediate characteristics between *Tr*Cel7A and *Tr*Cel7B. Conversely, truncation of B3 and B4 loops essentially maintained the characteristic CBH kinetics and promiscuity on the cellulose surface, indicating a different

functional role than the B2 loop.

A structural rationalization on the function of the B2 loop can be found by analysing the structure of *Tr*Cel7A in Michaelis complex with cellononaose (PDB 4C4C). The B2 loop covers the pyranose binding subsites -3 and -4 in *Tr*Cel7A (Bodenheimer et al. 2016)(Fig. 3.7) and plays a central role in interacting with the neighbouring loops A3 and B3 to create an tightly closed tunnel. Specifically, Y370 of the A3 loop stabilizes the B2 loop via interaction with N198 and N197 (gray structure in Fig. 3.7). The removal of these residues in the ΔB2 variants probably increases the overall flexibility of *Tr*Cel7A which means a more EG-like open cleft shape. Notably, N198, N197 and S196 are also naturally absent in *Phanerochaete chrysosporium* Cel7D (*Pc*Cel7D, yellow structure in Fig. 3.7), which is an atypical CBH because of its shorter A1, B2 and B3 loops, as shown in the crystal structure (PDB 1GPI, (Muñoz et al. 2001) and for its more pronounced EG-like kinetics compared to *Tr*Cel7A such as increased activity on amorphous cellulose (von Ossowski et al. 2003) and decreased processivity (Kurašin and Väljamäe 2011).

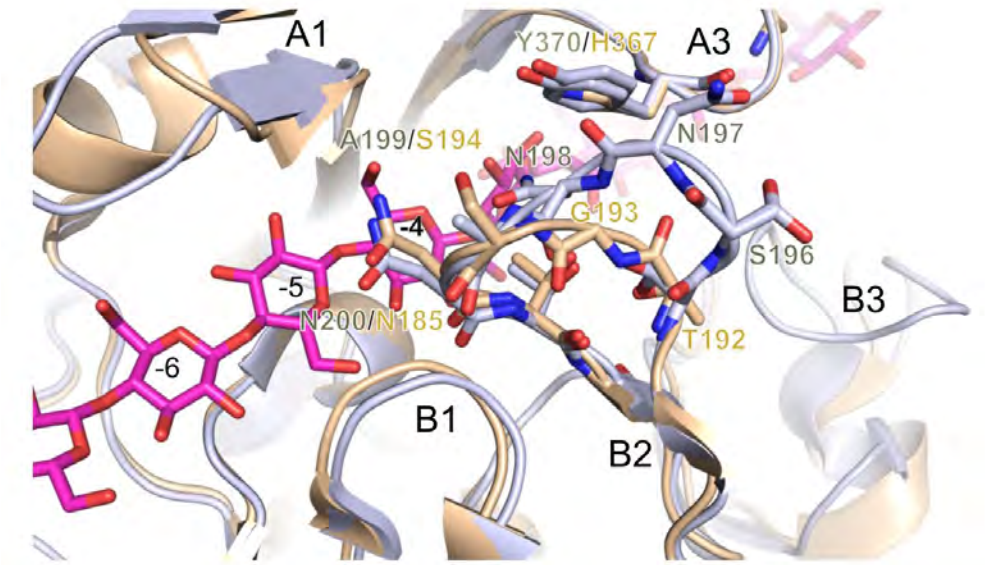


Figure 3.7. *The B2 loop of TrCel7A (gray, PDB 4C4C) and PcCel7D (yellow, PDB 1GPI) superimposed onto TrCel7A. The residue N200 of TrCel7A or N185 of PcCel7D are close to -5, -4 subsites for cellononaose (magenta sticks). The tip of the B2 loop in TrCel7A contains the residues A199, N198, N197 and S196, while in PcCel7D the loop shorter and contains S194, G193 and T192. The residues Y370 of the A3 loop of TrCel7A is substituted to H367 in PcCel7D. Figure created in Pymol (The PyMOL Molecular Graphics System, Version 2.3.2, Schrödinger, LLC).*

3.6 - Different enzymes, same specificity constant

For all the variants and the wild types investigated, we observed a strong linear relationship between ${}^{conv}V_{max}$ and ${}^{conv}K_M$ (Fig. 3.8) which resulted in almost identical values of the specificity constant η (Table 3.2).

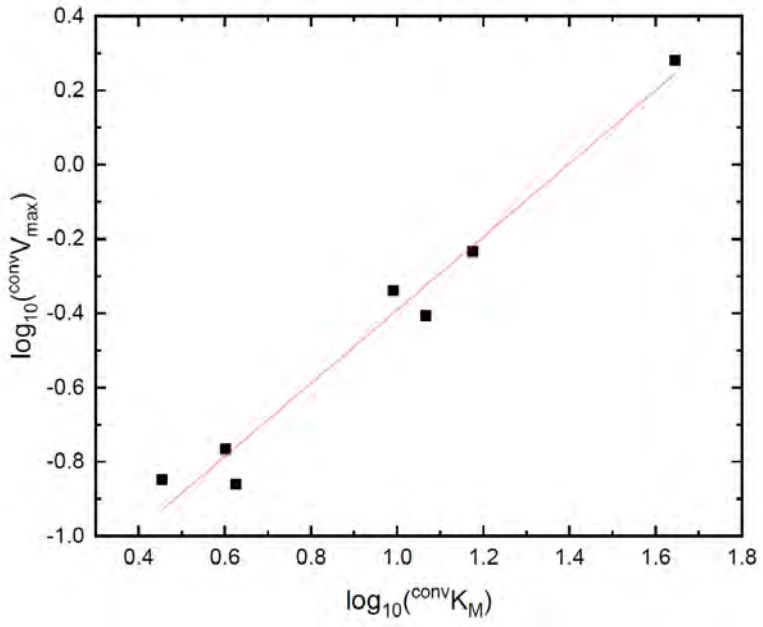


Figure 3.8. *Correlation between ${}^{conv}V_{max}/E_0$ and ${}^{conv}K_M$, data from Table 3.2 log-transformed. Pearson's r=0.89, n=7. The red line represents the best linear correlation from a least squares regression analysis of the data points. Figure created using OriginPro 2019 (OriginLab Corp., Northampton, MA, USA). Statistical analysis performed in JMP Version 14 (SAS Institute Inc., Cary, NC, 1989-2019).*

At least for the wild types, the similar specificity constants could have an evolutionary justification. It has been argued that enzymatic natural selection acts upon the specificity constant rather than the turnover rate (Bar-Even et al. 2011). As such, *Tr*Cel7A and *Tr*Cel7B might have evolved in a way that preserved similar specificity constants to ensure high efficiency, while in the meantime develop complementary kinetic properties which ensures good synergy between the enzymes (see also section 5.1 for synergy between *Tr*Cel7A and *Tr*Cel7B). Considering both wild type enzymes and variants, they follow a linear relationship where ${}^{conv}V_{max}$ and ${}^{conv}K_M$ changes concomitantly (Fig. 3.8). In other words, enzymes with low affinity such as *Tr*Cel7B have also high turnover number and *vice versa*, as in the case of *Tr*Cel7A. Intermediate characteristics are seen for the ΔB2 variants, with medium affinity and turnover. This suggest a limited functional plasticity

in GH7, which will be further rationalized in section 4.11 for the EGs.
Moreover, the linear scaling between K_M and V_{max} in Fig. 3.8 could have a more complex energetic interpretation, since the catalytic efficiency η can be related to the free energy difference between the free state (unbound enzyme) and the transition state governing the overall rate (Warshel 1998, Christensen et al. 2019). Since η remains very similar between the enzymes investigated, this free energy difference remains also quite similar, despite the affinity and turnover parameters vary. We will refrain from further expanding these thermodynamic interpretations, as more systematic works needs to be done to fully elucidate this aspect.

3.7 - Is there synergy between ΔB2-3 and *Tr*Cel7A?

As mentioned before, the variant ΔB2-3 showed the most intermediate kinetic properties between *Tr*Cel7A and *Tr*Cel7B, in particular a decrease in site promiscuity (Table 3.2), which could indicate a potential new synergy partner of *Tr*Cel7A. We thus investigated synergy mixtures of *Tr*Cel7A with ΔB2-3 under ^{inv}MM conditions, in a similar way as done in Article III, and compared the results with *Tr*Cel7A mixed with *Tr*Cel7B (Fig. 3.9). Materials and methods for these experiments are the same as described in Article III. The mix *Tr*Cel7A and *Tr*Cel7B showed high degree of synergy at all molar ratios investigated (see also section 5.1). Conversely, *Tr*Cel7A and ΔB2-3 showed no synergy. On the contrary, TrCel7A and ΔB2-3 showed lower activity then expected from the monocomponent curves, which indicate that the enzymes inhibit each another on the surface (we will return to this in section 5.3). This suggests that despite an increase in *endo*-like behaviour, ΔB2-3 still shares similar hydrolysable site preferences as *Tr*Cel7A.

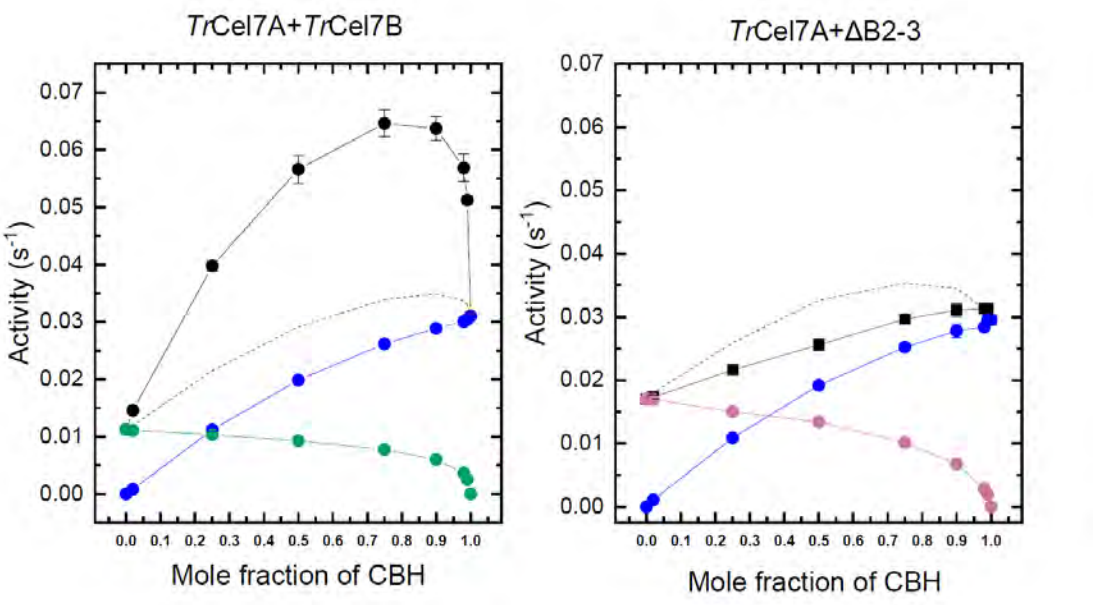


Figure 3.9. *Synergy between TrCel7A and TrCel7B (left panel), or TrCel7A with ΔB2-3 (right panel). Black squares correspond to the total activity in the synergy mixtures at different mole fraction of CBH. The monocomponent activity is indicated with blue circles for TrCel7A, green circles for TrCel7B and pink circles for ΔB2-3. Dashed lines indicate the sum of the monocomponent activities used in each mixture. Conditions: 25 °C, 10 g/L Avicel, total enzyme concentration 4 μM. For each panel, the x-axis represents the fraction of CBH used in the mixture, calculated as [CBH]/([CBH+EG]). Error bars indicate standard deviation from triplicate measurements. Materials and methods are the same as described in Article III. Figure created using OriginPro 2019 (OriginLab Corp., Northampton, MA, USA).*

3.8 - Product profile of *Tr*Cel7A, *Tr*Cel7B and loop deletion variants

We previously mentioned that processivity varies between CBHs and EGs, thus loop deletions in *Tr*Cel7A might have an effect also in this aspect. Indeed, Ossowski *et al.* reported a decrease in processivity when loop B3 was deleted in *Tr*Cel7A (von Ossowski et al. 2003). To investigate this, we analysed the wild types and deletion variants of Article I in terms of product profile on Avicel and measured the amount of soluble sugars glucose (G1), cellobiose (G2) and cellotriose (G3), produced. We then estimated processivity with two different approaches described below (see Fig. 3.10). Materials and methods of this part can be found in Appendix I. Fig. 3.10 shows the relative amount of G1, G2 and G3 produced by the different enzymes after 1 hour incubation at substrate-saturating conditions.

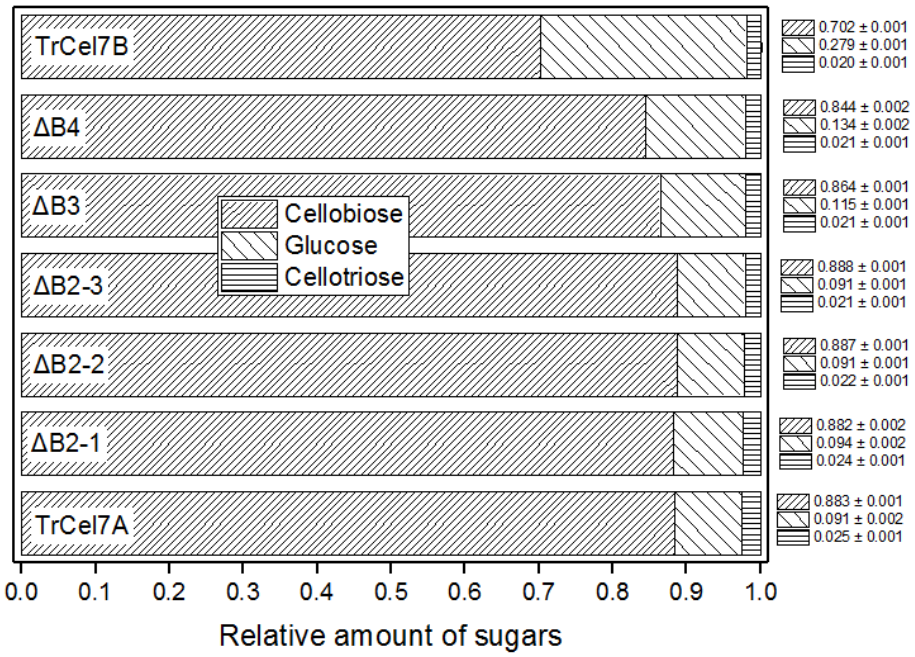


Figure 3.10. *Product profile of the enzymes studied in article I. Relative amount of glucose, cellobiose and cellotriose is shown. Conditions: 1h hydrolysis, 50 nM enzyme, 80 g/L Avicel. Figure created using OriginPro 2019 (OriginLab Corp., Northampton, MA, USA).*

The variants ΔB3 and ΔB4 showed an increase in production of G1 compared to *Tr*Cel7A, while all deletions in B2 loop seemed unaffected in the product profile, since the relative amount of sugars remained similar to *Tr*Cel7A. Based on this data, apparent processivity was calculated using two different methods: P=(G1+G2+G3)/G1 (Kari et al. 2017) or P=(G2/G3+G1) (von Ossowski et al. 2003) (Table 3.3)

Table 3.3. *Quantification of glucose (G1), cellobiose (G2) and cellotriose (G3) measured by HPAEC-PAD. Standard deviation from triplicate measurements is indicated. Estimated processivity values for TrCel7A, TrCel7B and the deletion variants of Article I using two different approaches is also shown.*

Enzyme	Sugars produced			Processivity	
	G1 (µM)	G2 (µM)	G3 (µM)	P=Gtot/G1*	P=(G2/G3+G1)**
*Tr*Cel7A	3.16±0.02	30.6±0.4	0.9±0.03	11	8
ΔB2-1	9.4±0.4	88±3	2.4±0.09	11	7
ΔB2-2	9.8±0.3	96±4	2.3±0.09	11	8
ΔB2-3	11.3±0.3	110±4	2.6±0.11	11	8
ΔB3	3.2±0.2	24±1.4	0.6±0.04	9	6
ΔB4	3.30±0.09	20.9±0.90	0.5±0.02	7	5
*Tr*Cel7B	85.0±0.9	214±1.6	6±0.4	4	2

*calculated as reported in (Kari et al. 2017).

**calculated as reported in (von Ossowski et al. 2003).

The processivity values are somewhat lower than previously reported for *Tr*Cel7A on Avicel (Kari et al. 2017), but it is interesting to analyse the results in a comparative way. ΔB3 showed a decrease in processivity (Table 3.3), in agreement with (von Ossowski et al. 2003). A structural explanation for this can be found in the fact that ΔB3 lacks T246, hydrogen bonding with the ligand in subsite +1 and Y247 in subsite -2 (Hobdey et al. 2016) and has less interactions with the neighbouring loops, which together hold together the closed tunnel shape necessary for the processive motion (Muñoz et al. 2001). Lower processivity is also observed for ΔB4 (Table 3.3). This effect could be related to the B4 loop close proximity to the product site. The B4 loop in *Tr*Cel7A appears to have fewer product site interactions compared to other CBHs because of the absence of an aspartate residue (for example, D347 in *Hir*Cel7A) which has been suggested being involved in aiding product expulsion (Momeni et al. 2013). Hence, a further deletion in this area could result in even fewer interactions with the product cellobiose, which might not diffuse away from the product site fast enough, thus reducing processivity (Hobdey et al. 2016). As for the ΔB2 variants, they showed almost identical processivity and product profile compared to *Tr*Cel7A (Table 3.3). The reason for this is not fully clear, but one could argue that the B2 loop plays a less important role in processivity compared to the loops forming the product site. The B2 loop could be more involved in binding to cellulose and transferring a cellulose strand from the crystal lattice into the active site, along with other entrance loops such as A1 and B1, while other loop interactions such as A3-B3 are more directly involved in processivity, as shown by a number of MD studies (Lin et al. 2013, Knott et al. 2014, Vermaas et al. 2019).
Finally, it is interesting to observe that *Tr*Cel7B shows some degree of processivity, as reported before (Murphy et al. 2012), also on EGs from other GH families (Kurašin and Väljamäe 2011), thus strengthening the idea that processivity is not limited to CBHs.

3.9 - The A1 loop of GH7 cellulases

The A1 loop is a highly variable region in GH7 enzymes, both in terms of amino acid composition and length, as shown by the multiple sequence alignment in Fig. 3.11A. The low sequence conservation suggest that this region is an important area for the evolution of different enzyme kinetic properties (Nestl et al. 2014). A structural alignment of all the GH7 crystal structures available from different organisms is shown in Fig 3.11B and gives an indication that the A1 loop sequence variability corresponds to different shapes and

lengths in the enzyme catalytic domain. Notably, in the structure of *Tr*Cel7A complexed with cellononaose (PDB 4C4C), Knott *et al.* observed that the A1 loop was in direct contact with the substrate, with Q101 interacting with the 6-hydroxyl of the glucose unit in subsite -6 (Fig. 3.12A) (Knott et al. 2013). This interaction might be conserved in other structures such as *Trichoderma atroviride* Cel7A, where Q is replaced by N, but lost in others like the CBHs from *Dictyostelium* species, where Q is substituted by G (Hobdey et al. 2016) (Borisova et al. 2018). Interestingly, some CBH structures contain a tyrosine at the tip of the A1 loop, as in the case of the CBHs from *Heterobasidion irregulare* (*Hir*Cel7A, Fig 3.12B), *Dapnhia pulex, Limnoria quadripunctata, Melanocarpus albomyces,* and *Myceliophthora thermophila,* where the tyrosine has been proposed interacting with the substrate at subsite -7 (Kern et al. 2013, Momeni et al. 2013)

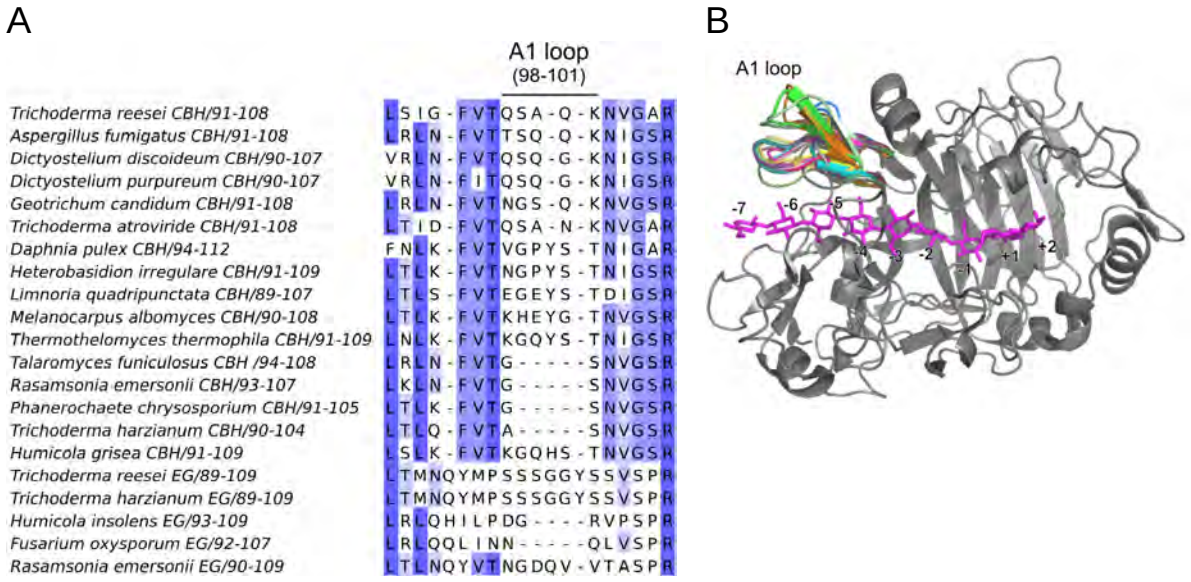


Figure 3.11*. The A1 loop of GH7 enzymes. (A) Portion of a multiple sequence alignment for the 21 known GH7 structures. The A1 loop sequence region is indicated on top of the alignment with residue numbering for TrCel7A. The sequence alignment was performed in Jalview (Waterhouse et al. 2009). A darker shade of blue indicates a higher percentage of sequence identity. (B) A1 loop superimposition of 21 GH7 structures (various colours) onto TrCel7A (gray, PDB 4C4C). Cellononaose is shown as magenta sticks, the subsite nomenclature is indicated. The structures overlap with root-mean-square differences (RMSD) values of ~0.3-0.8. The structural alignment was performed using Pymol (The PyMOL Molecular Graphics System, Version 2.3.2, Schrödinger, LLC).*

More differences in the A1 loop can be found in the presence or absence of an ion pair. In *Tr*Cel7A, K102 Fig. 3.12A at the C-terminal of the A1 loop is forming a salt bridge with E408 of the A2 loop (Bodenheimer and Meilleur 2016), and this interaction is possible for all the CBH structures containing a K at the C-terminal end of the A1 loop. Interestingly,

a number of structures show a short A1 loop, as in the case of *Talaromyces funiculosus/Penicillium funiculosum* (*Pf*Cel7A) and *Rasamsonia emersonii* (*Re*Cel7A) (Fig. 3.12C). Taylor and colleagues recently suggested that the differences in the A1 loop could explain the improved performance of *Pf*Cel7A on biomass compared to *Tr*Cel7A (Taylor et al. 2018). By molecular dynamics (MD) simulations, they proposed that the shorter A1 loop of *Pf*Cel7A results in the other tunnel enclosing loops to be more flexible and this could facilitate dissociation of stalled enzymes. A shorter A1 loop has also been connected to a wider entrance to the tunnel (Borisova et al. 2018) (Textor et al. 2013). Thus, this area appears to be an interesting region for further functional analysis in GH7.

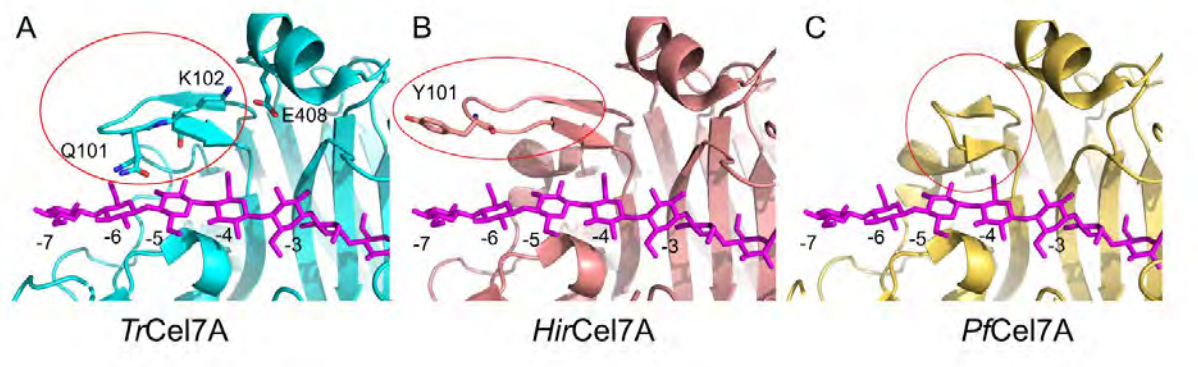


Figure 3.12. *The A1 loop in different GH7 enzymes. Cellononaose (magenta sticks) has been superimposed onto all the structures for clarity. (A) The A1 loop (red circle) in TrCel7A complexed with cellononaose (PDB 4C4C). The residue Q101 interacts with the cellulose chain, while K102 forms a salt bridge with E408 of the A2 loop. (B) The A1 loop in HirCel7A (PDB 2YG1) contains a tyrosine at the tip of the loop. (C) The short A1 loop of PfCel7A (PDB 4XEB). Figure created in Pymol (The PyMOL Molecular Graphics System, Version 2.3.2, Schrödinger, LLC).*

3.10 - A simple way to measure substrate accessibility in GH7

A number of studies (von Ossowski et al. 2003, Kurašin and Väljamäe 2011, Momeni et al. 2013, Sorensen et al. 2015, Taylor et al. 2018) including Article I appended to this thesis, suggest that in GH7 a more open tunnel/cleft shape is beneficial for high catalytic rates on cellulose, since a looser binding to cellulose can speed up the decomplexation rate. However, describing a cellulase in terms of being more "open" or "closed" is often based on structural comparisons between few enzymes and so far there has been no attempt to categorize the different GH7 structures based on their degree of "openness", which can be interpreted as solvent accessibility of the substrate upon binding. To overcome this, we propose a fast and quantitative way to measure the degree of openness between the GH7 enzymes, which is based on the PISA bioinformatic tool (www.ebi.ac.uk/pdbe/pisa/).

More details on the procedure can be found in Appendix II. Briefly, we superimposed the 21 available GH7 structures with the structure of *Tr*Cel7A complexed with cellononaose (PDB 4C4C) and measured an area (in Å^2) corresponding to how much of the total cellononaose is exposed to the solvent, hence called exposed surface area (ESA) of the substrate. This analysis was performed after an energy minimization procedure to ensure that the ligand was well accommodated in the different structures. The results are shown in Fig. 3.13.

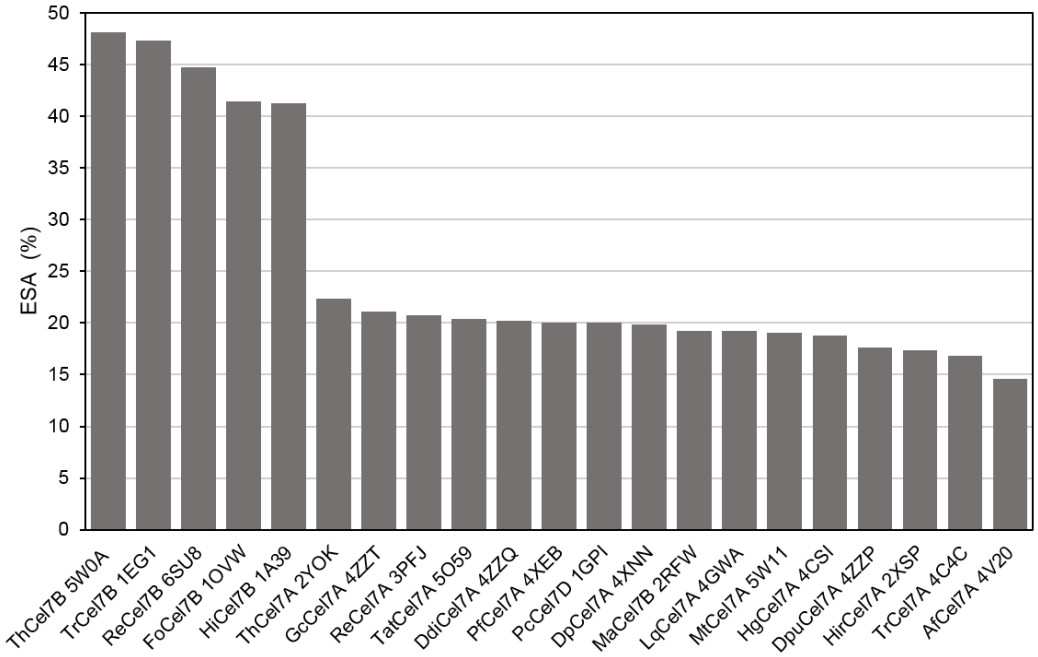


Figure 3.13. *Exposed surface area (ESA) of the ligand cellononaose for different GH7 enzymes. ESA is expressed as percentage of the total accessible surface area. Each enzyme is described by their short name followed by the PDB accession number used (see also Appendix II). Figure created in Microsoft Excel (2013).*

The structures were arranged by decreasing cellononaose ESA, and it was possible to observe a clear difference in terms of solvent exposure between the very open cleft of the EGs and a more closed tunnel of the CBHs. Interestingly, the scale is not continuous and there is a gap of about 300 Å^2 between the CBHs and EGs (Fig. 3.13 and Table A1, Appendix II), suggesting the catalytic domain the GH7 enzymes falls into a distinctive "tunnel" or "cleft" category, although finer differences can be observed within these groups.

From the scale of Fig. 3.13, it is possible to observe that among the CBHs with the most solvent-accessible substrate there is *Trichoderma harzianum* Cel7A (*Th*Cel7A) and

Rasamsonia emersonii Cel7A (*Re*Cel7A), both containing a very short A1 loop, as well as *Phaneochaete crysosporium* Cel7D (*Pc*Cel7D), which has been described being a very open CBH structure (von Ossowski et al. 2003). On the opposite end, the structure with the least exposed cellononaose appears to be *Aspergillus fumigatus* Cel7A (*Af*Cel7A) followed by *Tr*Cel7A (Fig. 3.13). Moreover, we note that *Heterobasidion irregulare* Cel7A (*Hir*Cel7A) has been suggested being more open than *Tr*Cel7A because of modifications in the A3-B3 loops (Momeni et al. 2013), which is also observed in Fig. 3.13. As for the EGs, *Trichoderma harzianum* (*Th*Cel7B) and *Tr*Cel7B are the most open structures, while *Humicola insolens Cel7B (HiCel7B) and Fusarium oxysporum* Cel7B (*Fo*Cel7B) are the most closed, while *Re*Cel7B is rather intermediate. As we will discuss in chapter IV, the more open architecture of the *Trichoderma* sp. EGs is mainly due to much shorter B3 and B4 loops compared to the other EG structures listed here.

Although this scale is not devoid of flaws, for example it does not take into account loop flexibility in solution and further structural changes occurring during catalysis, it offers a simple and objective overview of the "degree of openness" in GH7 structures and could be useful to find new correlations between rates, affinity and total surface interactions. Indeed, preliminary steady state kinetic analyses done on a number of enzymes illustrated in Fig. 3.13 show a positive correlation between either $^{conv}V_{max}/E_0$ or $^{conv}K_M$ and ESA (not shown), which could indicate that enzymes with low affinity can achieve higher turnover due to a more open structure. However, more extensive works (possibly including all 21 enzymes) are needed to fully explore these relationships.

Chapter IV – Rational design of cellulases for industrial applications

This chapter will focus on the structure and kinetic properties of thermostable cellulases for industrial purposes. A general introduction on the applicability of cellulases in industry was presented in section 1.5. Here we will highlight strategies and challenges related to improving cellulases' activity for high temperatures. One successful approach used in our research group is rational design, based on site directed mutagenesis and a detailed structural knowledge of the enzyme of interest. This approach is more difficult to implement in GH7 endoglucanases because of the paucity of structural and functional information available. In this chapter, a new crystal structure of a highly thermostable EG will be presented, along with kinetic and structural comparisons with homologous enzymes.

4.1 - The TEMPEN strategy to improve industrial cellulases

The studies performed during the course of this PhD work are part of the TEMPEN project (TEMPperature activation of ENzymes for biomass degradation), a collaboration between Roskilde University and Novozymes A/S, a major enzyme producing company. The focus of the project is to design new enzymes capable of effectively degrading lignocellulosic biomass at industrially-relevant temperatures. To achieve this, we apply a knowledge-based iterative workflow, schematized in Fig. 4.1. The workflow usually starts with studying a naturally occurring enzyme (phase 1), for example the thermophilic cellobiohydrolase *Re*Cel7A of Article III. The enzyme is tested at different conditions and we model the obtained results in order to identify the rate-limiting step of the process, detrimental in industrial conditions (phase 2) and formulate a hypothesis. In the case of *Re*Cel7A, the bottleneck identified was the decrease in affinity for the substrate at increasing temperatures, which we hypothesize can be solved by structural modifications (Article III). We initially test the hypothesis by creating *in silico* variants and we assess with different bioinformatic methods which modification on the structure could alleviate the problem (phase 3). In the *Re*Cel7A case, the introduction of a CBM domain could increase the adsorption at high temperatures. The variant(s) is then cloned, expressed, purified (phase 4) and tested to confirm/validate the hypothesis. Then, the workflow of

enzyme improvement continues for many iterations.

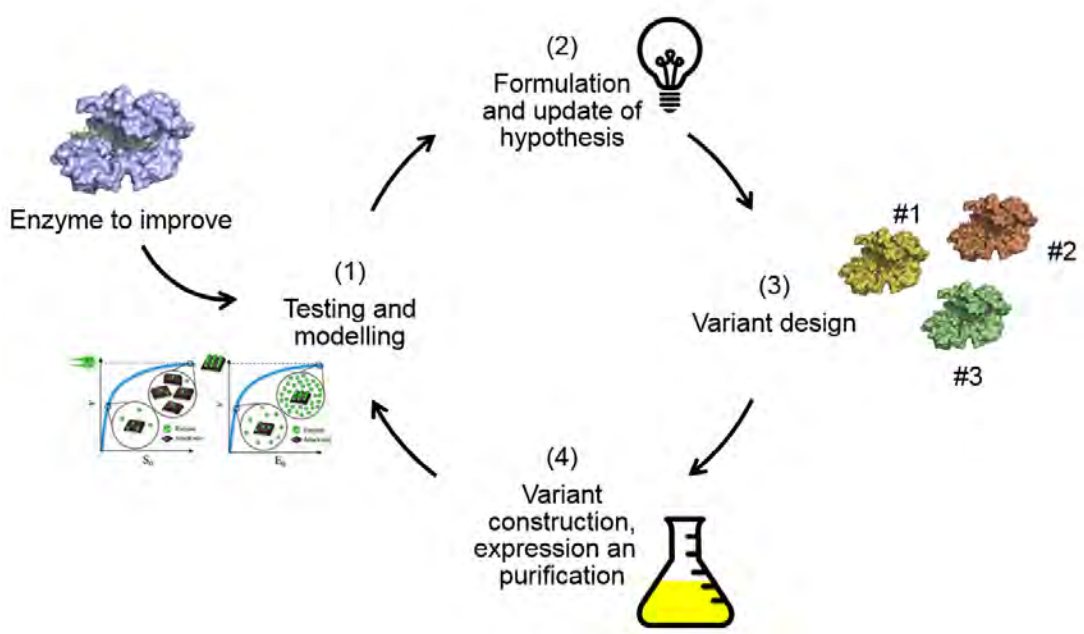


***Figure 4.1**. The TEMPEN workflow. See main text for more details.*

This workflow has been applied to modulate the activity of a number cellulases of different classes and families. For instance, >60 EG variants have been designed, produced and characterized using this iterative process during the course of this PhD. The obtained results allowed an analysis of the functional plasticity of the EGs at the end of this chapter (section 4.11).

4.2 - Protein engineering strategies for industrial cellulases

Enzymatic saccharification needs to be carried out at high substrate loads (20-40%) in order to render the process economically favourable. This both creates an increase in volumetric productivity and sugar concentration which reduces subsequent concentration costs (Modenbach et al. 2013). Due to the stability of cellulose (section 1.2), cellulases are relatively slow at degrading their substrate and a usual biomass saccharification process requires 2-5 days to achieve high (>80%) glucose yields (Xiao et al. 2004). As many industrial processes, high temperatures are preferred (for example, 50-55 °C for SHF) in order to accelerate the hydrolytic rates and decrease contaminations, but also to facilitate the recovery of the volatile product (ethanol) and lower costs of cooling after the pre-treatment (Verardi et al. 2012). That is why thermostable cellulases are the preferred

enzymes in the biorefinery sector. Such enzymes can be either isolated from natural sources, (*i.e.* from a thermophilic fungus) or can be variants produced via protein engineering (Kuhad et al. 2011). The latter is the preferred approach when a cellulase of interest is already known and shows high specific rates but not at the desired temperature. Different protein engineering strategies are available to modulate enzyme activity. High or medium throughput methods such as directed evolution (Eijsink et al. 2005) or SCHEMA recombination (Lutz 2010), have both been proven successful to improve the stability of GH7 CBHs (Goedegebuur et al. 2017) and GH6 cellulases (A (Heinzelman et al. 2009). However, some limitations arise from the expression host used to create the mutant libraries, which might not be optimal to obtain correct post translational modifications, or could give poor yields or insoluble expression of the enzyme of interest. Moreover, it is more difficult to understand the cause-effect relationship of a particular mutation. An alternative, low throughput approach extensively used also in our research group is rational design, which employs site-specific mutagenesis (Mohanram et al. 2013). The main advantages of this approach lies in the low number of mutants produced and that more mutations can be "rationally" combined together to obtain synergistic beneficial effects (Mitrovic et al. 2014). It requires, however, an extensive knowledge of the enzyme of interest to improve because the mutations applied are site-specific and based on structure and sequence considerations.

Few practical examples of rational design strategies for cellulases are listed below.

- The introduction of disulphide bonds by introducing new cysteine pairs increased the unfolding temperature in *Melanocarpus albomyces* cellobiohydrolase Cel7B (Voutilainen et al. 2009) and the endoglucanase *Tr*Cel7B (Zhang et al. 2014).
- A better packing of the hydrophobic core and the removal of surface cavities improved the stability of the endoglucanase Cel7B from *Hypocrea pseudokoningii* (Mitrovic et al. 2014).
- The introduction of prolines, with high conformational rigidity, reduce the entropy of unfolding (Li et al. 2005). Prolines have been introduced along with other mutations to create a thermostable variant of *Tr*Cel7A, which showed a striking increase in T_m of 10 °C compared to the wild type (Goedegebuur et al. 2017).

These strategies, along with many others not listed here, have been very successful to improve thermostability of mesophilic cellulases. However, high thermostability does not necessarily correlate with an increase activity. This aspect will be discussed in the next section.

4.3 - Thermoactivation of a GH7 cellobiohydrolase

The relationship between enzyme activity and temperature is the result of two independent processes: thermoactivation and enzyme inactivation. The former occurs when the enzyme is stable and the reaction rate increases exponentially with temperature according the Arrhenius equation. This can be described as:

$$\ln\frac{k_{T1}}{k_{T2}} = -\frac{E_a}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \quad \text{(Eq. 7)}$$

with k_{T1} and k_{T2} the rate constants at the temperatures T_1 and T_2, respectively, R is the gas constant, and E_a is the activation energy. Thermoactivation can be quantified by the Q_{10} value, the fractional rate change upon a 10 °C increment in temperature:

$$Q_{10} = \frac{k_{T2}}{k_{T1}}^{\frac{10}{T_2 - T_1}} \quad \text{(Eq. 8)}$$

At increasing temperatures, reversible or irreversible enzyme inactivation becomes relevant and activity declines rapidly. The result is a bell-shaped curve (Fig. 4.3) where the maximum in activity is defined by the temperature optimum (T_{opt}). Engineering a more thermophilic cellulase has the attractive advantage of shifting the T_{opt} upwards, thus further extending the exponentially increasing part of the activity curve. However, the degree of thermoactivation of cellulases on their natural substrate has been low compared to other non-interfacial enzymes (Elias et al. 2014). In Article III, we propose that the reason for this is a temperature-induced desorption of cellulases from the substrate surface. In other words, the higher the temperature, the lower is the binding of the cellulase to the substrate, which determines an increase in population of enzymes free in solution rather than bound on the cellulose interface (Fig. 4.2). Therefore, we propose that for industrial saccharification it is not only necessary to utilize thermostable cellulases, which ensure that the enzyme remains catalytically active for many days, but also that the enzymes remain associated with the substrate. This can be achieved by increasing its affinity properties via rational design.

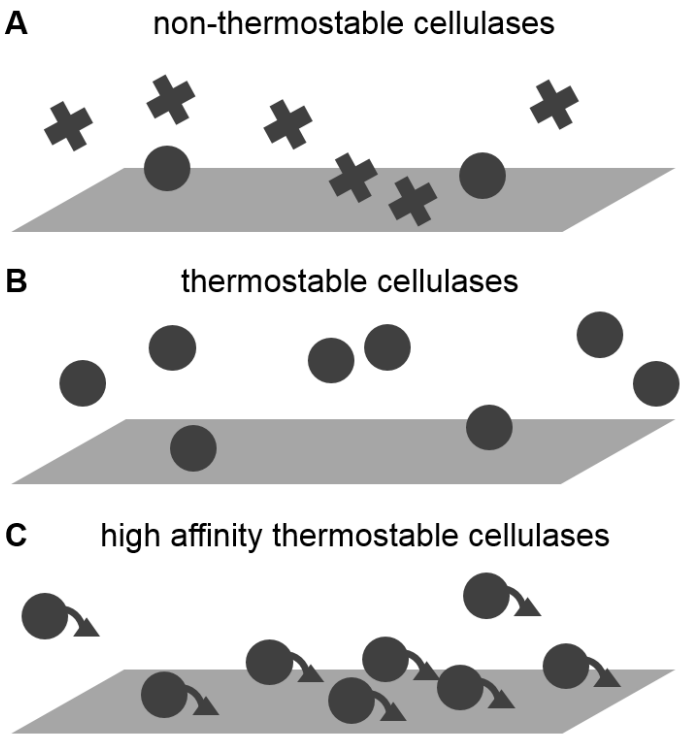


Figure 4.2. *Conceptual image of the temperature effect for different cellulases (black) acting on cellulose (gray) in industrial conditions. Three possible scenarios can occur when a cellulase is subjected to high temperatures of an industrial bioreactor. (A) Thermolabile cellulases are quickly destabilized by the high temperatures and become inactive (X symbol); (B) Thermostable cellulases will remain active for a prolonged period of time but not remain adsorbed onto the substrate surface; (C) Thermostable cellulases with improved affinity properties ensure temperature stability and association to the substrate.*

To test the validity of this hypothesis, we studied the thermostable cellobiohydrolase I from *Rasamsonia emersonii* (*Re*Cel7A) which is naturally composed of a single catalytic domain. We tested the activity of *Re*Cel7A at different temperatures and we observed that increase in temperatures resulted in a modest increase in activity (black points, Fig. 4.3), particularly in the industrially relevant range from 45 to 65 °C.

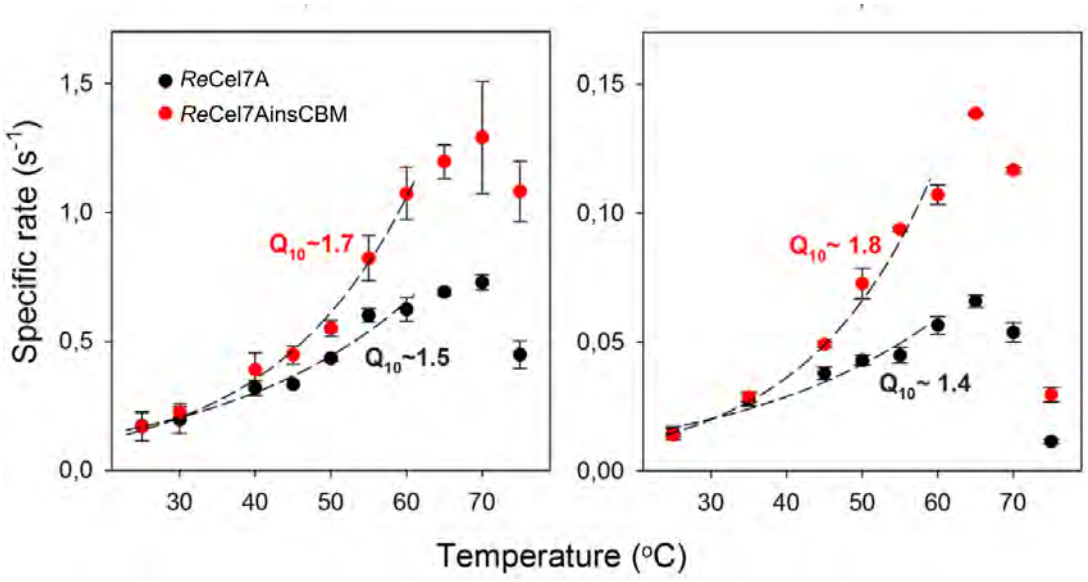


Figure 4.3. *Temperature-activity relationships for the enzymes ReCel7A and ReCel7AinsCBM measured on Avicel for 20 min. (left panel) or 24 h (right panel). Q_{10} values are calculated by fitting an exponential function in the temperature range 25-60 °C. Conditions: final enzyme concentration of 100 nM and a final substrate load of 10 g/L. Adapted from Article II.*

Using a rational design strategy we created a "high-affinity" variant of *Re*Cel7A, termed *Re*Cel7AinsCBM, where we added at the C-terminal of *Re*Cel7A the linker-CBM sequence of *Tr*Cel7A. This approach seemed promising because of the pivotal role of the CBM in improving cellulose affinity (chapter I). The temperature-activity relationships for the high-affinity variant are shown in red in Fig. 4.3 and activity has been compared to the wild type enzyme in assays at two different contact times, 20 min and 24 h. While close to 25-30 °C the two enzymes show almost identical specific rates, at increasing temperatures *Re*Cel7AinsCBM quickly becomes superior, and at 70 °C the activity is twice as high as *Re*Cel7A. By calculating Q_{10}, it was possible to measure that *Re*Cel7AinsCBM is better thermoactivated. The Q_{10} increases to ~1.7–1.8 compared to ~1.4–1.5 of the wild type (Fig. 4.3). To ensure that heat-induced dissociation from the substrate is not caused by enzyme unfolding, we also measured the distribution of free and substrate-bound enzyme in the same samples via two independent methods (not shown), where we observe no sign of enzyme inactivation between 25-55 °C and instead only a strong temperature-induced unbinding of *Re*Cel7A, while the activity shifts form surface to bulk. Taken together, these results show that the addition of a linker and CBM from another fungus (a mesophile no less) had the expected effect of increasing the population of adsorbed enzyme at high temperatures.

4.4 - Beyond CBHs: thermoactivation of an endoglucanase

Results from Article III and other comprehensive studies done on *Tr*Cel7A (Sorensen et al. 2015)) pointed out that decreased cellulose affinity at high temperatures might constitute a general limitation for industrial application of cellulases. In fact, also GH7 endoglucanases suffer from this phenomenon, as shown in the case of *Tr*Cel7B in Fig. 4.4.

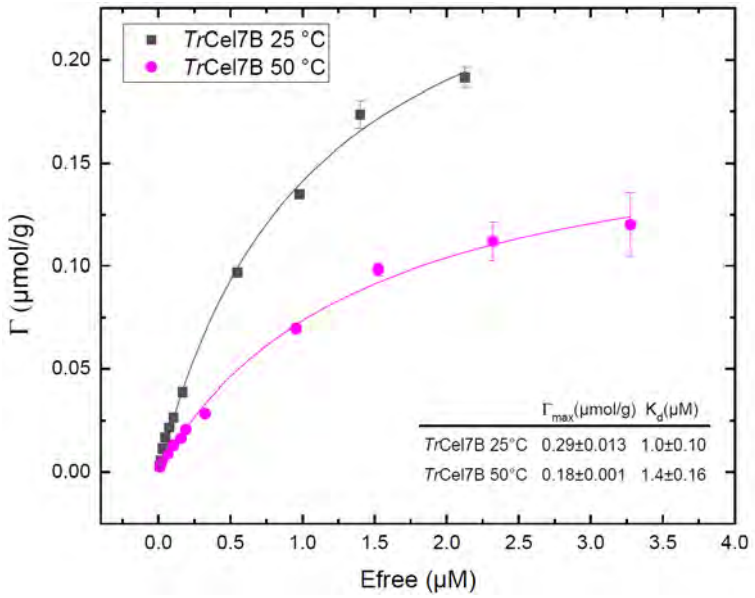


Figure 4.4. *Langmuir isotherms for TrCel7B at two temperatures, 25 and 50 °C. Less enzyme is bound when the temperature is increased. Conditions: Avicel load 15 g/L, enzyme concentration 0.05-5 μM. Materials and methods are the same as in article III. Error bars represent standard deviation from triplicates, lines are best fit of eq. 5. Binding parameters derived from the fit are shown in the insert. Figure and data analysis performed using OriginPro 2019 (OriginLab Corp., Northampton, MA, USA).*

At 50 °C, the binding to cellulose decreases significantly compared to 25 °C (Fig. 4.4). Considering that *Tr*Cel7B is stable at 50 °C for much longer than the 1 h investigated here (data not shown), this suggests that desorption at 50 °C is probably not ascribed to unfolding. Therefore, it is safe to hypothesize that a similar temperature-induced unbinding observed for the CBH *Re*Cel7A in the previous section, hampers also the activity of the EGs in industrial settings. We applied the same strategy described in Article II (*i.e.* increasing the activity of already thermostable enzymes by increasing their substrate affinity) using the paralog of *Re*Cel7A, the GH7 endoglucanase *Re*Cel7B (Article III). The enzyme showed very low binding at 50 °C, in fact it was not possible to reliably measure the amount of bound enzyme at the same conditions shown in Fig. 4.4. We thus

designed and produced a variant of *Re*Cel7B with added CBM from *Tr*Cel7A, *Re*Cel7BinsCBM. The temperature activity profiles of *Re*Cel7BinsCBM and *Re*Cel7B are shown in Fig. 4.5.

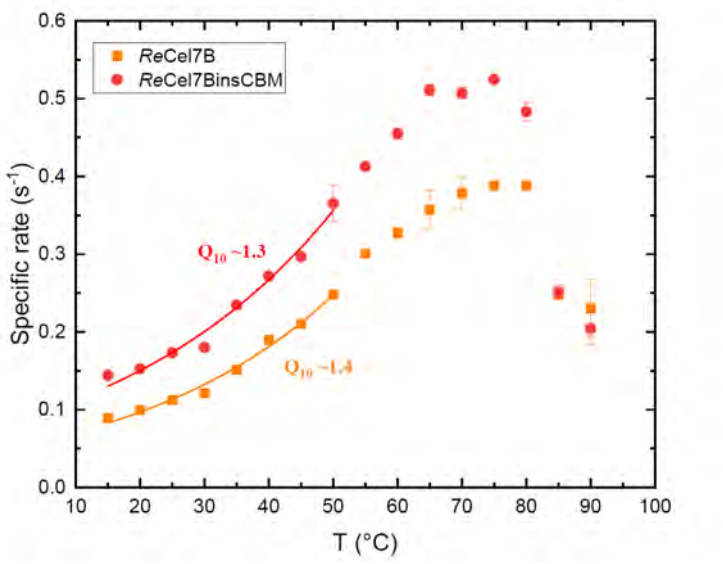


Figure 4.5. *Effect of temperature on the enzyme activity of ReCel7B (orange squares) and ReCel7BinsCBM (red circles). Conditions: Avicel load 90 g/L, enzyme concentration 0.5 μM, 1h contact time. Error bars are standard deviation from triplicates. Adapted from Article III.*

For the high affinity variant *Re*Cel7insCBM, we observe an increase in activity at all temperatures compared to *Re*Cel7B (Fig. 4.5). However, the T_{opt} shifted from about 80 °C of the wild type to 75 °C, in contrast with the CBH case observed in Fig. 4.3, and the Q_{10} value of *Re*Cel7BinsCBM remained essentially unchanged compared to *Re*Cel7B (Fig. 4.5). This led to the conclusion that the variant is not better thermoactivated at the condition tested. We don't have a clear explanation for this, and studying temperature effects in GH7 EGs is further complicated by the fact that it is not possible to reach full saturation under MM conditions, as shown in Fig. 4.10A, thus reliable kinetic parameters cannot be extracted. However, it is possible that good thermoactivation is not achieved because of the conditions, particularly the substrate load used, since it has been observed for the CBHs that strong temperature activation is achieved only at high substrate loads (Sorensen et al. 2015). Therefore, it is possible that thermoactivation is poor for the EGs because they are never fully saturated with the substrate. Nonetheless, the overall activity of *Re*Cel7BinsCBM is improved in the entire temperature range 15-80 °C compared to the natural enzyme, thus making the protein engineering strategy of section 4.3 effective also for an EG. However, rational design strategies such as the introduction of an exogenous

linker-CBM should follow careful consideration, and should take into account that that not only EGs have a different substrate promiscuity on cellulose compared to the CBHs (chapter II), but also that CBMs from family 1 show a range of different affinities (Linder et al. 1995, Lehtiö et al. 2003). In light of the results in Fig. 4.5 and Article II and III, designing new high affinity enzymes by testing the effect of different CBMs appears to be a promising strategy to improve the efficiency of EGs at high temperatures.

4.5 - Crystal structure of the endoglucanase *Re*Cel7B

In the beginning of the chapter, it was mentioned how a successful strategy of rational design strongly depends on a detailed structure-function understanding of the enzyme to improve (Zhang et al. 2013). While such knowledge is abundant for CBHs, since 15 unique structures are available from different species, it is rather limited for the GH7 endoglucanases. Probably due to the difficulties associated with purification and crystallization, only four structures of GH7 EGs are available in CAZy database (Lombard et al. 2013). In Article III, we report a fifth, new crystal structure and kinetic characterization of the highly thermostable EG, *Re*Cel7B. The enzyme is produced by the thermophilic ascomycete *Rasamsonia emersonii,* formerly known as *Talaromyces emersonii* (Houbraken et al. 2012), expressing under cellulolytic conditions a full set of cellulases of industrial relevance (Moloney et al. 1983): at least three forms of cellobiohydrolases, four endoglucanases and three β-glucosidases have been isolated from the culture broth, although their relative amount secreted is not known (Coughlan et al. 1984) (Tuohy et al. 2002).

The thermophilic origin of *Re*Cel7B is reflected in its biochemical characteristics. The enzyme shows a T_{opt} of about 80 °C at pH 5.0 (Fig. 4.5) and retains high T_m in a wide range of pH values, particularly towards the acidic range (Article III).

A phylogenetic analysis of the biochemically characterized GH7 extracted from CAZy database, indicates that *Re*Cel7B belongs to the ascomycotal EG clade, along with all the known GH7 EG structures (black stars in Fig. 4.6), namely *Humicola insolens* Cel7B (*Hi*Cel7B), *Fusarium oxysporium* Cel7B (*Fo*Cel7B), *Trichoderma reesei* Cel7B (*Tr*Cel7B) and *Trichoderma harzianum* Cel7B (*Th*Cel7B).

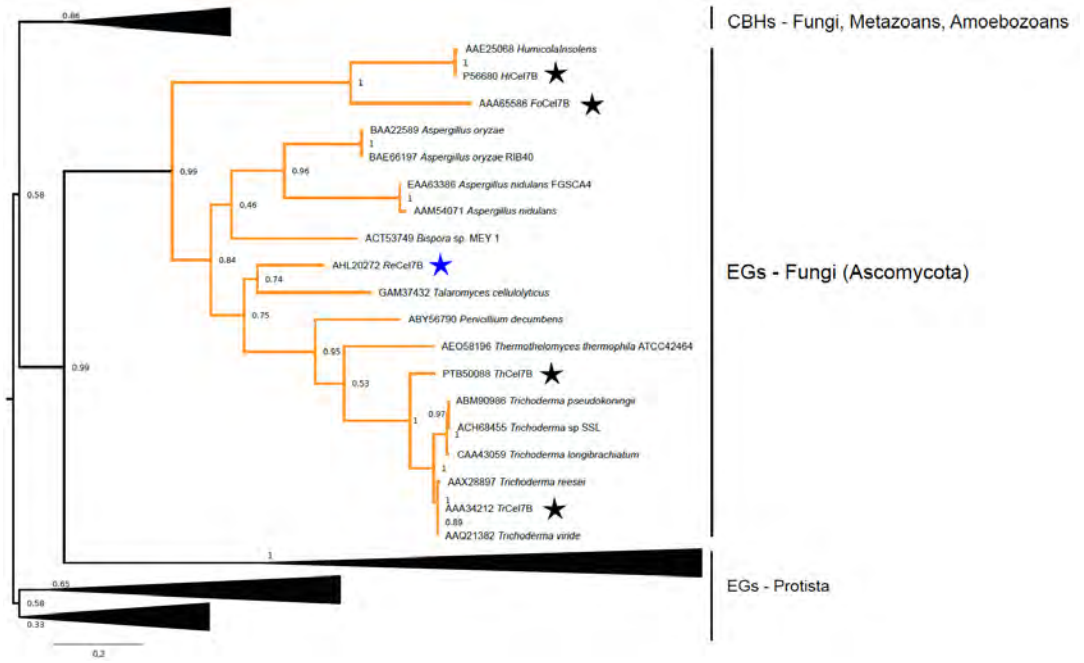


Figure 4.6. *The ascomycotal EG clade. A phylogenetic analysis of 87 biochemically characterized GH7 sequences allowed to study the evolutionary distribution of the ascomycotal EGs. The Genbank accession number and the organism of origin for each sequence is indicated. Only five EG structures have been solved to date (black stars) including the structure described in this thesis (blue star). Bootstrap values are shown at the nodes as percentage of 100 replicates. The scale bar is 0.2 amino acid substitutions per site. Adapted from Article III.*

As expected, the *Trichoderma* structures cluster together since they share high (87%) sequence identity. *Hi*Cel7B and *Fo*Cel7B belong to another cluster, and share lower (60%) sequence identity. *Re*Cel7B displays about 48-60% sequence identity amongst the four EGs considered and clusters with an EG from *Talaromyces cellulolyticus* (65% sequence identity). The sequence divergence of *Re*Cel7B suggests that it could harbour different structural features to expand the structure-function understanding of the GH7 EG subtype, thus it was chosen for crystallization experiments.

Crystals were obtained at high enzyme and salt concentration (>60 mg/mL enzyme and 2.4 M sodium malonate pH 7), after ~1.5 months. The crystals were approximately 200 µm in the longest direction (Fig. 4.7A) and belonged to space group $C222_1$. The structure of *Re*Cel7B is composed of a single catalytic domain and was solved at a resolution of 2.48 Å, with R-work and R-free factors of 0.18 and 0.2, respectively. The unit cell of *Re*Cel7B consists of a trimer (Fig. 4.7B), although the enzyme most likely functions as a monomer *in vivo*, as also shown by size-exclusion chromatography data (Article III). The glycan decorations (spheres in Fig. 4.7B) point towards the solvent, while the trimer interface is mainly formed by the flexible loops. It is interesting to observe that the solvent content

in the crystal is very high (about 78%) and very few crystal contacts are formed between the units.

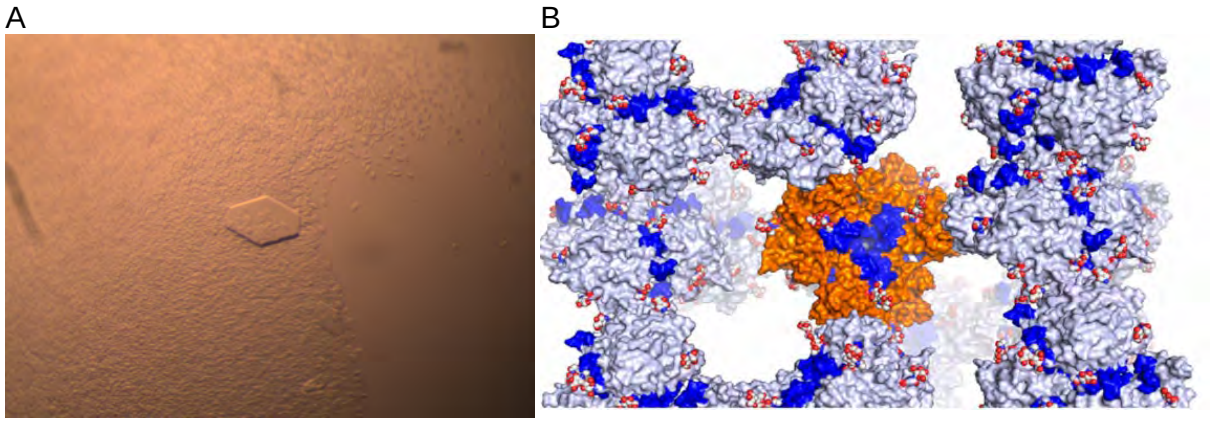


Figure 4.7. *Crystals and crystal lattice of ReCel7B. (A) Microphotography of the crystals obtained. The crystal in the centre has the approximate width of 200 nm. New nucleation points are visible in the area where a second crystal was destroyed. (B) The unit cell of ReCel7B is a trimer (orange) stabilized by loops (blue), while the extended glycans (gray spheres) face the bulk solvent. Adapted from Article III.*

4.6 - Structural comparisons between *Re*Cel7B and orthologs

In the following, we will show a detailed structural and sequence comparison of *Re*Cel7B with the EGs coming from three different fungal genera, *Tr*Cel7B, *Hi*Cel7B and *Fo*Cel7B. The analysis revealed different properties of the flexible loops forming the active site cleft (Fig. 4.8). The major loops investigated are termed A1-A3, B1 B3 and B4 following the same nomenclature proposed for the CBHs (Momeni et al. 2013). Loops S1 and S2 are secondary loops (Fig. 4.9B) stabilized by a glycosylation (section 4.8).

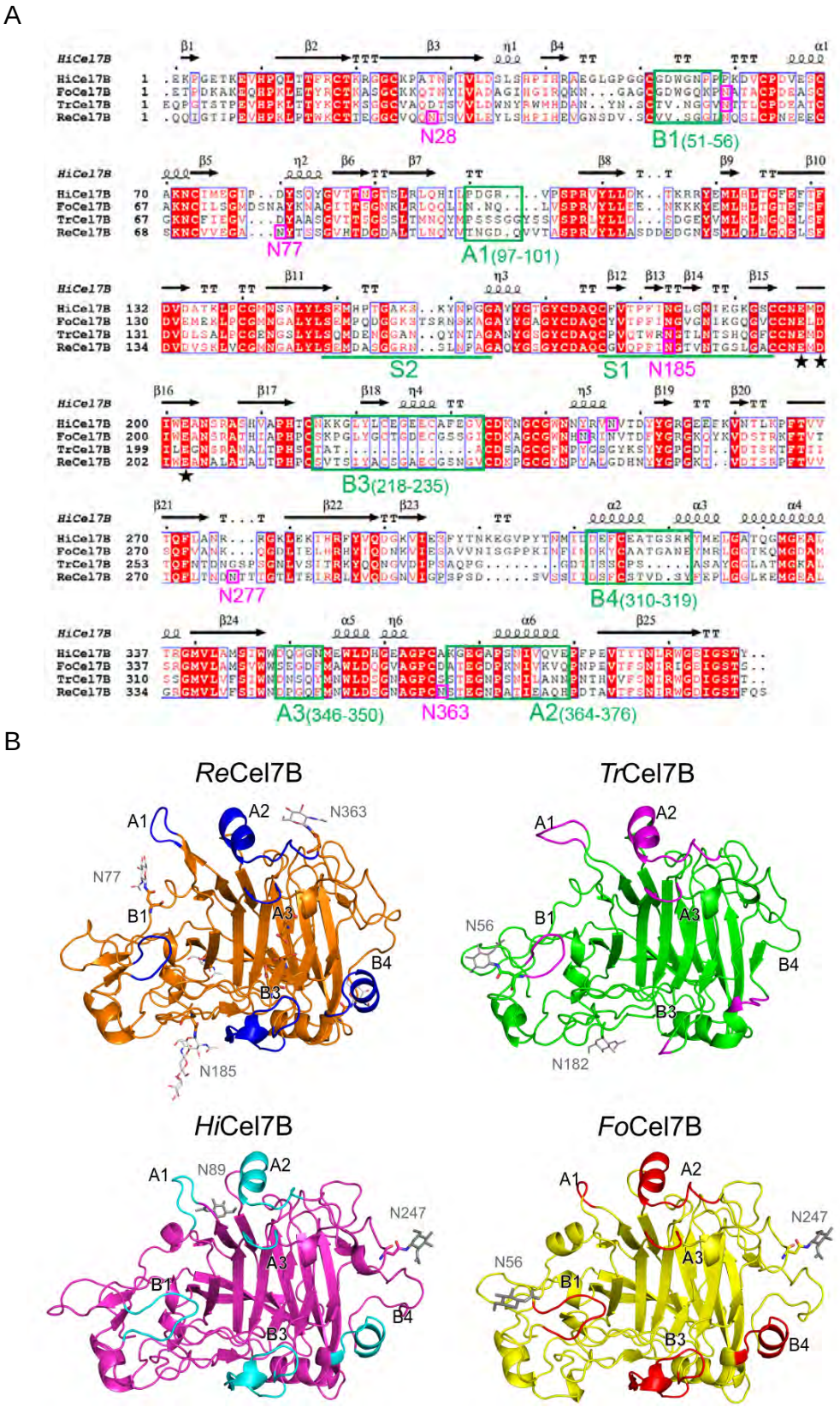


Figure 4.8. *Sequences and structures of four GH7 EGs. (A) Structure-based amino acid sequence alignment of ReCel7B (PDB 6SU8), TrCel7B (PDB 1EG1), FoCel7B (PDB 1OVW) and HiCel7B (PDB 2A39). Secondary structure elements of HiCel7B are shown on top of the alignment with the following figures and symbols: alpha helices (helices, α and η for a 3_{10}-helix), beta strands (arrows, β),*

turns (T, strict β-turns as TT and strict α-turns as TTT). Strictly identical residues are marked in white characters on a red background, while chemically similar residues are shown as red characters. Regions of conserved, highly similar residues are framed in blue boxes. Green frames indicate loop regions of interest, with loop nomenclature below and residue positions for ReCel7B. Black stars indicate the residues involved in the catalysis (E199, D201 and E204 in ReCel7B). Asparagine residues where N-glycosylation was identified in the different structures are framed in magenta, and the residue numbering for ReCel7B is indicated. Alignment created with Clustal Omega (Sievers et al. 2011) and rendered with ESPript 3.0 web server with default parameters (Robert and Gouet 2014). (B) Cartoon representation of the different structures shown in panel A. Loops A1-3, B1, B3 and B4 are highlighted in different colours. Asparagine residues with N-glycosylation are also indicated. Panel A adapted from Article III, panel B was created in Pymol (The PyMOL Molecular Graphics System, Version 2.3.2, Schrödinger, LLC).

As described in section 3.9, the A1 loop is heterogeneous in GH7, and this is also reflected in the EG structure and sequence alignment. While *Re*Cel7B and *Tr*Cel7B display a long A1 loop, in *Hi*Cel7B and *Fo*Cel7B it is rather short (Fig. 4.8B). The opposite is observed for the B1 loop, which is longer in *Hi*Cel7B and *Fo*Cel7B than *Re*Cel7B and *Tr*Cel7B. The presence of a longer A1 loop in *Re*Cel7B seems counter-intuitive in a thermostable enzyme, since this could increase fluctuations and destabilize the structure at high temperatures. Although we do not have a functional explanation for this, it is possible to note that the presence of charged and polar residues in *Re*Cel7B, namely D100 and Q101, could be involved in enhancing the hydrophilicity and solubility of the enzyme. In comparison, the A1 loop in *Tr*Cel7B is rich in serine and glycine, and we observe a π-stacking between Y102 of the A1 loop and P353 in the A2 loop, absent in *Re*Cel7B since tyrosine is replaced by valine (V102). No particular interactions are observed in *Fo*Cel7B between the A1 loop and the rest of the structure, whereas a salt bridge is observed in *Hi*Cel7B, between D101 and R103. The high sequence diversity of the A1 loop in the GH7 (Fig. 3.11) seems to indicate that less evolutionary pressure for structural stability is applied in this region. It is interesting to note that in the A1 loop of the EGs, no ion pair appear to be established with the neighbouring A2 loop, which was possible in some CBHs (section 3.9), suggesting an overall higher flexibility in the EG subtype. However, the fact that the A1 loop still exists and is not truncated in the EGs, prompts to speculate that it could play a role in the initial substrate binding. Indeed, MD simulations done by Lin *et al.* (Lin et al. 2013) suggests that the A1 and the B1 loops are involved in the initial binding and threading of the substrate in the active site of *Tr*Cel7B. Nonetheless, the role of the A1 loop in GH7 remains to be fully elucidated. As for the B1 loop, the most striking difference between *Re*Cel7B and *Tr*Cel7B compared to both *Fo*Cel7B and *Hi*Cel7B is the presence, in the latter two, of a tryptophan residue (W54 in *Hi*Cel7B and W51 in *Fo*Cel7B, which is π-stacking with F177 and Y177, respectively. The A2 loop is rather similar in the

EG structures (Fig. 4.8) although we observe that one glycosylation in *Re*Cel7B very close to this loop in position N363, which is a unique feature of *Re*Cel7B compared to the other EG structures.

The A3 loop is known to play a role in stabilizing neighbouring loops in the homologs CBHs. For example, *Tr*Cel7A contains two tyrosines (Y370 and Y371) in this region to effectively create a close tunnel shape because of the interaction with loop B2 and B3 (Momeni et al. 2013). In the EGs, loop B2 is completely absent and loop B3 is very short compared to the CBHs, therefore these interactions are not possible. Interestingly, in the A3 loop of *Re*Cel7B there is a proline, P347, absent in all the GH7 structures (Fig. 3.2), which potentially increases rigidity in the region.

The B3 loop also shows different lengths in the EGs, as seen for the A1 and B1 loops. It is longer in *Re*Cel7B, *Hi*Cel7B and *Fo*Cel7B, while it is absent in *Tr*Cel7B (Fig. 4.8). Moreover, the long B3 loop in the former enzymes is stabilized by an additional disulfide bridge (C225-230 in *Re*Cel7B), absent in *Tr*Cel7B. Interestingly, we observed a stabilizing interaction in this area due to the presence of the glycosylation site at N185 (section 4.8), which is not possible in *Hi*Cel7B and *Fo*Cel7B since they lack this glycosylation (Fig. 4.8 B). The B4 loop is located in the product site and is a short loop connecting two alpha helices (α2 and α3) in *Hi*Cel7B and *Fo*Cel7B (Fig. 4.8A). In *Re*Cel7B, one of these alpha helices is not formed due to the presence of another non-conserved proline (P322). On the other hand, *Tr*Cel7B completely lacks helix α2, thus making it the structure with the most open product site. Interestingly, we observe in *Re*Cel7B an acid pair between D317 and D237, absent in the other structures. The α-carboxylic acid groups of these residues are most likely deprotonated ($-COO^-$) in *Re*Cel7B crystal structure at pH 7. However, at acidic pH values optimal for *Re*Cel7B activity, they should be able to form a double hydrogen-bonded structure that increases stability in this region.

4.7 - Active site of *Re*Cel7B and other EGs

All GH7 enzymes display the conserved active site motif ExDxxE (Mertz et al. 2009), thus the proposed catalytic residues in *Re*Cel7B are E199 (nucleophile), D201 and E204 (acid/base), Fig. 4.9A. It has been suggested that the EGs contain at least 4 sub-sites to interact with the cellulose chain (Payne et al. 2015), and more have been proposed (Sonoda et al. 2019). Two highly conserved tryptophans in *Re*Cel7B, W344 and W353, are properly positioned to interact with the substrate in subsite -2 and +1 respectively, and other highly conserved residues important for substrate interactions were found in position R108, S342

and Y149 (Taylor et al. 2013) (Divne et al. 1998) (Parkkinen et al. 2008).
Interestingly, in the active site of both *Re*Cel7B and *Tr*Cel7B there is one major substitution compared to *Hi*Cel7B. The latter contains a histidine residue (H209) which has been ascribed as main responsible for the high pH optimum of *Hi*Cel7B, which is 7.5 (Schulein 1993). The residue H209 hydrogen bonds with the substrate in subsite +1. Studies done on an alanine to histidine mutant in *Tr*Cel7A highlighted the role of the histidine in providing protons to the active site glutamate (Becker et al. 2001). Histidine is absent both in *Tr*Cel7B and *Re*Cel7B and substituted to alanine (A211 in *Re*Cel7B) as seen for *Tr*Cel7A, thus strengthening the importance of alanine in this position to maintain an acidic pH optimum.
Moreover, it has been shown in the CBHs that the presence of two arginines (R251 and R394 in *Tr*Cel7A) interact with cellobiose in the product site (Payne et al. 2015). These residues are lost in the EGs, which gives a structural indication for the lower product inhibition and processivity observed in the EGs (Atreya et al. 2016).
Lastly, two tryptophans conserved in the CBHs structures, W40 and W38 (*Tr*Cel7A nomenclature) participate in ligand binding at position -4 and -7 respectively (Taylor et al. 2013). The aromaticity is conserved in *Tr*Cel7B with Y38 and W40 present at the equivalent positions (Fig. 4.9C). In the other EGs this interaction is lost, and both *Re*Cel7B and *Hi*Cel7B contain a serine and a proline in the corresponding positions (S37 instead of Y, P39 instead of W, *Re*Cel7B nomenclature) (Fig. 4.9C). Protein engineering approaches done by Davies *et al.* showed that additional sugar binding sites (+3 +4), might have been introduced in *Hi*Cel7B when mutating these residues to tryptophans in *Hi*Cel7B S37W/P39W double mutant (Davies et al. 1997). Since both *Hi*Cel7B and *Re*Cel7B are thermophilic enzymes -although *Hi*Cel7B has a much lower temperature optimum (Schulein 1993)- the presence of S37/P39 in *Re*Cel7B as well, suggests that these residues play a role in increasing structural rigidity, but coming at the price of a reduction in substrate interactions in the cleft. *Fo*Cel7B contains an isoleucine and a glycine at the corresponding positions, thus suggesting to be the most flexible in this region.

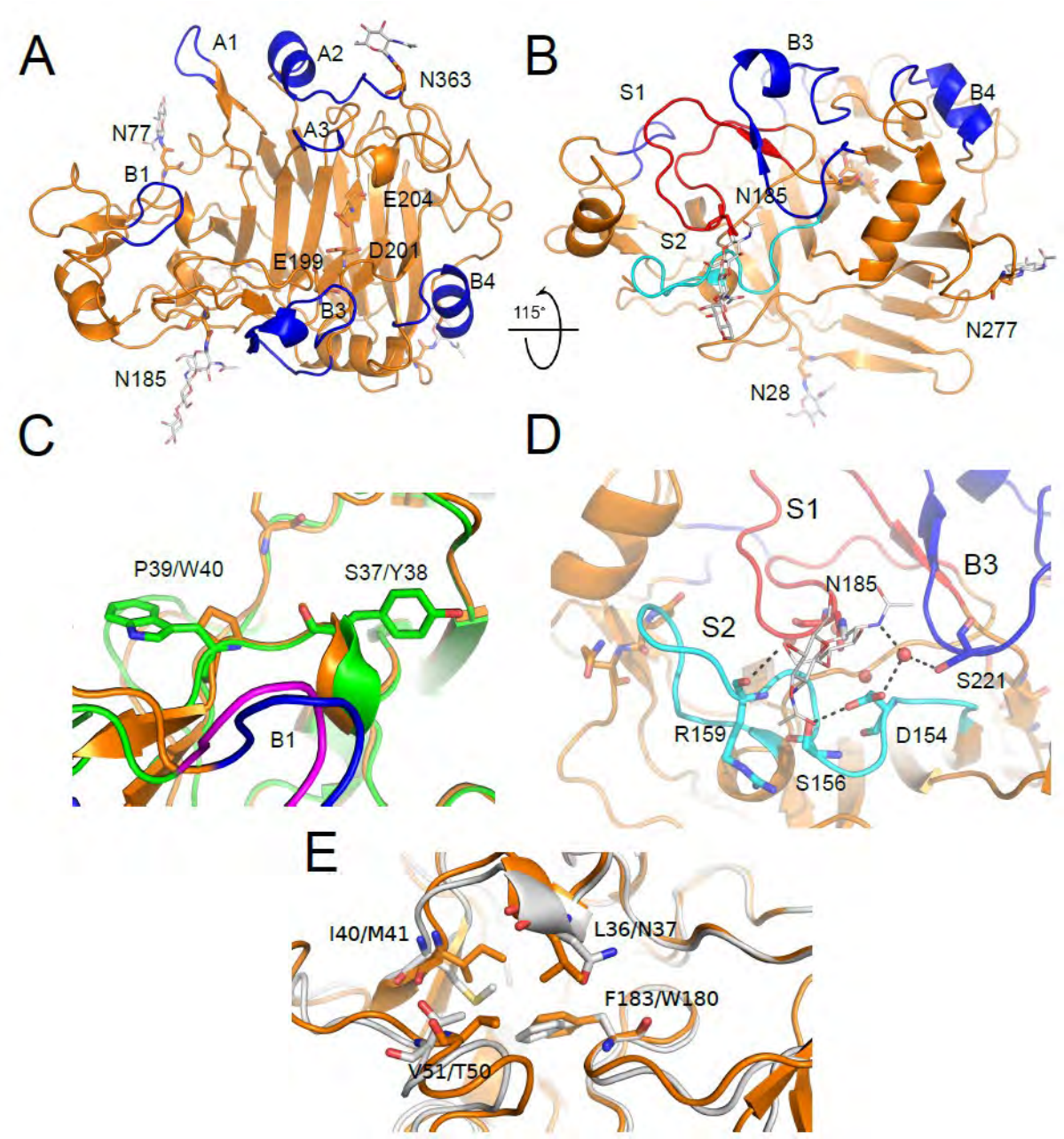


Figure 4.9. *The structure of ReCel7A (PDB 6SU8) and comparison with other EGs. (A) Cartoon representation of ReCel7B structure (orange) and flexible loops (blue). Active site residues E199, D201 and E204 are indicated, as well as the glycans (gray) at position N77, N185 and N363. (B) Rotated view of 115 ° on the horizontal axis of ReCel7B, with the glycosylations at N28 and N277 indicated. Loops S1 and S2 are highlighted in red and cyan respectively. (C) Superimposition with TrCel7B (green, PDB 1EG1). The residues S37 and P39 of ReCel7B are substituted by two aromatics in TrCel7B, Y38 and W40, respectively, which could be responsible for improving cellulose binding. (D) The glycosylation at position N185 of ReCel7B aids in stabilizing the close loops B3, S1 and S2 via hydrogen bonding (dotted lines). A water molecule stabilizes at the same time the N-acetyl group of the first N-acetylglucosamine attached to N185, the residue S221 in loop B3 and D154 in loop S2. The 6'O of the N-acetylglucosamine stabilizes R159. (E) Superimposition with the structure of Trichoderma pseudokoningii Cel7B (gray), created with homology modelling. In ReCel7B, the residues L36, I40, V51,*

F183 create a hydrophobic cavity. This cavity is less hydrophobic in Trichoderma pseudokoningii Cel7B, due to the presence of N37, M41, T50 and W180 in the corresponding positions. Figure adapted from Article III. Panel E was created in Pymol (The PyMOL Molecular Graphics System, Version 2.3.2, Schrödinger, LLC).

4.8 - Glycosylations and stabilizing interactions in *Re*Cel7B

The crystal structure of *Re*Cel7B revealed *N*-glycosylations at five positions (Fig. 4.9 A,B), further described at the end of this section. This prompted us to perform biochemical studies on the glycan composition of *Re*Cel7B to confirm the identity of the sugars. Intact mass spectrometry (MS) was performed both on the native enzyme and a deglycosylated version of *Re*Cel7B, which was treated with endoglycosidase H (Endo-H). Endo-H (EC 3.2.1.96) cleaves between the first two *N*-acetylglucosamine units attached to the asparagine. Thus, in deglycosylated *Re*Cel7B, one *N*-acetylglucosamine remains attached to each asparagine. Mass comparison between *Re*Cel7B and deglycosylated *Re*Cel7B revealed a mass shift of about 8 kDa (Article III), indicating extensive glycosylation in the native enzyme (the average size of a hexose is only 162.14 Da). In addition to this, monosaccharide composition analysis of the sugars extracted via acid hydrolysis revealed that the enzyme is mainly mannosylated (Table 4.1). Galactose was also identified in lower amount, as observed in other *Aspergillus oryzae* expressed enzymes (Nakao et al. 1987), and we believe it acts as capping substitution on the mannose chains to prevent further sugar additions (Deshpande et al. 2008).

Table 4.1. *Monosaccharide composition analysis of ReCel7B. Monosaccharides were determined by HPAEC-PAD on ReCel7B treated with Endo-H. The average of values from three hydrolysis experiments is given.*

Monosaccharide	**Amount (µg mg^{-1} protein)**
Mannose	85.0 ± 1.4
Galactose	27 ± 9
Glucosamine	42 ± 1
Galactosamine	nd*
Glucose	nd*

*nd:not detected

*Re*Cel7B crystal structure retained the native glycosylation since we did not perform any

glycan modifications prior to the crystallization trials. Of the five glycosylation sites observed, four are not observed in the other EG structures, namely N28, N77, N277 and N363 (Fig. 4.9 A,B). The fifth glycosylation, N185, is also present in *Tr*Cel7B (Fig. 4.8) and in its close ortholog *Th*Cel7B (Mackenzie et al. 1998, Sonoda et al. 2019). We observe that this specific glycosylation in *Re*Cel7B is involved in a complex stabilizing interaction between three adjacent loops B1, S2 and S1, described as follows. Residue N185 belongs to loop S1 (red, Fig. 4.9D). The 6'O of the first NAG attached to N185 is in hydrogen bond distance with the carbonyl in the backbone of R159 in loop S2 (cyan, Fig. 4.9D). Moreover, a water molecule is bridging between loop B3 (S221) and S2 (D154) via the acetyl group of NAG (red spheres, Fig. 4.9D). In loop S2, D154 is also involved in stabilizing another residue, S156 (Fig. 4.9D). We conclude that this modification is a key structural feature in enhancing the stability of *Re*Cel7B. To note, loop B3 is missing in *Tr*Cel7B (Fig 4.8), therefore we suggest that similar stabilizations do not occur in either *Tr*Cel7B or *Th*Cel7B.

The other glycan modifications present at positions N28, N77, N277 and N363 appear very solvent exposed and we do not find particular interactions with the peptide chain. Nonetheless, their functional role in stability cannot be excluded. As shown for example in *Hi*Cel7B, the partial removal of the glycans can result in a significant decrease in thermostability and pH stability (Hayashida et al. 1980). To further support the stability role of the glycans in *Re*Cel7B, we performed an enzymatic digestion of *Re*Cel7B with either Endo-H or Peptide-N-Glycosidase F (PNGase F, EC 3.5.1.52) at pH 7.0. Subsequent analysis via differential scanning fluorimetry (DSF) revealed a relative decrease in T_m of 2 and 4 °C after Endo-H and PNGase F treatment, respectively (Article III). In addition to this, the presence of numerous glycosylations in *Re*Cel7B could have other functions, such as favouring an increase in protease stability (Amore et al. 2017) and solubility (Beckham et al. 2012). On the other hand, the presence of extensive glycans could be detrimental for the activity and create steric hindrances that limit the cellulose binding and/or threading. This has been proposed for *Tr*Cel7A, in particular for glycans close to the area where initial cellulose threading can occur (Dotsenko et al. 2016).

In addition to glycosylations, a number of other stability promoting elements have been identified in *Re*Cel7B. Several prolines absent in *Tr*Cel7B were identified, namely P265/T248, P14/T15, P239/A222, P216/S213 (*Re*Cel7B/*Tr*Cel7B nomenclature), in addition to the aforementioned prolines in the B4 and A4 loops (section 4.6). Superficial ion pairs have been suggested being stability promoting elements especially at high temperatures, where the dielectric constant of water decreases (Taylor et al. 2010). Such pairs are found in *Re*Cel7B in D152-R335, D293-R264 and D136-K139, all absent in *Tr*Cel7B. Moreover, increased core hydrophobicity has been linked to the structure of

many thermostable proteins (Szilágyi et al. 2000). A better hydrophobic packing in *Re*Cel7B can be observed in the pocket created by I40, V51, F183 and L36 (Fig 4.9E). Semi-rational design approaches done by Mitrovic *et al.* indicated that a Met to Ile mutant (M63I) in an homologous EG from *Trichoderma pseudokiningii* significantly improved the thermostability (Mitrovic et al. 2014). The authors proposed that the presence of Ile increased the hydrophobicity in the surface cavity. This mutation correspond to I40 in *Re*Cel7B, therefore suggesting a stronger water repulsion in the cavity of *Re*Cel7B which aids thermostabilization.

4.9 - Kinetics of *Re*Cel7B and the role of the CBM

In addition to the structural analysis, *Re*Cel7B has been kinetically characterized on Avicel in comparison with *Tr*Cel7B. *Re*Cel7B, along with *Hi*Cel7B and *Fo*Cel7B, consists of a single catalytic domain whereas *Tr*Cel7B contains an additional linker-CBM domain (see scheme in Fig. 1.7). As seen in chapter I, most annotated GH7 do not possess a CBM, although it plays a fundamental role in cellulose targeting. To account for this difference in the kinetic comparison, in addition to the wild types we characterized two engineered variants of *Re*Cel7B and *Tr*Cel7B. One was *Tr*Cel7B deprived of linker-CBM (*Tr*Cel7BΔCBM), and the second one is *Re*Cel7B with an added linker-CBM from *Tr*Cel7A (*Re*Cel7BinsCBM). Results from ^{conv}MM, ^{inv}MM and binding isotherms on Avicel at 50 °C are illustrated in Fig. 4.10 and Table 4.2.

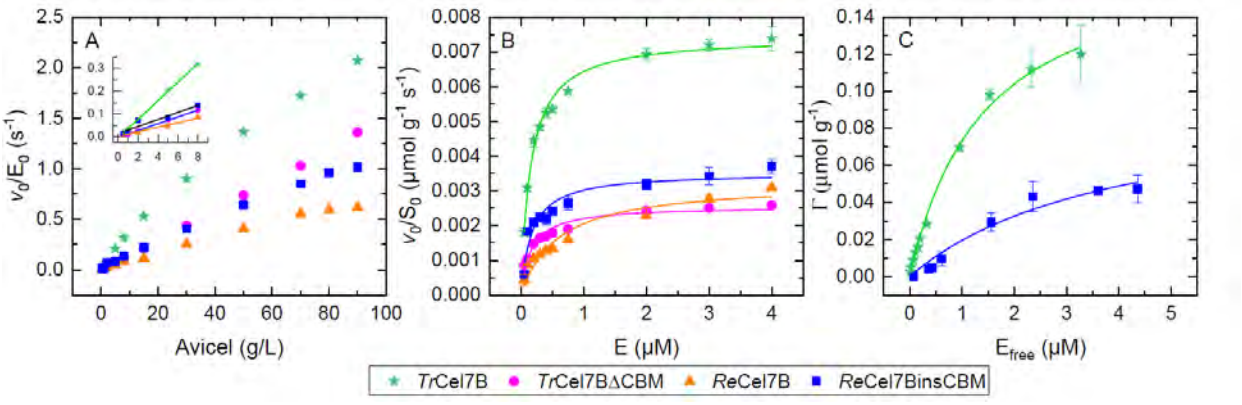


Figure 4.10. *Enzyme kinetics and binding isotherms of ReCel7B (orange up-pointing triangles), TrCel7B (green stars) and the variants ReCel7BinsCBM (blue squares) and TrCel7BΔCBM (pink circles). (A) Initial rates at low enzyme concentration and increasing substrate loads for ReCel7B, ReCel7BinsCBM, TrCel7B and TrCel7BΔCBM. Conditions: 100 nM enzyme, substrate loads 0.5-90 g/L. A linear fit of the initial values (insert) was performed to extract the specificity constant. (B)*

Inverse Michaelis-Menten approach for ReCel7B, ReCel7BinsCBM, TrCel7B and TrCel7BΔCBM. Conditions: substrate load of 8 g/L, enzyme concentrations 0.05-4 μM. (C) Binding isotherms for ReCel7BinsCBM and TrCel7B on Avicel 15 g/L. In all cases, symbols represent experimental data, error bars indicate standard deviation from triplicate measurements, and lines are a linear fit, best fits of Eq. 3 and Eq. 5 for panel A, B and C, respectively. Figure adapted from Article III.

Table 4.2. *Kinetic and adsorption parameters of TrCel7B, ReCel7B, TrCel7BΔCBM and ReCel7BinsCBM on Avicel at 50°C. Standard error is indicated.*

Enzyme	Specificity constant	Inverse Michaelis-Menten		Binding isotherms	
	η ($gL^{-1}s^{-1}$)$\times10^{-3}$	$^{inv}V_{max}/S_0$ ($\mu molg^{-1}s^{-1}$) $\times10^{-3}$	$^{inv}K_M$ (μM)	Γ_{max} (μmol/g)	K_d (μM)
*Tr*Cel7B	41 ± 1.2	7.5 ± 0.14	0.16 ± 0.013	0.18 ± 0.009	1.4 ± 0.16
*Tr*Cel7BΔCBM	15 ± 0.5	2.6 ± 0.10	0.16 ± 0.026	ND*	ND*
*Re*Cel7B	10 ± 0.9	3.3 ± 0.24	0.58 ± 0.127	ND*	ND*
*Re*Cel7BinsCBM	15 ± 2.8	3.5 ± 0.19	0.17 ± 0.037	0.10 ± 0.028	4.0 ± 1.99

*ND: not detectable

Under ^{conv}MM conditions, the enzymes investigated do not reach full saturation, therefore we only report the specificity constant η (Table 4.2). This value is much higher in *Tr*Cel7B than *Re*Cel7B (0.041 $gL^{-1}s^{-1}$ compared to 0.010 $gL^{-1}s^{-1}$). On the other hand, *Re*Cel7BinsCBM and *Tr*Cel7BΔCBM show similar η values of 0.015 $gL^{-1}s^{-1}$. For the ^{inv}MM, we observed that the inverse maximal specific rate of *Tr*Cel7B is about 2-fold higher than *Re*Cel7B (Table 4.2), indicating a higher ability to locate hydrolysable sites on the cellulose surface. As expected, the insertion of a CBM in *Re*Cel7BinsCBM increases the parameter $^{inv}V_{max}$ while the deletion of the CBM from *Tr*Cel7B dramatically decreases the activity to values lower than *Re*Cel7B (Table 4.2). This is also reflected in the binding isotherms, where the enzymes without CBM have very weak binding, while the addition of a CBM in *Re*Cel7B increases the binding from practically undetectable levels (<0.01 μmol/g) to 0.10 μmol/g (Table 4.2).

We conclude from these results that the catalytic performance of *Re*Cel7B was poorer than *Tr*Cel7B at this temperature, and this was particularly clear when only the catalytic domains were compared. This suggest that *Re*Cel7B has not evolved structural features that promotes a stronger ligand binding in the CD to compensate for the absence of a CBM, but rather that the structural modifications aimed to promote stability over activity and binding. Further functional considerations will be described in the next section.

4.10 - Synergy of *Re*Cel7B with cellobiohydrolases

The increased ability of *Tr*Cel7B to locate and hydrolyse more sites compared to *Re*Cel7B could also explain why synergy is overall higher for the former enzyme when mixed with different CBHs. We measured the degree of synergy (DS) of *Re*Cel7B in mixtures with CBHs, either *Re*Cel7A or *Tr*Cel7A, and compared the results with the synergistic capability of *Tr*Cel7B in combination with the same CBHs (Article III). The results (Fig. 4.11) indicate that DS is higher when the CBHs are mixed with *Tr*Cel7B, and distinctive synergy is observed at fractions containing as little as 1% *Tr*Cel7B.

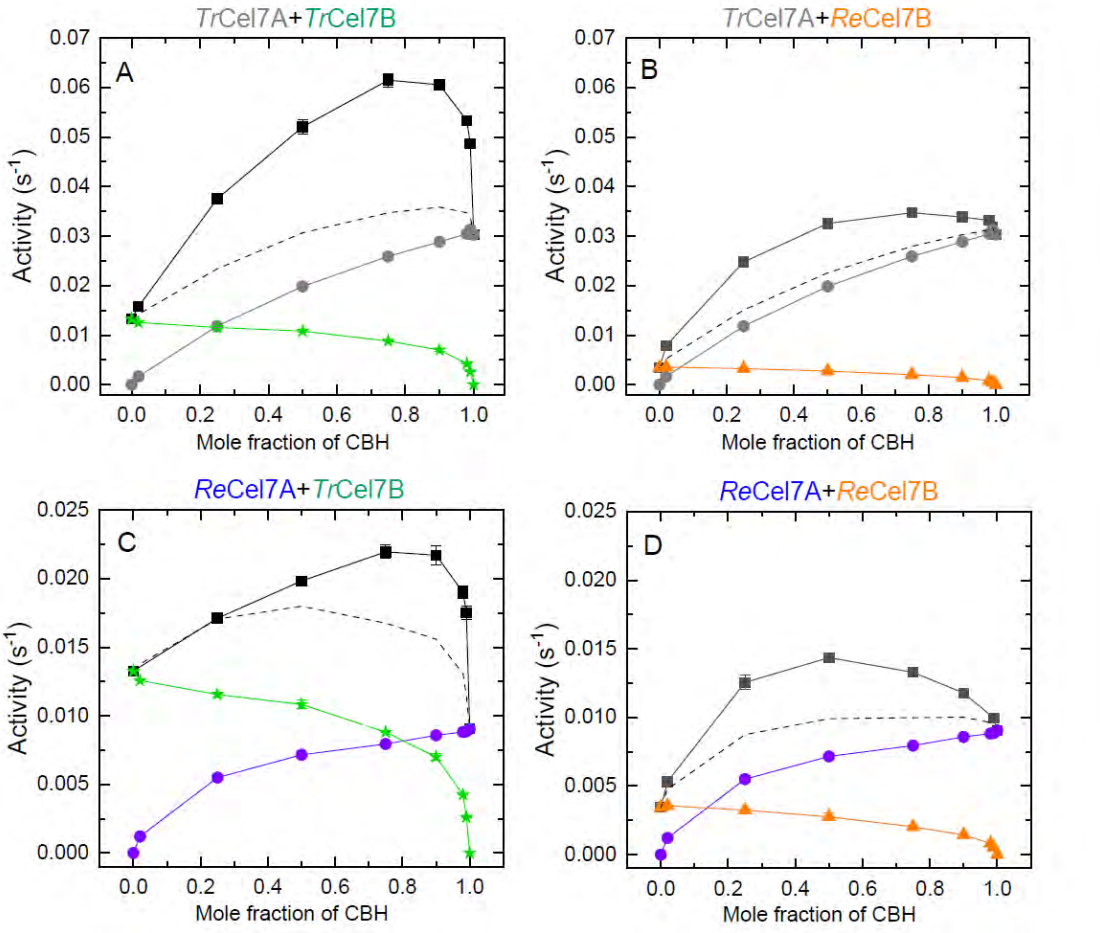


Figure 4.11. *Synergy curves for different EG-CBH mixtures. Activity data for synergy mixtures between (A) TrCel7A and TrCel7B, (B) TrCel7A and ReCel7B, (C) ReCel7A and TrCel7B and (D) ReCel7A and ReCel7B. Black squares correspond to the total activity in the synergy mixtures at different mole fractions of CBH. The monocomponent activity is indicated with gray circles for TrCel7A, orange triangles for ReCel7B, blue circles for ReCel7A and green stars for TrCel7B. Dashed lines indicate the sum of the activities of the monocomponents used in each mixture. Conditions: 25 °C, 10 g/L Avicel, total enzyme concentration 4 μM. The x-axis represents the fraction of CBH used in the mixture, calculated as [CBH]/([CBH+EG]). Adapted from Article III.*

These results allowed us to conclude that an EG with relatively high specificity constant and inverse maximal specific rate, in this case *Tr*Cel7B, is capable of synergizing more with a CBH compared to one where these parameters are lower (*Re*Cel7B). A second consideration derived from the synergy results is that the *Trichoderma reesei* EG-CBH system appears overall more efficient than the *Rasamsonia emersonii* one.

However, the low synergy observed in Fig. 4.11D might not reflect the real condition faced by the *Rasamsonia emersonii* cellulolytic system in nature. Although it is difficult to draw correlations between cellulases activity and fungal habitat, since there are no systematic studies available in literature, some considerations can be made. For instance, if *R. emersonii* cellulases have evolved to work at high substrate consistencies, then the evolutionary advantage of the CBM is reduced or absent, as proposed by Várnai et al. (2013), since the low water content of the environment enables physical proximity of enzymes and substrate. That could also explain why both *Re*Cel7A and *Re*Cel7B are naturally devoid of a CBM (Fig. 1.7). On the other hand, *T. reesei* could be more adapted to survive in habitats with high humidity, where the CBMs, present in both *Tr*Cel7A and *Tr*Cel7B, aid by concentrating the enzymes onto the substrate. Indeed, thermophilic ascomycetes similar to *R. emersonii* have been identified in aridland ecosystems, and are highly adapted to seasonal and day temperatures shifts (Powell et al. 2012). As such, they should not be considered as mere specialized organisms for high temperature niches (Maheshwari et al. 2000).

4.11 - The functional plasticity of *Re*Cel7B

Using a rational design approach, >50 enzyme variants of *Re*Cel7B were produced and characterized on Avicel. ^{conv}MM curves were created at 50 °C with the methods described in Article I, and the kinetic parameters $^{conv}V_{max}/E_0$ and $^{conv}K_M$ were extracted. A correlation analysis between these two parameters was performed on 35 of these variants and *Re*Cel7B, shown in Fig. 4.12. The identity and specific details on the differences between the variants will not be disclosed.

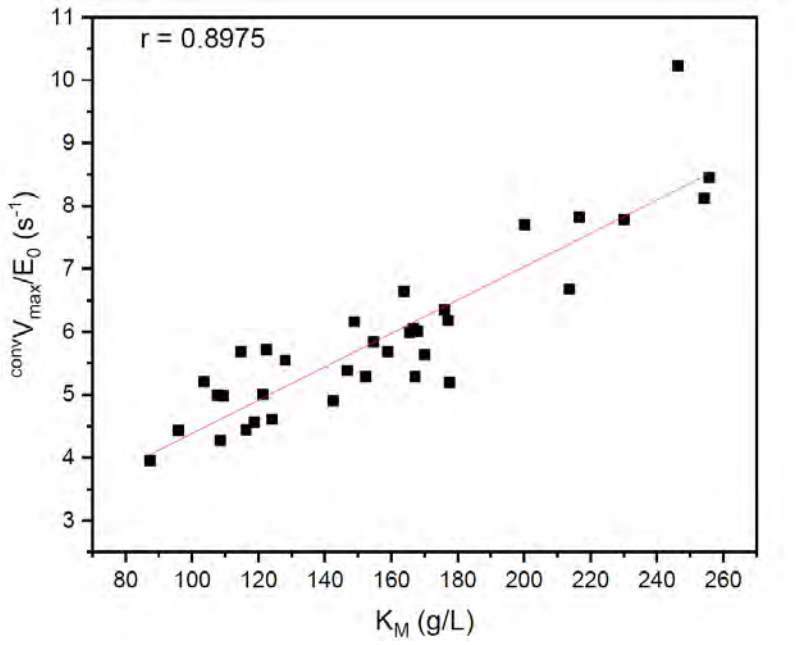


Figure 4.12. *Relationship between $^{conv}V_{max}$ and $^{conv}K_M$ in ReCel7B and 35 variants. Pearson's r correlation coefficient is indicated. The dotted line represents the best linear correlation from a least squares regression analysis. Graph created in OriginPro 2019 (OriginLab Corp., Northampton, MA, USA). Statistical analysis performed in JMP ver. 14 (SAS Institute Inc., Cary, NC, 1989-2019).*

Similar to the case of *Tr*Cel7A, *Tr*Cel7B and related variants described in chapter III, under ^{conv}MM conditions the conventional maximal specific rate has a significantly strong positive relationship (r=0.89, p<0.0001) with the affinity parameter $^{conv}K_M$ (Fig. 4.12). As a result, the specificity constant η remains very similar for all the investigated variants, shifting in an interval between 0.050 to 0.029 $gL^{-1}s^{-1}$. Possible interpretations of this relationship were mentioned in section 3.6, while here we will focus on considerations related to how to select the best enzyme for industrial conditions.

The observed positive correlation in Fig. 4.12 suggests that *Re*Cel7B has little functional

plasticity on cellulose, since every time the affinity properties are improved (K_M decreased) the turnover number concomitantly decreases, and it does not seem possible to achieve the ideal condition of a simultaneous high ${}^{conv}V_{max}$ and low ${}^{conv}K_M$. Considering the example of Fig. 4.12, when selecting suitable enzyme candidates for industrial conditions, one would be prone to select the ones with the highest maximal specific rate (at the right-hand of Fig. 4.12), and indeed if the substrate was to be plentiful (saturating conditions) then this enzyme should be preferred. In this case, the suitable variant would be the one with the maximal specific rate of 10 s^{-1}. However, this value will never be achieved in practice, because the substrate load used should be many times the K_M (in this case 5-10 times a substrate load of ~240 g/L). Thus, a more suitable parameter to select for the best enzyme is Γ_{attack}, which takes into account both how many sites are available on the cellulose surface and how quickly they get converted (chapter II). Obviously, this parameter should be taken into consideration along with other functional traits which are important but not mentioned in this section, such as the ability to synergize with other cellulases, chemical promiscuity and thermostability.

Chapter V – Enzyme-substrate interactions on the cellulose surface

This chapter focuses on kinetic interpretations of cellulases' activity when two different enzymes are mixed together. On one hand, cooperative interactions between cellulases enhance cellulose degradation. This will be described in the first part of the chapter, where synergy between *Tr*Cel7A and *Tr*Cel7B has been kinetically investigated. On the other hand, the activity of a cellulase can be inhibited by the presence of obstacles on the surface, created for instance by other enzymes that bind on the same productive sites. This is described in the second part of the chapter, where we discuss a new inhibition mechanism suitable for interfacial enzymes (article IV).

5.1 - *Endo-exo* synergy between CBHs and EGs

Enzymatic deconstruction of cellulose relies on the synergistic interaction of many different enzymes (Table 1.1). *Trichoderma reesei* cellulases are considered very efficient in such regard, and are implemented in enzymatic cocktails for saccharification processes (Bischof et al. 2016). Synergy occurs when the total enzyme activity of two (or more) enzymes mixed together is higher than the sum of the activities of the single, isolated elements (Ganner et al. 2012). The most simple and widely recognized model of synergistic enzymatic hydrolysis of cellulose is the so-called *endo-exo* model, which occurs between *exo-* (cellobiohydrolases) and *endo-* (endoglucanases) acting enzymes, and has been proposed already 40 years ago, despite the synergistic components were not identified yet (M. Wood et al. 1979). A simple explanation of the *endo-exo* synergy is the following. While *exo*-acting cellulases like *Tr*Cel7A hydrolyse primarily from the cellulose chain ends, and make several consecutive cuts before dissociating, *endo*-acting cellulases attack the internal cellulose surface (Fig. 5.1). Thus, *endo*-acting cellulases create new, free cellulose chain ends, which become new attack points for *exo*-cellulases. The *exo*-cellulases, in turn, will create more attack points for the *endo*-cellulases by disrupting the crystalline regions, thus making available accessible areas for *endo*-cellulases and therefore establishing a synergistic interaction (Boisset et al. 2001). Particularly for the *Tr*Cel7A-*Tr*Cel7B system, it is generally believed that *Tr*Cel7B can improve the activity of *Tr*Cel7A on the cellulose substrate by creating new attack points (Payne et al. 2015). It has also been discussed that *Tr*Cel7B acts by generating new exit points for stalled *Tr*Cel7A enzymes on the

cellulose surface, thus increasing the k_{off} and facilitating the release of unproductively bound CBH (Jalak et al. 2012). Whether there is a most prevalent mechanism is still unclear. Apart from kinetic interpretations, cellulase synergy models are further complicated by changes in substrate properties during conversion, as described in a number of works (reviewed in (Payne et al. 2015)).

The advantages of using GH7 endoglucanases in industrial enzyme mixtures are not only limited to their ability to boost CBHs activity. Different lignocellulose pretreatment methods can depolymerize hemicellulose, thus generating xylooligosaccharides and glucooligosaccharides which are potent inhibitors for *Tr*Cel7A (Kont et al. 2013). *Tr*Cel7B and other GH7 endoglucanases such as *Th*Cel7B and *Re*Cel7B showed side activities on hemicelluloses (Biely et al. 1991, Wang et al. 2014, Sonoda et al. 2019), thus becoming valuable also to reduce these inhibitory effects for the CBHs. Moreover, EGs have higher ability to degrade amorphous cellulose compared to CBHs (Payne et al. 2015), which becomes relevant for the degradation of pretreated biomass, where the crystallinity index is reduced, along with other factors (Karimi et al. 2016).

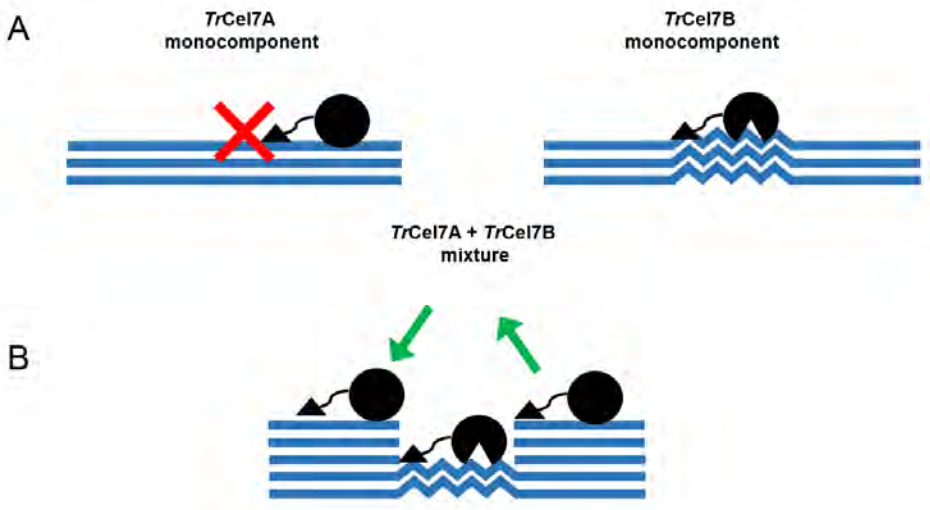


Figure 5.1. *Representation of the synergistic action of typical exo- and endo-cellulases. (A) TrCel7A and TrCel7B (black) are shown acting as monocomponent on their preferred regions on the cellulose surface (blue): a more crystalline region (parallel lines) for TrCel7A and a more amorphous one (zig-zag lines) for TrCel7B. An obstacle, such as irregularities on the cellulose surface, can block the action of the processive TrCel7A (red X symbol). (B) A mixture of TrCel7A and TrCel7B can synergize because TrCel7B creates new attack points for TrCel7A (green down-pointing arrow), and by removing obstacles for TrCel7A, thus creating new exit points for stalled enzymes (green up-pointing arrow).*

In the following, e*ndo-exo* synergism between *Tr*Cel7A and *Tr*Cel7B has been investigated using a systematic approach, based on the two steady-state kinetic models described in

chapter II, where initial rates are measured at short contact times and Avicel was used as model cellulosic substrate. Since the assays are performed at low conversion, no additional enzyme was necessary to relieve either *Tr*Cel7A or *Tr*Cel7B from product inhibition. In fact, the amount of product produced in the different assays remained in the low µM range, about one to three order of magnitude less than the reported inhibition constant values of either enzymes (Murphy et al. 2013, Teugjas et al. 2013, Olsen et al. 2016). Other types of synergism, for example between different types of *exo*-acting cellobiohydrolases, (Badino et al. 2017), inter-domain synergism between CD and CBM (Kont et al. 2016) and between non-hydrolytic enzymes and cellulases (Eibinger et al. 2014) will not be addressed here.

The next section describes unpublished results regarding a novel kinetic approach to study synergy between *Tr*Cel7A and *Tr*Cel7B. It is essentially based on creating inverse and conventional MM plots for different mixtures of *Tr*Cel7A and *Tr*Cel7B. The extracted kinetic parameters can then be compared with *Tr*Cel7A and *Tr*Cel7B acting alone on cellulose to obtain information on whether synergy is established and what are the conditions for obtaining the best improvement in parameters.

5.2 - Two kinetic approaches to study synergy between *Tr*Cel7A and *Tr*Cel7B

The ^{conv}MM and ^{inv}MM approaches described in chapter II allowed to directly compare the activity and cellulose site promiscuity of *Tr*Cel7A and *Tr*Cel7B. We used the same approaches to study *Tr*Cel7A and *Tr*Cel7B in mixtures. Material and methods for this section are found in Appendix III. Specifically, we combined *Tr*Cel7A and *Tr*Cel7B at eight different molar proportions (Table 5.1) while maintaining constant the total enzyme concentration in the different assays. For each combination, we created full ^{conv}MM and ^{inv}MM curves (Fig 5.2A and B, respectively). The data was fitted to eq. 1 and 3, respectively, to derive the apparent synergistic kinetic parameters (Table 5.2), which were then compared with the monocomponent parameters of *Tr*Cel7A and *Tr*Cel7B.

Table 5.1. *Proportions of TrCel7A and TrCel7B used, and the corresponding amount of TrCel7A present in the mixture expressed in percentage.*

Condition	Proportion of *Tr*Cel7B:*Tr*Cel7A	Mole percentage of *Tr*Cel7A (%)
1	0:1	100
2	1:100	99
3	1:50	98
4	1:10	90
5	1:4	75
6	1:1	50
7	4:1	25
8	10:1	10
9	50:1	2
10	1:0	0

Under ^{conv}MM conditions, it was clearly observed that the highest rate was found at 100% *Tr*Cel7B, and the lowest at 100% *Tr*Cel7A (compare black and red curves in Fig 5.2A). All the other activity values in the mixtures scaled with these two extremes. This was reflected in the extracted kinetic parameters (Table 5.2), which decreased linearly with the amount of *Tr*Cel7A present in the mixture (Table 5.2). This linear scaling is particularly evident when the kinetic parameters are plotted against the mole fraction of CBH used (Fig. 5.3). This linear relationship indicates that there is no synergy between *Tr*Cel7A and *Tr*Cel7B under these conditions. By only looking at the activity of *Tr*Cel7A alone (black curve in Fig. 5.2) one could be tempted to conclude that *Tr*Cel7A activity is boosted by the addition of increasing amount of *Tr*Cel7B (Fig. 5.2). However, since 100% *Tr*Cel7B corresponds to the highest value (red curve in Fig 5.2), this means that it is the increased amount of *Tr*Cel7B in the mixtures that changes the total activity, which is merely a reflection of its catalytic predominance. Since we do not find synergy under ^{conv}MM approach, we conclude that *Tr*Cel7A and *Tr*Cel7B do not react to each other's presence on the cellulose surface under these conditions. This is probably due to the high abundance of hydrolysable sites, which are not fully saturated with enzymes, therefore they do not become a limiting factor.

This conclusion is supported by the *endo-exo* synergy model proposed by Jalak and coworkers (Jalak et al. 2012). In this work, the authors study *endo-exo* synergy between *Tr*Cel7A and the GH5 endoglucanase *Tr*Cel5A on bacterial cellulose (BC). The authors propose that the rate-limiting step in cellulose hydrolysis for *Tr*Cel7A is the slow dissociation of the enzyme from the substrate, due to stalling in front of "obstacles" on the cellulose. Such obstacles were envisioned as irregularities present in BC. When the EG

is present, it can remove such obstacles effectively and thus create a synergistic interaction with *Tr*Cel7A. A similar situation can be imagined for the condition in Fig. 5.2A, since Avicel contains both of amorphous and crystalline cellulose (section 1.3). The aforementioned study also notes that synergy is significant only at high enzyme/substrate ratios (Jalak et al. 2012). In the opposite case of low enzyme/substrate ratios, the new chain ends generated by the EG are never encountered by *Tr*Cel7A. To obtain synergy by removing stalled *Tr*Cel7A, it is necessary that the new EG-generated attack points are in very close proximity to *Tr*Cel7A, possibly on the same cellulose chain where *Tr*Cel7A is acting upon (Jalak et al. 2012). Thus, in the conditions shown in Fig. 5.2A, the probability of encountering new EG-generated sites is very low, and synergy cannot be observed.

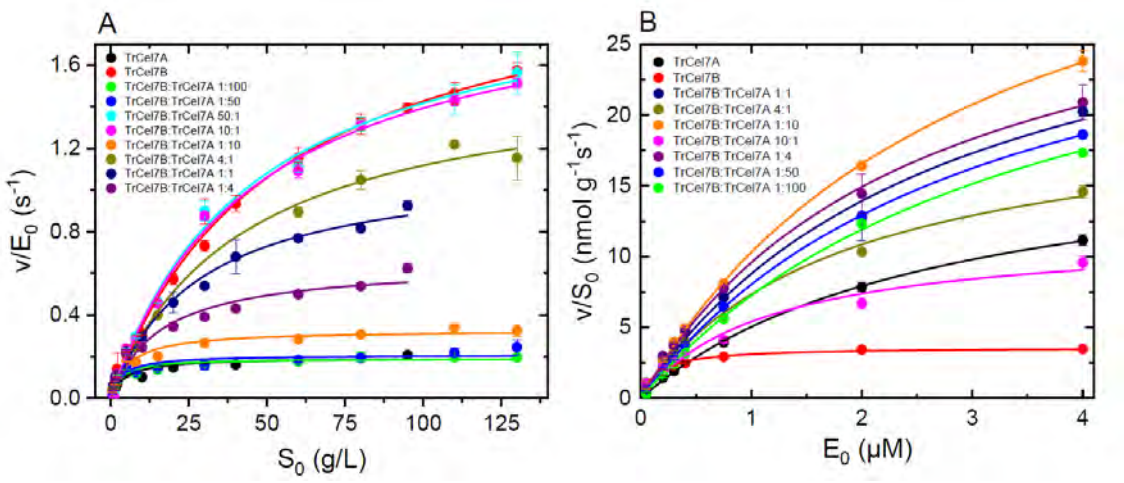


Figure 5.2. *Reaction rates of TrCel7A and TrCel7B acting alone or in mixtures on Avicel at two different conditions. (A) Conventional Michaelis-Menten approach (B) Inverse Michaelis-Menten approach. Colours vary according to the legend, and the proportion of TrCel7B and TrCel7A in the mixture vary according to Table 5.1. Continuous curves are fitting to Eq. 1 and 3 for panel A and B respectively. Error bars are standard deviations from triplicates. Experimental conditions: enzyme concentration 25 nM, Avicel load 5-130 g/L for panel A, enzyme concentration 0.05-4 μM, Avicel load 10 g/L for panel B. Figure and data analysis was performed in OriginPro 2019 (OriginLab Corp., Northampton, MA, USA).*

Table 5.2. *Kinetic parameters extracted from the conventional Michalis Menten and inverse Michaelis Menten plots shown in Fig. 5.2. Standard errors are indicated.*

Condition	Mole percentage of *Tr*Cel7A	^{conv}MM		^{inv}MM		Kinetic substrate accessibility
		$^{conv}\mathbf{V_{max}/E_0}$ (s^{-1})	$^{conv}\mathbf{K_M}$ (gL^{-1})	$^{inv}\mathbf{V_{max}/S_0}$ (nmol g^{-1} s^{-1})	$^{inv}\mathbf{K_M}$ (µM)	$\mathbf{\Gamma_{attack}}$ (µmol g^{-1})
1	100%	0.20 ± 0.01	4.5 ± 1.0	18.5 ± 0.55	2.7 ± 0.2	0.095 ± 0.006
2	99%	0.19 ± 0.008	2.8 ± 0.6	33 ± 1.6	3.5 ± 0.30	0.173 ± 0.010
3	98%	0.21 ± 0.01	3.4 ± 1.0	33 ± 0.4	3.1 ± 0.06	0.159 ± 0.009
4	90%	0.33 ± 0.01	5.7 ± 1.0	42.0 ± 0.9	3.1 ± 0.1	0.129 ± 0.006
5	75%	0.65 ± 0.04	16 ± 3	33.6 ± 1.4	2.5 ± 0.2	0.0518 ± 0.0040
6	50%	1.14 ± 0.07	28 ± 4	33.5 ± 3.8	2.8 ± 0.6	0.0295 ± 0.0040
7	25%	1.59 ± 0.08	42 ± 6.0	20.8 ± 1.6	1.8 ± 0.30	0.0131 ± 0.0012
8	10%	2.03 ± 0.09	46 ± 5	11.7 ± 1.1	1.2 ± 0.3	0.006 ± 0.0006
9	2%	2.06 ± 0.09	45 ± 5	-	-	-
10	0%	2.17 ± 0.07	53 ± 4	3.6 ± 0.10	0.2 ± 0.02	0.0017 ± 0.0001

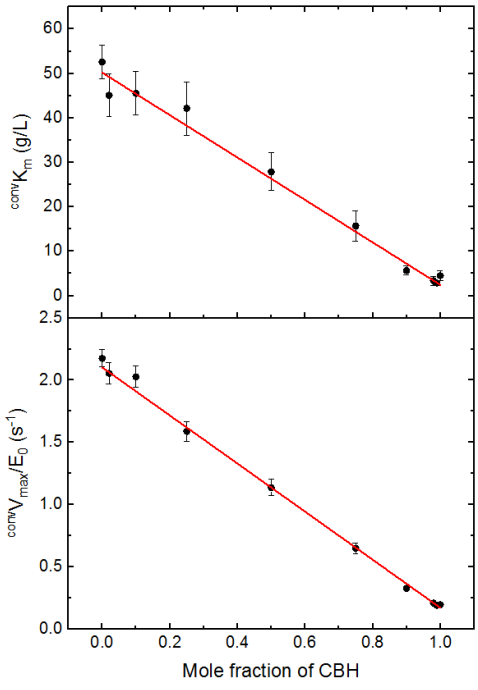


Figure 5.3. *The kinetic parameters $^{conv}V_{max}$ and $^{conv}K_M$ decrease linearly with increased amount of TrCel7A in the mixture. Data points and standard errors are taken from Table 5.2. Red line corresponds to a linear fit of the data points. Figure created with OriginPro 2019 (OriginLab Corp., Northampton, MA, USA).*

Considering the importance of a suitable enzyme/substrate ratio to obtain synergy, we studied synergy mixtures of *Tr*Cel7A and *Tr*Cel7B in a second condition, which is opposite from the conventional situation described before of enzymes saturated with substrate. This new condition investigated is the inverse MM approach (chapter II), applied for synergy mixtures. More specifically, we measured enzyme activity using a fixed, low substrate load (in this case Avicel 1%) and different total enzyme concentrations (see Appendix III for more details on materials and methods used). To measure activity in synergy mixtures, the total enzyme concentrations used were simply modified with different proportions of *Tr*Cel7A and *Tr*Cel7B (Table 5.1). The resulting curves for the different mixtures and monocomponents are shown in Fig. 5.2B. Apart from condition 8, synergy was present in all cases, as shown by the increase in activity compared to the monocomponents, and was measurable even the amount of *Tr*Cel7B added was very small (condition 2, Table 5.2).
To facilitate the interpretation of the results, each curve of Fig. 5.2B can be considered one "enzyme variant" rather than a mixture of two enzymes. As such, the parameters listed in Table 5.2 become comparisons between different enzymes.
As for the inverse maximal specific rate, the composition which gave the highest increase compared to *Tr*Cel7A and *Tr*Cel7B was condition 4, containing 90% *Tr*Cel7A and having an increase of $^{inv}V_{max}/S_0$ about 2.3-fold compared to *Tr*Cel7A acting alone (Table 5.2). This value remains still higher, but rather similar, for conditions containing 99%, 98%, 75%, 50% *Tr*Cel7A, while it worsen at 25% or 10% *Tr*Cel7A.
Combining the conventional and inverse maximal rates of Table 5.2 with eq. 4 allows to derive the attack site density values for the different combinations, $\Gamma_{attack.}$ As seen in chapter II, this value reflects the ability of an enzyme to locate and attack numerous sites on the cellulose surface, and it takes into account how quickly these sites get converted. As such, Γ_{attack} derived from synergy mixtures gives some indication on which combination is optimal to increase the overall efficiency on cellulose.
Combination 2, 3 and 4 containing respectively 99, 98 and 90% of *Tr*Cel7A, appear to be the best proportions to obtain the highest values of Γ_{attack} of 0.173, 0.159 and 0.129 µmol/g respectively (Table 5.2). Notably, by only adding 1% of *Tr*Cel7B (condition 2) the substrate accessibility increases 2-fold compared to *Tr*Cel7A acting alone. On the other hand, Γ_{attack} worsen at conditions containing 75% *Tr*Cel7A or less.
These results suggest that the best proportions of *Tr*Cel7A and *Tr*Cel7B to obtain efficient synergy under substrate-saturating conditions are in the range of 90-99% *Tr*Cel7A. Although this combination might not reflect the optimal condition on other substrates (Kostylev et al. 2012) we note that already in 1985 it was proposed by Henrissat *et al.* that the optimal proportion of *endo-exo* cellulases has to favour the *exo*-enzyme (Henrissat et al. 1985), and it is also consistent with what is found in *T. reesei* secretome, where the

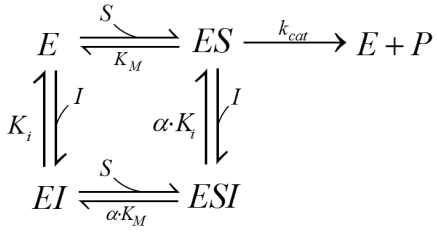
CBHs account for 80% of the total of the total cellulase protein secreted (Rosgaard et al. 2007) (see also Table 1.1).
Overall, these results indicate that not only the ^{inv}MM approach can be applied to investigate synergy mixtures, but it is also the suitable experimental condition where synergy can be measured. As described before for the opposite case, under inverse MM conditions the enzyme substrate-ratio is high, and the amount of attack sites can become saturated with enzymes (chapter II). This makes it possible for *Tr*Cel7A and *Tr*Cel7B to positively interact with each other on the substrate surface. As such, the presence of as little as 1% *Tr*Cel7B most likely facilitates dissociation of stalled *Tr*Cel7A, thus ensuring synergism (Jalak et al. 2012).

5.3 - Cellulases inhibition on the substrate surface

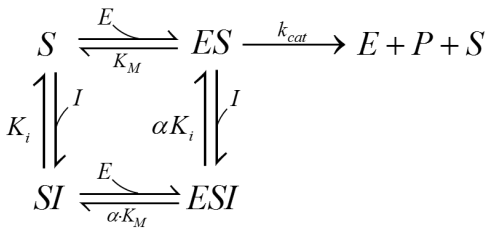
Apart from synergy, the presence of two or more enzymes acting simultaneously on the cellulose surface can have negative consequences on the overall efficiency of cellulose degradation. In some cases, non-synergistic interactions occur, as seen in the cases of EGs mixed together (Hu et al. 2015), or mixtures of two similar CBHs (Section 3.7). Competition in binding has also been reported (Kostylev and Wilson 2012).
In the following, we aim at exploring negative interactions on the cellulose surface by investigating the kinetic effects of *surface site inhibitors*. This term was utilized in Article IV to describe molecules that reduce the reactivity of the cellulose substrate by occupying potential attack sites for the cellulases by binding onto the surface. This negatively affects cellulases activity and the result is a measurable inhibitory effect, even though the inhibitor molecules do not directly interfere with the enzyme active site, as seen in normal cases of enzyme inhibition.
So far, there has been no attempt to model this type of enzyme inhibition using the recently developed kinetic tools available from interfacial enzymology outlined in chapter II. However, we note that a kinetic understanding of this type of inhibition could be useful to assess the potency of different types of surface site inhibitor molecules and to have a way to express quantitatively inhibition strength and mechanism. Notably, these analysis are fundamentally based on enzyme activity measurements and use standard equipment available in most enzymology laboratories, compared to more laborious (although insightful) techniques such as atomic force microscopy measurements (Igarashi et al. 2011) and single-molecule fluorescence imaging (Jung et al. 2013).

In Article IV, we kinetically describe this more unfamiliar type of inhibition, called *surface site inhibition*, where we focus on the fact cellulases activity can be affected by a number of molecules binding to the cellulose surface. As discussed below, we experimentally observe that catalytically inactive cellulases can behave as surface-site inhibitors, and several other potential surface site inhibitors could exist, and will be mentioned at the end of this section.

A surface-site inhibitor can be imagined as a molecule that lowers the accessibility of the enzyme to the substrate and/or lowers the reactivity of the enzyme-substrate complex. Taken as example the activity of *Tr*Cel7A, the former aspect is related to the creation of a physical obstacle onto the cellulose surface, which reduces the amount of available sites for catalysis. The latter element is related to the fact that a surface inhibitor could create obstacles when *Tr*Cel7A is already complexed with substrate, and thus hinder its processive movement.



Scheme 5.1. Conventional inhibition



Scheme 5.2. Surface-site inhibition

$$V_{ss} = \frac{E_0 \cdot S_0 \cdot k_{cat}}{S_0\left(1+\frac{I_0}{\alpha K_I}\right)+K_M\left(1+\frac{I_0}{K_I}\right)} \quad \text{(Eq. 5.1)}$$

$$V_{ss} = \frac{E_0 \cdot S_0 \cdot k_{cat}}{E_0\left(1+\frac{I_0}{\alpha K_I}\right)+K_M\left(1+\frac{I_0}{K_I}\right)} \quad \text{(Eq. 5.2)}$$

The conventional case used to describe bulk cellulases inhibition follows Scheme 5.1 (Nelson et al. 2008, Olsen et al. 2016) and it is shown as comparison with the surface-site inhibition mechanism of Scheme 5.2, introduced here. As in the conventional case, surface-site inhibition mechanisms can be always considered mixed type inhibition, with a competitive and uncompetitive component weighing differently according to the α factor of Scheme 5.2 (Cornish-Bowden 2013). The Scheme 5.2 corresponds to the inverse Michaelis Menten approach described by Kari *et al.* (Kari et al. 2017), but the reaction scheme is further expanded to take into account the presence of a surface-site inhibitor molecule. When α is very large, the surface-site inhibition can be considered pure competitive; when α is very small (close to 0) the surface-site inhibition model becomes

nearly identical to pure uncompetitive. Non-competitive surface-site inhibition can be seen a special case where α=1.

To experimentally demonstrate the applicability of the surface-site inhibition, we used the processive cellobiohydrolases *Tr*Cel7A and *Tr*Cel6A from *T. reesei* and used as surface-site inhibitor the catalytically inactive variants of these. In the following, I will only describe the results concerning *Tr*Cel7A, although in Article IV we explore the inhibition of *Tr*Cel6A as well. Specifically, *Tr*Cel7A$_{E212Q}$ and *Tr*Cel6A$_{D221N}$ are enzyme variants where we modified by site-directed mutagenesis one of the residues involved in the *retaining* and *inverting* mechanism, respectively. The variants were considered suitable inhibitor candidates because they possessed similar binding to their corresponding active version (Fig. 5.4 and Table 5.3), but no measurable activity.

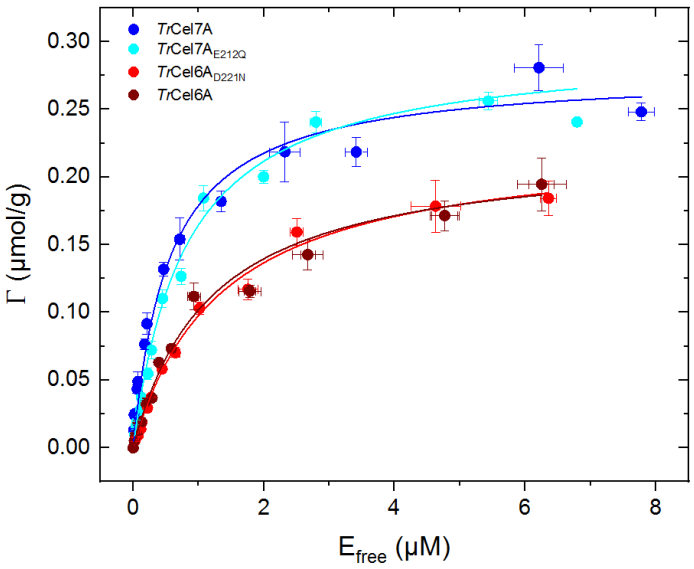


Figure 5.4. *Binding isotherms of TrCel7A (blue), TrCel7A$_{E212Q}$ (cyan), TrCel6A (dark red) and TrCel6A$_{D221N}$ (red), using Avicel as substrate. Symbols represent experimental data while solid lines are the best-fit to a one-site Langmuir binding isotherm. All experiments were performed at 25 °C. Error bars are standard deviations for triplicate measurements. Figure adapted from Article IV.*

Table 5.3. *Langmuir adsorption parameters for TrCel7A, TrCel7B and inactive variants TrCel7A, $TrCel7A_{E212Q}$ and $TrCel6A_{D221N}$, from data shown in Fig. 5.4. Standard errors are included.*

Protein	Γ_{max} (µmol/g)	K_d (µM)
*Tr*Cel7A	0.27 ±0.01	0.46 ±0.06
Tr$Cel7A_{E212Q}$	0.29 ±0.01	0.79 ±0.08
*Tr*Cel6A	0.22 ±0.01	1.13 ±0.14
Tr$Cel6A_{D221N}$	0.23 ±0.01	1.39 ±0.11

Moreover, both *Tr*$Cel7A_{E212Q}$ and *Tr*$Cel6A_{D221N}$ display reversible adsorption on cellulose, thus excluding the possibility of irreversible inhibition mechanisms (see Article IV). We created ^{inv}MM-type curves of *Tr*Cel7A either alone or in the presence of a fixed amount of surface site inhibitor (*i.e.* inactive enzyme) at different concentrations (Fig. 5.5). Notably, this condition of enzyme excess ensures that all the attack sites can be saturated with enzyme. As seen in chapter II and the previous section, the total amount of accessible sites is preserved during short initial conversions (Kari et al. 2017).

Six conditions were made for *Tr*Cel7A in the presence of *Tr*$Cel7A_{E212Q}$ (Fig. 5.5A) and five for *Tr*Cel7A with *Tr*$Cel6A_{D221N}$ (Fig. 5.5B). The obtained experimental data was fitted to eq. 5.2 using global non-linear regression analysis for the entire data set. Each time, we globally fitted simplified equations corresponding to four surface-site inhibition types (mixed, competitive, noncompetitive, uncompetitive), then, the most likely inhibition model was selected using the Akaike Information Criterion (Akpa et al. 2011) but also by simple visual inspection of Lineweaver-Burk type plots (more details are provided in the supplementary material of Article IV). The Akaike criterion provides an objective analysis of which model fits best the data with the least amount of parameters, and has been previously used to evaluate conventional inhibition mechanisms in *Tr*Cel7A (Olsen et al. 2016). In Table 5.4, the inhibition model that best describes the data is reported, along with the kinetic parameters derived from the best fit.

Results in Fig. 5.5 clearly shows that the presence of either *Tr*$Cel7A_{E212Q}$ or *Tr*$Cel6A_{D221N}$ was inhibitory for *Tr*Cel7A activity in all cases, indicating that the molecules are reversible obstacles for *Tr*Cel7A. We were expecting a decrease in activity for *Tr*Cel7A in the presence of *Tr*$Cel7A_{E212Q}$ not only because the enzymes show similar cellulose binding, but because they probably also retained similar attack site preference. Inhibition of *Tr*Cel7A in the presence *Tr*$Cel6A_{D221N}$ was also somewhat expected, despite the non-reducing end specificity of *Tr*Cel6A, considering that *Tr*Cel7A and *Tr*Cel6A are able to synergize (Badino et al. 2017). Synergy implies the possibility that a fraction of cellulose attack sites

are shared between the enzymes (see also previous section). Nonetheless, inhibition in this case could have also be envisioned since the inactive enzyme is a considerable size compared to a cellobiose unit, thus could physically cover potential cellulose attack sites and create steric hindrance for *Tr*Cel7A molecules.

As for the inhibition strength and mechanism, *Tr*Cel7A in the presence of *Tr*Cel7A$_{E212Q}$ showed surface site mixed type inhibition (Fig. 5.5A and Table 5.4). The inhibition constant K_i corresponds to 0.74 µM (Table 5.3). As mentioned before, in mixed-type surface site inhibition, both a competitive and an uncompetitive component co-exist. Nonetheless, the competitive component seems to be the most prevalent, since $\alpha \approx 5$ (Table 5.3). On the other hand, in the case of *Tr*Cel7A in the presence of *Tr*Cel6A$_{D221N}$ (Fig. 5.5B), the surface-site inhibition mechanism observed is of the competitive type, with K_i value of 0.47 µM (Table 5.4).

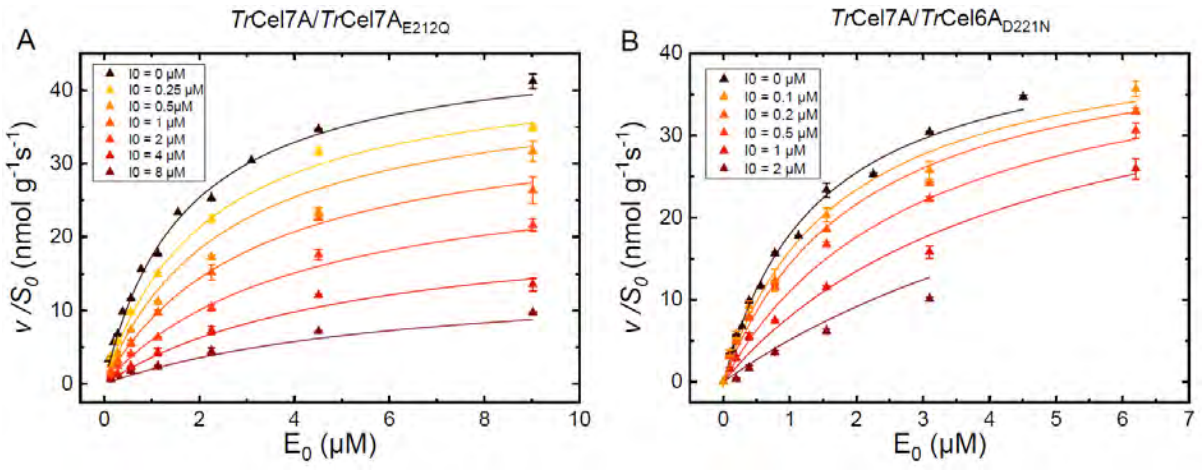


Figure 5.5. *Inverse inhibition kinetics. Steady-state reaction rates for the hydrolysis of Avicel to soluble sugars under different inhibitor concentrations. (A) Inhibition of TrCel7A by TrCel7A$_{E212Q}$. (B) Inhibition of TrCel7A by TrCel6A$_{D221N}$. The legend indicates the (constant) inhibitor concentration used in each condition. Symbols are experimental data points and lines are the best-fit results from the global, non-linear regression analysis to eq. 5.2 as described in the text. Conditions: Avicel 2 g/L, 25 °C. Adapted from Article IV.*

Table 5.4. *Kinetic parameters and surface site inhibition mechanism and strength for TrCel7A in the presence of either TrCel7A$_{E212Q}$ or TrCel6A$_{D221N}$, from the best fit shown in Fig 5.5. Standard errors are included.*

Enzyme / Inhibitor	$^{inv}V_{max}/S_0$ (nmol g^{-1} s^{-1})	$^{inv}K_M$ (µM)	K_i (µM)	α	Surface-site inhibition mechanism
*Tr*Cel7A/ *Tr*Cel7A$_{E212Q}$	46.8 ±1.1	1.69 ±0.1	0.74 ±0.08	5.11 ±1.37	Mixed
*Tr*Cel7A/ *Tr*Cel6A$_{D221N}$	43.7 ±1.3	1.43 ±0.11	0.47 ±0.04	-	Competitive

These results can be interpreted on the basis of the known kinetic and binding properties of *Tr*Cel7A. The enzyme action on the substrate surface is limited to a one-dimensional movement, that is, the enzyme usually starts from the reducing end of a cellulose chain and processively sweeps along one direction to hydrolyse the glyosidic bonds. This makes *Tr*Cel7A, as well as other interfacial enzymes, more vulnerable to obstacles on the surface. This phenomenon has been documented in a number of works (Jalak and Väljamäe 2010, Kurašin and Väljamäe 2011, Cruys-Bagger et al. 2012), where *Tr*Cel7A is described being unproductively stuck behind obstacles, until either enzyme (or obstacle) dissociate. Moreover, Igarashi *et al.* performed detailed AFM studies on *Tr*Cel7A and *Tr*Cel7A$_{E212Q}$ and showed that the inactive enzyme cannot perform any processive movement but rather binds to the cellulose surface, in this way becoming a reversible obstacle for *Tr*Cel7A (Igarashi et al. 2009, Igarashi et al. 2011).

These observations can be analysed in light of the inhibition studies of Fig. 5.5A. The surface-site inhibition shown in the presence of *Tr*Cel7A$_{E212Q}$ is mixed type, thus *Tr*Cel7A$_{E212Q}$ is both an obstacle for the free enzyme and for the ES complex (schematically illustrated in Fig. 5.6A). The former aspect refers to the fact that the inhibitor binds to cellulose equally strongly as *Tr*Cel7A (Table 5.3), and reduces the amount of attack sites available for *Tr*Cel7A on the surface, when the enzyme is still free. However, *Tr*Cel7A$_{E212}$ also acts as obstacle for the ES complex. This can be envisioned as *Tr*Cel7A, once bound and threaded a cellulose chain into the active site, encounters obstacles for the processive movement due to *Tr*Cel7A$_{E212Q}$ molecules bound (Fig. 5.6A). Thus, the uncompetitive component of the surface site inhibition can be interpreted as *Tr*Cel7A forming an unproductive complex with *Tr*Cel7A$_{E212Q}$ and the substrate. This is indeed supported by the aforementioned AFM studies that also shows that the immobility of *Tr*Cel7A give rise to "traffic jams" on the cellulose surface, at high enzyme concentrations as in the case explored here (Igarashi et al. 2011).

Interestingly, inhibition of *Tr*Cel7A switched to pure competitive type when the inhibitor is *Tr*Cel6A$_{D221N}$ (Fig. 5.5B) Even though this surface-site inhibition type is rather different from the mixed type case of *Tr*Cel7A with *Tr*Cel7A$_{E212Q}$, we observe that *Tr*Cel6A$_{D221N}$ is a less strong inhibitor for *Tr*Cel7A (Table 5.4). Moreover, *Tr*Cel6A$_{D221N}$ bind less strongly than *Tr*Cel7A on cellulose (Table 5.3), thus the main hindrance faced by *Tr*Cel7A in the presence of this inhibitor is the apparent decrease in cellulose attack sites (Fig. 4.6B). The uncompetitive component of the surface site inhibition is practically zero, thus turning a potentially mixed-type inhibition to a competitive type. Such surface-site inhibition type also implies that the inhibitory effect of *Tr*Cel6A$_{D221N}$ can be eventually overcome by increasing the amount of *Tr*Cel7A.

An interesting structural consideration can be done by comparing the inhibition between *Tr*Cel7A with *Tr*Cel7A$_{E212Q}$ where the uncompetitive component is more prevalent than the case *Tr*Cel7A with *Tr*Cel6A$_{D221N}$. It has been previously shown that *Tr*Cel7A binds more strongly that *Tr*Cel6A on cellulose ((Badino et al. 2017) and Table 5.3). Although binding to cellulose is determined by a number of factors, it is generally believed that the stronger binding of *Tr*Cel7A comes from the additional amount of substrate binding sites in the active site tunnel compared to *Tr*Cel6A (Payne et al. 2015). Thus, inhibitors with higher substrate affinity (*Tr*Cel7A$_{E212Q}$) occupy more sites on the cellulose surface, which could increase the strength of the uncompetitive component.

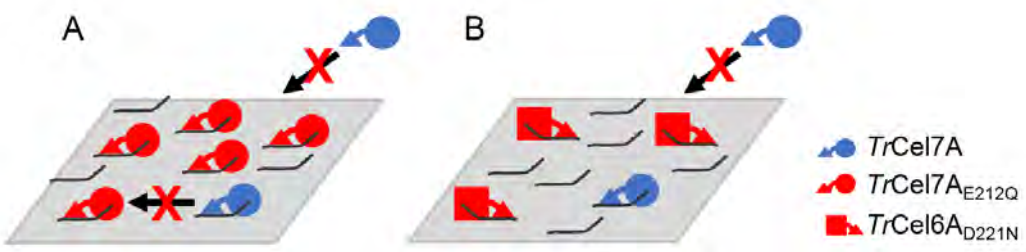


Figure 5.6. *Graphical interpretation of the surface site inhibition mechanisms shown in Fig. 5.5. (A) Mixed surface site inhibition of TrCel7A (blue) by TrCel7A$_{E212Q}$ (red). The enzyme activity is hindered both with a reduction of accessible sites (black lines) and because of stalling while in complex with the substrate (B) Competitive surface site inhibition of TrCel7A (blue) by TrCel6A$_{D221N}$ (red). The enzyme activity is hindered by a reduction of accessible sites.*

The potential of the surface-site inhibition kinetic is yet to be fully explored. In principle, this approach can be used to understand the kinetic modifications of cellulases in the presence of a number of different molecules that bind to the substrate surface and interfere with cellulases activity. Such potential surface-site inhibitor candidates could be hemicelluloses such as xylan, which has been shown to bind to cellulose (Busse-Wicher et al. 2014) or the complex lignin polymer, which remains bound to cellulose in the pretreatment step prior to enzyme addition (Pareek et al., 2013; Kumar and Wyman, 2014). However, the role that lignin plays in enzymatic hydrolysis is very complex and lignin and its derivates appear to have a dual role of both inhibitors and deactivators (Kim et al. 2011). Moreover, surface site inhibition kinetics could also shed light on the disputed role of surfactants on cellulases activity (Eriksson et al. 2002). Lastly, surface-site inhibition could be also useful to gain deeper insights into synergy, which in light of this chapter can be seen as a combination of both promoting and unfavourable contributions.

Concluding remarks

In this thesis, a detailed kinetic, structural and biochemical characterization of fungal GH7 cellulases has been performed. These enzymes are commonly employed by lignocellulose-degrading fungi to attack the plant material and obtain energy. In industrial bioconversions, they are widely implemented in biorefineries to produce sustainable fuels and chemicals. The two main subgroups in GH7 are cellobiohydrolases (CBHs) and endoglucanases (EGs). We use the enzymes *Tr*Cel7A and *Tr*Cel7B from *Trichoderma reesei* as representative members of the former and latter subgroup, respectively. Two steady-state models allowed to describe the kinetic differences between *Tr*Cel7A and *Tr*Cel7B. The EG *Tr*Cel7B displays much higher turnover frequency compared to *Tr*Cel7A under enzyme-saturating conditions. While this is an indication of catalytic superiority, *Tr*Cel7B simultaneously shows limited ability to find numerous hydrolysable sites (or attack sites) on cellulose, which is reflected in a low value of the kinetic substrate accessibility parameter. On the other hand, *Tr*Cel7A is more promiscuous on the cellulose surface due to its ability to hydrolyse about half the sites the enzyme binds to. The site selectivity of the EGs appears to be a major limitation for their activity improvement.

These kinetic differences are ascribed to structural differences between GH7 CBHs and EGs. Four flexible loops covering the substrate-binding tunnel in the CBHs are absent in the EGs. We directly investigated the functional roles of these loops by systematic deletions in *Tr*Cel7A and measuring the changes in kinetic parameters, binding and activity on different cellulosic substrates. This analysis allowed us to conclude that the loop denoted as B2 is a key determinant for the differences between CBH- or EG-like activity. Among the variants investigated, ΔB2-3 was the one with the most pronounced intermediate characteristics between *Tr*Cel7A and *Tr*Cel7B.

In industrial settings, several challenges are encountered by cellulases to perform saccharification. One of the main hurdles we observed is a temperature-induced desorption of cellulases from their substrate. With a protein engineering strategy called rational design, we improved the activity of the thermostable CBH *Re*Cel7A via the addition of a CBM domain from family 1, which facilitates cellulose targeting. Similar strategies appear to be useful to improve industrial performance of thermostable EGs. However, we also observe an interdependence between high affinity and low turnover which limits the functional plasticity of the EGs.

Structural information is fundamental to understand activity, affinity and thermostability properties of cellulases. This is of particular relevance for rational design

strategies, where the cause-effect relationship of mutations is explored. We solve the crystal structure of the highly thermophilic EG from *Rasamsonia emersonii*, *Re*Cel7B at 2.4 Å resolution. Detailed structural and kinetic characterization of *Re*Cel7B in comparison with homologous EGs allowed us to identify a number of amino acidic interactions which stabilize and rigidify loop regions, and elucidated the role of glycosylations as stability promoting elements. High structural stabilization comes at the price of reduced substrate binding, which limits the performance of *Re*Cel7B compared to the mesophilic ortholog *Tr*Cel7B.

Positive interactions between CBHs and EGs on cellulose were investigated with a kinetic approach based on creating enzymes mixtures of *Tr*Cel7A and *Tr*Cel7B and studying the effects under conditions of either substrate or enzyme excess. The results showed that the two enzymes display high synergy only when the attack sites are saturated with enzymes, and the optimal synergy condition is found at high proportions (90-99%) of *Tr*Cel7A.

On the other hand, decrease in activity due to surface-site inhibition occurs when *Tr*Cel7A is mixed with catalytically inactive cellulases. These molecules hamper the activity of *Tr*Cel7A with different mechanisms and inhibition strengths, in a manner highly dependent on the cellulose binding strength of the inhibitor.

Overall, this work provides new insights on the structure-function differences between GH7 cellulases, which paves the way for further development of this industrially relevant enzyme family.

Reference list

Abraham, M. J., T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl (2015). "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers." SoftwareX **1**: 19-25.

Agger, J. W., P. K. Busk, B. Pilgaard, A. S. Meyer and L. Lange (2017). "A new functional classification of glucuronoyl esterases by peptide pattern recognition." Frontiers in microbiology **8**: 309.

Akpa, O. M. and E. I. Unuabonah (2011). "Small-Sample Corrected Akaike Information Criterion: An appropriate statistical tool for ranking of adsorption isotherm models." Desalination **272**(1-3): 20-26.

Amore, A., B. C. Knott, N. T. Supekar, A. Shajahan, P. Azadi, P. Zhao, L. Wells, J. G. Linger, S. E. Hobdey, T. A. Vander Wall, T. Shollenberger, J. M. Yarbrough, Z. Tan, M. F. Crowley, M. E. Himmel, S. R. Decker, G. T. Beckham and L. E. Taylor, 2nd (2017). "Distinct roles of N- and O-glycans in cellulase activity and stability." Proc Natl Acad Sci U S A **114**(52): 13667-13672.

Artzi, L., E. A. Bayer and S. Morais (2017). "Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides." Nature Reviews Microbiology **15**(2): 83.

Atreya, M. E., K. L. Strobel and D. S. Clark (2016). "Alleviating product inhibition in cellulase enzyme Cel7A." Biotechnol Bioeng **113**(2): 330-338.

Badino, S. F., S. J. Christensen, J. Kari, M. S. Windahl, S. Hvidt, K. Borch and P. Westh (2017). "Exo-exo synergy between Cel6A and Cel7A from Hypocrea jecorina: Role of carbohydrate binding module and the endo-lytic character of the enzymes." Biotechnology and bioengineering **114**(8): 1639-1647.

Badino, S. F., J. Kari, S. J. Christensen, K. Borch and P. Westh (2017). "Direct kinetic comparison of the two cellobiohydrolases Cel6A and Cel7A from Hypocrea jecorina." Biochimica et Biophysica Acta -Proteins Proteomics **1865**(12): 1739-1745.

Bar-Even, A., E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik and R. Milo (2011). "The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters." Biochemistry **50**(21): 4402-4410.

Becker, D., C. Braet, H. Brumer, M. Claeyssens, C. Divne, B. R. Fagerstrom, M. Harris, T. A. Jones, G. J. Kleywegt and A. Koivula (2001). "Engineering of a glycosidase Family 7 cellobiohydrolase to more alkaline pH optimum: the pH behaviour of Trichoderma reesei Cel7A and its E223S/A224H/L225V/T226A/D262G mutant." Biochemical Journal **356**(1): 19-30.

Beckham, G. T., Z. Dai, J. F. Matthews, M. Momany, C. M. Payne, W. S. Adney, S. E. Baker and M. E. Himmel (2012). "Harnessing glycosylation to improve cellulase activity." Current opinion in biotechnology **23**(3): 338-345.

Beckham, G. T., J. F. Matthews, Y. J. Bomble, L. Bu, W. S. Adney, M. E. Himmel, M. R. Nimlos and M. F. Crowley (2010). "Identification of amino acids responsible for processivity in a Family 1 carbohydrate-binding module from a fungal cellulase." The Journal of Physical Chemistry B **114**(3): 1447-1453.

Biely, P., M. Vršanska and M. Claeyssens (1991). "The endo-1, 4-β-glucanase I from Trichoderma reesei: Action on β-1, 4-oligomers and polymers derived from d-glucose and d-xylose." European journal of biochemistry **200**(1): 157-163.

Bischof, R. H., J. Ramoni and B. Seiboth (2016). "Cellulases and beyond: the first 70 years of the enzyme producer Trichoderma reesei." Microbial cell factories **15**(1): 106.

Bissaro, B., A. Varnai, Å. K. Røhr and V. G. Eijsink (2018). "Oxidoreductases and reactive oxygen species in conversion of lignocellulosic biomass." Microbiol. Mol. Biol. Rev. **82**(4): e00029-00018.

Bodenheimer, A. M. and F. Meilleur (2016). "Crystal structures of wild-type Trichoderma reesei Cel7A catalytic domain in open and closed states." FEBS letters **590**(23): 4429-4438.

Boisset, C., C. Pétrequin, H. Chanzy, B. Henrissat and M. Schülein (2001). "Optimized mixtures of recombinant Humicola insolens cellulases for the biodegradation of crystalline cellulose." Biotechnology bioengineering **72**(3): 339-345.

Borisova, A. S., E. V. Eneyskaya, S. Jana, S. F. Badino, J. Kari, A. Amore, M. Karlsson, H. Hansson, M. Sandgren and M. E. Himmel (2018). "Correlation of structure, function and protein dynamics in GH7 cellobiohydrolases from Trichoderma atroviride, T. reesei and T. harzianum." Biotechnology for biofuels **11**(1): 5.

Busse-Wicher, M., T. C. Gomes, T. Tryfona, N. Nikolovski, K. Stott, N. J. Grantham, D. N. Bolam, M. S. Skaf and P. Dupree (2014). "The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of Arabidopsis thaliana." The Plant Journal **79**(3): 492-506.

Cannella, D. and H. Jørgensen (2014). "Do new cellulolytic enzyme preparations affect the industrial strategies for high solids lignocellulosic ethanol production?" Biotechnology bioengineering **111**(1): 59-68.

Chen, L., M. R. Drake, M. G. Resch, E. R. Greene, M. E. Himmel, P. K. Chaffey, G. T. Beckham and Z. Tan (2014). "Specificity of O-glycosylation in enhancing the stability and cellulose binding affinity of Family 1 carbohydrate-binding modules." Proceedings of the National Academy of Sciences **111**(21): 7612-7617.

Christensen, S. J., J. Kari, S. F. Badino, K. Borch and P. Westh (2018). "Rate-limiting step and substrate accessibility of cellobiohydrolase Cel6A from Trichoderma reesei." The FEBS journal **285**(23): 4482-4493.

Christensen, S. J., K. B. R. M. Krogh, N. Spodsberg, K. Borch and P. Westh (2019). "A biochemical comparison of fungal GH6 cellobiohydrolases." Biochemical Journal **476**(15): 2157-2172.

Cornish-Bowden, A. (2013). Fundamentals of enzyme kinetics, John Wiley & Sons.

Coughlan, M. P., M. A. Folan, A. Mchale, P. J. Considine and A. P. Moloney (1984). "The Talaromyces emersonii enzyme system." Applied Biochemistry and Biotechnology **9**(4): 331-332.

Cragg, S. M., G. T. Beckham, N. C. Bruce, T. D. Bugg, D. L. Distel, P. Dupree, A. G. Etxabe, B. S. Goodell, J. Jellison and J. E. McGeehan (2015). "Lignocellulose degradation mechanisms across the Tree of Life." Current Opinion in Chemical Biology **29**: 108-119.

Crooks, G. E., G. Hon, J.-M. Chandonia and S. E. Brenner (2004). "WebLogo: a sequence logo generator." Genome research **14**(6): 1188-1190.

Cruys-Bagger, N., J. Elmerdahl, E. Praestgaard, K. Borch and P. Westh (2013). "A steady-state theory for processive cellulases." FEBS J **280**(16): 3952-3961.

Cruys-Bagger, N., J. Elmerdahl, E. Praestgaard, H. Tatsumi, N. Spodsberg, K. Borch and P. Westh (2012). "Pre-steady-state kinetics for hydrolysis of insoluble cellulose by cellobiohydrolase Cel7A." J Biol Chem **287**(22): 18451-18458.

Cruys-Bagger, N., H. Tatsumi, G. R. Ren, K. Borch and P. Westh (2013). "Transient kinetics and rate-limiting steps for the processive cellobiohydrolase Cel7A: effects of substrate structure and carbohydrate binding domain." Biochemistry **52**(49): 8938-8948.

Cruys-Bagger, N., J. Elmerdahl, E. Praestgaard, K. Borch and P. Westh (2013). "A steady-state theory for processive cellulases." The FEBS journal **280**(16): 3952-3961.

Dana, C. M., A. Dotson-Fagerstrom, C. M. Roche, S. M. Kal, H. A. Chokhawala, H. W. Blanch and D. S. Clark (2014). "The importance of pyroglutamate in cellulase Cel7A." Biotechnol Bioeng **111**(4): 842-847.

Davies, G. J., V. Ducros, R. J. Lewis, T. V. Borchert and M. Schülein (1997). "Oligosaccharide specificity of a family 7 endoglucanase: insertion of potential sugar-binding subsites." Journal of biotechnology **57**(1-3): 91-100.

Dean, R., J. A. Van Kan, Z. A. Pretorius, K. E. Hammond-Kosack, A. Di Pietro, P. D. Spanu, J. J. Rudd, M. Dickman, R. Kahmann, J. Ellis and G. D. Foster (2012). "The Top 10 fungal pathogens in molecular plant pathology" Molecular plant pathology **13**(4): 414-430.

Deshpande, N., M. R. Wilkins, N. Packer and H. Nevalainen (2008). "Protein glycosylation pathways in filamentous fungi." Glycobiology **18**(8): 626-637.

Divne, C., J. Stahlberg, T. Reinikainen, L. Ruohonen, G. Pettersson, J. K. Knowles, T. T. Teeri and T. A. Jones (1994). "The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from Trichoderma reesei." Science **265**(5171): 524-528.

Divne, C., J. Ståhlberg, T. T. Teeri and T. A. Jones (1998). "High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from Trichoderma reesei1." Journal of molecular biology **275**(2): 309-325.

Dotsenko, A. S., A. V. Gusakov, P. V. Volkov, A. M. Rozhkova and A. P. Sinitsyn (2016). "N-linked glycosylation of recombinant cellobiohydrolase I (Cel7A) from Penicillium verruculosum and its effect on the enzyme activity." Biotechnology bioengineering **113**(2): 283-291.

Eibinger, M., T. Ganner, P. Bubner, S. Rošker, D. Kracher, D. Haltrich, R. Ludwig, H. Plank and B. Nidetzky (2014). "Cellulose surface degradation by a lytic polysaccharide monooxygenase and its effect on cellulase hydrolytic efficiency." Journal of Biological Chemistry **289**(52): 35929-35938.

Eijsink, V. G., S. Gåseidnes, T. V. Borchert and B. van den Burg (2005). "Directed evolution of enzyme stability." Biomolecular engineering **22**(1-3): 21-30.

Elias, M., G. Wieczorek, S. Rosenne and D. S. Tawfik (2014). "The universality of enzymatic rate–temperature dependency." Trends in Biochemical Sciences **39**(1): 1-7.

Eriksson, T., J. Börjesson and F. Tjerneld (2002). "Mechanism of surfactant effect in enzymatic hydrolysis of lignocellulose." Enzyme and Microbial Technology **31**(3): 353-364.

Eveleigh, D. E. and B. S. Montenecourt (1979). Increasing Yields of Extracellular Enzymes. Advances in Applied Microbiology. D. Perlman, Academic Press. **25:** 57-74.

Forte, A., A. Zucaro, S. Faugno, R. Basosi and A. Fierro (2018). "Carbon footprint and fossil energy consumption of bio-ethanol fuel production from Arundo donax L. crops on marginal lands of Southern Italy." Energy **150**: 222-235.

Ganner, T., P. Bubner, M. Eibinger, C. Mayrhofer, H. Plank and B. Nidetzky (2012). "Dissecting and reconstructing synergism in situ visualization of cooperativity among cellulases." Journal of Biological Chemistry **287**(52): 43215-43222.

Girfoglio, M., M. Rossi and R. Cannio (2012). "Cellulose degradation by Sulfolobus solfataricus requires a cell-anchored endo-β-1-4-glucanase." Journal of bacteriology **194**(18): 5091-5100.

Goedegebuur, F., L. Dankmeyer, P. Gualfetti, S. Karkehabadi, H. Hansson, S. Jana, V. Huynh, B. R. Kelemen, P. Kruithof and E. A. Larenas (2017). "Improving the thermal

stability of cellobiohydrolase Cel7A from Hypocrea jecorina by directed evolution." Journal of Biological Chemistry **292**(42): 17418-17430.

Gruno, M., P. Väljamäe, G. Pettersson and G. Johansson (2004). "Inhibition of the Trichoderma reesei cellulases by cellobiose is strongly dependent on the nature of the substrate." Biotechnology bioengineering **86**(5): 503-511.

Guvench, O., S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison and A. D. MacKerell Jr (2011). "CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling." Journal of chemical theory computation **7**(10): 3162-3180.

Hamed, S. A. M. (2013). "In-vitro studies on wood degradation in soil by soft-rot fungi: Aspergillus niger and Penicillium chrysogenum." International Biodeterioration Biodegradation **78**: 98-102.

Hangasky, J. A., A. T. Iavarone and M. A. Marletta (2018). "Reactivity of O2 versus H2O2 with polysaccharide monooxygenases." Proceedings of the National Academy of Sciences **115**(19): 4915-4920.

Harman, G. E. and C. P. Kubicek (1998). Trichoderma And Gliocladium, Volume 2: Enzymes, Biological Control and commercial applications, CRC Press.

Hayashida, S. and H. J. A. Yoshioka (1980). "The role of carbohydrate moiety on thermostability of cellulases from Humicola insolens YH-8." Agriculturalbiological Chemistry Hoppe-Seyler **44**(3): 481-487.

Heinzelman, P., C. D. Snow, I. Wu, C. Nguyen, A. Villalobos, S. Govindarajan, J. Minshull and F. H. Arnold (2009). "A family of thermostable fungal cellulases created by structure-guided recombination." Proceedings of the National Academy of Sciences **106**(14): 5610-5615.

Henrissat, B., H. Driguez, C. Viet and M. Schülein (1985). "Synergism of cellulases from Trichoderma reesei in the degradation of cellulose." Bio/technology **3**(8): 722.

Himmel, M. E., S. Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady and T. D. Foust (2007). "Biomass recalcitrance: engineering plants and enzymes for biofuels production." Science **315**(5813): 804-807.

Hobdey, S. E., B. C. Knott, M. Haddad Momeni, L. E. Taylor, 2nd, A. S. Borisova, K. K. Podkaminer, T. A. VanderWall, M. E. Himmel, S. R. Decker, G. T. Beckham and J. Stahlberg (2016). "Biochemical and Structural Characterizations of Two Dictyostelium Cellobiohydrolases from the Amoebozoa Kingdom Reveal a High Level of Conservation between Distant Phylogenetic Trees of Life." Appl Environ Microbiol **82**(11): 3395-3409.

Houbraken, J., H. Spierenburg and J. C. Frisvad (2012). "Rasamsonia, a new genus comprising thermotolerant and thermophilic Talaromyces and Geosmithia species." Antonie Van Leeuwenhoek **101**(2): 403-421.

Hu, J., K. Gourlay, V. Arantes, J. Van Dyk, A. Pribowo and J. N. Saddler (2015). "The accessible cellulose surface influences cellulase synergism during the hydrolysis of lignocellulosic substrates." ChemSusChem **8**(5): 901-907.

Igarashi, K., A. Koivula, M. Wada, S. Kimura, M. Penttilä and M. Samejima (2009). "High speed atomic force microscopy visualizes processive movement of Trichoderma reesei cellobiohydrolase I on crystalline cellulose." Journal of biological chemistry **284**(52): 36186-36190.

Igarashi, K., T. Uchihashi, A. Koivula, M. Wada, S. Kimura, T. Okamoto, M. Penttilä, T. Ando and M. Samejima (2011). "Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface." Science **333**(6047): 1279-1282.

Jalak, J., M. Kurasin, H. Teugjas and P. Valjamae (2012). "Endo-exo synergism in cellulose hydrolysis revisited." J Biol Chem **287**(34): 28802-28815.

Jalak, J. and P. Väljamäe (2010). "Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis." Biotechnology and bioengineering **106**(6): 871-883.

Jalak, J. and P. Väljamäe (2014). "Multi-mode binding of cellobiohydrolase Cel7A from Trichoderma reesei to cellulose." PloS one **9**(9): e108181.

Jeoh, T., M. J. Cardona, N. Karuna, A. R. Mudinoor and J. Nill (2017). "Mechanistic kinetic models of enzymatic cellulose hydrolysis-A review." Biotechnol Bioeng **114**(7): 1369-1385.

Jung, J., A. Sethi, T. Gaiotto, J. J. Han, T. Jeoh, S. Gnanakaran and P. M. Goodwin (2013). "Binding and movement of individual Cel7A cellobiohydrolases on crystalline cellulose surfaces revealed by single-molecule fluorescence imaging." Journal of Biological Chemistry **288**(33): 24164-24172.

Kari, J., M. Andersen, K. Borch and P. Westh (2017). "An Inverse Michaelis–Menten Approach for Interfacial Enzyme Kinetics." ACS Catalysis **7**(7): 4904-4914.

Kari, J., S. J. Christensen, M. Andersen, S. S. Baiget, K. Borch and P. Westh (2019). "A practical approach to steady-state kinetic analysis of cellulases acting on their natural insoluble substrate." Analytical Biochemistry: 113411.

Kari, J., R. Kont, K. Borch, S. Buskov, J. P. Olsen, N. Cruyz-Bagger, P. Valjamae and P. Westh (2017). "Anomeric Selectivity and Product Profile of a Processive Cellulase." Biochemistry **56**(1): 167-178.

Karimi, K. and M. J. Taherzadeh (2016). "A critical review of analytical methods in pretreatment of lignocelluloses: composition, imaging, and crystallinity." Bioresource technology **200**: 1008-1018.

Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass and M. J. Sternberg (2015). "The Phyre2 web portal for protein modeling, prediction and analysis." Nature protocols **10**(6): 845.

Kern, M., J.E. McGeehan, S.D. Streeter, R. N. Martin, K. Besser, L. Elias, W. Eborall, G.P. Malyon, C.M. Payne, M.E. Himmel and K. Schnorr (2013). "Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance." Proccedings of the National Academy of Science **110**(25): 10189-10194.

Kim, H.-W. and K. Ishikawa (2011). "Functional analysis of hyperthermophilic endocellulase from Pyrococcus horikoshii by crystallographic snapshots." Biochemical Journal **437**(2): 223-230.

Kim, T. H. and Y. Lee (2007). Pretreatment of corn stover by soaking in aqueous ammonia at moderate temperatures. Applied Biochemistry and Biotecnology, Springer**:** 81-92.

Kim, Y., E. Ximenes, N. S. Mosier and M. R. Ladisch (2011). "Soluble inhibitors/deactivators of cellulase enzymes from lignocellulosic biomass." Enzyme Microbial Technology **48**(4-5): 408-415.

Klein-Marcuschamer, D., P. Oleskowicz-Popiel, B. A. Simmons and H. W. Blanch (2012). "The challenge of enzyme cost in the production of lignocellulosic biofuels." Biotechnology bioengineering **109**(4): 1083-1087.

Kleywegt, G. J., J. Y. Zou, C. Divne, G. J. Davies, I. Sinning, J. Stahlberg, T. Reinikainen, M. Srisodsuk, T. T. Teeri and T. A. Jones (1997). "The crystal structure of the catalytic core domain of endoglucanase I from Trichoderma reesei at 3.6 A resolution, and a comparison with related enzymes." J Mol Biol **272**(3): 383-397.

Knott, B. C., M. F. Crowley, M. E. Himmel, J. Ståhlberg and G. T. Beckham (2014). "Carbohydrate–protein interactions that drive processive polysaccharide translocation in enzymes revealed from a computational study of cellobiohydrolase processivity." Journal of the American Chemical Society **136**(24): 8810-8819.

Knott, B. C., M. Haddad Momeni, M. F. Crowley, L. F. Mackenzie, A. W. Götz, M. Sandgren, S. G. Withers, J. Ståhlberg and G. T. Beckham (2013). "The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies." Journal of the American Chemical Society **136**(1): 321-329.

Kont, R., J. Kari, K. Borch, P. Westh and P. Väljamäe (2016). "Inter-domain synergism is required for efficient feeding of cellulose chain into active site of cellobiohydrolase Cel7A." Journal of Biological Chemistry **291**(50): 26013-26023.

Kont, R., M. Kurašin, H. Teugjas and P. Väljamäe (2013). "Strong cellulase inhibitors from the hydrothermal pretreatment of wheat straw." Biotechnology for biofuels **6**(1): 135.

Koshland Jr, D. (1953). "Stereochemistry and the mechanism of enzymatic reactions." Biological reviews **28**(4): 416-436.

Kostylev, M. and D. Wilson (2012). "Synergistic interactions in cellulose hydrolysis." Biofuels **3**(1): 61-70.

Krah, F.-S., C. Bässler, C. Heibl, J. Soghigian, H. Schaefer and D. S. Hibbett (2018). "Evolutionary dynamics of host specialization in wood-decay fungi." BMC evolutionary biology **18**(1): 119.

Kraulis, P. J., G. M. Clore, M. Nilges, T. A. Jones, G. Pettersson, J. Knowles and A. M. Gronenborn (1989). "Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from Trichoderma reesei. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing." Biochemistry **28**(18): 7241-7257.

Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." Journal of molecular biology **372**(3): 774-797.

Kuhad, R. C., R. Gupta and A. Singh (2011). "Microbial cellulases and their industrial applications." Enzyme research **2011**.

Kurašin, M. and P. Väljamäe (2011). "Processivity of cellobiohydrolases is limited by the substrate." Journal of Biological Chemistry **286**(1): 169-177.

Laskowski, R. A., E. G. Hutchinson, A. D. Michie, A. C. Wallace, M. L. Jones and J. M. Thornton (1997). "PDBsum: a Web-based database of summaries and analyses of all PDB structures." Trends in biochemical sciences **22**(12): 488-490.

Lehtiö, J., J. Sugiyama, M. Gustavsson, L. Fransson, M. Linder and T. T. Teeri (2003). "The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules." Proceedings of the National Academy of Sciences **100**(2): 484-489.

Leschine, S. B. (1995). "Cellulose degradation in anaerobic environments." Annual review of microbiology **49**(1): 399-426.

Li, W., X. Zhou and P. Lu (2005). "Structural features of thermozymes." Biotechnology advances **23**(4): 271-281.

Li, X.-Y. and B.-J. Tang (2017). "Incorporating the transport sector into carbon emission trading scheme: an overview and outlook." Natural Hazards **88**(2): 683-698.

Lin, Y., G. T. Beckham, M. E. Himmel, M. F. Crowley and J.-W. Chu (2013). "Endoglucanase Peripheral Loops Facilitate Complexation of Glucan Chains on Cellulose via Adaptive Coupling to the Emergent Substrate Structures." The Journal of Physical Chemistry B **117**(37): 10750-10758.

Linder, M., G. Lindeberg, T. Reinikainen, T. T. Teeri and G. Pettersson (1995). "The difference in affinity between two fungal cellulose-binding domains is dominated by a single amino acid substitution." FEBS letters **372**(1): 96-98.

Linder, M., M. L. Mattinen, M. Kontteli, G. Lindeberg, J. Ståhlberg, T. Drakenberg, T. Reinikainen, G. Pettersson and A. Annila (1995). "Identification of functionally important amino acids in the cellulose-binding domain of Trichoderma reesei cellobiohydrolase I." Protein Science **4**(6): 1056-1064.

Lombard, V., H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat (2013). "The carbohydrate-active enzymes database (CAZy) in 2013." Nucleic acids research **42**(D1): D490-D495.

Lutz, S. (2010). "Beyond directed evolution—semi-rational protein engineering and design." Current opinion in biotechnology **21**(6): 734-743.

M. Wood, T. and S. I. McCrae (1979). Synergism Between Enzymes Involved in the Solubilization of Native Cellulose.

Mach, R. L., B. Seiboth, A. Myasnikov, R. Gonzalez, J. Strauss, A. M. Harkki and C. P. Kubicek (1995). "The bgl1 gene of Trichoderma reesei QM 9414 encodes an extracellular, cellulose-inducible β-glucosidase involved in cellulase induction by sophorose." Molecular microbiology **16**(4): 687-697.

Mackenzie, L. F., G. Sulzenbacher, C. Divne, T. A. Jones, H. F. Wöldike, M. Schülein, S. G. Withers and G. Davies (1998). "Crystal structure of the family 7 endoglucanase I (Cel7B) from Humicola insolens at 2.2 Å resolution and identification of the catalytic nucleophile by trapping of the covalent glycosyl-enzyme intermediate." Biochemical journal **335**(2): 409-416.

Madeira, F., J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. Tivey, S. C. Potter, R. D. Finn and R. Lopez (2019). "The EMBL-EBI search and sequence analysis tools APIs in 2019." Nucleic acids research.

Maheshwari R., Bharadwaj G. and Bhat M. K. (2000). "Thermophilic fungi: their physiology and enzymes." Microbiol. Mol. Biol. Rev. **64**(3): 461-488.

Makela, M., N. Donofrio and R. de Vries (2014). "Plant biomass degradation by fungi." Fungal Genet Biol **72**: 1.

Mandels, M. and J. Weber (1969). The Production of Cellulases. Cellulases and Their Applications, AMERICAN CHEMICAL SOCIETY. **95:** 391-414.

Marriott, P. E., L. D. Gomez and S. J. McQueen-Mason (2016). "Unlocking the potential of lignocellulosic biomass through plant science." New Phytol **209**(4): 1366-1381.

Mattinen, M. L., M. Linder, T. Drakenberg and A. Annila (1998). "Solution structure of the cellulose-binding domain of endoglucanase I from Trichoderma reesei and its interaction with cello-oligosaccharides." European journal of biochemistry **256**(2): 279-286.

Merino, S. T. and J. Cherry (2007). Progress and challenges in enzyme development for biomass utilization. Biofuels, Springer**:** 95-120.

Mertz, B., X. Gu and P. J. Reilly (2009). "Analysis of functional divergence within two structurally related glycoside hydrolase families." Biopolymers **91**(6): 478-495.

Mitrovic, A., K. Flicker, G. Steinkellner, K. Gruber, C. Reisinger, G. Schirrmacher, A. Camattari and A. Glieder (2014). "Thermostability improvement of endoglucanase Cel7B from Hypocrea pseudokoningii." Journal of Molecular Catalysis B: Enzymatic **103**: 16-23.

Modenbach, A. A. and S. E. Nokes (2013). "Enzymatic hydrolysis of biomass at high-solids loadings–a review." Biomass Bioenergy **56**: 526-544.

Mohanram, S., D. Amat, J. Choudhary, A. Arora and L. Nain (2013). "Novel perspectives for evolving enzyme cocktails for lignocellulose hydrolysis in biorefineries." Sustainable Chemical Processes **1**(1): 15.

Moloney, A. P., P. J. Considine and M. P. Coughlan (1983). "Cellulose hydrolysis by the cellulases produced by Talaromyces emersonii when grown on different inducing substrates." **25**(4): 1169-1173.

Momeni, M. H., F. Goedegebuur, H. Hansson, S. Karkehabadi, G. Askarieh, C. Mitchinson, E. A. Larenas, J. Stahlberg and M. Sandgren (2014). "Expression, crystal structure and cellulase activity of the thermostable cellobiohydrolase Cel7A from the fungus Humicola grisea var. thermoidea." Acta Crystallogr D Biol Crystallogr **70**(Pt 9): 2356-2366.

Momeni, M. H., C. M. Payne, H. Hansson, N. E. Mikkelsen, J. Svedberg, A. Engstrom, M. Sandgren, G. T. Beckham and J. Stahlberg (2013). "Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus Heterobasidion irregulare." J Biol Chem **288**(8): 5861-5872.

Mosier, N., C. Wyman, B. Dale, R. Elander, Y. Lee, M. Holtzapple and M. Ladisch (2005). "Features of promising technologies for pretreatment of lignocellulosic biomass." Bioresource technology **96**(6): 673-686.

Muñoz, I. G., W. Ubhayasekera, H. Henriksson, I. Szabó, G. Pettersson, G. Johansson, S. L. Mowbray and J. Ståhlberg (2001). "Family 7 cellobiohydrolases from Phanerochaete

chrysosporium: crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 Å resolution and homology models of the isozymes11Edited by R. Huber." Journal of Molecular Biology **314**(5): 1097-1111.

Murphy, L., C. Bohlin, M. J. Baumann, S. N. Olsen, T. H. Sorensen, L. Anderson, K. Borch and P. Westh (2013). "Product inhibition of five Hypocrea jecorina cellulases." Enzyme Microb Technol **52**(3): 163-169.

Murphy, L., N. Cruys-Bagger, H. D. Damgaard, M. J. Baumann, S. N. Olsen, K. Borch, S. F. Lassen, M. Sweeney, H. Tatsumi and P. Westh (2012). "Origin of initial burst in activity for Trichoderma reesei endo-glucanases hydrolyzing insoluble cellulose." J Biol Chem **287**(2): 1252-1260.

Nakamura, A., H. Watanabe, T. Ishida, T. Uchihashi, M. Wada, T. Ando, K. Igarashi and M. Samejima (2014). "Trade-off between processivity and hydrolytic velocity of cellobiohydrolases at the surface of crystalline cellulose." Journal of the American Chemical Society **136**(12): 4584-4592.

Nakao, Y., Y. Kozutsumi, I. Funakoshi, T. Kawasaki, I. Yamashina, J. H. Mutsaers, H. Van Halbeek and J. F. Vliegenthart (1987). "Structures of oligosaccharides on β-galactosidase from Aspergillus oryzae." The journal of biochemistry **102**(1): 171-179.

Nelson, D. L., A. L. Lehninger and M. M. Cox (2008). Lehninger principles of biochemistry, Macmillan.

Nestl, B. M. and B. Hauer (2014). "Engineering of Flexible Loops in Enzymes." ACS Catalysis **4**(9): 3201-3211.

Nikolaidis, N., N. Doran and D. J. Cosgrove (2014). "Plant expansins in bacteria and fungi: evolution by horizontal gene transfer and independent domain fusion." Mol Biol Evol **31**(2): 376-386.

Oksanen, T., J. Pere, J. Buchert and L. Viikari (1997). "The effect of Trichoderma reesei cellulases and hemicellulases on the paper technical properties of never-dried bleached kraft pulp." Cellulose **4**(4): 329-339.

Olsen, J. P., K. Alasepp, J. Kari, N. Cruys-Bagger, K. Borch and P. Westh (2016). "Mechanism of product inhibition for cellobiohydrolase Cel7A during hydrolysis of insoluble cellulose." Biotechnol Bioeng **113**(6): 1178-1186.

Parkkinen, T., A. Koivula, J. Vehmaanpera and J. Rouvinen (2008). "Crystal structures of Melanocarpus albomyces cellobiohydrolase Cel7B in complex with cello-oligomers show high flexibility in the substrate binding." Protein Sci **17**(8): 1383-1394.

Pauly, M. and K. Keegstra (2008). "Cell-wall carbohydrates and their modification as a resource for biofuels." The Plant Journal **54**(4): 559-568.

Payne, C. M., B. C. Knott, H. B. Mayes, H. Hansson, M. E. Himmel, M. Sandgren, J. Stahlberg and G. T. Beckham (2015). "Fungal cellulases." Chem Rev **115**(3): 1308-1448.

Pearl, F. M., C. Bennett, J. E. Bray, A. P. Harrison, N. Martin, A. Shepherd, I. Sillitoe, J. Thornton and C. A. Orengo (2003). "The CATH database: an extended protein family resource for structural and functional genomics." Nucleic acids research **31**(1): 452-455.

Pellegrini, V. O., N. Lei, M. Kyasaram, J. P. Olsen, S. F. Badino, M. S. Windahl, F. Colussi, N. Cruys-Bagger, K. Borch and P. Westh (2014). "Reversibility of substrate adsorption for the cellulases Cel7A, Cel6A, and Cel7B from Hypocrea jecorina." Langmuir **30**(42): 12602-12609.

Peterson, R. and H. Nevalainen (2012). "Trichoderma reesei RUT-C30 – thirty years of strain improvement." **158**(1): 58-68.

Powell, A. J., K. J. Parchert, J. M. Bustamante, J. B. Ricken, M. I. Hutchinson and D. O. Natvig (2012). "Thermophilic fungi in an aridland ecosystem." Mycologia **104**(4): 813-825.

Prates, E. T., X. Guan, Y. Li, X. Wang, P. K. Chaffey, M. S. Skaf, M. F. Crowley, Z. Tan and G. T. Beckham (2018). "The impact of O-glycan chemistry on the stability of intrinsically disordered proteins." Chem Sci **9**(15): 3710-3715.

Puranen, T., M. Alapuranen and J. Vehmaanperä (2014). Trichoderma enzymes for textile industries. Biotechnology and biology of Trichoderma, Elsevier**:** 351-362.

Ragauskas, A. J., C. K. Williams, B. H. Davison, G. Britovsek, J. Cairney, C. A. Eckert, W. J. Frederick, J. P. Hallett, D. J. Leak and C. L. Liotta (2006). "The path forward for biofuels and biomaterials." science **311**(5760): 484-489.

Robert, X. and P. Gouet (2014). "Deciphering key features in protein structures with the new ENDscript server." Nucleic acids research **42**(W1): W320-W324.

Rosgaard, L., S. Pedersen, J. Langston, D. Akerhielm, J. R. Cherry and A. S. Meyer (2007). "Evaluation of minimal Trichoderma reesei cellulase mixtures on differently pretreated barley straw substrates." Biotechnology progress **23**(6): 1270-1276.

Rytioja, J., K. Hildén, J. Yuzon, A. Hatakka, R. P. de Vries and M. R. Mäkelä (2014). "Plant-polysaccharide-degrading enzymes from basidiomycetes." Microbiol. Mol. Biol. Rev. **78**(4): 614-649.

Saloheimo, M. and T. M. Pakula (2012). "The cargo and the transport system: secreted proteins and protein secretion in Trichoderma reesei (Hypocrea jecorina)." Microbiology **158**(1): 46-57.

Sandgren, M., J. Ståhlberg and C. Mitchinson (2005). "Structural and biochemical studies of GH family 12 cellulases: improved thermal stability, and ligand complexes." Progress in biophysics molecular biology **89**(3): 246-291.

Schulein, M. (1993). "Humicola insolens alkaline cellulases." Found. Biotech. Ind. Ferment. Res. **8**: 109-116.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert and J. Söding (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." Molecular systems biology **7**(1): 539.

Silveira, R. L. and M. S. Skaf (2014). "Molecular dynamics simulations of family 7 cellobiohydrolase mutants aimed at reducing product inhibition." The Journal of Physical Chemistry **119**(29): 9295-9303.

Somerville, C. (2006). "Cellulose synthesis in higher plants." Annu. Rev. Cell Dev. Biol. **22**: 53-78.

Sonoda, M. T., A. S. Godoy, V. O. A. Pellegrini, M. A. S. Kadowaki, A. S. Nascimento and I. Polikarpov (2019). "Structure and dynamics of Trichoderma harzianum Cel7B suggest molecular architecture adaptations required for a wide spectrum of activities on plant cell wall polysaccharides." Biochimica et Biophysica Acta (BBA) - General Subjects **1863**(6): 1015-1026.

Sorensen, T. H., N. Cruys-Bagger, M. S. Windahl, S. F. Badino, K. Borch and P. Westh (2015). "Temperature Effects on Kinetic Parameters and Substrate Affinity of Cel7A Cellobiohydrolases." J Biol Chem **290**(36): 22193-22202.

Ståhlberg, J., G. Johansson and G. Pettersson (1993). "Trichoderma reesei has no true exo-cellulase: all intact and truncated cellulases produce new reducing end groups on cellulose." Biochimica et Biophysica Acta **1157**(1): 107-113.

Sun, Y. and J. Cheng (2002). "Hydrolysis of lignocellulosic materials for ethanol production: a review." Bioresource technology **83**(1): 1-11.

Szilágyi, A. and P. Závodszky (2000). "Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey." Structure **8**(5): 493-504.

Taylor, C. B., C. M. Payne, M. E. Himmel, M. F. Crowley, C. McCabe and G. T. Beckham (2013). "Binding site dynamics and aromatic–carbohydrate interactions in processive and non-processive family 7 glycoside hydrolases." The Journal of Physical Chemistry B **117**(17): 4924-4933.

Taylor, L. E., 2nd, B. C. Knott, J. O. Baker, P. M. Alahuhta, S. E. Hobdey, J. G. Linger, V. V. Lunin, A. Amore, V. Subramanian, K. Podkaminer, Q. Xu, T. A. VanderWall, L. A. Schuster, Y. B. Chaudhari, W. S. Adney, M. F. Crowley, M. E. Himmel, S. R. Decker and G. T. Beckham (2018). "Engineering enhanced cellobiohydrolase activity." Nature communications **9**(1): 1186-1186.

Taylor, T. J. and I. I. Vaisman (2010). "Discrimination of thermophilic and mesophilic proteins." BMC structural biology **10**(1): S5.

Teugjas, H. and P. Väljamäe (2013). "Product inhibition of cellulases studied with 14 C-labeled cellulose substrates." Biotechnology for biofuels **6**(1): 104.

Textor, L. C., F. Colussi, R. L. Silveira, V. Serpa, B. L. de Mello, J. R. C. Muniz, F. M. Squina, N. Pereira Jr, M. S. Skaf and I. Polikarpov (2013). "Joint X-ray crystallographic and molecular dynamics study of cellobiohydrolase I from Trichoderma harzianum: deciphering the structural features of cellobiohydrolase catalytic activity." **280**(1): 56-69.

Tuohy, M. G., D. J. Walsh, P. G. Murray, M. Claeyssens, M. M. Cuffe, A. V. Savage and M. P. Coughlan (2002). "Kinetic parameters and mode of action of the cellobiohydrolases produced by Talaromyces emersonii." Biochimica et Biophysica Acta - Protein Structure Molecular Enzymology **1596**(2): 366-380.

Uhlig, H. (1998). Industrial enzymes and their applications, John Wiley & Sons.

Väljamäe, P., V. Sild, A. Nutt, G. Pettersson and G. Johansson (1999). "Acid hydrolysis of bacterial cellulose reveals different modes of synergistic action between cellobiohydrolase I and endoglucanase I." European journal of biochemistry **266**(2): 327-334.

van den Brink, J. and R. P. de Vries (2011). "Fungal enzyme sets for plant polysaccharide degradation." Applied microbiology biotechnology **91**(6): 1477.

Várnai, A., M. Siika-aho and L. Viikari (2010). "Restriction of the enzymatic hydrolysis of steam-pretreated spruce by lignin and hemicellulose." Enzyme and Microbial Technology **46**(3): 185-193.

Várnai, A., M. Siika-aho and L. Viikari (2013). "Carbohydrate-binding modules (CBMs) revisited: reduced amount of water counterbalances the need for CBMs." Biotechnology for biofuels **6**(1): 30.

Verardi, A., I. De Bari, E. Ricca and V. Calabrò (2012). "Hydrolysis of lignocellulosic biomass: current status of processes and technologies and future perspectives." Bioethanol: 95-122.

Vermaas, J. V., R. Kont, G. T. Beckham, M. F. Crowley, M. Gudmundsson, M. Sandgren, J. Ståhlberg, P. Väljamäe and B. C. Knott (2019). "The dissociation mechanism of processive cellulases." Proceedings of the National Academy of Sciences.

von Ossowski, I., J. Ståhlberg, A. Koivula, K. Piens, D. Becker, H. Boer, R. Harle, M. Harris, C. Divne, S. Mahdi, Y. Zhao, H. Driguez, M. Claeyssens, M. L. Sinnott and T. T. Teeri (2003). "Engineering the Exo-loop of Trichoderma reesei Cellobiohydrolase, Cel7A. A comparison with Phanerochaete chrysosporium Cel7D." Journal of Molecular Biology **333**(4): 817-829.

Voutilainen, S. P., H. Boer, M. Alapuranen, J. Jänis, J. Vehmaanperä and A. Koivula (2009). "Improving the thermostability and activity of Melanocarpus albomyces cellobiohydrolase Cel7B." Applied microbiology biotechnology Advances **83**(2): 261.

Wang, K., H. Luo, P. Shi, H. Huang, Y. Bai and B. Yao (2014). "A highly-active endo-1, 3-1, 4-β-glucanase from thermophilic Talaromyces emersonii CBS394. 64 with application potential in the brewing and feed industries." Process Biochemistry **49**(9): 1448-1456.

Warshel, A. (1998). "Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites." Journal of Biological Chemistry **273**(42): 27035-27038.

Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer and L. Bordoli (2018). "SWISS-MODEL: homology modelling of protein structures and complexes." Nucleic acids research **46**(W1): W296-W303.

Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp and G. J. Barton (2009). "Jalview Version 2—a multiple sequence alignment editor and analysis workbench." Bioinformatics **25**(9): 1189-1191.

Xiao, Z., X. Zhang, D. J. Gregg and J. N. Saddler (2004). "Effects of sugar inhibition on cellulases and beta-glucosidase during enzymatic hydrolysis of softwood substrates." Appl Biochem Biotechnol **113-116**: 1115-1126.

Zhang, S., Y. Wang, X. Song, J. Hong, Y. Zhang and L. Yao (2014). "Improving Trichoderma reesei Cel7B thermostability by targeting the weak spots." Journal of chemical information modeling **54**(10): 2826-2833.

Zhang, W., Y. Kou, J. Xu, Y. Cao, G. Zhao, J. Shao, H. Wang, Z. Wang, X. Bao and G. Chen (2013). "Two major facilitator superfamily sugar transporters from Trichoderma reesei and their roles in induction of cellulase biosynthesis." Journal of Biological Chemistry **288**(46): 32861-32872.

Zhang, X. Z., Zhang Y. H. P. (2013). "Cellulases: characteristics, sources, production and applications." Bioprocessing technologies in biorefinery for sustainable production of fuel, chemicals and polymers **1**: 131-146.

Zhang, Y.-H. P., J. Cui, L. R. Lynd and L. R. J. B. Kuang (2006). "A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure." **7**(2): 644-648.

Zhang, Y., S. Yan and L. Yao (2013). "A mechanistic study of Trichoderma reesei Cel7B catalyzed glycosidic bond cleavage." The Journal of Physical Chemistry B **117**(29): 8714-8722.

Zhang, Y. H. and L. R. Lynd (2004). "Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems." Biotechnol Bioeng **88**(7): 797-824.

Appendix I

Product profile of the enzymes studied in Article I

The production of glucose (G1) cellobiose (G2) and cellotriose (G3) was measured for *Tr*Cel7A, *Tr*Cel7B and the variants ΔB2-1, ΔB2-2, ΔB2-3, ΔB3 and ΔB4 at 25 °C. Enzymes and substrate (Avicel PH-101) were mixed to a final concentration of 50 nM and 80 g/L respectively. The reaction was performed in Eppendorf tubes at a final volume of 300 µL in triplicates. The tubes were incubated in Eppendorf thermomixers operating at 1100 rpm. After 1 h, the reactions were quenched with 300 µL 0.2 M NaOH and centrifuged for 3 min. The supernatant was then transferred in glass vials and the amount of sugars produced was measured by High-Performance Anion-Exchange Chromatography Coupled with Pulsed Electrochemical Detection (HPAEC-PAD) using a Dionex ICS-3000 ion chromatograph (Thermo Fisher Scientific, Waltham, MA) equipped with an electrochemical detector a CarboPac PA-10 column. Sugar standards at 9 concentrations, G1 500-1.95 µM, G2 500-1.95 µM and G3 300-1.17 µM, were also quenched in the same way and analysed in duplicates, before and after the sample analysis to ensure no spontaneous degradations of the sugars caused by the NaOH treatment occurred. Samples were eluted using the following multistep gradient program at a flow rate 1 mL/min: 100 mM NaOH, 25 mM sodium acetate (0-4.75 min), 190 mM NaOH, 425 mM sodium acetate (4.75-5 min), 100 mM NaOH, 25 mM sodium acetate (5.25-11.5 min). Data was analysed using the instrument software Chromeleon 7.1. An example of the obtained chromatograms in shown in Fig. A1.

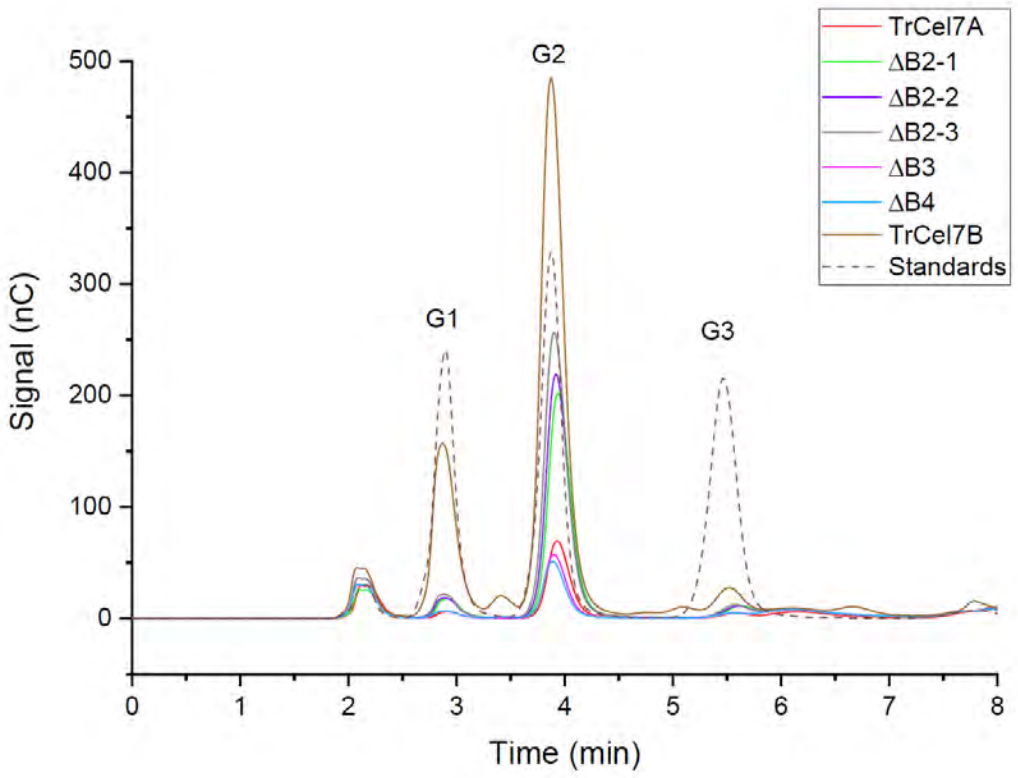


Fig. A1. *HPAEC-PAD chromatogram profiles of the variants discussed in Article I. Peak identity of the sugar standards is indicated. The unassigned peak at 2 min is associated with the standard buffer used in all the measurements.*

Appendix II

Measuring the exposed surface area of cellononaose in the different GH7 structures

Protein interfaces, surfaces and assemblies' service (PISA) at the European Bioinformatics Institute (www.ebi.ac.uk/pdbe/pisa/) can be used to measure the accessible surface area of a protein bound to a ligand (Krissinel et al. 2007) . As a start, I analysed with PISA the structure of *Tr*Cel7A bound to the substrate cellononaose (PDB 4C4C). I extracted two values, one corresponding to the total accessible surface area (ASA) of the free ligand cellononaose, the second was the buried surface area (BSA). By subtracting ASA-BSA, it was possible to extract a third value, here named exposed surface area (ESA), which corresponds to the solvent-exposed surface area of the ligand. I performed the same analysis with all the available GH7 structures by placing the ligand cellononaose via superimposition onto 4C4C using Pymol. After superimposition, the structures underwent an energy minimization procedure performed by Kay Schaller (DTU Bioengineering) with GROMACS 2018 (Abraham et al. 2015). The minimizations were performed in a steepest-descent over 10000 iterations or until convergence was reached. The CHARMM36 force field was used to describe the system (Guvench et al. 2011). The structures were visualized and processed using PyMOL Version 2.3 (Schrödinger, LLC). Details of the results are shown in Table A1.

Appendix II

Table A1. *The values of cellononaose free accessible surface area (ASA) and buried surface area (BSA) when complexed with the different GH7 enzymes. The solvent exposed surface area (ESA) was calculated by subtracting ASA and BSA. The PDB entry used for each analysis and the organism of origin is indicated.*

Enzyme	Organism	PDB accession number	Cellononaose solvent accessible surface area (ASA), Å^2	Cellononaose buried surface area (BSA) upon interface, A^2	Cellononaose exposed surface area (ESA) upon interface, A^2	Exposed surface area, (% of ASA)
*Th*Cel7B	*Trichoderma harzianum EG*	5W0A	1667.24	865.55	801.69	48.08
*Tr*Cel7B	*Trichoderma reesei EG*	1EG1	1671.67	881.11	790.56	47.29
*Re*Cel7B	*Rasamsonia emersonii EG*	6SU8	1673.61	924.78	748.83	44.74
*Fo*Cel7B	*Fusarium oxisporium EG*	1OVW	1664.91	975.72	689.19	41.40
*Hi*Cel7B	*Humicola insolens EG*	1A39	1669.88	980.54	689.34	41.28
*Th*Cel7A	*Trichoderma harzianum CBH*	2YOK	1671.32	1297.58	373.74	22.36
*Gc*Cel7A	*Geotrichum candidum CBH*	4ZZT	1673.37	1319.57	353.80	21.14
*Re*Cel7A	*Rasamsonia emersonii CBH*	3PFJ	1676.45	1328.88	347.57	20.73
*Tat*Cel7A	*Trichoderma atroviride CBH*	5O59	1668.32	1328.56	339.76	20.37
*Ddi*Cel7A	*Dictyostelium discoideum CBH*	4ZZQ	1672.71	1334.56	338.15	20.22
*Pf*Cel7A	*Talaromyces funiculosus CBH*	4XEB	1665.27	1331.29	333.98	20.06
*Pc*Cel7D	*Phanerochaete crysosporium CBH*	1GPI	1662.44	1329.59	332.85	20.02
*Dp*Cel7A	*Daphnia pulex CBH*	4XNN	1670.64	1339.23	331.41	19.84
*Ma*Cel7B	*Melanocarpus albomyces CBH*	2RFW	1697.58	1371.64	325.94	19.20
*Lq*Cel7A	*Limnoria quadripunctata CBH*	4GWA	1666.51	1345.97	320.54	19.23
*Mt*Cel7A	*Myceloptora termophila CBH*	5W11	1677.63	1358.03	319.60	19.05
*Hg*Cel7A	*Humicola grisea CBH*	4CSI	1671.18	1357.77	313.41	18.75
*Dpu*Cel7A	*Dictyostelium purpureum CBH*	4ZZP	1682.94	1386.35	296.59	17.62
*Hir*Cel7A	*Hetetorbasidion irregulare CBH*	2XSP	1668.02	1378.81	289.21	17.34
*Tr*Cel7A	*Trichoderma reesei CBH*	4C4C	1673.81	1391.8	282.01	16.85
*Af*Cel7A	*Aspergillus fumigatus CBH*	4V20	1670.79	1426.88	243.91	14.60

Appendix III

Activity assays under enzyme saturating and substrate saturating conditions

All activity assays were performed in a 50 mM sodium acetate buffer pH 5.0, henceforth called standard buffer, at 25°C. Avicel (PH101, Sigma Aldrich 11365) was used as substrate, after being washed five times in deionized water and two times in buffer to remove any soluble sugar contaminants. For the enzyme saturation approach, the total enzyme concentration used was 25 nM, while the Avicel load varied between 5 and 130 g/L. For the substrate saturation approach, the substrate load was 10 g/L and the total enzyme concentration varied between 0.05 and 4 µM. Suitable blanks were included in all experiments. In the synergy experiments, *Tr*Cel7A and *Tr*Cel7B were mixed together in different proportions to obtain the molar ratios listed in Table 5.1 of the thesis main text. For the reference experiments, only one enzyme, either *Tr*Cel7A or *Tr*Cel7B was used. The final reaction volume was 250 µL in all cases. Enzymes and substrate were mixed with a contact time of 1 hour in 96-well microtiter plates (96F 269620, Thermo Scientific) and then placed in Eppendorf Thermomixer operating at 1100 rpm. The reactions were stopped by a three-minute centrifugation at 2000 ×g. The amount of soluble reducing ends produced was quantified by the *para*-hydroxybenzoic acid hydrazide (PAHBAH) method (Lever 1973). From the centrifuged plates, 60 µL of the supernatant were retrieved and transferred to PCR plates. Then, 90 µL of a 15 mg/mL solution of PAHBAH in 0.18 M potassium sodium tartrate tetrahydrate (P2347 Sigma) and 0.5 M NaOH (106469 Merck) were addded. The plates were incubated at 95 °C for 10 min and 5 ° C for 5 min in a T100 Thermal Cycler (BioRad). A six-point cellobiose (Fluka, 22150) standard curve, with concentrations varying between 0.032 and 1 mM, was included in all experiments. After incubation, 100 µL were transferred into new plates and the absorbance at 405 nm was recorded using a plate reader (SpectraMax M2e, Molecular Devices).