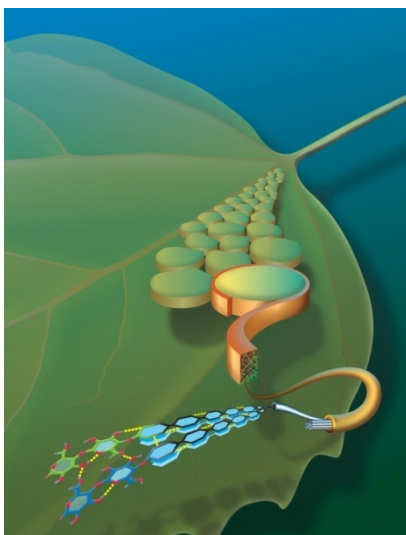# DOCTORAL THESIS

Construction of microbial platform for the production of thermophilic enzymes implicated in lignocellulose degradation

Anthi C. Karnaouri

National Technical University of Athens, Greece

January, 2015

Thesis for the degree of Doctor of Philosophy

Construction of microbial platform for the production of thermophilic enzymes implicated in lignocellulose degradation

Anthi C. Karnaouri

January, 2015

Thesis

submitted in fulfillment of the requirements for the degree of doctor

at National Technical University of Athens

by the authority of Prof. Paul Christakopoulos,

in the presence of the Thesis Committee,

to be defended in public

on Thursday, January 8th, 2015.

Thesis committee

Thesis supervisor

Prof. Paul Christakopoulos,

National Technical University of Athens / Luleå University of Technology

*Thesis co-supervisors*

Prof. Dimitrios Kekos, National Technical University of Athens

Assist. Prof. Evangelos Topakas, National Technical University of Athens

Other members

Prof. Fragkiskos Kolisis, National Technical University of Athens

Prof. Miltiadis Typas, University of Athens

Assist. Prof. Dimitrios Xatzinikolaou, University of Athens

Prof. Ulrika Rova, Luleå University of Technology
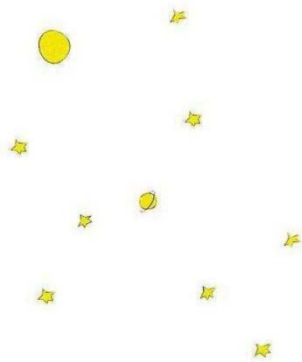
"What makes the desert beautiful is that somewhere it hides a well."

"Little Prince", Antoine De Saint-Exupery (1900–1944)

The present work represents the result of my doctoral thesis conducted between July 2010 and December 2014. The main part of the studies, covering the molecular cloning, production and characterization of the enzymes used in this thesis, was carried out at Biotechnology Laboratory, National Technical University of Athens in Greece. Fermentations and production of the enzymes at large scale, as well as hydrolysis experiments with agricultural and forest materials were conducted at Luleå University of Technology of Sweden. With great pleasure I would like to express my gratitude to all who have contributed to this thesis.

First, I would very much like to thank my supervisor, Prof. Paul Christakopoulos for his valuable guidance, inspiration, and support throughout the whole period. Secondly, I would like to thank my assistant supervisor Assist. Prof. Evangelos Topakas for all the fruitful discussions and his major contribution to this thesis by his holistic view and by ability to see things from a different angle. I am very grateful to Prof. Ulrika Rova for giving me the opportunity to work at Luleå University of Technology and helping me with difficulties aroused.

I would also like to thank all my former and present colleagues at NTUA Biotechnology Laboratory and LTU Biochemical Process Engineering Group for making the working days enjoyable during those four very good and pleasant years. It has been a pleasure to be a part of this open-minded, international environment. Special thanks to Maria Moukouli, Maria Dimarogona, Thomas Paschos and Marianna Charavgi for interesting and inspiring brain-stormings and lots of help in the lab. To Magnus Sjöblom and Madhu Nair Muraleedharan for their help with fermentation processes and to Leonidas Matsakas for his help at the statistical analysis of the hydrolysis experiments results.

The State Scholarship's Foundation (I.K.Y.) in Greece is gratefully acknowledged for financially supporting this work. During this doctoral thesis I received a Grant Scholarship for postgraduate studies, as well as a Fellowship for a Short Term Scientific Mission to Luleå University of Technology by the same Foundation.

I would also like to thank my dear friends both in Greece and Sweden for their enormous enthusiasm, always positive thinking, and encouragement when the things got rough. They made my life inside and outside the university very enjoyable.

Το θερμότερο ευχαριστώ το οφείλω στην οικογένειά μου, τους γονείς μου Χριστόφορο και Φωτεινή και τον αδερφό μου Γιάννη, για την ηθική συμπαράσταση, την ενθάρρυνση και την οικονομική υποστήριξη που μου παρείχαν. Κλείνοντας, ευχαριστώ ολόψυχα το Δημοσθένη, για την υπομονή, την κατανόηση και τη συνεχή στήριξή του.

Anthi Karnaouri
Athens, January 2015

List of Publications

I. Cloning, expression, and characterization of a thermostable GH7 endoglucanase from *Myceliophthorat hermophila* capable of high-consistency enzymatic liquefaction.

    Karnaouri AC, Topakas E, Christakopoulos P.

    Appl Microbiol Biotechnol. 2014 Jan;98(1):231-42

II. Cloning, expression and characterization of an ethanol tolerant GH3 β-glucosidase from *Myceliophthora thermophila.*

    Karnaouri A, Topakas E, Paschos T, Taouki I, Christakopoulos P.

    PeerJ. 2013 Feb 26;1:e46

III. Genomic insights into the fungal lignocellulolytic system of *Myceliophthora thermophila.*

    Karnaouri A, Topakas E, Antonopoulou I, Christakopoulos P.

    Front Microbiol. 2014 Jun 18;5:281

IV. Optimization of tailor-made enzyme cocktail for deconstruction of agricultural and forest residues (*in preparation)*

Conferences

I.   Καρναούρη, Α.Χ., Τόπακας, Ε., Χριστακόπουλος, Π., Παραγωγή και χαρακτηρισμός μιας θερμοσταθερήςενδογλουκανάσης για την ρευστοποίηση λιγνινοκυτταρινούχου βιομάζας, Προφορική ανακοίνωση, 9ο Πανελλήνιο Επιστημονικό Συνέδριο Χημικής Μηχανικής, Αθήνα, 2013

II.   Karnaouri A, Topakas E, Christakopoulos P, A thermostable GH7 endoglucanase from _Myceliophthorathermophila_ capable of high-consistency enzymatic liquefaction, Poster Presentation, 10th Carbohydrate Bioengineering Meeting (CBM10), Prague, Czech Republic, 2013

III.   Karnaouri A, Topakas E, Christakopoulos P, Molecular cloning, expression and characterization of a thermostable GH7 endoglucanase from _Myceliophthorathermophila_ with high-consistency liquefaction properties, Oral Presentation, Bio4energy Research School, Skellefteå, Sweden, 2013

IV.   Karnaouri A, Topakas E, Christakopoulos P, Cloning, heterologous expression and characterization of cellulolytic enzymes from the thermophilic fungi _Myceliophthorathermophila_, Poster Presentation, Plant and Seaweed Polysaccharides Symposium (PSP 2012), Nantes, France, 2012

V.   Karnaouri A, Topakas E, Christakopoulos P, Cloning, heterologous expression and characterization of a family 7 thermophilicendoglucanase from _Myceliophthorathermophila_ (_St_EG7), Poster Presentation, 9th Carbohydrate Bioengineering Meeting (CBM9), Lisbon, Portugal, 2011

Cellulose, the most abundant polysaccharide on Earth, is a remarkable pure organic polymeric component of plant material, consisting solely of 1,4-linked *β*-D-glucopyranose units held together in a giant straight chain molecule. In nature, a variety of microorganisms are known for producing a set of enzymes, referred to as cellulases, that are capable of degrading this insoluble polymer to soluble sugars, primarily cellobiose and glucose. The application and interest in cellulases has particularly increased in recent years with the utilization of the enzymes in the production of bioethanol from lignocelluloses. *Myceliophthora thermophila* (synonym *Sporotrichum thermophile)* is a thermophilic filamentous fungus, isolated from soil in eastern Russia, classified as an ascomycete, and constitutes an exceptionally powerful cellulolytic organism; it synthesizes a complete set of enzymes necessary for the breakdown of cellulose. The genome of this fungus has been recently sequenced and annotated, allowing systematic examination and identification of enzymes required for the decomposition of lignocellulosic biomass. In this thesis, the genes encoding five cellulases, including two *endoglucanases* belonging to glycoside hydrolase families GH5 and GH7, two *cellobiohydrolases* belonging to the families GH6 and GH7 and one β-*glycosidase* belonging to the family GH3, were cloned and expressed in methylotrophic yeast *P. pastoris*, and their properties were investigated. In addition, the enzymes were produced in high cell density cultures, in the controlled environment of fermenters. The protein's overexpression in a host suitable for industrial production is important in order to achieve low-cost and highly efficient production. The enzymes were purified to their homogeneity and were used for the development of tailor-made enzyme mixtures targeted towards particular feedstocks, including agricultural and forest residues, where they were tested for their ability to maximize hydrolysis yields.

Η κυτταρίνη αποτελεί το αφθονότερο οργανικό πολυμερές που συναντάται στη φύση, γεγονός που την καθιστά κατάλληλη ως φτηνά αξιοποιήσιμη πηγή άνθρακα και πρώτη ύλη για ποικίλες βιοτεχνολογικές εφαρμογές. Ως εκ τούτου, η απομόνωση νέων ενζύμων με δράση που στοχεύει στην αποικοδόμηση ή την τροποποίηση των κυτταρινούχων υλικών κρίνεται απαραίτητη. Οι θερμόφιλοι οργανισμοί αποτελούν σημαντική πηγή κυτταρινολυτικών ενζύμων με πολλές βιοτεχνολογικές εφαρμογές, καθώς οι διεργασίες αυτές συνήθως απαιτούν συνθήκες υψηλής θερμοκρασίας. Ο *Myceliophthora thermophila* είναι ένας θερμόφιλος αερόβιος μύκητας, ο οποίος αυξάνεται με μέγιστο ρυθμό σε θερμοκρασίες 45 – 50 ^{o}C. Ο μύκητας παράγει πολλά θερμοσταθερά ένζυμα τα οποία έχουν απομονωθεί και χαρακτηριστεί και χρησιμοποιούνται στη βιομηχανία σε βιοδιεργασίες που απαιτούν υψηλές θερμοκρασίες. Σκοπός της παρούσας εργασίας ήταν η απομόνωση και ετερόλογη έκφραση πέντε γονιδίων από το γονιδίωμα του συγκεκριμένου μύκητα που κωδικοποιούν πρωτεΐνες που εμπλέκονται στην αποικοδόμηση της λιγνοκυτταρίνης, καθώς και η παραγωγή και ο καταλυτικός χαρακτηρισμός του ανασυνδυασμένων ενζύμων. Οι νουκλεοτιδικές αλληλουχίες δύο ενδογλουκανασών των οικογενειών GH5 και GH7, δυο κελλοβιοϋδρολασών των οικογενειών GH6 και GH7 και μιας β-γλυκοσιδάσης της οικογένειας GH3 ανασύρθηκαν από τη βάση δεδομένων Genome Portal, ενισχύθηκαν, κλωνοποιήθηκαν στον κατάλληλο φορέα και χρησιμοποιήθηκαν για το μετασχηματισμό και την ετερόλογη έκφραση στο ζυμομύκητα *P. pastoris*. Ακολούθησε καθαρισμός των παραγόμενων πρωτεϊνών, τα οποία εν συνεχεία χρησιμοποιήθηκαν για τη δημιουργία ενζυμικών μιγμάτων και δοκιμάστηκαν για την υδρολυτική τους ικανότητα έναντι φυσικών υποστρωμάτων αγροτικής και δασικής προέλευσης.

Table of contents

CHAPTER 1

Introduction

Cellulose, the most abundant polysaccharide on Earth, is a remarkable pure organic polymeric component of plant material, consisting solely of 1,4-linked β-D-glucopyranose units held together in a giant straight chain molecule. Cellulose has long been harvested as commercial fibers from the seed hairs of cotton (over 94% cellulose), bast fibers (60–80% cellulose) from flax, hemp, jute and ramie or wood (40–55% cellulose), which is a common building material or is used as a source for purified cellulose. Wood represents a composite material with cellulose as a major part combined in excellent form with lignin and hemicelluloses, creating a unique high-strength and durable material, and recently came again into focus as a renewable energy resource. In nature, a variety of microorganisms are known for producing a set of enzymes capable of degrading this insoluble polymer to soluble sugars, primarily cellobiose and glucose. Enzymes involved in these processes are called cellulases and are consisting of at least three classes of enzymes, namely, *endoglucanases* (EG), *cellobiohydrolases* (CBH) and *β-glucosidases* (BG). Cellulases can be used in the variety of applications within food, vine, animal feed, textile and pulp and paper industry (Bhat, 2000). The application and interest in cellulases has particularly increased in recent years with the utilization of the enzymes in the production of bioethanol from lignocellulose (Sun and Cheng, 2002). It has been proposed that agricultural wastes, such as wheat straw, and forest residues may be economically converted to bioethanol.

Bioethanol is derived from alcoholic fermentation of sucrose or simple sugars that are produced from biomass and can be used as a petrol additive/substitute. When blended at low concentrations (5%) with petrol (gasoline) or diesel for use in today's vehicles, under the EU quality standard EN 228, no engine modification is required; it is covered by vehicle warranties and considered to be a sustainable transportation fuel. Alternatively, with engine modification and adopted vehicles, bioethanol can be used at higher levels, for example, E85 (85% bioethanol). The main motivation for investments in research and process development concerning bioethanol production is environmental concern related to global warming. The focus has, in particular, been

turned towards the reduction of emission of greenhouse gases (GHG), such as carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O), as well as VOC (volatile organic compounds) and other particles arising from fossil fuel combustion and land-use change as a result of human activities.

According to the Renewable Energy Directive (RED) 2003/30/EC of the European Parliament, the European Union (EU) creates a Community framework to promote the use of biofuels. The ultimate goal is to reduce dependency on the use of oil-based fuels, a significant cause for concern in terms of the environment and security of supply. The Directive sets a minimum percentage of biofuels (10%) to replace diesel or petrol for transport purposes in each Member State and proposes an action plan aimed at increasing the share of biofuels to more than 20 % of European petrol and diesel consumption by 2020. According to analyses of the Member States' National Renewable Energy Action Plans (NREAPs), biomass will make up 19 per cent of total renewable electricity in the year 2020, 78 per cent of total renewable heating and cooling in 2020 and 89 per cent of total renewable energy in transport. Altogether, bioenergy is expected to contribute over 50 per cent of total renewable energy use.

Sustainable bioethanol production would decrease the dependency on the traditional, natural oil, reserves, which can due to their restricted geographical localization cause political tension and economic instability. Under EU proposal 0547 from November 7, 2001, a series of goals were set for member states to introduce biofuels for diesel and gasoline. By 2005, 2 % of transport fuel should be accounted for by biofuels; by 2020, the goal is 20 %. There are several reasons for biofuels to be considered as relevant technologies by both developing and industrialized countries. These include energy security, environmental concerns, foreign exchange savings, and socioeconomic issues related to the rural sector (Demirbas 2007).

State of the art in biofuel production and current study

Currently, transportation fuels based on biomass (*i.e.* biofuels) are identified as 1st and 2nd generation biofuels. *First-generation biofuels* are produced from sugar, starch, vegetable oil or animal fats using conventional technologies. The basic feedstocks are often seeds and grains such as wheat, corn and rapeseed. The most

common first generation biofuels are bioethanol, biodiesel and starch-derived biogas, but also straight vegetable oils, biomethanol and bioethers may be included in this category (Cherubini and Jungmeier 2010). The main advantages of first generation biofuels are due to the high sugar or oil content of the raw materials and their easy conversion into biofuel. Nevertheless, concerns exist about the sourcing of feedstocks, including the impact it may have on biodiversity and land use, and competition with food crops (Naik *et al.*, 2010). These limitations are expected to be partially overcome by developing the so-called 2nd generation biofuels (Cherubini *et al.*, 2009). *Second generation biofuels* are made from non-edible feedstocks, such as waste from agriculture, forestry and industry and dedicated lignocellulosic crops. Contrarily to first generation biofuels, where the utilized fraction (grains and seeds), represents only a small portion of the above-ground biomass, second generation biofuels can rely on the whole plant for bioenergy production and they offer the opportunity for coproduction of valuable biofuels and value-added chemical compounds. Following this procedure will lead to better energy, environmental and economic performances through the development of so-called biorefinery concept (Kamm *et al.*, 2006). *Third-generation biofuels* are fuels that would be produced from algal biomass, which has a very distinctive growth yield as compared with classical lignocellulosic biomass (Brennana and Owendea, 2010). Production of biofuels from algae usually relies on the lipid content of the microorganisms, which can be processed via transesterification to produce biodiesel, but still scaling-up faces challenges with several geographical and technical limitations associated with algal biomass production (Lee and Lavoie, 2013).
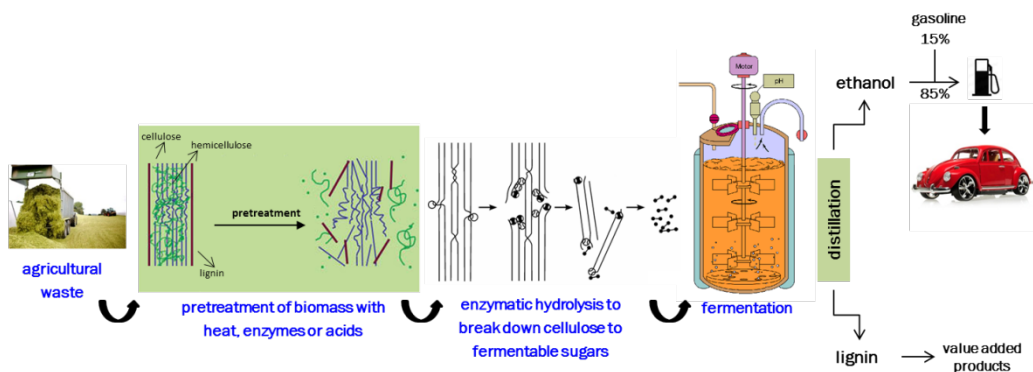
Figure 1.1. Schematic presentation of the conventional biochemical platform featuring enzymatic hydrolysis of lignocellulosic-based material.

Apart from biofuels, industrial bioproducts – chemicals and materials produced from biomass – play a key role in the so-called "*biorefinery concept*" for fostering a new bioindustry. There is a tremendous potential to supplement and supplant the petroleum resources used today to manufacture billions of pounds of important chemical products, most of which are value-added products, such as sugar and starch bioproducts, oil- and lipid- based bioproducts, cellulose derivatives, fibers and plastics, gum and wood chemicals (Cherubini 2010; Paster *et al.*, 2003).

In this thesis the focus is on the second-generation biofuels technology and more precisely on the process of enzymatic hydrolysis of cellulose. In general terms, the production of bioethanol from lignocellulose involves a degradation of the polymeric compounds, primarily cellulose and hemicellulose, to sugars, which are then subsequently fermented by microorganisms to ethanol. The process can be performed in a number of different ways (Olsson *et al.*, 2004). **Figure 1.1** shows an example of the process steps used for the conversion of lignocellulosic-based waste material to bioethanol. The procedure involves four key steps: pretreatment of the material, hydrolysis, fermentation and distillation for product recovery. *Pretreatment* targets at the removal of hemicellulose and lignin parts and isolation of cellulose from the lignocellulosic biomass. Many processes have been considered, including classical pulping processes (Jin *et al.*, 2010), steam explosion (Lavoie *et al.*, 2010), and organosolv processes (Brosse *et al.*, 2009). Isolation of cellulose is a technological challenge because it has to produce the highest purity of cellulose to remove most inhibitors without consuming too much energy or too many chemicals. The degradation of cellulose to glucose is accomplished in the next step, *hydrolysis / saccharification.* Two approaches are generally used for saccharification of cellulose: either enzymatic (Sun and Cheng, 2002) or by chemical hydrolysis using acids (Chornet *et al.*, 2010). In both cases, there are some limitations to the processes, such as the price of the enzymes or the high amount of harmful by-products that can hamper the subsequent steps. The discovery of new enzymes with high rates of catalytic activity that can be produced easily in large scale could reduce the cost. This topic is the main focus of this thesis; the production and mode of action of effective cellulolytic monenzymes and enzyme mixtures will be described in the following Chapters. After hydrolysis, sugars are produced from hydrolysis as reactive intermediates for subsequent fermentation to fuels and chemicals. Microorganisms (*e.g.* yeast) can ferment sugars to ethanol, which is further

on distillated, and mixed with gasoline to obtain blends such as E85 (a mixture of 15 % gasoline and 85 % ethanol).

In **Chapter 2** of this thesis, the composition of lignocellulose (plant cell wall polysaccharides) is introduced, and moreover, the characteristics of cellulose from plant materials (used in the industrial processes), and cellulose from model substrates, readily used in research, are discussed and compared. Cellulolytic enzymes and their main characteristics, as well as cooperative action between the different enzyme classes are presented. A brief description of the properties of the thermophilic fungi *Myceliophthora thermophila* and the heterologous expression tool organism, *Pichia pastoris* is given.

In **Chapter 4**, an overview of the cellulolytic and hemicellulolytic potential of *Myceliophthora thermophila* is described regarding the degradation of plant cell wall material (**Paper I**). The genome of this fungus has been recently sequenced and annotated, allowing systematic examination and identification of enzymes required for the decomposition of lignocellulosic biomass.

In **Chapters 5-7**, the successful heterologous expression and characterization of five genes isolated from *M. thermophila*'s genome is described in detail. The recombinant enzymes, including two endoglucanases belonging to glycoside hydrolase families GH5 and GH7, two cellobiohydrolases belonging to the families GH6 and GH7 and one β-glucosidase belonging to the family GH3, were expressed in methylotrophic yeast *P. pastoris*, and their properties were investigated. In addition, the enzymes were produced in high cell density cultures, in the controlled environment of fermenters. The enzyme's overexpression in a host suitable for industrial production is important in order to achieve low-cost and highly efficient production (**Papers II & III**).

In **Chapter 8**, tailor-made enzyme mixtures targeted towards particular feed stocks were tested for their ability to maximize hydrolysis yields. Four "core" cellulases, in the presence of other three "accessory" enzymes, all isolated from *M. thermophila*, were tested against agricultural and forest residues. Synergistic interactions among different enzymes were determined through various mixture optimization experiments.

References

Brennana, L., and P. Owendea. 2010. Biofuels from microalgae—A review of technologies for production, processing, and extractions of biofuels and coproducts. *Renewable Sustainable Energy Rev.* 14:557–577.

Brosse, N., Sannigrahi, P., Ragauskas, A. 2009. Pretreatment of *Miscanthus giganteus* using the ethanol organosolv process for ethanol production. *Ind. Eng. Chem. Res.* 48:8328–8334.

Cherubini, F. and Jungmeier, G. 2010. LCA of a biorefinery concept producing bioethanol, bioenergy, and chemicals from switchgrass. *International Journal of Life Cycle Assessment* 15:53-66.

Cherubini, F. 2010. The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. *Energ. Convers. Managem.* 51:1412–1421.

Cherubini, F., Bird, N., Cowie, A., Jungmeier, G., Schlamadinger, B., Woess-Gallasch, S. 2009. Energy and GHG-based LCA of biofuel and bioenergy systems: key issues, ranges and recommendation. *Resour. Conserv. Recycl. 53*, 434–47.

Chornet, M., Chornet, E., Lavoie, G.M., 2010. Conversion of cellulosic biomass to sugar. US Provisional Patent application #80685 2.

Demirbas, A. 2007. Progress and recent trends in biofuels. *Progress in Energy and Combustion Science 33*: 1–18.

Jin, Y., H. Jameel, H.-M. Chang, and R. Phillips. 2010. Green liquor pretreatment of mixed hardwood for ethanol production in a repurposed kraft pulp mill. *J. Wood Chem. Technol.* 30:86–104.

Kamm, B., Kamm, M., Gruber, P.R., Kromus, S. 2006. Biorefinery systems – an overview. In: Kamm B, Gruber PR, Kamm M, editors. Biorefineries – industrial processes and products (status quo and future directions), vol. 1. Wiley-VCH.

Lavoie, J.-M., Capek-Menard, E., Gauvin, H., Chornet, E. 2010. Production of pulp from *Salix viminalis* energy crops using the FIRSST process. *Bioresour. Technol.* 101:4940–4946.

Lee, R.A. and Lavoie, J.M. 2013. From first- to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity. *Animal Frontiers. 3*: 6-11.

Naik, S.N., Goud, V.V., Rout, P.K., Dalai, A.K. 2010. Production of first and second generation biofuels: A comprehensive review. *Renew. Sustain. Energy Rev. 14*, 578–597.

Olsson, L., Jørgensen, H., Krogh, K.B.R., Roca, C. 2004. Bioethanol production from lignocellulosic material. In: *Polysaccharides: structural diversity and functional versatility*. New York: Marcel Dekker Inc. p 957-993.

Paster, M., Pellegrino, J.L., Tracy M.C. 2003. Industrial Bioproducts: Today and Tomorrow, Prepared by Energetics Inc., Columbia, Maryland for the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Office of the Biomass Program Washington, D.C.

Sun, Y., and J. Cheng. 2002. Hydrolysis of lignocellulosic materials for ethanol production: A review. *Bioresour. Technol.* 83:1–11.

CHAPTER 2

2.1 Cell-wall degrading enzymes

2.1.1 Components and Morphology of the cell wall

Even though the relative amounts and structures of the cell-wall polysaccharides may vary greatly among species and from tissue to tissue within a plant., the fundamental mechanisms of interaction between these polysaccharides are highly similar (Carpita and McCann, 2000). They are the most abundant organic compounds found in nature and are conventionally divided into three groups: cellulose, hemicellulose (e.g., xyloglucans, xylans, and mannans), and pectins. Morphologically, the cell wall can be distinguished in three different parts (**Figure 2.1**). The middle lamella is synthesized as the first layer.
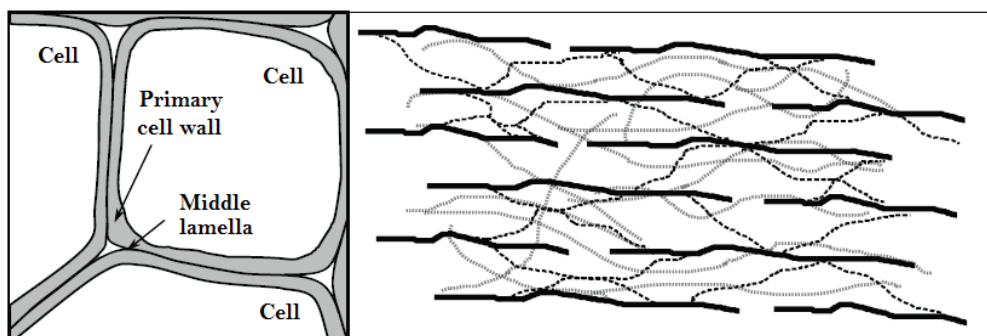
Figure 2.1: Model of the plant cell wall. *(left)* Schematic view on a single cell surrounded by the primary cell wall; the middle lamella of the cell walls are thought to connect and interlink neighboring cells (McCann *et al.*, 1992). *(right)* Simple model of the primary cell wall of flowering plants (Carpita and Gibeaut, 1993). Cellulose microfibrils (black horizontal rods) are linked by hemicellulose chains (dotted black lines). The cellulose-hemicellulose network is embedded in a matrix of pectins (gray dotted lines).

The cell wall is pectin-rich, has a high water-holding capacity and is a deposit for sugars, water and ions (Reiter, 2002). The middle lamella is very flexible allowing

the cells to quickly expand (Verhertbruggen and Knox, 2006). It is the outermost layer of the cell wall and is, in later stages of development, responsible for cell-cell signaling and pathogen-attack response mechanisms (Cosgrove, 2005). After the middle lamella the primary plant cell wall is synthesized. It mainly consists of celluloses and hemicelluloses forming a more rigid but still flexible network (McNeil *et al.*, 1984). When cell growth ceases the secondary cell wall is formed. The hemicellulose-cellulose network is fortified by the secretion and polymerization of phenols, leading to the formation of the stiff, extremely robust lignocellulose network.

Cellulose

Cellulose is a homopolymer of β-1,4-linked glucose units that is synthesized at the plasma membrane by the cellulase-synthase complex. The ascending chains are released apoplastically, where they group to cellulose microfibrils (Mutwil *et al.*, 2008). Glucose chains are tightly bound to each other by van der Waals forces and hydrogen bonds into crystalline structures called elementary fibril (consisting of around 40 glucan chains), about 40 Å wide, 30 Å thick and 100 Å long (Bidlack *et al.*, 1992). Aggregates of elementary fibrils, of essentially an infinite length, and a width of approximately 250 Å, are called microfibrils (Fan, *et al.*, 1982).These microfibrils are bundles of 36 cellulose chains of approximately 200 nm length, that provide mechanical strength to the cell wall. It consists of amorphous regions of loosely arranged fibers and crystalline regions that widely persist enzymatic degradation. On average, cellulose constitutes 10–30 w/w% of the total cell wall (McNeil *et al.*, 1984).

Regions within the microfibrils with high order are termed *crystalline*, and less ordered regions are termed *amorphous*. The term "*amorphous*" cellulose is widely accepted even though it can be contradictory. Amorphous material is defined as material which is formless or lacks definite shape, however, amorphous cellulose probably still possesses some degree of order (O'Sullivan, 1997). Larsson, *et al.* (1997), investigated molecular ordering of cellulose and reported that most of the amorphous regions correspond to the chains that are located at the surface, whereas crystalline components occupy the core of the microfibril (**Figure 2.3.A**). A different molecular architecture of crystalline and amorphous cellulose is suggested by Moiser *et al.* (1999) and Tenkanen *et al.* (2003). They describe cellulose as being *semi-crystalline*, with

regions of high crystallinity averaging approximately 200 glucose residues in length separated by amorphous regions, as illustrated at **Figure 2.3.B**.
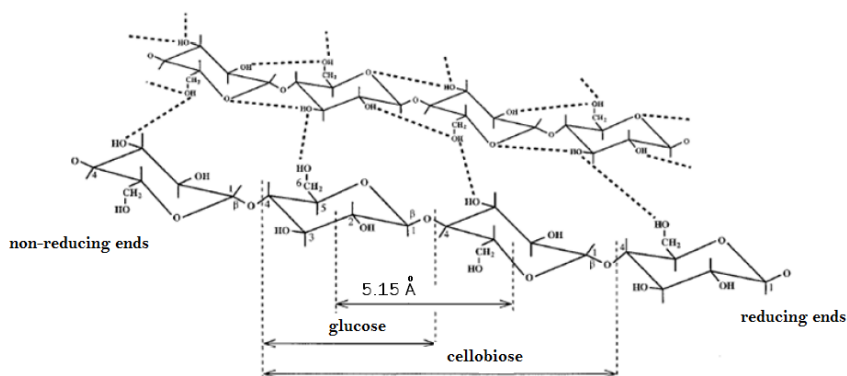
Figure 2.2: Chemical structure of cellulose. Linear β-1,4-linked glucose is the chemical repeating unit, while the structural repeat is β-cellobiose, and consequently each glucoside is oriented at 180° in respect to its neighbors (Hilden and Johansson, 2004).
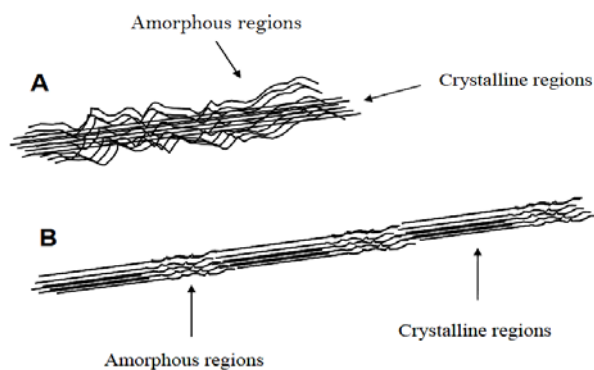
Figure 2.3: Two different views on how crystalline and amorphous cellulose is distributed within the microfibril. **A:** Crystalline cellulose is in the core of the microfibril, and it is surrounded by amorphous substrate. **B:** Crystalline and amorphous regions are being repeated in horizontal dimension.

Cellulose exists in seven crystal structures (polymorphs) designated as celluloses Iα, Iβ, II, IIII, IIIII, IVI, and IVII (O'Sullivan, 1997). In nature, cellulose Iα and Iβ are the most abundant crystal forms. Iα polymorph is meta-stabile, and thus, more reactive

than Iβ. The co-existence of two polymorphs of native cellulose, which have different stabilities, may imply that the part of the Iα polymorph within the microfibril is most prone to the enzymatic attack.

Hemicelluloses

Hemicelluloses are a heterogenic group of polysaccharides that share a β-1,4-linked backbone as a structural motif (Scheller and Ulvskov, 2010). The backbone can be composed of glucose (in xyloglucans and β-glucans), xylose (in xylans, arabinoxylans and glucuronoarabinoxylans), mannose (in mannans and galactomannans) or mannose and glucose (in galactoglucomannans). In contrast to cellulose, the backbone of hemicelluloses is usually substituted. In some plants, xylans can be substituted with arabinose, xylose, glucuronic acid, acetic acid, coumaric acid, ferulic acid or more complex oligomers of these substituents. Because of their cellulose-like backbone structure, hemicelluloses, like xyloglucan and xylan, can bind non-covalently to cellulose (Levy *et al.*, 1997).
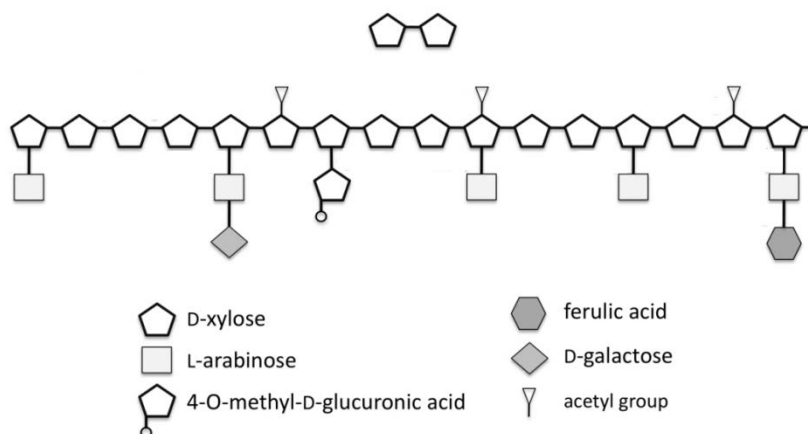
Figure 2.4. Schematic structure of hemicellulose.

Lignin

Lignin is probably the most complex and the least characterized molecular group among the wood components. Its primary purpose is to give strength and water permeability to plants, but also to protect plants from pathogen infections. Lignin is

composed of *p*-hydroxyphenoyl, guaiacyl and/or syringyl monomers linked in three dimensions. These three monomers differ in the methoxylation pattern of the aromatic ring (Douglas, 1996). As it is the case for hemicellulose, the composition and amount of lignin present in the woody material varies according to species, cell type and stage of tissue development. Lignin accounts for approximately 20-35 % of wood structure (Fan *et al.*, 1982).

2.1.2. Cell-wall degrading enzymes

Types of cell-wall degrading enzymes

Microorganisms like bacteria and fungi are a rich source of cell-wall degrading enzymes. Generally, cell-wall degradation requires two types of enzymes; *exo-enzymes* remove single sugars or small oligomers from the ends of a polymer and hydrolyze oligomers down to monomers, whereas *endo-enzymes* cleave linkages within a polymer backbone. Endo-action leads to a rapid decrease in average molecular mass of the substrate and to the formation of new ends that may be attacked by exo-enzymes. Most enzymes are rather *substrate-specific*, which means that each cell-wall polysaccharide requires a specific set of glucosidases for its degradation. In addition to glucosidases, different esterases are required for the removal of non-sugar substituents, like methylesters, acetyl groups and feruloyl groups. Auxiliary enzymes, like the recently identified lytic-polysaccharide monooxygenases, have been shown to significantly increase the accessibility of cellulases to cellulose by the oxidative cleave/disruption of crystalline cellulose region (Quinlan *et al.*, 2011; Phillips *et al.*, 2011).

Classification of cell-wall degrading enzymes

Cell-wall degrading enzymes can be classified into three main classes of enzymes: *Glycoside hydrolases* (E.C. 3.2.1), *Polysaccharide lyases* (E.C. 4.2.2) and *Carbohydrate esterases* (E.C. 3.1.1, Webb, 1992). Hydrolases cleave the glycosidic linkage between two sugar moieties with the addition of one water molecule. Lyases cleave the glycosidic linkage by introducing a double bond. The carbohydrate esterases are a heterogeneous group of enzymes that contain pectin methyl esterases, pectin and rhamnogalacturonan acetyl esterases and hydroxycinnamic acid esterases, like ferulic

acid esterase. An attempt has been made to classify carbohydrate active enzymes based on their enzymatic cleavage mechanism and according to their amino acid sequence. The Carbohydrate-Active enZYme (CAZy) database, therefore, subdivides the enzyme classes glycosyl hydrolase (GH), glycosyl lyase and carbohydrate esterase into different families (Coutinho and Henrissat, 1999). The members of one family have similar structural motifs, but they may have different substrate specificities and modes of action. Analogously, enzymes with the same substrate specificities and modes of action may belong to different families. A very detailed data collection is provided by the enzyme database (BRENDA, http://www.brendaenzymes.info) that intents to summarize all accessible data of known enzymes (Scheer *et al.*, 2011).

Cellulose-degrading enzymes

Cellulases are modular enzymes that are composed of independently folded, structurally and functionally discrete units, referred to as either domains or modules (Henrissat *et al.*, 1998). Most commonly, cellulases consist of one *catalytic domain* (CD) and one *carbohydrate binding module* (CBM), which is usually joined to the CD by a relatively long (30-44 amino acids), often glycosylated, *linker peptide.*

By definition, a CBM is a contiguous amino acid sequence within a carbohydrate active enzyme with a restrained fold and independent carbohydrate-binding activity. It is generally accepted that the primary role of CBM is to accommodate physical contact of the enzyme to the cellulose, increasing at the same time both the effective concentration of the enzyme, but also the time the enzyme will spend in the near proximity of the substrate. CBMs are currently distributed within 49 families, ranging from small (30-40 amino acids), family 1, peptides, to modules consisting of over 200 residues (in families 11 and 17). All fungal CBMs (relevant for the enzymes used during this thesis) belong to family 1. Those peptides primarily demonstrate affinity for crystalline cellulose. The cellulose binding surface has been shown to be a planar surface with three aromatic amino acids and few conserved polar residues (Linder *et al.*, 1995; Mattinen *et al.*, 1998). Such a binding specificity implies that, in perfect cellulose crystals, the surface area of the proposed binding site for the CBMs is very limited. Lehtio *et al.*, 2003 observed fully reversible binding of family 1 CBM to crystalline

cellulose at 4°C. Reversibility of CBM binding to the cellulose is an important issue as it will promote the hydrolysis reaction to proceed from another point on the crystal, i.e. enzyme loss due to unproductive binding is minimized. Enzymes lacking CBM, i.e. only having one module (catalytic domain), have been shown to still have the ability to absorb to cellulose, but often with lower affinity compared to the full length enzyme (Karlsson *et al.*, 2002). CD and CBM are connected by linker peptide. The linker sequences from different enzymes rarely share any apparent sequence homology, but their amino acid composition is typically rich in proline and hydroxyl amino acids (Gilkes *et al.*, 1991). It has been suggested that linkers represent extended, flexible hinges between the two domains facilitating their independent function (Burton *et al.*, 1989; Bushuev *et al.*, 1989).
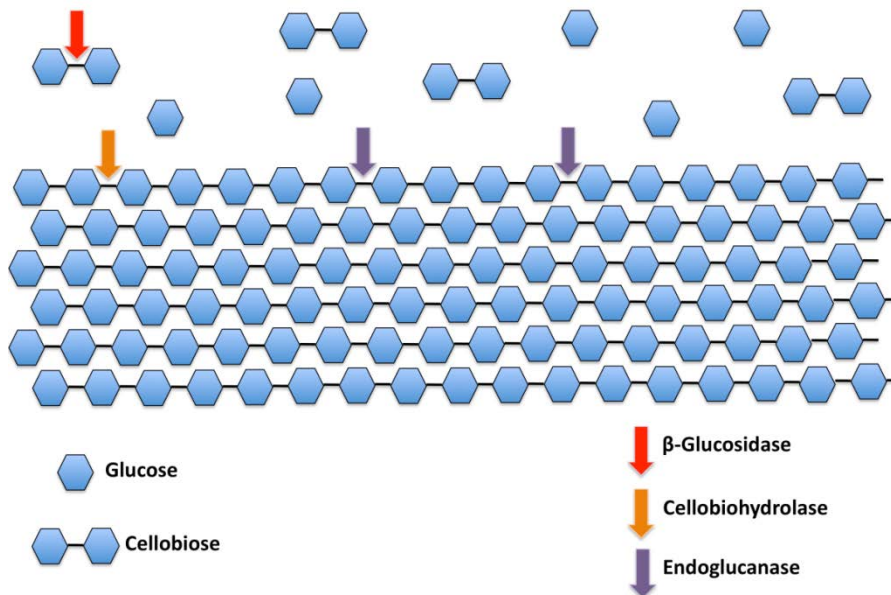
Figure 2.5. Enzymatic activities associated with cellulose deconstruction. The final product is glucose, the main carbon source readily metabolized by fungi (de Souza, 2013).

Cellulose degradation is attributed to the synergistic action of three complementary enzyme activities: (1) *endoglucanases* (EGs, EC 3.2.1.4); (2) *exoglucanases*, including cellodextrinases (EC 3.2.1.74) and cellobiohydrolases (CBHs, EC 3.2.1.91 for

the non-reducing end acting CBHs and EC 3.2.1.176 for the reducing end acting ones) and (3) *β-glucosidases* (BGs, EC 3.2.1.21) (Lynd *et al.*, 2002). Amorphous regions of the polysaccharide chain are cleaved randomly by EGs, while CBHs remove processively cellooligosaccharides from chain ends. The latter are the most abundant enzymes in the secretome of cellulolytic fungi (Jun *et al.*, 2011; Ribeiro *et al.*, 2012). Their main representatives are GH family 7 (CBH I) that attack the reducing end of a cellulose chain and GH family 6 (CBH II) that are specific toward the non-reducing end of the chain. Until very recently, CBHs were considered as the main degraders of the crystalline part of cellulose (Sweeney and Xu, 2012). Cellulose degradation may also require the presence of non-catalytic proteins, such as expansins, to make cellulose more accessible (amorphous) to enzymes (Cosgrove, 2000).

EGs are widespread among GH families, with examples described for families 5–9, 12, 44, 45, 48, 51, and 74 on the continually updated CAZy database (http://www.cazy.org/; Lombard *et al.*, 2014). Most of them show optimal activity at neutral or acidic pH and at temperatures below 50°C (Maheshwari *et al.*, 2000). Exo-glucanases (or CBHs) act in a processive manner (Davies and Henrissat, 1995) and are classified only to two families, as referred previously. One of the important features of all CBHs is that they can act on microcrystalline cellulose (Terri, 1997). BGs include enzymes of GH1 and GH3 families that hydrolyze cellobiose and short (soluble) cellooligosaccharides to glucose that could subsequently fermented to ethanol; e.g., the hydrolysis reaction is performed in the liquid phase, rather than on the surface of the insoluble cellulose particles, such as EGs and CBHs. The removal of cellobiose is an important step of the enzymatic hydrolysis process, as it assists in reduction of the inhibitory effect of cellobiose on EG and CBH. BG activity has often been found to be rate-limiting during enzymatic hydrolysis of cellulose (Duff and Murray, 1996; Tolan and Foody, 1999), and due to that the commercial cellulase enzyme preparations are often supplemented with BG activity.

Hemicellulose-degrading enzymes

Hemicellulose polymers have a much more diverse structure than cellulose and consequently several enzymes are needed to completely degrade the polysaccharide into monosaccharides. Xylan that is the major component of hemicellulose in the plant

cell wall, is consisted of a β-D-(1,4)-linked xylopyranosyl backbone, which, depending on the origin, can be substituted with arabinofuranosyl, 4-0-methylglucopyranosyl, feruloyl and acetyl groups (Shibuya and Iwasaki, 1985). Feruloyl groups can form strong networks through peroxidase-catalyzed oxidative coupling forming diferuloyl bridges (Topakas *et al.*, 2007). The main enzymes needed for depolymerization are xylanases, assisted by accessory enzymes such as β-xylosidases and different arabinofuranosidases making the xylan backbone more accessible (Sørgensen *et al.*, 2007). Other accessible enzymes that enhance xylan degradation are acetyl-xylanesterases (Poutanen *et al.*, 1990), ferulic acid esterases (Topakas *et al.*, 2007), and α-glucuronidases (De Vries *et al.*, 1998).
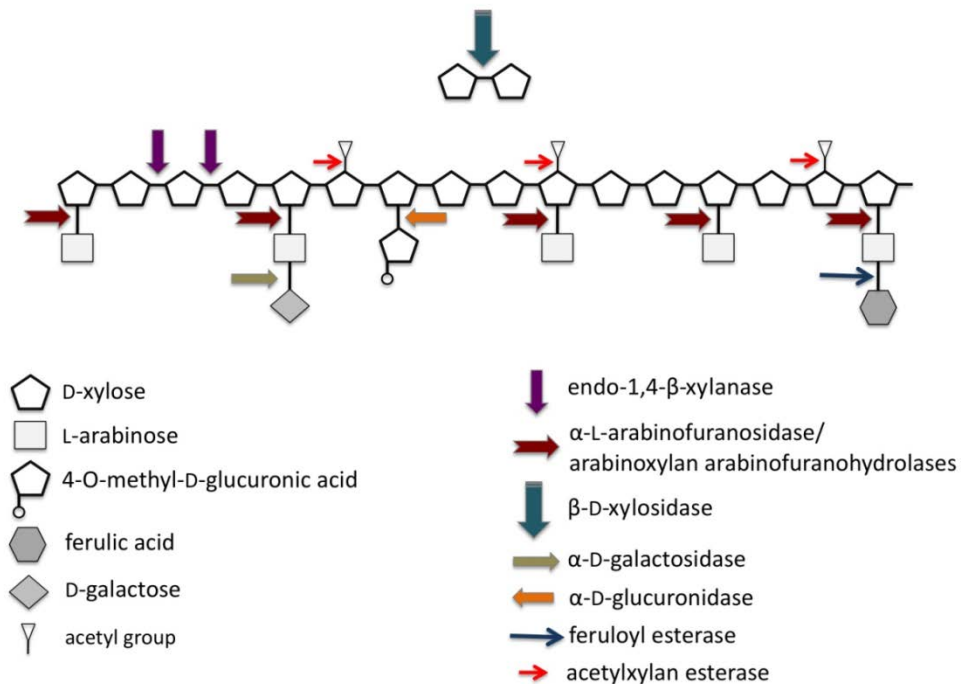
Figure 2.6. Enzymatic activities associated with hemicellulose deconstruction.. The arrows represent each enzyme active for a determined substrate (de Souza, 2013).

Xyloglucan specific exo-β-1,4-glucanase (Xgl74A; EC 3.2.1.155) is classified to GH74 family and catalyzes the hydrolysis of (1-4)-D-glucosidic linkages in xyloglucans aiming in the successful removal of oligosaccharides from the chain end (Grishutin *et*

al., 2004). Xyloglucan is a major structural polysaccharide found in the primary cell walls of higher plants that interact with cellulose microfibrils via hydrogen bonds to form a structural network that is assumed to play a key role in cell wall integrity. It consists of a cellulose-like backbone of β-1,4-linked D-glucopyranose (D-Glcp) residues, which most of them are substituted at C-6 with α-d-Xylp-(1→6) residues, to which other saccharides may be attached (most frequently, d-Galp and l-Fucp).

Xylan backbone is degraded by endo-xylanases and β-xylosidases. Xylanases (endo-1,4-β-xylanases, EC 3.2.1.8) are enzymes hydrolyzing β-1,4-glycosidic linkages in the backbone of xylans, while most of them belong to GH family 10 or 11 based on amino acid similarities and structural features (Henrissat, 1991). GH10 xylanases exhibit less substrate specificity than GH11 enzymes and can hydrolyze different types of decorated xylans, while GH11 xylanases are highly specific and do not tolerate many decorations on the xylan backbone (Biely *et al.*, 1997). β-Xylosidases (EC 3.2.1.37) hydrolyze the soluble xylo-oligosaccharides and xylobiose from the non-reducing end liberating xylose, produced by the activity of xylanases. These enzymes play an important role in xylan degradation by relieving the end product inhibition of endoxylanases (Knob *et al.*, 2010).

Mannan polymer primarily consists of a backbone structure composed of β-1,4-bound mannose residues or combination of glucose and mannose residues and can be hydrolyzed to its monomers with the synergistic action of β-mannanases (EC 3.2.1.78), β-mannosidases (EC 3.2.1.25), α-galactosidases (EC 3.2.1.22), and acetylmannan esterases (E.C. 3.1.1.6) (McCleary, 1988). The backbone of glucomannans can be attacked endoglucanases as well.

Pectin-degrading enzymes

Homogalacturonan degradation requires the action of an exo-polygalacaturonase and an endo-polygalacturonase (endoPG). *Rhamnogalacturonan I* is degraded by RG-lyases or RG-hydrolases (Mutter *et al.*, 1998). Acetyl groups are removed by pectin acetyl esterases. For the degradation of *galactan and arabinogalactan I*, endo-galactanase, β-galactosidase and arabinofuranosidase activities are needed.

Arabinan is degraded by a number of different enzymes that belong to the CAZy glycoside hydrolase (GH) families 3, 27, 43, 51 54, 62 and 93 (Coutinho and Henrissat, 1999, http://www.cazy.org). α-L-arabinofuranosidases (AFase; EC 3.2.1.55) are enzymes that release arabinofuranose residues substituted at position O-2 or O-3 of mono or di-substituted xylose residues (Gruppen *et al.*, 1993). Apart from that, AFases act in synergism with other arabinohydrolases, endo-(1,5)-α-L-arabinanases (ABNase; EC 3.2.1.99) for the decomposition of arabinan, a major pectin polysaccharide. Arabinan consists of a backbone of α-(1,5)-linked L-arabinofuranosyl residues, some of which are substituted with α-(1,2)- or α-(1,3)-linked arabinofuranosides (Weinstein and Albersheim, 1979). Degradation of arabinan polymer to arabinose sugars is driven by the synergistic action of two major enzymes, AFases and ABNases (Kim, 2008). AFases specifically catalyze the hydrolysis of terminal non-reducing L-arabinofuranosyl residues from arabinan, while the resulting debranched backbone could be efficiently hydrolyzed by endo-acting ABNases, thus generating a variety of arabino-oligosaccharides with an inverting mode of action (Beldman *et al.*, 1997).

The role of auxiliary enzymes

Until recently, only hydrolytic enzymes were thought to play a role in the degradation of recalcitrant cellulose and hemicelluloses to fermentable sugars. Recent studies demonstrate that enzymes from the GH family 61 show lytic polysaccharide monooxygenase activity (LPMO) and have an enhancing cellulolytic effect when combined with common cellulases (Horn *et al.*, 2012). Together with cellobiose dehydrogenase (CDH; EC 1.1.99.18), an enzymatic system capable of oxidative cellulose cleavage is formed, which increases the efficiency of cellulases and boosts the enzymatic conversion of lignocellulose. It has long been thought that the proteins of GH family 61 are accessory proteins enhancing cellulose decomposition. They were thus frequently referred to as the "cellulose enhancing factors" (Harris *et al.*, 2010) and previously thought to have no or only weak endoglucanase activity (Karlsson *et al.*, 2001). Now, these enzymes are now reclassified to AA9 family of CAZy database and their mode of action provide a new dimension to the classical concept of cellulose degradation, as recently reviewed by Dimarogona *et al.* (2013).
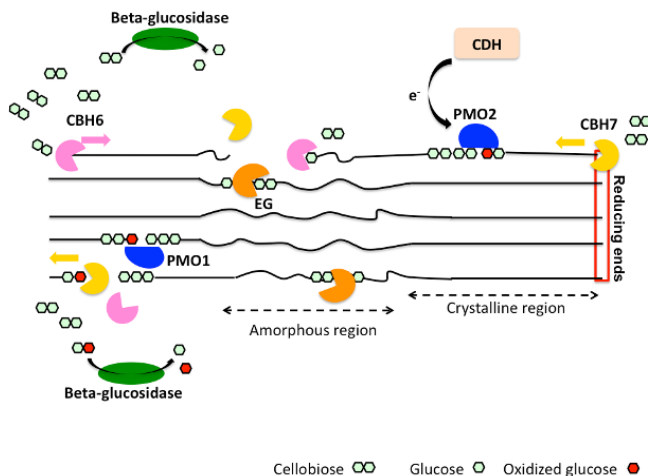
Figure 2.7. Current view of the components that are involved in the enzymatic degradation of cellulose, involving *cellobiohydrolases* (CBH), *endoglucanases* (EG), *type1 and type 2 PMOs* (PMO1 hydroxylating the C1 position of the glucose moiety and PMO2 being specific for C4, respectively). *Cellobiose dehydrogenase* (CDH) is a potential electron donor for PMOs. (Dimarogona *et al.*, 2012).

These copper-dependent enzymes were shown to cleave cellulose by an oxidative mechanism provided that reduction equivalents from CDH or low molecular weight reducing agents (e.g., ascorbate) are available (Langston *et al.*, 2011). In some genomes, AA9 genes even outnumber cellulose genes. It remains to be elucidated whether all of these encoded enzymes have PMO activity, but their large number emphasizes the importance of oxidative cellulose cleavage. *M. thermophila*'s genome has 25 AA9 genes, encoding putative proteins acting as accessory LPMOs enzymes (Berka *et al.*, 2011). This number is outstanding in comparison to common lignocellulolytic organisms, as *A. niger* (seven sequences) and *T. reesei* (nine sequences). This difference can explain the high efficiency of hydrolysis of *Myceliophthora* in nature substrates and reveals the crucial role of these enzymes in the whole procedure.

References

Beldman, G., Schols, H. A., Pitson, S. M., Searle-van Leeuwen, M.F., Voragen, A.G.J. 1997. Arabinans and arabinan degrading enzymes, in *Advances in Macromolecular Carbohydrate Research*, Vol. 1, ed Sturgeon R., editor. (Stamford, CT: JAI Press Inc), 1–64.

Berka, R.M., Grigoriev, I.V., Otillar, R., Salamov, A., Grimwood, J., Reid, I., *et al.* 2011. Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris. Nat. Biotechnol.* 29, 922–927.

Bidlack, J., Malone, M., Benson, R. 1992. Molecular structure and component integration of secondary cell wall in plants. *Proc. Okla. Acad. Sci.* 72: 51-56.

Biely, P., Vrsanská, M., Tenkanen, M., Kluepfel, D. 1997. Endo-b-1,4-xylanase families: differences in catalytic properties. *J. Biotechnol.* 57, 151–166.

Burton, J., Wood, S.G., Pedyczak, A., Siemieon, I.Z. 1989. Conformational preferences of sequential fragments of the hinge region of human IgA1 immunoglobulin molecule: II. *Biophys. Chem.* 33: 39-45.

Bushuev, V.N., Gudkov, A.T., Liljas, A., Sepetov, N.F. 1989. The flexible region of protein L12 from bacterial ribosomes studied by proton nuclear magnetic resonance. *J. Biol. Chem.* 264: 4498-4505

Carpita, N. C. and Gibeaut, D. M., 1993. Structural models of primary cell walls in flowering plants: Consistency of molecular structure with the physical properties of the walls during growth. *Plant J. 3* (1), 1–30.

Cosgrove, D.J., 2000. Loosening of plant cell walls by expansins. *Nature* 407, 321–326.

Cosgrove, D.J., 2005. Growth of the plant cell wall. Nat. Rev. *Mol. Cell. Biol.* 6, 850–861.

Coutinho, P.M. and Henrissat, B., 1999. Carbohydrate-active enzymes: An integrated database approach. In: Gilbert, H. J., Davies, G. J., Svensson, B., Henrissat, B. (Eds.), Recent advances in carbohydrate engineering. *Royal Soc. Chem.*, pp. 3–12.

Davies, G. and Henrissat, B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure. 3*, 853–859.

de Souza, W.R. 2013. Microbial Degradation of Lignocellulosic Biomass, Sustainable Degradation of Lignocellulosic Biomass - Techniques, Applications and Commercialization, Chapter 9, ISBN: 978-953-51-1119-1.

de Vries, R.P., Poulsen, C.H., Madrid, S., Visser, J. 1998. AguA, the gene encoding an extracellular a-glucuronidase from *Aspergillus tubingensis*, is specifically induced on xylose and not on glucuronic acid. *J. Bacteriol.* 180, 243–249.

Dimarogona, M., Topakas, E., Christakopoulos, P. 2012. Cellulose degradation by oxidative enzymes, *Comput Struct Biotechnol J*;2:e201209015.

Dimarogona, M., Topakas, E., Christakopoulos, P. 2013. Recalcitrant polysaccharide degradation by novel oxidative biocatalysts. *Appl. Microbiol. Biotechnol.* 97, 8455–8465.

Douglas, C.J. 1996. Phenylpropannoid metabolism and lignin biosynthesis: from weeds to trees. *Trends Plant Sci.* 1: 171-178.

Duff, S.J.B. and Murray, W.D. 1996. Bioconversion of forest products industry waste cellulosics to fuel ethanol: a review. *Biores. Technol.* 55, 1–33.

Fan, L.T., Lee, Y-H., Gharpuray, M.M. 1982. The nature of lignocellulosics and their pretreatments for enzymatic hydrolysis. *Adv. Bichem. Eng/Biotechnol* 23: 158-187.

Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C., Warren, R.A.J. 1991. Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. *Microbiol. Rev. 55*: 303-315.

Grishutin, S. G., Gusakov A. V., Markov A. V., Ustinov B. B., Semenova M. V., Sinitsyn A. P. (2004). Specific xyloglucanases as a new class of polysaccharide-degrading enzymes. *Biochim. Biophys. Acta* 1674, 268–281.

Gruppen, H., Kormelink, F.J.M., Voragen, A.G.J. 1993. Water-unextractable cell wall material from wheat flour. A structural model for arabinoxylans. *J. Cer. Sci.* 18, 111–128.

Harris, P.V., Welner, D., McFarland, K.C., Re, E., Poulsen, J.C.N., Brown, K., *et al.* 2010. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* 49, 3305–3316.

Henrissat, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 280, 309–316.

Henrissat, B. 1998. Enzymatic cellulose degradation. *Cellulose Commun.* 5: 84-90.

Hildén, L. and Johansson, G. 2004. Recent developments on cellulases and carbohydrate binding modules with cellulose affinity. *Biotechnology Letters* 26: 1683-1693.

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

Jun, H., Kieselbach, T., Jönsson, L. 2011. Enzyme production by filamentous fungi: analysis of the secretome of *Trichoderma reesei* grown on unconventional carbon source. *Microb. Cell Fact.* 10, 68.

Karlsson, J., Saloheimo, M., Siika-aho, M., Tenkanen, M., Penttila, M., Tjerneld, F. 2001. Homologous expression and characterization of Cel61A (EG IV) of *Trichoderma reesei. Eur. J. Biochem.* 268, 6498–6507.

Karlsson, J., Siika-aho, M., Tenkanen, M., Tjerneld, F. 2002. Enzymatic properties of the low molecular mass endoglucanases Cel12A (EG III) and Cel45A (EG V) of *Trichoderma reesei. J. Biotechnol.* 99: 63-78.

Kim, T.J. 2008. Microbial exo- and endo-arabinosyl hydrolases: structure, function, and application in L-arabinose production, in Carbohydrate-Active Enzymes, ed Park K. H., editor. (Cambridge: CRC Press; Woodhead Publishing Ltd.), 229–257.

Knob, A., Terrasan, C.R.F., Carmona, E.C. 2010. β-Xylosidases from filamentous fungi: an overview. *World J. Microbiol. Biotechnol.* 26, 389–407.

Langston, J.A., Shaghasi, T., Abbate, E., Xu, F., Vlasenko ,E., Sweeney, M.D. 2011. Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl. Environ. Microbiol.* 77, 7007–7015.

Larsson, P.T., Wickholm, K., Iversen, T. 1997. A CP/MAS 13C NMR investigation of molecular ordering in celluloses. *Carbohydr. Res.* 302: 19-25.

Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., Teeri, T.T. 2003. The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *PNAS.* 100: 484-489.

Levy, S., Maclachlan, G., Staehelin, L. A., 1997. Xyloglucan sidechains modulate binding to cellulose duringin vitro binding assays as predicted by conformational dynamics simulations. *Plant J.* 11 (3), 373–386.

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

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B. 2014. The Carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495.

Lynd, L.R., Weimer, P.J., van Zyl, W.H., Pretorius, I.S. 2002. Microbial cellulose utilization: fundamentals and biotechnology *Microbiol molec biol rev* 66: 506-77.

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

Mattinen, M.L., Linder, M., Drakenberg, T., Annila, A. 1998. Solution structure of the cellulosebinding domain of endoglucanase I from *Trichoderma reesei* and its interaction with cellooligosaccharides. *Eur. J. Biochem.* 256: 279-286.

McCann, M. C., Wells, B., Roberts, K., 1992. Complexity in the spatial localization and length distribution of plant cell-wall matrix polysaccharides. *J. Microscopy* 166 (1), 123–126.

McCleary, B.V. 1988. β-D-Mannanase. *Meth. Enzymol.* 160, 596–610.

McNeil, M., Darvill, A. G., Albersheim, P., 1984. Structure of plant cell walls: X. Rhamnogalacturonan I, a structurally complex pectic polysaccharide in the walls of suspension cultured sycamore cells. *Plant Physiol.* 66 (6), 1128–1134.

Moiser, N.S, Hall, P., Ladisch, C.M., Ladisch, M.R. 1999, Reaction kinetics, molecular action and mechanisms of cellulosic proteins. *Adv. Biochem. Eng. Biotehnol.* 65: 23-40.

Mutter, M., Renard, C. M. G. C., Beldman, G., Schols, H. A., Voragen, A. G. J., 1998. Mode of action of RGhydrolase and RG-lyase toward rhamnogalacturonan oligomers. Characterization of degradation products using RG-rhamnohydrolase and RG-galacturonohydrolase. *Carbohyd. Res.* 311 (3), 155–164. 20.

Mutwil, M., Debolt, S., Persson, S., 2008. Cellulose synthesis: A complex complex. *Curr. Op. Plant Biol.* 11, 252–257. 12

O'Sullivan, 1997. Cellulose: the structure slowly unravels. *Cellulose* 4: 173-207.

Phillips, C.M., Beeson, W.T., Cate, J.H., Marletta, M.A. 2011. Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa. Acs Chemical Biology* 6:1399–1406.

Poutanen, K., Sundberg, M., Korte, H., Puls, J. 1990. Deacetylation of xylans by acetyl esterases of *Trichoderma reesei. J. Appl. Microbiol. Biotechnol.* 33, 506–510.

Quinlan, R.J., Sweeney, M.D., Lo Leggio, L., Otten, H., Poulsen, J.C.N., Johansen, K.S., Krogh, K.B.R.M., Jorgensen, C.I., Tovborg, M., Anthonsen, A., *et al.* 2011. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci USA* 108:15079–15084.

Reiter,W. D. 2002. Biosynthesis and properties of the plant cell wall. *Curr. Opin. Plant Biol.* 5 (6), 536–542.

Ribeiro, D.A., Cota, J., Alvarez, T.M., Brüchli, F., Bragato, J., Pereira, B.M., *et al.* 2012. The *Penicillium echinulatum* secretome on sugar cane bagasse. PLoS ONE. 7:e50571.

Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Söhngen, C., Stelzer, M., Thiele, J., Schomburg, D., 2011. BRENDA, the enzyme information system in 2011. *Nucl. Acids Res. 39* (suppl 1), D670–D676.

Scheller, H.V. and Ulvskov, P., 2010. Hemicelluloses. *Ann. Rev. Plant Physiol.* 61, 263–289.

Shibuya N. and Iwasaki T. 1985. Structural features of rice bran hemicellulose. *Phytochemistry* 30, 488–494.

Sørensen, H. R., Pedersen, S., Jørgensen, C. T., Meyer, A. S. (2007). Enzymatic hydrolysis of wheat arabinoxylan by a recombinant "minimal" enzyme cocktail containing a-xylosidase and novel endo-1,4-a-xylanase and a-L-arabinofuranosidase activities. *Biotechnol. Prog. 23*, 100–107.

Sweeney, M.D. and Xu, F. 2010. Biomass converting enzymes as industrial biocatalysts for fuels and chemicals: recent developments. *Catalysts 2*, 244–263.

Tenkanen, M, Niku-Paavola M-L, Linder M, Viikari L. 2003. Cellulase in food processing Iin *Handbook of food enzymology.* New York: Marcel Dekker Inc. p 879-915.

Terri, T.T. 1997. Crystalline cellulose degradation: a new insight into the function of cellobiohydrolases. *Trends Biotechnol.* 15, 160–167.

Tolan, J.S. and Foody, B. 1999. Cellulases from submerged fermentation. *Adv. Biochem. Eng. Biotechnol.* 65, 41–67.

Topakas, E., Vafiadi, C., Christakopoulos, P. 2007. Microbial production, characterization and applications of feruloyl esterases. *Proc. Biochem.* 42, 497–509.

Verhertbruggen, Y. and Knox, J.P., 2006. Pectic polysaccharides and expanding cell walls. Vol. 5. Springer-Verlag, Berlin, Heidelberg, Ch. 5, pp. 139–158.

Webb, E.C. 1992. Enzyme nomenclature 1992: Recommendations of the NC-IUBMB on the nomenclature and classification of enzymes. *Academic Press.*

Weinstein, L. and Albersheim, P. 1979. Structure of plant cell walls: IX. Purification and partial characterization of a wall-degrading endo-arabanase and an arabinosidase from *Bacillus subtilis. Plant Physiol.* 63, 425–432.

2.2 *Myceliophthora thermophila* as a thermophilic fungus

Myceliophthora thermophila (synonym *Sporotrichum thermophile)* is a thermophilic filamentous fungus, isolated from soil in eastern Russia, classified as an ascomycete, and constitutes an exceptionally powerful cellulolytic organism; it synthesizes a complete set of enzymes necessary for the breakdown of cellulose. The cell density and the exponential growth rate of the microorganism on cellulose (0.09 to 0.16 h^{-1}) is similar to that on glucose (0.1 h^{-1}), revealing its remarkable ability to utilize cellulose as efficiently as glucose (Bhat, 1987). Several isolates of *M. thermophila* can grow on cellulose-rich material and can decompose complex substrates such as birch chips, wood pulp and wheat straw (Bhat & Maheshwari 1987). Colonies of *M. thermophila* grow rapidly and initially appear cottony-pink, but then turn cinnamon-brown and granular in texture. Microscopic examination reveals septate hyphae with several obovoidal to pyriform conidia arising singly or in small groups from conidiogenous cells. Conidia are typically 3.0 − 4.5 μm x 4.5 - 11.0 μm in size, hyaline, smooth, and thick-walled. Occasionally a secondary conidium can form at the distal tip of primary conidium (van Oorschot 1977).
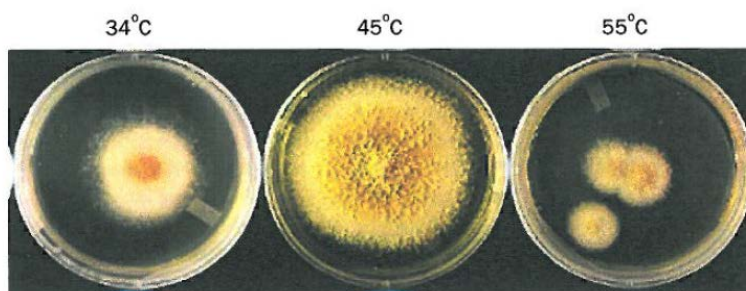
Figure 2.8. Colonies grow rapidly at 45 °C in comparison to 30°C or 55°C and appear dry, thin, broadly spreading with a surface texture that varies from floccose or cottony to granular or powdery. The color is at first white, then pink, buff, and finally fulvous or cinnamon brown. Their hyphae are colorless, about 2μm broad (Morgenstern *et al .,* 2012).

Myceliophthora thermophila has a wide range of synonyms over the history of its classification and distinction of sexual states. This fungus was originally described as *Sporotrichum thermophilum* (Apnis, 1963) but it was later found that the species lacked

clamp connections, a typical characteristic of the basidiomycetous genus, *Sporotrichum*. It was then reclassified to the ascomyceteous genus, *Chrysosporium*, and became known as C. *thermophilum* (von Klopotek, 1974). Two years later Klopotek described *Thielavia heterothallica* as the teleomorph of C. *thermophilum* (von Klopotek 1976) before the genus *Corynascus* was introduced by von Arx in 1983. The genus *Myceliophthora* was not used to describe the anamorph until 1977 (van Oorschot 1977). The current name for the teleomorph is *Corynascus heterothallicus*, which has been observed through phylogenetic analysis to bear very strong DNA sequence homology to *M. thermophila* (van den Brink *et al.*, 2012). Recently, the C1 wild-type strain VKM F-3500-D, extensively studied by Dyadic International Inc., initially classified as *Chrysosporlum lucknowense*, was turn out to be a *Myceliophthora thermophila* isolate, based on molecular studies (Berka *et al.*, 2011).

Kingdom	*Fungi*
Phylum	*Ascomycota*
Subphhylum	*Pezlzomycotlna*
Class	*Sordarlomycetes*
Subclass	*Sordarlomycetidae*
Order	*Sordarlales*
Family	*Chaetomlaceae*
Genus	*Myceliophthora*
Species	*Myceliophthora thermophila*

Species	*Sporotrichum thermophile*, A. E. Aipinis 1963
Synonyms	*Crysosporium thermophilum*. A.E. Alpinls von Klopotek 1974
	Thlelalvia heterothallica, A.E. Alpinis C.A. van Oorschot 1977
	Corynascus heterothallicus, J.A. von Arx 1983
	Myceliopthora thermophila, A.E. Aplinis, C.A. van Oorschot 197

Table 2.1. Taxonomy of *Myceliopthora thermophila* and synonyms of the species (Uniprot taxonomy ID: 78579).

The 38.7 Mbp genome of *M. thermophila*, comprising about 9500 genes, organized in 7 chromosomes, has been sequenced and annotated (Joint Genome Institute. University of California, http://genome.jgi-psf.org; Berka *et al.*, 2011). It revealed a large number of genes putatively encoding industrially important enzymes, such as carbohydrate-active enzymes (CAZy), proteases, oxido-reductases and lipases, while more than 200 sequences have been identified exclusively for plant cell-wall-degrading enzymes. These sequences encode a large number of glycoside hydrolases (GH) and polysaccharide lyases, covering the most of the recognized families, as summarized in **Table 2.2**. In addition, *M. thermophila* was developed into a proprietary mature enzyme production system with easy scaling (C1 strain; Visser *et al.*, 2011). The main features of C1 are the high production levels (up to 100 g/l protein), as well as the maintenance of low viscosity levels at the culture medium, thus enabling fermentation process to reach very high densities.

M. thermophila exhibits an impressing number of accessory enzymes belonging to AA9 (previously described as GH61) and family 1 carbohydrate binding modules (CBM), which are the highest found in fungi (Berka *et al.*, 2011). Family 1 CBM presents a cellulose-binding function and is almost exclusively found in enzymes of fungal origin (http:;;www.cazy.org; Guillen *et al.*, 2009). In addition, M. *thermophila* distinguishes itself from other cellulolytic fungi, such as *Aspergillus niger* and *Trichoderma* reesei by the presence of a relatively high number of (glucurono) arabinoxylan degrading enzymes (Hinz *et al.*, 2009). Eleven putative xylanases were found that belong into GH 10 and 11 families compared to five in both *A. niger* (Broad Institute of Harvard and MIT, http://www.broadinstitute.org) and *T*. reesei (Joint Genome Institute, University of California, http://genome.jgi-psf.org), while fourteen arabinofuranosidases belonging to GH 43 51 and 62 families were found compared to thirteen in A. *niger* and three in *T*. reesei, rendering *M. thermophila* a promising source of hemicellulolytic enzymes. Studying the secretome of *M. thermophila* after 30 h of growth in barley and alfalfa straws, it was found to comprise of 683 predicted proteins, 230 of which are proteins with unknown function (Berka *et al.*, 2011). Based on transcriptome analysis, many of these proteins were found at the secretome, including accessory enzymes, hypothetical proteins and proteins with unknown function that were upregulated when the fungus is grown up in more complex substrates, such as

agricultural straws, compared to glucose. This is an indication of their crucial role in lignocellulose degradation (Berka *et al.*, 2011). *M. thermophila* grows in temperatures between 25 and 55°C, while a relative growth performance study on mycobroth agar plates indicated that the optimum condition is at 45°C (Morgenstern *et al.*, 2012). At a suboptimal temperature (30°C), when grown on cellulose, the conidia tend to form a very limited mycelium that precociously develops asexual reproductive structures (microcycle conidiation). Although the mechanism of this cellular response is not understood, microcycle conidiation (**Figure 2.9**) may be a survival strategy of producing propagules in the shortest possible time under suboptimal conditions (Maheshwari *et al.*, 2000). The temperature optima for several enzymes with the same specific activity that are produced and have been characterized from *M. thermophila*, range from 50°C to 70°C. For example, StEG5 endoglucanase, expressed in *A. niger*, exhibits a T_{opt} of 70°C (Tambor *et al.*, 2012), while recombinant *Mt*EG7 expressed in *Pichia pastoris* exhibited an optimal temperature of 60°C (Karnaouri *et al.*, 2014). The same characteristic is also observed for *M. thermophila* xylanases expressed in *A. niger*, showing optimal activity at temperatures between 50°C and 70°C (Berka *et al.*, 2011), underpinning the enzymatic potential that is not only diverse in catalytic activities, but also in properties increasing its efficiency in various temperatures.

Apart from the consortium of enzymes that are encoded by *M. thermophila'*s genome for the degradation of lignocellulosic biomass, a number of thermostable enzymes with important industrial applications are also produced. Among them are broad-specificity HAP- phytases (myoinositol hexakisphosphate phosphohydrolases) that are efficient in breaking down phytic acid and can be used for supplementing livestock feed with phosphorus (Singh & Satyanaraxana, 2010) and laccases that can act as clean substitutes for harmful chemical reagents used in the paper and pulp industry and textile dyes (Berka *et al.*, 1997). Laccases are also useful in ecological restoration through soil bioremediation and ability to degrade rubber as well as to polymerize lignin from waste material from the kraft process, which can be further burned for energy production or used as raw materials for a variety of new products (Gargulak and Lebo, 2000) obtained by processes involving enzymatic or chemical modifications (Sena-Martins *et al.*, 2008). *M. thermophila* is also able to produce a polygalacturonase (PGase), an enzyme with pectinase activity which hydrolyze complex polysaccharides mainly composed of D-galacturonic acid, and has applications in the food and textile

industries (Kaur *et al.*, 2004). Biosynthesis of fructo-oligosaccharides (FOS), such as 1-kestose, 6-kestose and neokestose, has been reported during growth of the fungus on media containing high sucrose concentrations (Katapodis *et al.*, 2004). These oligosaccharides are currently used as feed supplements (Kim *et al.*, 2000) and once they are consumed, they stimulate the growth of *Bifidobacteria* in the gastrointestinal tract of humans and animals with important associated (Kunz & Rudloff, 1993). They comprise a safe food for diabetics, since they are hardly hydrolyzed by the digestive enzymes (Yun, 1996) and they have been related to the decrease in total cholesterol, triglyceride and phospholipid levels in the serum (Rivello-Urgell & Santamaria-Orleans, 2001).

Specific activity		CAZy module(s)	No id. seq.
Cellulases	endoglucanases	GH 5, 7, 12, 45	8
	cellobiohydrolases	GH 6, 7	7
	β-glucosidases	GH 1,3	8
Xylanases	xylanases	GH 10, 11	12
	xylosidases	GH 3, 43	4
Arabinases	endoarabinases	GH 43	3
	exo-arabinases / arabinofuranosidases	GH 43, 51, 62	11
Mannanases	endomannanases	GH 5, 26	3
	mannosidases	GH 2	2
Pectinases	polygalacturonases	GH28	2
	rhamnosidases	GH78	1
	pectin lyases	PL1, PL3, PL4, PL20	8
	pectin esterases	CE 8, 12	4
Esterases	feruloyl esterases	CE 1	4
	acetyl esterases	CE 3, 5, 16	8
	acetylmannanesterases	CE 12	2
	glycuronoyl esterases	CE 15	2

Table 2.2. Number of predicted CAZymes encoded in the genome of *M. thermophila.* **GH**s, Glycoside hydrolases; **CE**s ,carbohydrate esterases and **PL**s, polysaccharide lyases are included, covering the most of the recognized families.
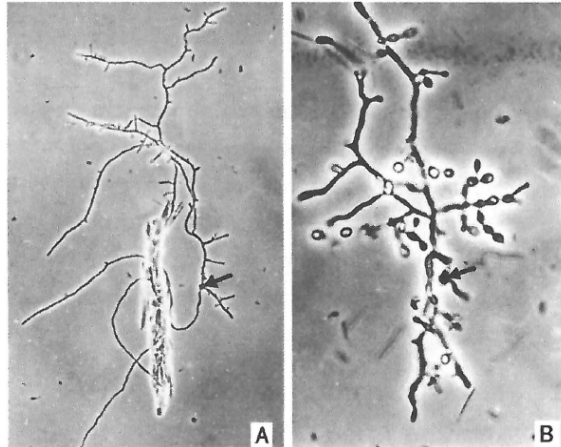
Figure 2.9. Microcycle conidiation in *Myceliophthora thermophila* at 30°C. (**A**) Oval asexual spores in a germ ling, after 24 hour incubation of the fungus in shake cultures with shredded Whatman filter paper as the carbon source. The insoluble particle is a piece of cellulose fiber. (**B**) Precocious differentiation of asexual spores after 72 hours. The germinated conidium is indicated by an arrow (phase-contrast micrographs). Conidia appear on short protrusions, singly or in small groups on ampulliform swellings, obovoidal to pyriform, hyaline, with the size between 4.5-11.0 x 3.0-4.5 µm. They are rough-walled initially and become smooth walled in old cultures, with thick walls. (Maheshwari *et al*., 2000).

References

Apnis, A.E. 1963. Occurrence of thermophilic microfungi.in certain alluvid soils near Nottingham. *Nova. Hed.* 5, 57-78.

Berka R.M., Grigoriev I.V., Otillar R., Salamov A., Grimwood J., Reid I., *et al.* 2011. Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris. Nat. Biotechnol.* 29, 922–927.

Berka R. M., Schneider P., Golightly E. J., Brown S. H., Madden M., Brown K. M., *et al.* 1997. Characterization of the gene encoding an extracellular laccase of *Myceliophthora thermophila* and analysis of the recombinant enzyme expressed in *Aspergillus oryzae. Appl. Environ. Microbiol.* 63, 3151–3157.

Bhat, K.M., Maheshwari, R. 1987. *Sporotrichum thermophile* growth, cellulose degradation and cellulase activity. *Appl Environ Microbial.* 53, 2175·82.

Gargulak, J.D., Lebo, S.E. 2000. Commercial use of lignin-based materials. Historical, Biological, and Materials Perspectives, American Chemical Society, Washington, 305-320.

Hinz, S.W.A., Pouvreau, L., Joosten, R., Bartels, J., Jonathan, M.C., Wery, J., Schols, H.A. 2009. Hemicellulase production in *Chrysosporium* lucknowense *C1. J. Cereal Sci.* 50, 318-323.

Karnaouri, A.C., Topakas, E., Christakopoulos, P. 2014. Cloning, expression, and characterization of a thermostable GH7 endoglucanase from *Myceliophthora thermophila* capable of high-consistency enzymatic liquefaction. *Appl. Microbial. Biotechnol.* 98, 231-42.

Katapodis, P., Kalogeris, E., Kekos, D., Macris, J., Christakopoulos, P. 2004. Biosynthesis of fructo-oligosaccharides by *Sporotrichum thermophile* during submerged batch cultivation in high sucrose media. *Appl. Microbial. Biotechnol.* 63, 378-382.

Kim, B.W., Kwon, H.J., Park, H.Y., Nam; S.W., Park, J.P., Yun, J.W. 2000. Production of a novel transfructosylating enzyme from *Bacillus* macerans EG-6. *Bioprocess Eng* 23, 11-16.

Kunz, C., Rudloff, S. 1993. Biological functions of oligosaccharides in human milk. *Acta Paediatr.* 82, 903-912.

Morgenstern, I., Pawlowski, J., Ishmael, N., Darmond, C., Marqueteau, S., Moisan, M.C., Quenneville, G., Tsang, A. 2012. A molecular phylogeny of thermophilic fungi. *Fung. Bioi.* 116, 489-502.

Rivero·Urgell, M., Santamaria-Orleans, A. 2001. Oligosaccharides: application in infant food. *Early Hum. Dev.* 65, [Supp]:S43-S52.

Sena-Martins, G., Almeida-Vara, E., Duarte, J.C. 2008. Eco-friendly new products from enzymatically modified industrial lignins. *Ind. Crops Prod.* 27, 189-195.

Singh, B. and Satyanarayana, T. 2010. Plant growth promotion by an extracellular HAP-Phytase of a thermophilic mold *Sporotrichum thermophile. Appl Biochem Biotechnol.* 160, 1267-1276.

Tambor, J.H., Ren. H., Ushinsky, S., Zheng, Y., Riemens, A., St-Francois, C., Tsang, A., Pawlowski, J., Storms, R. 2012. Recombinant expression, activity screening and functional characterization identifies three novel endo-1,4-13-glucanases-that efficiently hydrolyse cellulosic substrates. *Appl. Microbial. Biotechnol.* 93,203-14.

van den Brink, J., Samson, R.A., Hagen, F., Boekhout, T., de Vries, R.P. (2012). Phylogeny of the industrial relevant, thermophilic genera *Myceliophthora* and *Corynascus Fung. Divers.* 52, 197-207.

van Oorschot, C.A.N. 1977. The genus *Myceliophthora. Persoonia* 9, 404–409.

van Oorschot, C.A.N. 1980. A revision of *Chrysosporium* and allied genera. Studies in *Mycology* 20.

Visser, H., Joosten, V., Punt, P.J., Gusakov, A.V., Olson, P.T., Joosten, R., Bartels, J., Visser, J., Sinitsyn, A.P., Emalfarb, M.A., Verdoes, J.C., Wery, J. 2011. Development of a mature fungal technology and production platform for industrial enzymes based on a *Myceliophthora thermophila* isolate, previously known as *Chrysosporium lucknowense* C1. *Ind. Biotechnol.* 7, 214-223.

von Klopotek, A. 1974. Revision of thermophilic *Sporotrichum* species: *Chrysosporium thermophilum* (Apinis) comb. nov. and *Chrysosporium fergusii* spec. nov. equal status conidialis of *Corynascus thermophil us* Fergus and (Sinden) comb. nov. *Arch. Microbial.* 98, 365-369.

von Klopotek, A. 1976. *Thielavia heterothallica* spec. nov., die perfekte form von *Chrysosporium thermophilum. Arch. Microbial.* 107, 223.

Yun, J.W. 1996. Fructo oligosaccharides- occurrence, preparation, and application. *Enzyme Microb. Technol.* 19,107- 117.

2.3 *Pichia pastoris* as a methylotrophic yeast

The ability of certain yeast species to utilize methanol as a sole source of carbon and energy was discovered less than 40 years ago by Koichi Ogata. Several years later, in the early 1980s, Phillips Petroleum with the Salk Institute Biotechnology/Industrial Associate Inc. (SIBIA), a biotechnology company located in La Jolla, Calif., developed and established *Pichia pastoris* as a heterologous gene expression system. Nowadays, the methylotrophic yeast *P. pastoris* is a well established eukaryotic host for the production of heterologous proteins preferentially secreted into the medium to simplify further down-stream procedures (Cereghino *et al.*, 2002; Cereghino & Cregg, 2000).

P. pastoris became popular due to the availability of strong inducible and constitutive promoter systems for recombinant protein expression, strong tendency for respiratory growth as opposed to fermentative growth and extremely high cell density fermentation in excess of 100 g/l dry weight (Macauley-Patrick *et al.*, 2005). This methylotrophic yeast is a eukaryote, so its intracellular environment is generally more suitable for correct folding of eukaryotic proteins and it has the ability of post-translational disulfide bond formation, proteolytic processing and processing of signal sequences, disulfide bridge formation, certain types of lipid addition, and *O*- and *N*-linked glycosylation, which may be crucial for biological activity. Moreover, when used as heterologous host, the stable Integration of expression plasmids at specific sites in *P. pastoris* genome, and the simplicity of techniques needed for the molecular genetic manipulation, such as DNA-mediated transformation, gene targeting, gene replacement, and cloning by functional complementation, render it a promiscuous candidate for heterologous protein expression (Higgins 1998; Cereghino & Cregg, 2000).

P. pastoris is a member of *Ascomycete* yeasts, a monophyletic lineage with a single order of about 1000 known species. Asexual production of this genus is described by yeast cells arising by multilateral budding on a narrow base. *P. pastoris* remains haploid unless forced to mate. Strains with complementary markers can be mated by subjecting them to a nitrogen-limited medium. Conjugation between a cell and its bud or between independent cells may occur, leading to the formation of ascosporogenous four-hat-shaped ascospores, indicating this species is homothallic. After 3 days incubation at

25°C, plate culture on agar consists of cells spherical to ovoid in shape, 2-4 x 2.2 - 5.8 µm which occur singly or in pairs, cream colored, dull to faintly glistening and butyrous. Neither pseudohyphae nor true hyphae are produced (Kurtzman and Robnett, 1998). *P. pastoris'* generation time is approximately 90 min in liquid medium containing yeast extract, peptone and dextrose (YPD), and approximately 3 h in defined medium containing yeast nitrogen base with ammonium sulfate and without amino acids and dextrose (YPD). With methanol as sole carbon source and a defined culture medium, the generation time is around 5 h (Cregg *et al.*, 2009). *Pichia pastoris* is phylogenetically distinct from other methanol yeasts and has been reassigned to the newly described genus *Komagataella* following phylogenetic analysis of gene sequences (Yamada *et al.*, 1995). The type strain of *Komagataella pastoris* (synonym *Pichia pastoris*) was isolated from a chestnut tree in France (Guilliermond 1919), but the majority of *P. pastoris* strains have been isolated from exudates and rotten wood from a variety of trees growing in Europe (Diauchy *et al.*, 2003).

Kingdom	*Fungi*
Phylum	*Ascomycota*
Subphylum	*Saccharomycotina*
Class	*Saccharomycetes*
Order	*Saccharomycetales*
Family	*Phaffomycetaceae*
Genus	*Komagataella*
Species	*Komagataella pastoris*

Species	*Komagataella pastoris* **(Gullllennond) Y.Yemada, Mmude, Maeda & Mlkata (1995)**
Synonyms	*Zygosaccharomyces pastoris* Guilllermond (1919)
	Saccharomyces pastori (Gulll iermond) Lodder & Kreger van Rij (1952)
	Petasospora pastori (Guilllermond) Boldin & Abadie (1954)
	Pichia pastoris **(Gullllerrnond) Phaff (1958)**
	Zygowilia pastori (Guiiliermond) Kudryavtsev (1960)
	Zymopichia pastori (Guilliermond) Novak & ZSolt (1961)

Table 2.3. Taxonomy of *Komagataella pastoris* (synonym *P. pastoris*) and synonyms of the species (UniProt taxonomy ID: 4922).
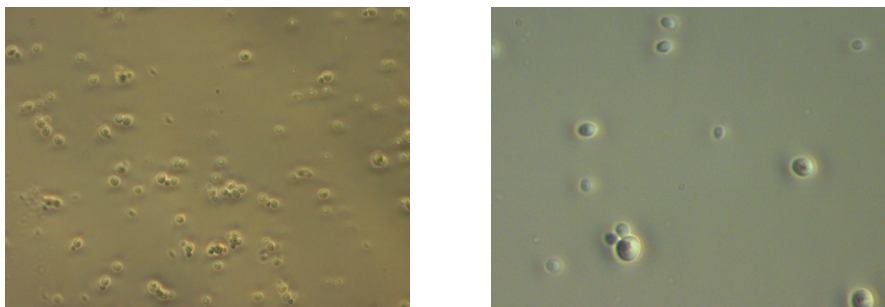
Figure 2.10. *Pichla pastoris* X33 cells growing at fermentor, showing normal budding phenotype, as well as two buds attached to the mother cell (Mickey Mouse-like appearance). Basal Salts Medium, pH 5.0, methanol used as carbon source, magnitude x40 (left) and x100 (right).

2.3.1 Methanol and Glycerol utilization Pathway at *Pichia pastoris*

Methylotrophic yeasts, belonging to genera of *Pichia*, *Hansenula*, *Torulopsis* and *Candida*, possess a general methanol utilization pathway (MUT pathway) that is highly compartmentalized in methanol-induced micro bodies, peroxisomes, and cytoplasm (Faber *et al.*, 1995). This metabolic pathway involves several unique enzymes, which are present at substantial levels only when the cells are grown on methanol. In this metabolic pathway, peroxisomes play an indispensable role since they harbor the three key enzymes, alcohol oxidase. catalase and dihydroxyacetone synthase. The subsequent reactions in methanol assimilation and dissimilation pathway are localized in the cytosol (Veenhuis *et al.*, 1983).

Methanol enters the peroxisome and the enzyme alcohol oxidase (AOX) catalyzes the first step in the dissimilation pathway, the oxidation of methanol to formaldehyde and hydrogen peroxide, utilizing molecular oxygen as an electron acceptor. AOX is sequestered within the peroxisome along with catalase, which further decomposes hydrogen peroxide to oxygen and water. Catalases can either catalatically or peroxidately; which reaction predominates is difficult to establish (Rose & Harrison, 1989). The formaldehyde generated from the methanol oxidation enters both the dissimilatory pathway to yield energy and the assimilatory pathway for generation of biomass. In the dissimilatory pathway, a portion of the formaldehyde generated by AOX enters the cytosol, where it forms a complex with reduced glutathione, S-

formylglutathlone, and is further oxidized to carbon dioxide by two subsequent NAD^+-dependent dehydrogenase reactions which are a source of energy for cells growing on methanol. In the first step, formaldehyde dehydrogenase catalyzes the production of formate, from which carbon dioxide is generated by the action of formate dehydrogenase (Subramani 1998).

Mode of action	Reaction catalysed
General mode	$H_2O_2 + RH_2 \rightarrow 2\ H_2O + R$
Catalatic	$H_2O_2 + H_2O_2 \rightarrow 2\ H_2O + O_2$
Peroxidatic	$H_2O_2 + CH_3OH \rightarrow HCHO + 2\ H_2O$
	$H_2O_2 + HCHO \rightarrow HCOOH + H_2O$
	$H_2O_2 + HCOOH \rightarrow CO_2 + 2H_2O$

Table 2.4. Possible reactions catalysed by catalase during methanol metabolism in yeast (Veenhuis *et al.*, 1983). The reaction starts with the formation of a primarily complex between H_2O_2 and Fe^{+3} of the catalase prosthetic group, which reacts either with another H_2O_2 molecule for the catalatic reaction, or with a hydrogen donor for the peroxidatic reaction.

In the assimilatory pathway, formaldehyde that remains in the peroxisome reacts with xylulose-5-phosphate (Xu5P). In this reaction catalyzed by dlhydroxyacetone synthase, two C3 compounds, dihydroxyacetone (DHA) and glyceraldehyde-3-phosphate (GAP), are produced, in a transketolase reaction between formaldehyde and Xu5P (van Dljken *et al.*, 1978). Dihydroxyacetone is phosphorylated by cytosolic dihydroxyacetone kinase. In a subsequent aldolase reaction with GAP, it forms fructose-1,6-blphosphate which is converted to fructose-6-phosphate by a phosphatase. These compounds are further metabolized in the cytosol to eventually regain xylulose-5-phosphate in a cyclic pathway including transaldolase, transketolase, pentose phosphate isomerase and epimerase reactions. One-third of the glyceraldehyde-3-phosphate produced becomes available for central metabolism and the generation of biomass. The flux of formaldehyde generated from methanol over the oxidative or the biosynthetic pathways is strictly regulated. Regeneration of Xu5Pin the assimilation pathway needs ATP, so in case of intracellular energy limitation, there is no available substrate for dihydroxyacetone synthase, resulting in an increase of formaldehyde

available for energy production (Rose & Harrison, 1989). The final reaction during the assimilatory pathway is described below:

3 HCHO +3ATP → 1GAP+3AOP+ 2 Pi
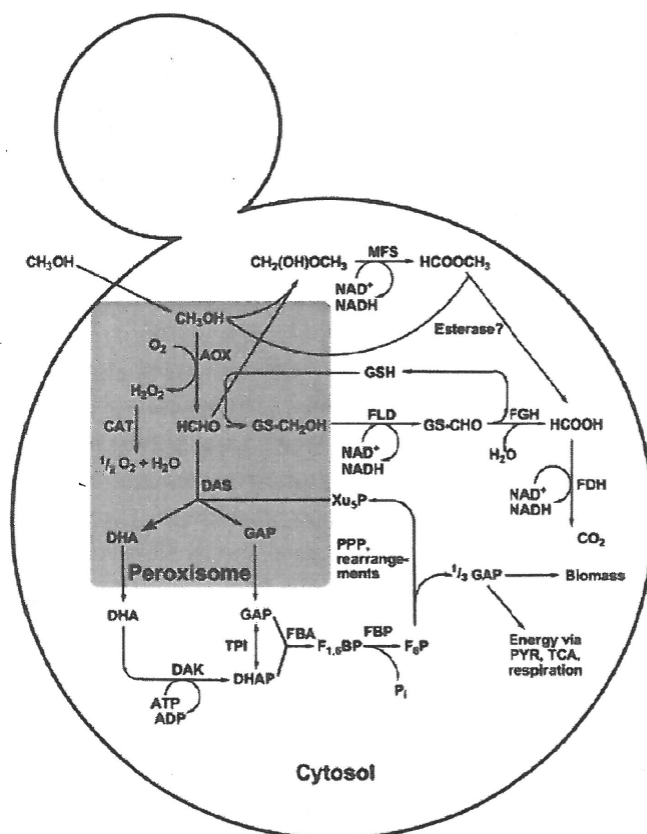
Figure 2.11. Methanol metabolism pathways and their compartmentation in methylotrophic yeasts. (Hartner & Glleder, 2006).

AOX: alcohol oxidase (EC 1.1.3.13), CAT: catalase (EC 1.11.1.6), FlD: formaldehyde dehydrogenase (EC 1.2.1.1), **FGH**: S-formylglutathione hydrolase (EC 3.1.2.12), FOH: formate dehydrogenase (EC 1.2.1.2), **DAS**: dlhydroxyacetone synthase (EC 2.2.1.3), **TPI**: triosephosphate isomerase (EC 5.3.1.1), **OAK**: dihydroxyacetone kinase (EC 2.7.1.29), **FBA**: fructose 1,6-bisphosphate aldolase (EC 4.1.21.13), **FBP**: fructose 1,6-bisphosphatase (EC 3.1.3.11), **MFS**: methylformate synthase: **DHA**: dihydroxyacetone, **GAP**: glyceraldehyde 3 -phosphate, **DHAP**: dihydroxyacetone phosphate, F1, **6BP**: fructose 1,6-bisphosphate, **F6P**: fructose 6-phosphate, PI: phosphate, **XUSP**: xylulose 5-phosphate, **GSH**: glutathione, **PVR**: pyruvate; **PPP**: pentose phosphate pathway, **TCA**: tricarboxylic acid cycle.

Regulation of methanol metabolism in yeast is a very complex process including control of synthesis and activation of the corresponding enzymes as well as their degradation. Synthesis of methanol metabolizing enzymes is induced by methanol, formaldeyde, and formate and is repressed by glucose, ethanol and glycerol (Egli *et al.*, 1980; Unrean 2014).

Glycerol is utilized as a carbon source under aerobic condition by methylotrophic yeast. The catabolic pathway involves passive diffusion across the plasma membrane, phosphorylation by a glycerol kinase, and oxidation by a mitochondrial glycerol phosphate ubiquinone oxireductase (Gancedo *et al.*, 1968). Glycerol enters glycolysis after its conversion to glyceraldehyde 3-phosphate, and requires respiration to dispose of NADH in order to serve as an energy source. Previous studies (Chiruvolu, 1998) revealed that during the batch, fed-batch, or induction phase, the use of glycerol results in ethanol production, resulting subsequently to repression of the AOX1 promoter.

2.3.2. *Pichia pastoris* expression system

Advantages of the *Pichia* expression system, as described above in detail, include not only growth to very high cell densities in simple defined medium, but also strongly inducible promoters (Cregg *et al.*, 1993). The importance of this expression system is underscored by its commercialization by Invitrogen as the *Pichia pastoris* Expression System (Invitrogen Corporation, Sarlsbad, CA, USA). Expression of any heterologous gene in *P. pastoris* includes primarily the insertion of the gene of interest into an appropriate expression vector, following by the introduction of the recombinant vector into the *P. pastoris* genome and finally, the evaluation and phenotype analysis of the strains for the foreign gene product. *P. pastoris'* genome structure allows the homologous recombination between genomic and artificially introduced DNAs. Cleavage of a *P. pastoris* vector within a sequence shared by the host genome stimulates homologous recombination events that efficiently target integration of the vector to that genomic locus (Cregg and Madden, 1987). A variety of *P. pastoris* expression vectors and host strains are available. All *P. pastoris* expression strains are derived from NRRL-Y 11430 (Northern Regional Research Laboratories, Peoria, IL). Most of them have one or more auxotrophic mutations which allow for selection of expression

vectors containing the appropriate selectable marker gene upon transformation. All of these strains grow on complex media but require supplementation with the appropriate nutrient for growth on minimal media. Several protease-deficient strains have been shown to be effective in reducing degradation of some foreign proteins (White *et al.*, 1995). This is especially noticeable in fermenter cultures, because the combination of high cell density and lysis of a small percentage of cells results in a relatively high concentration of these vacuolar proteases (Goodrick *et al.*, 2001).

The AOX1 promoter has been the most widely reported and utilized of all the available promoters for *P. pastoris* (Cereghino *et al.*, 2001). AOX1 and AOX2 are two genes with 97% homology that encode alcohol oxidase (AOX, EC 1.1.3.13) activity in the cell and catalyze the oxidation of methanol to formaldehyde, generating hydrogen peroxide within peroxisome. Alcohol oxidase has a poor affinity for oxygen, and *P. pastoris* compensates for this by generating large amounts of the enzyme. Both AOX1 and AOX2 are regulated in a similar manner. The regulation involves both repression-derepression and induction mechanisms (Cregg *et al.*, 1989). The AOX1 promoter regulates 85% of the alcohol oxidase activity in the cell, but is repressed at transcription level by unlimited growth on ethanol, glycerol or glucose. In cells fed with methanol at growth-limiting rates in fermenter cultures, AOX levels are dramatically induced, constituting >30% of total soluble protein (Couderc & Baratti, 1980). The majority of heterologous protein production in *P. pastoris* is based on the fact that enzymes required for the metabolism of methanol are only present when cells are grown on methanol (Egli *et al.*, 1980), so AOX promoter has been the most successful system reported for this organism. The 'AOX1 promoter-Gene X' expression cassette is inserted into the *Pichia* genome along an existing marker for selection of transformed cells. Existing markers include genes that encode for enzymes of a biosynthetic pathway (HIS4 for histidinol dehydrogenase, ARG4 for argininosuccinate lyase, URA3 for orotidine 5P-phosphate decarboxylase) or antibiotic resistance genes, such as the *Sh ble* gene from *Streptoalloteichus hindustanus* which confers resistance to the bleomycin-related antibiotic zeocin (Cregg 1985; Higgins 1998; Cregg and Madden 1989; Cereghino *et al.*, 1999). Zeocin causes cell death by intercalating into DNA and induces double strand breaks of the DNA. Cu^{+2} give the blue color. In a zeocin resistant strain, when *Sh ble* gene is expressed, the produced polypeptide bounds the drug, thus transforming it into an inactive form.

The strong inducibility of AOX1 promoter, the alternative weaker AOX2-mediated methanol oxidation and the tight regulation of AOX1 promoter variants (Hartner and Glieder, 2006) allow the design of different expression strains with specific properties. Apart from the wild strain X33, possessing functional AOX1 and AOX2 genes (Mut$^+$ strain, methanol utilization plus), several other phenotypes have been constructed and are used for the production of heterologous proteins in case, depending on the product. Mut^S strain (methanol utilization slow) is generated when AOX1 is knocked out and Mut$^-$ (methanol utilization minus) strain is unable to grow on methanol as the sole carbon source, as both the AOX1 and AOX2 genes are disrupted (Cregg *et al.*, 1989).
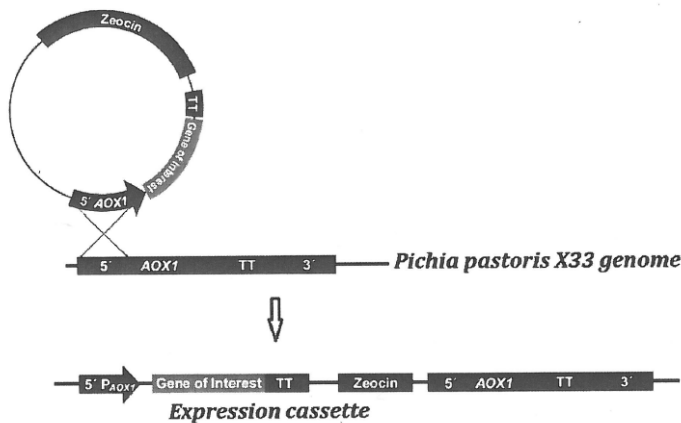
Figure 2.12. Insertion of the plasmid 5' to the intact AOX1 locus of *Pichia pastoris* X33 (Mut$^+$) results in Integration of AOX1 promoter, foreign gene of interest and *Sh ble* zeocin resistance gene. TT represents the AOX1 transcription termination region (TT). In addition, for secretion of foreign proteins, most vectors possess a secretion signal, such as *S. cerevislae* α-factor. The *S. cerevislae* α-factor prepro-signal is the most widely used and successful secretion signal, being in some cases better than the leader sequence of the native heterologous protein. All expression vectors have been designed as *Escherichia coli* / *P. pastoris* shuttle vectors, containing an origin of replication for plasmid maintenance in *E. coli* and markers functional in one or both organisms.

Apart from the AOX promoters that have therefore been the most widely utilized, other promoter options are available for the production of foreign proteins in *Pichia*, and these are discussed comprehensively by Cereghino & Cregg, 2000. Alternative promoters are the *P. pastoris* constitutively expressed GAP promoter, the FLD1 promoter which offers the flexibility to induce expression using either methanol or methylamine, an inexpensive nontoxic nitrogen source, and the moderately expressing promoters PEX8 and YPT1, used in order to prevent proteins from being misfolded or unprocessed during post-translational modifications under high levels of expression (Brierley, 1998).

Post translational modification: Glycosylation and Proteolysis

Glycosytatlon patterns

As a eukaryote, *P. pastoris* exhibits a variety of post-translational modifications; *O*- and *N*-linked glycosylation is one of the most common performed by this yeast. Glycosylation occurs in the lumen of the endoplasmic reticulum after protein translation. A variety of native proteins are glycosylated, so it is necessary for them to have the appropriate g1ycosylatlon patterns when expressed heterologously, in order to ensure and maintain their biological activity. Unlike *Saccharomyces cereviseae* that tends to hyperglycosylate *N*-linked sugars of expressed proteins (50 - 150 mannose units), *P. pastoris* glycosylation is characterized by shorter patterns, more similar to that of higher eukaryotes with 8-14 mannose residues per chain (Bretthauer & Castellino, 1999).

P. pastoris, like other yeasts and fungi, add *O*-oligosaccharides to the hydroxyl side groups (-OH) of serine and threonine of secreted proteins. These are composed only of mannose residues and tend to have more simple structure than those of higher eukaryotic cells. *Pichia* is possible to glycosylate heterologous proteins that are not normally glycosylated by the native host or, it may glycosylate proteins on different serine and threonine residues than the post translational mechanism of the native host (Cereghino and Cregg, 2000). Glycosylation may increase the molecular weight of the expressed protein up to 20 kDa higher than the theoretical one, as in case of glucoamylase catalytic domain from *Aspergillus awamori*; glycosylation patterns may

alter among heterologous proteins and in some cases and they may be responsible for modifying the function of the molecule (Mochizuki *et al.*, 2001).

Apart from *O* - glycoylation, *P. pastoris* adds *N* - oligosaccharides to the $-NH_2$ side group of *Asn* residues of the proteins. A pre-assembled oligosaccharide unit consisting of N-acetylglucosamine (GlcNAc), mannose (Man) and glucose (Glc), (Glc3Man9GlcNAc2), is transferred from dolichyl pyrophosphate to amide groups of asparagine residues, in the lumen of the ER by the enzyme UDP- GlcNAc : dolichol PGlcNAc - transferase (Dennis *et al.*, 1999). Prior to transfer, the oligosaccharide is bound to the dolichole molecule through a high energy purophosphoric bond and this is driving force for the transfer of it. The majority of glycosylation patterns result from downstream subsequent modification. In the secretory pathway, the oligosaccharide is trimmed further by the removal of the three glucose residues by glucosidase II and I. The α-1,2-linked mannose residue is also removed by α-1,2-mannosidases to give Man8GlcNAc2 (Lis & Sharon, 1993). Further processing to *cis* - Golgi involves mannoses are added and the oligomannose units can be linked α-1,6 to the α-1,3 mannose in the Man α-1,3-Man-β-1,4-GlcNAc2 inner core (Herscovics & Orlean, 1993).
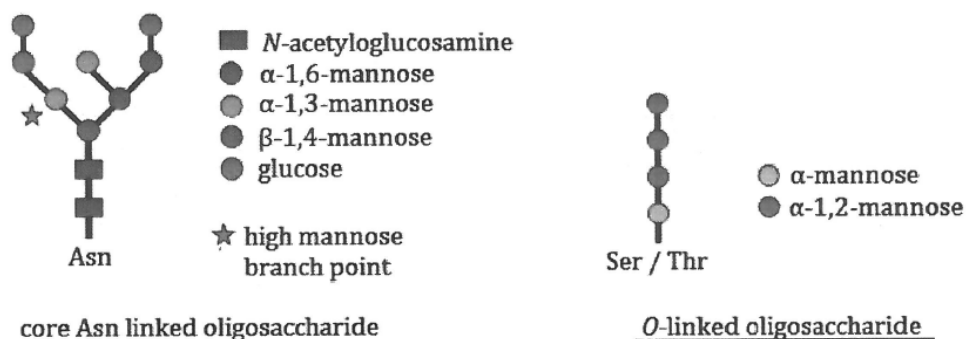
Figure 2.13. Representative Structure of the core pre-assembled oligosaccharide transferred to asparagine residues and the common *O*-linked glucans (Man4 –Ser/Thr). The branch point where further oligosaccharides are added during further processing of *N*-glycosytated proteins to *cis* - Golgi is indicated by the star. High mannose content is typical of yeast glycosylation.

Heterogeneity in glycosylation is a result of differences in oligosaccharide-protein populations in terms of the type, length and Identity of the oligosaccharides added to Asn-X-Ser/Thr sequons, resulting in protein products whose micro-properties, such as the isoelectric point, vary slightly. The glycoprotein variations may be dependent on the surrounding residues (Lis & Sharon, 1993). Even in the same glycosylation Golgi machinery of one cell, it is possible for two proteins to be differently glycosylated and has been proposed that this heterogeneity is related to specific sequences recognized by the various glycosylation systems. It has been observed that some proteins heterologously expressed in *P. pastoris* vary considerably in terms of the number of the mannose units added to the same polysaccharide core (Daly and Hearn 2005).

Proteolysis

Proteolysis is of indispensable importance for the control and execution of many cellular events. Proteolytic enzymes are designed to act either in the intracellular or the extracellular environment. The former are involved in processes such as removal of signal peptides after translocation of proteins through membranes, processing of Inactive precursors (zymogens/proenzymes) to generate fully active proteins, inactivation of short-lived regulatory molecules and degradation of proteins that arise from mutations, misfolding, transcription or translation errors. The latter are secreted to the culture medium and digest proteins for supplementation of peptides and amino acids for nutrition. Based on the mode of action, proteases are divided into those hydrolyzing peptide bonds within polypeptide chain, endoproteinases, and those acting on the carboxyl· or amino- terminal region, exopeptidases. Based on the features of active site, proteolytic enzymes belong to four major classes: serine, metallo-, thiol and aspartyl proteases. Serine proteases often have a higher pH optimum, whereas aspartyl proteases a lower pH optimum (Dery *et al.*, 1998; Tang & Wong, 1987).

The protease system that occurs in *P. pastoris* is similar to this described in *S. cereviseae* and is organized in three main groups of enzymes: vacuolar (lysosomal) proteases, proteases located along with the secretory pathway and the cytosolic vacuoles contain various proteases whose levels vary according to the nutritional conditions. A total of seven vacuolar proteases are known: two endoproteinases,

proteinase A (PrA) and proteinase B (PrB); two carboxypeptidases, carboxypeptidases Y and S (CpY, CpS); two aminopeptidases, aminopeptidases I and Co (Apl, ApCo) and dipeptidyl aminopeptidase B (DPAPB). All the above enzymes are soluble within the vacuole, except DPAP-B which found in the vacuolar membrane. The aspartyl protease proteinase A. encoded by the gene *PEPE4* is capable of self-activation and subsequent activation of other vacuolar proteases, such as carboxypeptidase Y and proteinase B. Proteinase B is encoded by the *PRB1* gene and is a member of the subtilisin family of serine proteases. The proteases of the secretory pathway, which are mainly located in the Golgi apparatus and the plasmatic membrane, act to process precursors to one or more secreted peptides (Flores *et al.*, 1999). This group consists of signal peptidase, Kex2 endoprotease and Kex1 carboxypeptidase, dlpeptidyl aminopeptidase A and yeast aspartyl protease Ill (Yap3 protease) (Zhang *et al.*, 2007). The cytosolic proteasome of *Pichia pastoris* is a multicatalytic proteinase system, responsible for the selection and rapid degradation of short-lived proteins and proteins that are detrimental to the cell growth. The interplay between vacuolar and proteasome proteolysis is a key regulatory cell process, particularly in response to stress (van den Hazel *et al.*, 1996).

The most prominent proteolytic enzymes are of vacuolar origin and they comprise the major source of problems with degradation of secreted recombinant proteins expressed in *P. pastoris*, particularly under conditions of nutrient deprivation. These enzymes are non-specific and do not require ATP. Problems that can be caused involve either total degradation of product and reduction of yield or loss of biological activity when the product is truncated. Moreover, degradation intermediates can cause contamination in downstream processing, because of their similar physico-chemical and affinity characteristics (Macauley-Patrick *et al.*, 2005). In general, the secreted recombinant proteins can potentially be proteolytically degraded in the culture medium by extracellular proteases, cell-bound proteases (Kang *et al.*, 2000) and/or by intracellular proteases from lysed cells. Signs of proteolysis include low recombinant protein levels and active or immunoreactive products that are smaller than the full-length product. Degraded protein can also run as a "smear" after SDS-PAGE, running from approximately the correct size of the product to smaller sizes. Although secreted recombinant proteins do not go to the vacuole, they can contact proteases in the culture medium from the lysis of a small portion of the cells. In fermenters, the concentration of other cellular materials, such as proteases, increases as well in high-

cell density culture. Proteolytic degradation has been a perpetual problem since yeast has been employed to express recombinant proteins, especially in fermenters (Van Den Hazel *et al.*, 1996). Peroxisome biogenesis has been shown a high dependency on methanol consumption. The proliferation of peroxisomes is inhibited in a medium containing glucose and stimulated in the methanol phase under a carbon-limited fed-batch culture condition, which can cause problems at initial inducing phase of fermentation (Kim *et al.*, 2013).

Several strategies, based on engineering of the protein structure to resist the protease (protein level strategies), modification of cultivation conditions and selection of protease deficient host strains, are used to overcome proteolytic degradation. Protein level strategies are based on the molecular engineering of amino acid sequences non-essential for the function of the protein, in order to modify or delete sequences recognized by proteases (Gustavsson *et al.*, 2001).

Cultivation level strategies include changing the medium composition or the incubation conditions (Jahic *et al.*, 2003). *P. pastoris* is capable of growing across a relatively broad pH range from 3.0 to 7.0. This range has little or no effect on the growth rate, allowing adjustment of pH to one that causes limited protease activity. Different pH values were found to be optimal regarding production levels and protein stability. For example, pH 6.0 was found to be optimal for production recombinant mouse epidermal factor (Clare *et al.*, 1991), but pH 3.0 was optimal in case of Insulin-like growth factor I (Brierley, 1998). Medium composition and addition of amino acid rich supplements, such as peptone or casamino acids, to the culture medium has been shown to reduce protein degradation. The peptone can act as alternative and competing substrate for proteases and can also repress protease Induction caused by nitrogen limitation (Macauley-Patrick *et al.*, 2005). Addition of protease inhibitors in bioreactor has also been reported to result in higher yields (Holmquist *et al.*, 1997). There is a positive effect of reduced growth temperature on the secretion of heterologous proteins into the culture supernatant. Even though 20 and 30°C temperature set-points correspond to 60 and 100% of the maximum specific growth rate (μ_{max}) respectively, a reduction of temperature from 30 to 20°C has been reported to result in a 3-fold increase of the specific productivity of the expressed recombinant protein (Dragosits *et al.*, 2009). A significant amount of proteins, including mainly protein fragments and proteins, which are involved in amino acid metabolism, RNA

and ribosome biogenesis, genome maintenance and proteasomal components, have been shown to expressed in increased levels at lower cultivation temperature (20°C). Lower cultivation temperature can influence yields of recombinant proteins since the rate of proteolysis is lowered for the pure thermodynamic reason. Moreover, prevents from cell lysis and release of protease activity in fermentation media, increases cell viability and enhanced protein folding (Hong *et al.*, 2002).

Cell level strategies include the use of host strains that are defective in the vacuolar proteases. SMD1163 (*his4 pep4 prb1*), SMD1165 (*his4 prb1*), SMD1168 (*his4 pep4*) and SMD1168H (*his4*) are protease deficient strains (Goodrick *et al.*, 2001). The *pep4* and *prbl* genes encode proteinase A and B respectively, as mentioned above. Proteinase A is responsible for the activation of other vacuolar proteases. Proteinase B, prior to its processing and activation of proteinase A, exhibits about half the activity of the processed enzyme and carboxypeptidase Y appears to be completely inactive prior to proteinase A-mediated proteolytic activation. Therefore, pep4 mutants display a substantial decreased or eliminated activity of proteinase A and carboxypeptidase Y, and a partial reduction of proteinase B activity. In the *prb1* mutant, only proteinase B activity is eliminated, whereas *pep4 prb1* double mutants show a substantial reduction or elimination of all three of these protease activities. Protease-deficient strains, combined with other strategies to reduce proteolysis, have been invaluable in the production of laccase (Jonsson *et al.*, 1997) and therapeutic cytokine human granulocyte macrophage colony stimulating factor (Babu *et al.*, 2008). Despite the advantages that they offer, these strains are not as vigorous and robust as wild-type strains. In addition to lower viability, they exhibit lower growth rate, are more difficult to transform and require greater care in growth and storage.

2.3.3. Fermentation of *Pichia pastoris* using the P_{AOX1}

P. pastoris exhibits a preference for respiratory growth, a major advantage relative to *S. cerevisiae*. In high cell density cultures, ethanol (the product of *S. cerviseae* fermentation) may reach toxic levels, acting as a limiting factor for further growth and recombinant protein production. In contrast, *P. pastoris* (Mut[+]) can be cultured aerobically (DOT > 20% air saturation) at extremely high densities (100 g/l dry weight) in the controlled environment of the fermentor with high oxygen transfer

capabilities and efficient cooling systems are crucial (Schilling *et al.*, 2001). Bioreactor controlled cultivation is also required to maintain the methanol concentration at a specified level, not exceeding the maximum methanol consumption rate of the cells (Jahic *et al.*, 2006).

Fermentation protocols for *Pichia* generally include three separate phases, glycerol batch phase, glycerol fed-batch phase and methanol fed-batch phase. First is the glycerol batch phase, in which cells are initially grown in a batch mode (in a simple, defined medium) with the repressing carbon source, glycerol in order to accumulate biomass, but no recombinant protein. Repression of the P_{AOX1} ensures that a high cell density is reached before the beginning of the recombinant protein expression and any adverse effect that may be caused by methanol. Maximum specific growth rate on glycerol for *P. pastoris* (Mut+) has been reported to be 0.26 h^{-1} compared to 0.14 h^{-1} for methanol (von Stockar *et al.*, 2007). In addition, the cell viability is higher for cells grown on glycerol compared to those grown on methanol which have the additional benefit of reducing release of degrading proteases from cell lysis in the cultivation medium.

In the second phase, the glycerol fed-batch phase, a limited glycerol feed is initiated following exhaustion of the glycerol, and cell mass is increased to a desired concentration level prior to methanol induction. The exponential feed of glycerol results in a slight increase in alcohol oxidase activity and, a subsequent expression of low amounts of recombinant protein. This is attributed to the fact that the P_{AOX1} promoter is de-repressed during this phase due to the absence of excess glycerol produced, allowing the cells to tolerate higher initial feed rates of methanol in the methanol fed-batch phase. The transition to the methanol feeding has been reported to be smoother when this short glycerol phase was inserted, compared to the case when no exponential glycerol feed was applied, when a considerable cell death and a much longer adaptation to methanol were observed (Jahic *et al.*, 2002).

Finally, in the methanol fed-batch Induction phase, methanol is added to the culture to induce expression of the recombinant protein. This step is typically started with very low feed rates of methanol to allow the cells fully adapt to methanol metabolism. During the adaptation period the cells generate peroxisomes and the adequate enzymes required for methanol metabolism (van der Klei *et al.*, 2006). When

the shift of carbon sources has been completed, the methanol feed rate is increased upward exponentially in order to maintain constant specific growth rate until the oxygen transfer ultimately limits further increase. Careful control of the methanol concentration in bioreactor cultivations is crucial for process development, as too high concentrations will be toxic to the cells, and too low levels may not be enough to induce AOX transcription (Cos *et al.*, 2006).
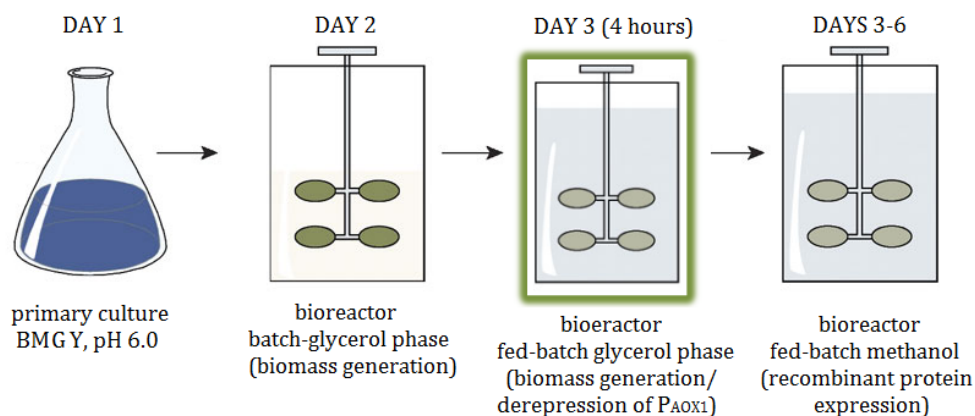
Figure 2.14. For scale-up cultivation of *P. pastoris*, a three-phase fermentation scheme is employed. Inoculum from primary culture in defined medium (BMMY, buffered complex glycerol medium, as recommended by *Invitrogen*) is inoculated to bioreactor. After 24 hours of batch glycerol phase. A 4 hour limited glycerol feed phase starts, followed by a fed -batch methanol induction phase.

Several limitations to standard *P. pastoris* methanol fed-batch cultivation arise and most of them are attributed to catabolite repression of AOX promoter and proteolysis of recombinant proteins (Zhang *et al.*, 2010). AOX1 is catabolite-repressed at non-growth-limiting rate, and its activity is much higher at the initial phase of methanol induction than later in cultivation. Lysis of cells during cultivation results in vacuolar proteases release, as described above. Utilization of methanol as a carbon source puts particularly high demands on the oxygen and heat transfer capabilities of the bioreactor system, due to high oxygen consumption rates of methanol grown cells and high heat generation related to methanol dissimilation (Schilling *et al.*, 2001). Moreover, when recombinant proteins are produced at exaggerate levels, it is more

likely that proteins with aberrant folding properties or altered glycosylation patterns may provoke inefficient product secretion or endoplasmic retention of these molecules. This leads to the activation of unfolded protein response (UPR) or endoplasmic reticulum associated degradation (ERAD) cell mechanisms, causing lower yield of biologically active protein expressed (Vanz *et al.*, 2014). Lower temperature during cultivation (20°C) may lead to generally higher stability of proteins and reduced degradation mechanisms, resulting in a higher secretion capacity.

References

Babu, S.K., Muthukumaran A.A., Meenakshisundaram S. 2008. Construction of intein-mediated hGMCSF expression vector and its purification in *Pichia pastoris. Protein Expr Purif.* 57, 201-5.

Bretthauer, R.K., Castellino, F.J. 1999. Glycosylation of *Pichia pastoris*-derived proteins, *Biotechnol App Biochem* 30, 193-200.

Brierley, R.A. 1998. Secretion of recombinant human insulin-like growth factor I (IGF-I). *Methods Mol Biol.* 103:149-77.

Cereghino GL, Lim M, Johnson MA, Cereghino JL, Sunga AJ, Raghavan D, Gleeson M, Cregg JM. *Gene* 2001; 263: 159–169.

Cereghino, G.P., Cereghino, J.L., Ilgen C., Cregg, J.M. 2002. Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris .Cur Op Biotechnol* 13:329–332.

Cereghino, G.P.L., Cregg, J.M., 1999. Applications of yeast in biotechnology: protein production and genetic analysis. *Curr Opin Biotechnol.* 10, 422-427.

Cereghino, J.L. and Cregg, J.M. 2000. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris. FEMS Microbiology Reviews*, 24, 45-66.

Chiruvolu, V., Cregg, J.M., Meagher, M.M. 1997. Recombinant protein production in an alcohol oxidase-defective strain of *Pichia pastoris* in fedbatch fermentations. *Enz Microb Technol*, 21, 277-283.

Clare, J.J., Romanos, M.A., Rayment, F.B., Rowedder, J.E., Smith, M.A., Payne, M.M., Sreekrishna, K., Henwood, C.A. 1991. Production of mouse epidermal growth-factor in yeast - High-level secretion using *Pichia pastoris* strains containing multiple gene copies. *Gene.* 105, 205-212.

Cos. O., Ramón, R., Montesinos, J.L., Valero, F. 2006. Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: a review. *Microb Cell Fact.* 6, 5:17.

Couderc, R. and Baratti, J. 1980. Oxidation of methanol by the yeast, *Pichia pastoris* -Purification and properties of the alcohol oxidase. *Agricult Biol Chem.* 44, 2279-2289.

Cregg, J.M. and K.R Madden, K.R. 1987. Development of yeast transformation systems and construction of methanol-utilization-defective mutants of *Pichia pastoris* by gene disruption. *Biological Research on Industrial Yeasts*, CRC Press, Boca Raton, FL 1–18.

Cregg, J.M. and Madden, K.R. 1989. Use of site-specific recombination to regenerate selectable markers. *Mol Gen Genet.* 219, 320-3

Cregg, J.M., Barringer, K.J., Hessler, A.Y., Madden, K.R. 1985. *Pichia pastoris* as a host system for transformations .*Mol. Cell. Biol.*, 5, 3376–3385.

Cregg, J.M., Madden ,K.R., Barringer, K.J., Thill, G.P., Stillman, C.A. 1989. Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris.* Mol Cel Biol 9, 1316–23.

Cregg, J.M., Tolstorukov, I., Kusari, A., Sunga, J., Madden, K., Chappell, T. 2009. Expression in the yeast *Pichia pastoris. Methods Enzymol.* 463, 169-89.

Daly, R. and Hearn, M.T. 2005. Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit.* 18:119-38.

Dennis, J.W., Granovsky, M., Warren, C.E. 1999. Protein glycosylation in development and disease *Bioessays* 21: 412–421.

Déry, O., Corvera, C.U., Steinhoff, M., Bunnett, N.W. 1998. Proteinase-activated receptors: novel mechanisms of signaling by serine proteases. *Am J Physiol.* 274 (6 Pt 1), C1429-52.

Diauchy, D., Tornai-Lehoczki, J., Peter, G. 1999. Restriction enzyme analysis of PCR amplified rDNA as a taxonomic tool in yeast identification. *Syst. Appl. Microbiol.* 22, 445-453.

Dragosits, M., Stadlmann, J., Albiol, J., Baumann, K., Maurer, M., Gasser, B., Sauer, M., Altmann, F., Ferrer, P., Mattanovich, D. 2009. *The effect of temperature on the proteome of recombinant Pichia pastoris. J Proteome Res.* 8, 1380-92.

Egli, T., van Dijken, J.P., Veenhuis, M., Harder,W., Fietcher, A. 1980. Methanol metabolism in yeasts: regulation of the synthesis of catabolic enzymes. *Arch Microbiol* 124, 115-121.

Faber, K.N., Haima, P., Gietl, C., Harder, W., Ab, G., Veenhuis, M. 1995. The methylotrophic yeast *Hansenula polymorpha* contains an inducible import pathway for peroxisomal matrix proteins with an N-terminal targeting signal (PTS2 proteins). *Proc Natl Acad Sci U S A.* 91, 12985–12989.

Flores, A., Briand, J. F., Gadal, O., Andrau, J. C., Rubbi, L., Van Mullem, V., Boschiero, C., Goussot, M., Marck, C., Carles, C. *et al.* 1999. A protein-protein interaction map of yeast RNA polymerase III. *Proc. Natl. Acad. Sci. USA* 96, 7815-7820.

Gancedo, C., Gancedo, J.M., Sols, A. 1968. Glycerol metabolism in yeasts. Pathways of utilization and production. *Eur J Biochem.* 5, 165-72.

Goodrick, J.C., Xu, M., Finnegan, R., Schilling, B.M., Schiavi, S., Hoppe ,H., Wan, N.C. 2001. High-level expression and stabilization of recombinant human chitinase produced in a continuous constitutive *Pichia pastoris* expression system. *Biotechnol Bioeng.* 74, 492-7.

Guilliermond, A.1919. Sur une nouvelle levure acopulation heterogamique. *C.R.Soc.Biol.*82, 466-470.

Gustavsson, M., Lehtio, J., Denman, S., Teeri, T.T., Hult, K., Martinelle, M. 2001. Stable linker peptides for a cellulose-binding domain-lipase fusion protein expressed in *Pichia pastoris. Protein Eng*, 14, 711-5.

Hartner, F.S. and Glieder, A. 2006. Regulation of methanol utilisation pathway genes in yeasts. *Microb. Cell Fact* 5, 1–29.

Hartner, F.S. and Glieder, A. 2006. Regulation of methanol utilization pathway genes in yeasts. *Microb Cell Fact* 14, 5–39.

Higgins, D.R., Cregg, J.M. 1998. Methods in Molecular Biology: *Pichia* Protocols. Humana: Totowa, NJ.

Holmquist, M., Tessier, D.C., Cygler, M. 1997 High-level production of recombinant *Geotrichum candidum l*ipases in yeast *Pichia pastoris. Protein Expr Purif.* 11, 35-40.

Hong, F., Meinander, N.Q., Jonsson, L.J. 2002. Fermentation strategies for improved heterologous expression of laccase in *Pichia pastoris. Biotechnol Bioeng* 79: 438–449.

Jahic, M., Gustavsson, M., Jansen, A.K., Martinelle, M., Enfor, S.O. 2003. Analysis and control of proteolysis of a fusion protein in *Pichia pastoris* fed-batch processes. *J Biotechnol*, 102, 45-53.

Jahic, M., Rotticci-Mulder, J.C., Martinelle, M., Hult, K., Enforset, S-O. 2002. Modeling of growth and energy metabolism of *Pichia pastoris* producing a fusion protein. *Bioprocess Biosyst Eng* 24, 385-393.

Jahic, M., Veide, A., Charoenrat, T., Teeri, T., Enfors, S-O. 2006. Process technology for production and recovery of heterologous proteins with *Pichia pastoris. Biotechnol Prog* 22, 1465–73.

Jones, E.W. 1991. Three proteolytic systems in the yeast *Saccharomyces cereviseae. J Biol Chem.* 266, 7963-7966.

Jönsson, L.J., Saloheimo, M., Penttilä, M. 1997. Laccase from the white-rot fungus *Trametes versicolor*: cDNA cloning of lcc1 and expression in *Pichia pastoris. Curr Genet.* 32, 425-30.

Kang, H.A., Choi, E.S., Hong, W.K., Kim, J.Y., Ko, S.M., Sohn, J.H., Rhee, S.K. 2000. Proteolytic stability of recombinant human serum albumin secreted in the yeast *Saccharomyces cerevisiae, Appl Microb Biotechn* 53, 575-582.

Kim, S., Warburton, S., Boldogh, I., Svensson, C., Pon, L., d'Anjou, M., Stadheim, T.A., Choi, B.K. 2013. Regulation of alcohol oxidase 1 (AOX1) promoter and peroxisome biogenesis in different fermentation processes in *Pichia pastoris. J Biotechnol.* 166, 174–81.

Kurtzman, C.P. and Robnett, C.J. 1998. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie van Leeuwenhoek.* 73, 331-371.

Lis, H. and Sharon, N. 1993. Protein glycosylation—structural and functional aspects. *Europ J Biochem* 218: 1–27.

Macauley-Patrick, S., Fazenda, M.L., McNeil, B., Harvey, L.M. 2005. Heterologous protein production using the *Pichia pastoris* expression system. *Yeast*, 22, 249-270.

Mochizuki, S., Hamato, N., Hirose, M. 2001. Expression and characterisation of recombinant human antithrombin III in *Pichia pastoris Protein Expression and Purification* 23: 55–65.

Rose, A.H. and Harrison, J.S. 1989. The Yeasts, Vol. 3, Academic Press, London.

Schilling, B.M., Goodrick, J.C., Wan, N.C. 2001. Scale-up of a high cell density continuous culture with Pichia pastoris X-33 for the constitutive expression of rhchitinase. *Biotechnol Prog* 17, 629-633.

Subramani, S. 1998. Components involved in peroxisome import, biogenesis, proliferation, turnover and movement. *Physiol Rev* 7, 171-188.

Tang, J. and Wong, R.N. 1987, Evolution in the structure and function of aspartic proteases. *J Cell Biochem.* 33, 53-63.

Unrean P. 2014. Pathway analysis of *Pichia pastoris* to elucidate methanol metabolism and its regulation for production of recombinant proteins. *Biotechnol Prog.* 30, 28-37.

Van Den Hazel, H.B., Kielland-Brandt, M.C., Winther, J.R.1996. Review: biosynthesis and function of yeast vacuolar proteases. *Yeast* 12, 1-16.

van der Klei, I.J., Yurimoto, H., Sakai, Y., Veenhuis, M. 2006. The significance of peroxisomes in methanol metabolism in methylotrophic yeast. *Biochim Biophys Acta.* 1763, 1453-62.

Van Dijken, J.P., Harder, W., Beardsmore, A.J., Quayle, J.R. 1978. Dihydroxyacetone: an intermediate in the assimilation of methanol by yeasts? *FEMS Microbiol Lett* 4, 97-102.

Vanz, A.L., Nimtz, M., Rinas, U. 2014. Decrease of UPR- and ERAD-related proteins in *Pichia pastoris* during methanol-induced secretory insulin precursor production in controlled fed-batch cultures *Microb Cell Fact* 13:23.

Veenhuis, M., van Dijken, J.P., Harder, W. 1983. The significance of peroxisomes in the metabolism of one-carbon compounds in yeast. *Adv Microb Physiol* 24, 1-82.

von Stockar, U., Schenk, J., Marison, I.W. 2007. *A simple method to monitor and control methanol feeding of Pichia pastoris fermentations using mid-IR spectroscopy. J Biotechnol*, 128, 344–353.

White, C.E., Hunter, M.J., Meininger, D.P., White, L.R., Komives, E.A. 1995. Large-scale expression, purification and characterization of small fragments of thrombomodulin: the roles of the sixth domain and of methionine 388. *Protein Eng.* 8, 1177-87.

Yamada, Y., Matsuda, M., Maeda, K., Mikata, K. 1995. The phylogenetic relationships of methanol-assimilating yeasts based on the partial sequences of 18S and 26S ribosomal RNAs: the proposal of *Komagataella* gen. nov. (*Saccharomycetaceae*). *Biosci Biotechnol Biochem.* 59, 439-44.

Zhang, P., Zhang, W., Zhou, X., Bai, P., Cregg, J.M., Zhang, Y. 2010. Catabolite repression of Aox in Pichia pastoris is dependent on hexose transporter PpHxt1 and pexophagy. *Appl Environ Microbiol.* 76, 6108-18.

Zhang, Y.W., Liu, R.J., Wu, X.Y. 2007. The proteolytic systems and heterologous proteins degradation in the methylotrophic yeast *Pichia pastoris, Ann Microbiol* 57, 553-560.

CHAPTER 3

Materials and Methods

3.1. Equipment - Chemicals

Equipment available at National Technical University of Athens (NTUA) and Luleå Tekniska Universitet (LTU) used in this thesis is listed below

- pH-meter 537 (WTW, Germany)
- Systec V-100 (Systec GmbH, Germany) vertical Top Loading Autoclave
- Eppendorf Thermomixer Comfort Incubator (Eppendorf, Germany)
- Temperature and agitation controlled water-baths and table incubators
- Ostwald glass viscometer
- Anton Paar Physica MCR rheometer (Anton Paar Gmbh, Austria), with a parallel plate system with roughened plates
- Mini-PROTEAN 3 protein electrophoresis cell (BIORAD, U.S.)
- Floor model incubator shakers ZHWY-211C, ZHICHENG Analytical Instruments Manufacturing Co. Ltd (China).
- Fraction collector (Waters, Millipore, U.S.)
- Gradient thermal cycler TC-512 TECHNE (U.S.)
- Low speed lab shaker Orbit LS, Labnet (U.K.)
- Freeze dryer, Christ ALPHA 1-4, B.Braun Biotec. International, Melsungen, (Germany).
- Electroporation cell Micropulser™ BIORAD (U.S.)
- Agarose electrophoresis cell Easigel H1-set, Scie-plas (U.K.)
- Amicon Ultrafiltration apparatus Stirred Cell Model 8400, 400 mL and PM-10 membranes (Amicon, Millipore, U.S.)
- Gel imaging and analysis system InGenius BioImaging, Syngene (U.K.) equipped with GeneSnap v6.05 and GeneTools v3.06 software.
- Ultrapure Water Systems Direct-Q (Millipore, U.S.)
- Low-pressure protein purification system ECONO gradient pump, BIORAD, U.S.) equipped with UV detector (280 nm)

- ÄKTA Prime Plus liquid chromatography system, equipped with Prime View 5.31. software (GE Healthcare Life Sciences)

- 3 lt glass autoclavable Applikon bioreactors, equipped with an *ez*-Control system (Applikon Biotechnology B.V., Netherlands)

- High-performance anion exchange chromatography (HPAEC), with a pulsed amperometric detector equipped with a gold electrode (PAD), equipped with a CarboPac PA-1 (4 x 250 mm, Dionex) column, Jasco PU 1580i pump (Japan) and Clarity Version 2.3.3.124 software (DataApex, Czech Republic)

- Refrigerated floor model Centrifuge J-20XP, J2-21 and TJ-6, Beckman Coulter (U.S.), ALC 4239R (U.K.) και Eppendorf Brinkmann micro-entrifuge Eppendorf 3200 (Germany)

- Tunable Microplate Reader SPECTRAmax 250 (Molecular Devices, U.S.)

- Spectrophotometer Hitachi UV 2000

- Novex® XCell SureLock™ Mini-Cell system (Invitrogen, Carlsbad, U.S.) with pre-casted Nu-PAGE® Novex 4-12% Bis-Tris gels

Chemicals, which were used in this thesis, were obtained from Merck (Darmstadt, Germany), Fluka (Buchs, Switzerland), Sigma-Aldrich (St. Louis, U.S.) and AppliChem (Germany). Plastic and glass consumables were obtained from Greiner-Bio One (Germany), Sterilin Limited (U.K.), SCHOTT AG (Germany), Eppendorf (Gemany), Whatman (U.K.), Millipore (U.S.) και ROTH (U.S.). The reagents for molecular biology experiments (enzymes, buffer solutions, nucleotides) were obtained from Fermentas (U.S.) New England Biolabs® Inc., (U.S.), Takara BIO Inc., (Japan), Invitrogen (U.S.) Invivogen (France), Novagen (U.S.), Clontech (U.S.), EMD4Biosciences (Germany), Stratagene (U.S.), BIORAD Laboratories (U.S.). DNA and RNA Purification and Elution Kits were from Sigma-Aldrich (U.S.) and Qiagen (Hilden, Germany).

All primers were ordered from VBC Genomics (Vienna, Austria). According to the manual the lyophilised primers were dissolved with sterile water. After incubation for 15 minutes at 65°C the primer stock solution was diluted 1:10 to a concentration of 10 mM before it was used for further experiments. All Primers were stored at -20°C.

All buffers listed below were prepared with distilled water, apart from those used in molecular biology experiments that were used with ultrapure water MilliQ (R=18.2 MΩcm^{-1}).

- **DNA extraction solutions**

DNA extraction solution: 0.1 M Tris, pH 7.5, 0.7 M NaCl, 10 mM EDTA, 1% (w/v) CTAB, 1%(v/v) *β*-mercaptoethanol

TE solution: 10 mM Tris-HCl, 1 mM EDTA, pH 8.0

- **Agarose gel buffers**

TBE (10 X): 108 g/l Tris base, 55 g/l boric acid, 40 mL/l 0.5 M EDTA, pH 8.0

Loading Buffer: 900 μL glycerol 50% (v/v), 100 μL bromophenol blue 10X

- ***P. pastoris* cultivation media and buffers**

Potassium phosphate buffer (1M): 132 mL of 1M K_2HPO_4 and 868 mL of 1M KH_2PO_4 were combined. The pH was adjusted to 6.0 and was filter sterilized. It was stored at 4°C

10X YNB: 134g yeast nitrogen base with ammonium sulfate and without amino acids was dissolved in 1 liter water.

500X biotin: 20 mg biotin was dissolved in100 water and filter was sterilized. It was store at 4°C.

10X Glycerol: 100 mL glycerol in 900 mL ddH_2O

PTM1 trace salts: 6.0 g cupric sulfate-5H_2O, 0.08 g sodium iodide, 3.0 g manganese sulfate-H_2O, 0.2 g sodium molybdate-2H_2O, 0.02 g boric acid, 0.5 g cobalt chloride, 20.0 g zinc chloride, 65.0 g ferrous sulfate-7H_2O, 0.2 g biotin, 5.0 mL sulfuric acid, dH_2O to 1 liter

- **SDS – PAGE Buffers**

Sample buffer: 3.55 mL 0.25M Tris Base pH 6.8, 1.8 mL glycerol 50% (v/v), 1.8 mL *β*–mercaptoethanol, 0.71 gr SDS, 2.85 mL bromophenol blue0.1 % (w/v)

Gel Running buffer: 3.03 g/l Tris-base, 14.4 g/l glycine, 1 g/l SDS, pH 8.3

Staining solution: 0.4% (w/v) Coomassie G-250, 400 mL methanol, 100 mL acetic acid, 500 mL ddH$_2$O

Destaining solution: 200 mL methanol, 100 mL acetic acid, 700 mL ddH$_2$O

- **IEF – PAGE buffers**

Staining solution: 0.02% (w/v) Phast Gel Blue R, 0.1% (w/v) CuSO$_4$, 300 mL methanol, 100 mL acetic, 600 mL ddH$_2$O

Destaining solution: 300 mL methanol, 100 mL acetic acid, 600 mL ddH$_2$O

3.3. Microbial strains

1) The thermophilic fungi *Sporotrichum thermophile* ή *Myceliophthora thermophila* ATCC 42464 (DSM No:1799, Germany) as the primary source of genes.

2) Bacterial strains *Esherichia coli* **TOP10** [F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G] και *Esherichia coli* **TOP10F'** [F´{*lac*Iq Tn*10* (TetR)} *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L *end*A1 *nup*G] from Invitrogen. *E. coli* TOP10 strain was used for cloning of pCR® Blunt plasmid (lnvitrogen), whereas, TOP10F' strain for cloning of pPICZαC plasmid (Invitrogen).

3) The methylotrophic yeast *Pichia pastoris*, X33 (WT strain, Mut⁺ phenotype) (Invitrogen) for the heterologous expression of recombinant proteins.

3.4. Cultivation media and stategies

The components were dissolved in deionised water and autoclaved at 121°C and 0.1 mPa for 20 minutes. When the amount of dextrose was higher than 0,4%, it was autoclaved separately. Antibiotics and vitamins were filter sterilized (0.2 μm filter

pore diameter) first and added to the media when it had cooled down to a temperature of about $55 - 60°C$. All media or plates were stored at 4°C.

- ### *M. thermophila* cultures

Shake flasks liquid medium: 1 g/l K_2HPO_4, 0.5 g/l $MgSO_4·7H_2O$, 0.1 g/l $CaCl_2·2H_2O$, 10 g/l yeast extract, 1% (w/v) glucose, pH 5.0

Agar medium: 30 g/l malt extract, 3 g/l peptone, 1.5% (w/v) agar, pH 5.6

Liquid submerged cultures of *M. thermophila* were grown in Erlenmeyer flasks, volume 250 ml, $45°C$, under 180 rpm agitation and incubation for as long as needed in order to ensure the efficient biomass production. Cultivation of fungal mycelia on agar media was done in slants, at $45°C$, for 5-6 days.

- ### *E. coli* culture cultures

LB (Luria Bertani) ± kanamycin: 1% (w/v) NaCl, 1% (w/v) tryptone, 0,5% (w/v) yeast extract, pH 7.4 ± 50 µg/mL kanamycin

LB-agar ± kanamycin: LB with 1.5% (w/v) agar ± 50 µg/mL kanamycin

LS-LB (low salt LB) ± zeocin: 0.5% (w/v) NaCl, 1% (w/v) tryptone, 0.5% (w/v) yeast extract, pH 7.4 ± 25 µg/mL zeocin

LS LB–agar ± zeocin: LB with 1.5% (w/v) agar ± 25 µg/mL zeocin

E. coli TOP10 and TOP10F' were cultured at $37°C$ in LB, when kanamycin was used, or in LS − LB, when zeocin was used, as this antibiotic may be inactivated by high salinity conditions. After sterilization, LB / LS-LB medium was cooled down in room temperature and then was supplemented with antibiotics under sterile conditions. To prepare LB agar plates, 15 g L^{-1} agar was added to the above medium prior to autoclave. The LB agar plates containing antibiotic were store at $4°C$.

- ### *P. pastoris* culture media

YPD (Yeast extract Peptone Dextrose medium) ± zeocin: 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose ± 25 µg/mL zeocin

YPD-agar ± zeocin: YPD with 2% (w/v) agar ± 25 µg/mL zeocin

YPDS (Yeast extract Peptone Dextrose Sorbitol medium) ± zeocin: 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose, 1 M sorbitol ± 25 μg/mL zeocin

YPDS-agar ± zeocin: YPDS με 1.5% (w/v) agar ± 25 μg/mL zeocin

MD (Minimal Dextrose medium)-agar: 1.34% (w/v) yeast nitrogen base (YNB), 2% (w/v) dextrose, 4×10^{-5} % (w/v) biotin και 1.5% (w/v) agar

MM (Mininal Methanol medium)-agar: 1.34% (w/v) YNB, 4×10^{-5} % (w/v) biotin, 0.5% (v/v) methanol και 1.5% (w/v) agar

BMGY (Buffered Glycerol complex medium): 1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM phosphate buffer, pH 6.0, 1.34% (w/v) YNB: 4×10^{-5}% (w/v) biotin, 1% (v/v) glycerol

BMMY (Buffered Methanol complex medium): 1% (w/v) yeast extract, 2% (w/v) peptone, 4×10^{-5}% (w/v) biotin, 0.5% (v/v) methanol

BSM (Fermentation Basal Salts Medium): 85% (26.7 ml) phosphoric acid, 0.93 g calcium sulfate, 18.2 g potassium sulfate, 14.9 g magnesium sulfate-$7H_2O$, 4.13 g potassium hydroxide, 30 g glycerol, dH_2O to 1 liter + 4.35 ml PTM_1 trace salts
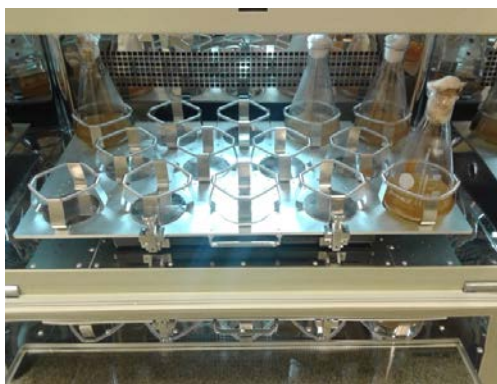
Figure 3.1. Cultivation of *P. pastoris* in shake flasks

Cultivation of *P. pastoris* was performed under continuous agitation (200 rpm), at 28 - 30°C, in YPD, in Erlenmeyer shake flasks (liquid cultures) or in YPD-agar plates. After electroporation, the cells were streaked out on YPDS-agar plates containing 100 μg mL^{-1} of zeocin and sorbitol which increases the viability of cells, and incubated at 30°C for 72 hours. MM and MD media were used for screening of

transformants. Small scale liquid cultures were performed in BMGY and BMMY media. One single *P. pastoris* colony harboring the recombinant gene was cultivated in BMGY medium for 18-24 hours, at 30°C in a shaker (200 rpm) and then inoculated into the production medium BMMY reaching $OD_{600}=1$. When *stock cultures* were prepared, *P. pastoris* strains containing the recombinant genes were grown in YPD medium over night. Then, culture was centrifuged, cells were collected and resuspended in YPD containing 15% (v/v) glycerol, to a final $OD_{600} = 50\text{-}100$ (approximately $2.5 - 5.0 \times 10^9$ cells/mL). The mixture was frozen in liquid nitrogen and stored at -80°C.

Cultivation of *P. pastoris* cells in *high cell-density fermentation* was performed in the basal salts medium (BSM), as described in the *Pichia* fermentation guidelines provided by Invitrogen (Invitrogen, *Pichia* Fermentation Process Guidelines). The engineered P. pastoris from the frozen glycerol stock was inoculated into BMGY medium in a baffled flask containing a total of 5-10% of the initial fermentation volume. The culture was grown overnight at 30°C and 200 rpm. The fermenter was sterilized with the fermentation Basal salts medium containing 3% glycerol. After sterilization and cooling, temperature was set to 28°C, agitation and aeration were adjusted to operating conditions, 800 rpm and 4 vvm , respectively. pH of the fermentation medium was adjusted to 5.0 with 28% ammonium hydroxide (undiluted ammonium hydroxide). The fermentation medium was supplemented aseptically with 4.35 mL of PTM1 trace salts per liter of the fermentation medium.
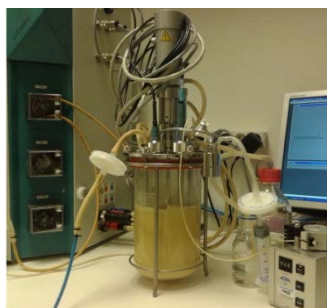
Figure 3.2. Fermentation of *Pichia pastoris* recombinant strains in Applikon bioreactor, equipped with an *ez*-Control system (Applikon Biotechnology B.V., Netherlands). Glycerol and methanol feeding was performed with *Alitea* XV pump with *MasterFlex* 96400 L/S 14 silicon tube (glycerol) and *PharmaMed* 0.51mm (methanol).

The fermenter was inoculated with the culture generated in the propagation shake flask to 10% of initial fermentation volume. After 24 hours of batch fermentation in glycerol medium, the end of glycerol batch was indicated by a sharp increase in the dissolved oxygen (DO) tension. This stage was followed by a 5-hour step of fed-batch glycerol one; during this step 50% w/v glycerol, with PTM_1 salts was fed at an initial flow-rate of 12 mL/h/lt of culture medium and was reduced gradually until it was fully consumed. At the same time, temperature was reduced from to 28^oC to 25^oC and finally to 23^oC and 2 mL of methanol were added manually in small aliquots with syringe. Total consumption of glycerol was again indicated by a spike in the DO. At the onset of methanol fed-batch phase, casamino acids solution was added at a final concentration of 3 g/lt and then, a feed of 100% MeOH, with PTM_1 was initiated at a flow rate of 1.9 mL/h/lt. The methanol consumption rate was monitored indirectly by stopping the feed and checking the "*lag phase*", while increasing the methanol feed rate manually. After 8h, feed rate was adjusted to a maximum of 5,46 mL/h/lt and maintained for ~20h, causing extracellular expression of the recombinant enzyme into the supernatant.. Then, the temperature was decreased to 21^oC and pure oxygen supply was set to maintain dissolved oxygen levels between 60-30 %. Induction time lasted approximately 145-170h in total. Sampling was performed at the end of each stage and twice daily. 10 mL samples were taken for each time point. Samples were analyzed for cell growth (OD_{600} and wet cell weight), protein concentration and enzyme activity.
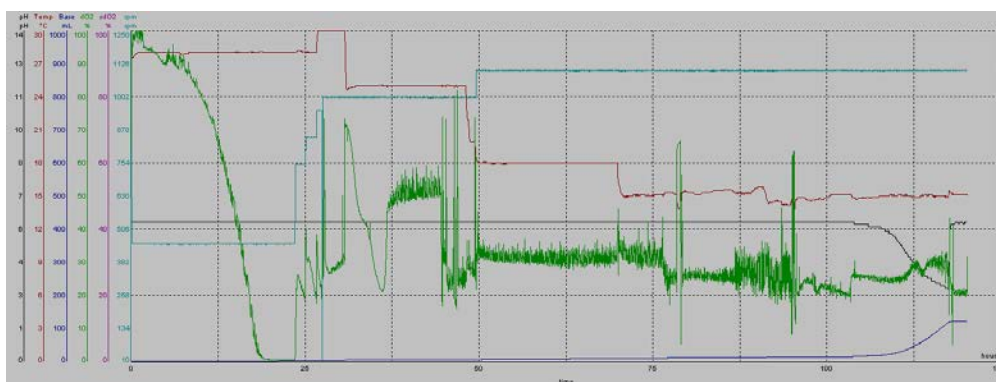
Figure 3.3. Typical schematic representation of parameters during *P. pastoris* fermentation (*green*: dissolved oxygen, *red*: temperature, light blue: agitation).

In this study, plasmid vector pCR® Blunt (Invitrogen) was used for transformation of bacterial cells *E. coli* TOP10 and vector pPICZaC (Invitrogen) for transformation of *E. coli* TOP10F' and *P. pastoris* yeast cells.

- **pCR® Blunt (Invitrogen)**

The plasmid vector pCR® Blunt (3512 bp) enables the cloning of DNA fragments carrying blunt ends, so it is appropriate for the direct cloning of PCR products. Insertion place is located between *LacZa* and *ccdB* genes. The *ccdB* (control of cell death) gene is interrupted when the gene of interest is inserted, enabling only the recombinant bacterial cells containing that gene to grow after transformation. The expression vector contains a kanamycin and a zeocin resistance gene that both allow selection for positive transformants that are resistant to zeocin.
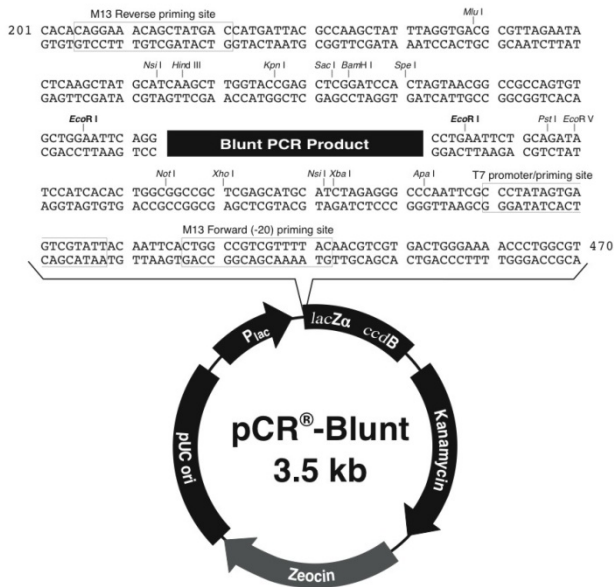
Figure 3.4. Gene map of the plasmid vector pCR® Blunt (3512 bp).

- **pPICZαC (Invitrogen)**

Among all commercial vectors that are used for expression of recombinant protein in *P. pastoris*, pPICZαC was selected to transfer the agu1 gene into *P. pastoris* The pPICZαC expression cassette contains a copy of α-mating factor (α- MF) signal sequence before the multicloning site, which leads to secretion of the recombinant protein. This vector also carries the 5′ AOX1 promoter The pPICZαC expression vector contains a zeocin resistance gene that allows selection for positive transformants that are resistant to zeocin. The short length of the pPICZαC (6kb) allows for efficient transformation an contributes to prepare stable expression strains compared to larger expression vectors developed for transformation into the *P. pastoris* (Daly and Hearn, 2005).

The alpha–mating factor pre-pro leader sequence (α-MF) consists of two regions, namely the pre and pro sequences. The pre-sequence contains a 19 amino acid signal peptide and the prosequence consists of a 60 amino acids. Upon translation, the pro-protein is translocated into the endoplasmic reticulum, while the signal peptidase cleaves the signal peptide sequence off the protein. For further processing the pro-protein is carried to Golgi where *kex2* protease removes the pro-sequence prior to secretion of the mature protein into the extracellular media (Brake, *et al.*, 1984).
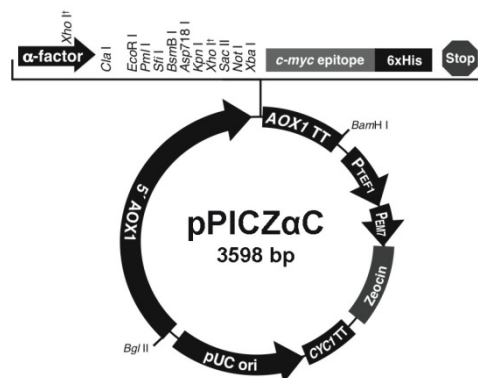
Figure 3.5. Gene map of the plasmid vector pPICZaC (3598 bp).

5′ AOX1 is a 942 bp fragment containing the AOX1 promoter that allows methanol-inducible, high-level expression in *Pichia* cells, target plasmid integration to the AOX1 locus and allows the insertion of expression cassette via homologous recombination.

Multiple cloning site with 10 unique restriction sites allows insertion of gene of interest into the expression vector

C-terminal myc-epitope tag (Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu-Asn) permits detection of the fusion protein by the respective antibody, in case this is necessary.

C-terminal polyhistidine tag permits rapid purification of recombinant fusion protein through one step affinity chromatography on metal-chelating resin

AOX1 Transcription Termination (TT) is the native transcription termination and polyadenylation signal from AOX 1 gene (~260 bp) that permits efficient 3′ mRNA processing, including polyadenylation, for increased mRNA stability

TEF1 promoter is referred to transcription elongation factor 1 gene promoter from *Saccharomyces cerevisiae* that drives expression of the *Sh ble* gene in Pichia, conferring Zeocin™ resistance (GenBank Acc. no. D12478,D01130).

EM7 (synthetic prokaryotic promoter) is a constitutive promoter that drives expression of the *Sh ble* gene in *E. coli* , conferring Zeocin™ resistance *Sh ble* gene (*Streptoalloteichus hindustanus ble* gene) Zeocin™ resistance gene for selection in E. coli

CYC1 transcription termination region is the 3′ end of the *Saccharomyces cerevisiae* CYC1 gene that allows efficient 3′ mRNA processing of the *Sh ble* gene for increased stability (GenBank Acc. no. M34014)

pUC origin allows replication and maintenance of the plasmid in *E. coli*

SacI, PmeI, BstX I sites represent unique **restriction sites** that permit linearization of the vectors at the AOX1 locus for efficient integration into the *Pichia* genome

3.6. DNA molecular techniques

3.6.1. DNA extraction methods

At the end of each procedure, DNA concentration was determined by measuring the OD260 nm (Sambrook *et al.*, 1989) and foolowing the equation:

$$C_{DNA} = [(OD_{260nm} \times 50 \times dilution) / 1000] \, (\mu g/\mu L)$$

as DNA concentration equal to 50 μg DNA/mL of solution corresponds to $OD_{260nm}=1$. In order to determine the purity of DNA sample, the ratio OD260nm/ OD280 nm was calculated. OD280nm measurement is indicative of protein content of a sample, so increased ratio corresponds to protein free samples.

- **DNA extraction from *M. thermophila***

    DNA extraction from therophilic fungus M. thermophila was performed according to protocol instructions from Sigma – Aldrich, GenElute ™ Plant, Genomic DNA Miniprep Kit. Submerged fungal cultures were incubated for 48h in glucose containing medium, then biomass was collected via filtration, freeze-dried, was ground vigorously and milled 0.1 g of powder were used for DNA extraction.

- **Plasmid DNA extraction**

    Plasmid DNA extraction was performed according to protocol instructions from Fermentas, GeneJET™ Plasmid Miniprep Kit, which is based on binding of plasmid DNA on the silica membrane in the spin column, removal of contaminants and subsequent elution with a small volume of the 10 mM Tris-HCl, pH 8.5.

- **DNA extraction from agarose gel**

    DNA extraction from agarose gel is a procedure necessary when PCR reaction product contains more than one sequences, so the gene of interest must be separated from the mixture. It includes the removal of agarose particles and isolation of the nucleic sequence for the subsequent cloning and other molecular techniques. In this study, DNA extraction from agarose gel was performed following the instructions of Macherey-Nagel, Nucleospin Gel Clean up kit. For better results, high-purity agarose gel was used (Seakem® Gold Agarose, Cambrex Bio Science Rockland, Inc., Danemark).

3.4 Polymerase Chain Reaction-PCR

    In the present thesis, PCR reactions were performed at TC-512 thermocycler (TECHNE, U.S.). Annealing temperature was set according to nucleic sequences melting temperature (T_m), following the equation $T_m = 2 \times (A+T) + 4 \times (G+C)$, where A,G,C,T represent adenine, guanine, cytocine and thymine. Primers Tm was defined by Eurofins MWG Operon (Germany). High-fidelity KOD Hot Start (Novagen) polymerase from *Thermococcus kodakarensis* was used. The enzyme exhibits proofreading activity to avoid mismatches.

	< 500 bp	500 – 1000 bp	1000 – 3000 bp	> 3000 bp
Polymerase activation	95ᵒC, 2 min	95ᵒC, 2 min	95ᵒC, 2 min	95ᵒC, 2 min
Denaturation	95ᵒC, 20 sec	95ᵒC, 20 sec	95ᵒC, 20 sec	95ᵒC, 20 sec
Annealing	58ᵒC, 10 sec	58ᵒC, 10 sec	58ᵒC, 10 sec	58ᵒC, 10 sec
Extension	72ᵒC, 10 s/kb	72ᵒC, 15 s/kb	72ᵒC, 20 s/kb	72ᵒC, 25 s/kb
Final extension	72ᵒC, 1 min	72ᵒC, 1 min	72ᵒC, 1 min	72ᵒC, 1 min

Table 3.1. Polymerase chain reaction conditions depending on the length of the target gene. The enzyme used was high-fidelity KOD Hot Start (Novagen), from *Thermococcus kodakarensis*. PCR reaction takes place for 20 - 40 cycles. When genomic DNA is used as template, increased number of cycles is necessary, as the target gene is contained in lower amount, whereas when plasmid DNA is used, fewer cycles can be used. In case of overlapping PCR, more cycles ensure efficient product concentration, which is crucial for the successful subsequent transformation experiments.

Polymerase	1 μL (1 U/μL)
DNA	5 μL (10-500 ng)
Primer 1	1 μL (50 pmol)
Primer 2	1 μL (50 pmol)
dNTPs mix	5 μL (8 mM)
MgSO₄ (25mM)	3 μL
Buffer solution	5 μL (10×)
Ultrapure H₂O	29 μL

Table 3.2. Composition of Polymerase Chain Reaction mixture for the amplification of DNA fragments. Final reaction volume is 50 μL.

3.6.2. Restriction enzymes

Restriction enzymes for double-cut restriction analysis, as well as linearization of plasmid vectors that were used in this thesis are listed below. Digestion and linearization reactions were performed according to Takara's protocols (Japan).

Restriction enzyme	Source organism	Restriction site
ClaI	*Caryophanon latum*	5' T CTAGA 3' 3' AGATC T 5'
XbaI	*Xanthomonas badrii*	5' AT CGAT 3' 3' TAGC TA 5'
SacI	*Streptomyces achromogenes*	5' GAGCT C 3' 3' C TCGAG 5'
PmeI	*Pseudomonas mendocina*	5' GTTT AAAC 3' 3' CAAA TTTG 5'

Table 3.3. Restriction enzymes that were used in this thesis and their action sites.

Restriction enzymes *ClaI* and *XbaI*		Restriction enzyme *SacI* and *PmeI*	
Plasmid vector / gene	5 µL	Plasmid vector / gene	80 µL
buffer (10x)	2 µL	buffer (10x)	10 µL
Restriction enzyme	1 µL	Restriction enzyme	5 µL
Ultrapure water	12 µL	Ultrapure water	5 µL

Table 3.4. Composition of Digestion and Linearization reaction mixtures. Digestion with *ClaI* and *XbaI* was performed to provide sticky complementary ends for the subsequent ligation of the DNA gene of interest and the appropriate plasmid vector. The enzymes were also used for restriction analysis in order to confirm the insertion of gene of interest after ligation. *SacI* and *PmeI* enzyme were used for the linearization of plasmid vectors before transformation of *P. pastoris* cells through electroporation.

Digested DNA fragments were mixed with the linearized vector in the presence of T4 DNA ligase from Fermentas (Maryland, USA). The ligation reaction set up was according the manufacturer's instructions. The ligation mixtures were incubated overnight at room temperature. Ligation products were used for transformation into pCR® Blunt or pPICZaC vectors according to Zero Blunt® PCR Cloning Kit and pPICZa Cloning kit (Invitrogen).

pCR® – Blunt (25 ng)	1 μL
PCR product	5 μL
Buffer (10×)	1 μL
Ultrapure water	2 μL
Ligase (4U/μL)	1 μL

Table 3.5. Composition of Ligation reaction mixture for the insertion of gene to the plasmid vector pCR® Blunt. Incubation takes place at 16^0C for 1-2 hours. Final reaction volume is 10 μL.

	1:3	**1:5**	**1:7**	**Control 1**	**Control 2**
pPICZaC (100 - 200 ng)	1 μL	1 μL	1 μL	1 μL	1 μL
DNA fragment	3 μL	5 μL	7 μL	---	---
Buffer (10×)	2 μL	2 μL	2 μL	2 μL	2 μL
Ultrapure water	13 μL	11 μL	9 μL	16 μL	17 μL
Ligase (4U/μL)	1 μL	1 μL	1 μL	1 μL	---

Table 3.6. Composition of Ligation reaction mixture for the insertion of gene to the plasmid vectorpPICZaC. Incubation takes place at 16^0C for 3-4 hours. Final reaction volume is 20 μL.

- **Transformation of *E. coli* competent cells**

500 μL of competent cells *E.coli* were supplemented with 1 μl of the appropriate vector. The mixture was incubate for 30 min on ice before thermal heat shock at 42^oC for 90 sec. Immediately, the transformation tube was incubated on ice for additional 10 minutes and 1mL LB medium was added. Cells were incubated at 37^oC and 200 rpm for 1 hour and then streaked on LB agar plates containing 50 / 100 μg mL^{-1} kanamycin / zeocin respectively. Plates were incubate at 30^oC overnight and positive transformant cells were selected.

- **Transformation of *P. pastoris* by electroporation**

Electric pulses of intensity in kilovolts per centimeter and of duration in microseconds to milliseconds cause a temporary loss of the semi -permeability of cell membranes, thus leading to ion leakage, escape of metabolites, and increased uptake by cells of drugs, molecular probes, and DNA. A generally accepted term describing this phenomenon is "*electroporation.*" (Tsong, 1991). Electroporation has found many applications, such as introduction of plasmids or foreign DNA into living cells for gene transfections, fusion of cells to prepare heterokaryons, hybridoma etc., insertion of proteins into cell membranes, improving drug delivery and hence effectiveness in chemotherapy of cancerous cells and alteration of genetic expression in living cells (Dower *et al.,* 1988). In *P. pastoris,* electroporation is used for the insertion of recombinant plasmid vectors carrying the gene of interest, causing the subsequent production of the heterologous protein from the yeast cells. This method is more effective than other methods of chemical transformation and leads to higher yield of living transformed cells ($10^3 - 10^4$ cells/μg linearized DNA).

P.pastoris X33 was grown overnight, in shake flasks, in YPD medium at 30^oC and 180 rpm in a rotary shaker. Cells were harvested for electroporation according to the Invitrogen protocol (Cat. no.V195-20, version F) and resuspended in 1 mL of 1 M ice cold sorbitol. The linearized vectors were mixed with the resuspended cells and then transferred to an electroporation cuvette, which was then incubated on ice for 5 min. The BIO-RAD electroporation device's parameters were

adjusted to *P. pastoris* set up. Cells were pulsed and then 1 mL of 1 M ice cold sorbitol was added to the electroporation cuvette. Contents of the cuvette were transferred to a 15 mL Falcon tube and incubated at without shaking for 2 hours. Finally the cells were streaked out on YPDS agar plates containing 100 μg mL-1 of zeocin and incubated at for 72 hours. Zeocin resistant colonies were selected and grown overnight, in shake flasks, in YPD medium containing 100 μg mL^{-1} of zeocin at 30°C. Cells were harvested at 12000 rpm and their genomic DNA were isolated by phenol chloroform method and amplified by PCR through AOX forward and reverse primers to verify the recombination of the expression cassette into the X33 strains genome. Then the PCR products were ran on gel electrophoresis.

A   **B**   **C**
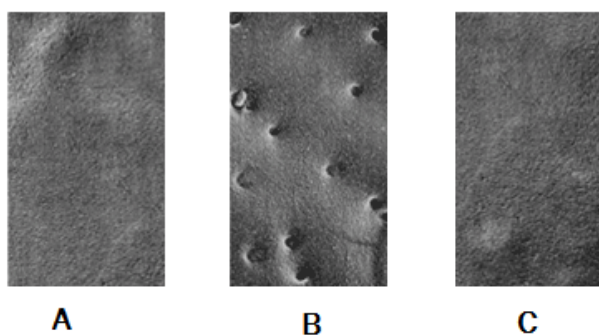
Figure 3.6. Schematic representation of plasma cell membrane during electroporation. Application of electric pulse leads to increase of permeability through small diameter pores, that close immediately after allowing the insertion of linearized plasmid DNA.

4.1.4. Immobilized Metal Affinity Chromatography, IMAC

IMAC was introduced in 1975 as a group-specific affinity technique for separating proteins (Porath *et al.*, 1975). The principle is based on the reversible interaction between various amino acid side chains and immobilized metal ions. Depending on the immobilized metal ion, different side chains can be involved in the adsorption process. Most notably, histidine, cysteine, and tryptophan side chains have been implicated in protein binding to immobilized transition metal ions and zinc

(Sulkowski, 1985; Zhao *et al.*, 1991). TALON Resin is cobalt-based IMAC resin designed to purify recombinant polyhistidine-tagged proteins (Bush *et al.*, 1991); it is compatible with many commonly used reagents and allows protein purification under native or denaturing conditions.
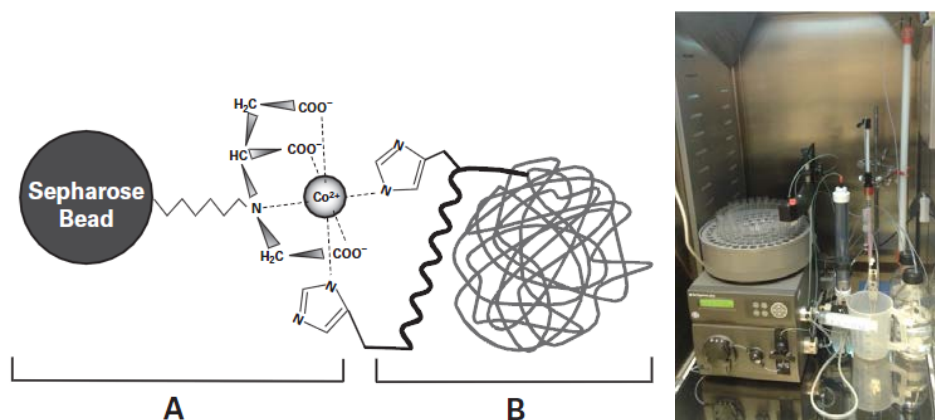
Figure 3.7. Schematic diagram of the TALON IMAC System. (*left*) **Part A.** TALON Metal Affinity Resin; A Sepharose bead bearing the tetradentate chelator of the Co2+ metal ion. **Part B.** The polyhistidine-tagged recombinant protein binds to the resin. (*right*) ÄKTA Prime Plus liquid chromatography system, equipped with Prime View 5.31. software (GE Healthcare Life Sciences).

In this thesis, IMAC column containing 15 mL Talon® Metal Affinity Resin (Clontech, U.S.) was initially equilibrated with 300 mL Talon buffer 1X with flow rate 2.5 mL/min. Sample containing the histidine-tagged protein in excess was loaded onto the column at a flow rate 1 mL/min. Column was washed with 300 mL Talon buffer 1X, or until the OD_{280nm} reached the baseline indicating that all non-histidine tagged proteins had been eluted. Then, a linear gradient from 0 to 100 mM imidazole in 20 mM Tris-HCl buffer containing 300 mM NaCl (60 ml, pH 8.0) was applied at a flow rate of 2 ml/min. Fractions (2 ml) containing the protein of interest were concentrated and the homogeneity of was examined on a SDS-PAGE.

References

Brake, A. J. *et al.*, 1984. Alpha-factor-directed synthesis and secretion of mature foreing proteins in *Saccharomyces cerevisiae. Proc Natl Sci USA*, 81, 4642-4646.

Daly, R. and Hearn, M.T., 2005. Expression of heterologous proteins in *Pichia pastoris*: a usefull experimental tool in protein engineering and production. *Journal Molecul Recognit*, 18, 119-138.

Dower, W.J., Miller, J.F., Ragsdale, C.W. 1988. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic acids Research* 16, 6127-6145.

Porath, J., Carlsson, J., Olsson, I. & Belfrage, G. 1975 Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* 258:598-599.

Sambrook, J., Fritisch, E.F., Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual, 2nd Ed., N. Y.: Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sulkowski, E. 1985. Purification of proteins by IMAC. *Trends Biotechnol.* 3,1-7.

Tsong, T.Y. 1991. Electroporation of cell membranes. *Biophys J.* 60. 297–306.

Zhao, Y.J., Sulkowski, E., Porath, J. 1991. Surface topography of histidine residues in lysozymes. *Eur. J. Biochem.* 202, 1115-1119.

CHAPTER 4

The lignocellulolytic system of *Myceliophthora thermophila* (PAPER I)

The filamentous fungus *Myceliophthora thermophila* constitutes an exceptionally powerful cellulolytic microorganism that synthesizes a complete set of enzymes necessary for the breakdown of plant cell wall. The genome of this fungus has been recently sequenced and annotated, allowing systematic examination and identification of enzymes required for the degradation of lignocellulosic biomass. In this chapter, an overview of the cellulolytic and hemicellulolytic potential of this fungus regarding the degradation of plant cell wall material will be given, revealing the existence of an expanded enzymatic repertoire including numerous enzymes with auxiliary activities, covering the most of the recognized CAZy families (**Table 4.1***). All sequences used in this study were extracted from Genome Portal database (http://genome.jgi-psf.org) and the continually updated CAZy database (http://www.cazy.org/; Lombard *et al.*, 2014). The conserved domains were found with Pfam/InterProscan (Punta *et al.*, 2012; http://pfam.sanger.ac.uk/), while the theoretical molecular mass and isoelectric point for each protein were calculated using the ProtParam tool of ExPASY (http://web.expasy.org/protparam/). Predicted secretome was extracted using SignalP v4.0 (http://www.cbs.dtu.dk/services/SignalP/). Post-translational glycosylation sites were predicted with NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) and NetOGlyc 3.1 server (http://www.cbs.dtu.dk/services/NetOGlyc/). These data offer a better understanding of activities embedded in fungal lignocellulose decomposition mechanisms and suggest that *M. thermophila* could be made usable as an industrial production host for cellulolytic and hemicellulolytic enzymes.

* Tables are provided at the end of this Chapter.

4.1. The Cellulolytic system of *M. thermophila*

The cellulolytic system of *M. thermophila* consists of a repertoire of enzymes with endoglucanase (EG), cellobiohydrolase (CBH) and β-glycosidase (BGL) activity. Throughout the genome of this fungus, there are eight sequences encoding EG activity, seven sequences of CBH activity and nine sequences of BG activity (**Figure 4.1**; **Table 4.2***). The theoretical average molecular weight of the translated proteins is calculated at 51.05 ± 16.2 kDa and the theoretical pI at 5.58 ± 0.3. EGs are distributed to families GH5, 7, 12, and 45, all predicted to possess a secretion signal and several *N*- and *O*-glycosylation sites. Only two of them exhibit a CBM that belong in family 1. CBHs represent three non-reducing acting enzymes of GH6 family and four reducing-end acting enzymes of GH7 family. All of these enzymes seem to be targeted to secretion pathway and modified with glycans during post-translational modifications. BGs are classified to GH3 family, except one GH1 sequence, while none of them exhibit a CBM, as expected. Four are secreted and have potential *N*- and *O*-glycosylation sites, showing the highest molecular weight compared to the other cellulases, with a theoretical average value of 85.18 ± 3.2 kDa.
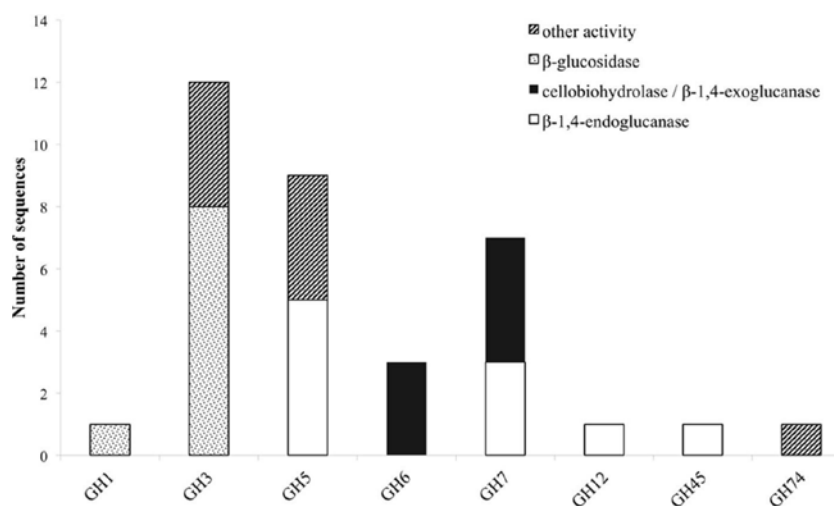
Figure 4.1. Distribution of cellulolytic enzymes of *M. thermophila* throughout eight GH families. Other activities refer to β-xylosidase (GH3), β-1,6-galactanase, β-1,3-glucanase, endo-1,4-beta-mannosidase or putative proteins with unknown function (GH5). GH74 represents xyloglucan specific 1,4-endoglucanase/xyloglucanase.

* Tables are provided at the end of this Chapter.

Totally, 12 cellulases have been isolated and characterized (**Table 4.3***). The group of Bukhtojarov *et al.* (2004) investigated the properties of individual cellulases from the multienzyme complex produced by a mutant strain of *M. thermophila* C1 (Visser *et al.*, 2011). Among EGs, the highest saccharification activity was displayed by EG60 and EG51, representing enzymes of 60 and 51 kDa, respectively, which exhibited pI values of 3.6 and 5.0, respectively. It has been shown later that the EG51 and EG60 represent the GH5 and GH7 EGs from *M. thermophila*, respectively, (Gusakov *et al.*, 2011). A different EG (*St*Cel5A) displays a typical GH5 domain, exhibiting optimal activity at pH 6.0 and 70°C and retained greater than 50% of its activity following 2 h of incubation at 55°C, diluted in 10 mM citrate buffer pH 4.5 (Tambor *et al.*, 2012). Two EG genes, belonging to GH7 and GH5 families were functionally expressed in methylotrophic yeast *P. pastoris* and subsequently characterized, as described in **Chapter 5** and **Chapter 7** of this thesis respectively. Substrate specificity analysis revealed that the GH7 EG enzyme is one of the most thermostable fungal enzymes reported up to now and exhibits high activity on substrates containing β-1,4-glycosidic bonds as well as activity on xylan-containing substrates (Karnaouri *et al.*, 2014). Moreover, *Mt*EG7a was proved to liquefy rapidly and efficiently pretreated wheat straw, indicating EGs' key role to the initial step of hydrolysis of high-solids lignocellulose substrates (Karnaouri *et al.*, 2014). This change in viscosity of these substrates is probably due to the gradual reduction of the average chain length of cellulose polysaccharides by endo-acting enzymes, such as endoglucanases.

Totally, four CBHs and two BGs have been isolated from *M. thermophila* crude supernatant and studied. CBH IIb is the product of MYCTH_66729 gene that represents an enzyme of GH6 family, which is attached to polysaccharide substrate through a CBM and exhibits high levels of activity in comparison to other CBHs (Gusakov *et al.*, 2007). In the same study, the isolation of CBH Ib, a GH7 family enzyme (MYCTH_2140736) is reported, which acts mainly against microcrystalline cellulose and CMC. Bukhtojarov *et al.* (2004) studied the properties of CBH Ia and CBH IIa, which are classified to GH7 and GH6 family, respectively. CBH Ia is the product of MYCTH_109566 gene, and seems to be expressed in two isoforms with distinct molecular weights, one exhibiting the catalytic domain owing a CMB and the

* Tables are provided at the end of this Chapter.

other only the catalytic domain and part of the linker, after proteolysis. This enzyme is produced as a major protein of fungi's secretome (20–25% of the total extracellular protein) and adsorbed strongly on microcrystalline cellulose. It has been shown that there is a significant synergism between CBH IIb and CBH Ia enzymes during substrate hydrolysis (Gusakov *et al.*, 2007). The MYCTH_66729 and MYCTH_109566 genes encoding the two enzymes belonging to GH6 and EG7 families respectively were functionally expressed in methylotrophic yeast *P. pastoris* and subsequently characterized, as described in **Chapter 7** of this thesis.

Apart from the enzymes with cellulolytic activity, *M. thermophila* was found to produce an exo-β-1,4-glucanase (Xgl74A) that catalyzes the hydrolysis of (1-4)-D-glucosidic linkages in xyloglucans aiming in the successful removal of oligosaccharides from the chain end (Grishutin *et al.*, 2004). The enzyme exhibits high specific activity toward tamarind xyloglucan, and very low or absent activity against carboxymethylcellulose (CMC) and barley β-glucan. Due to its unique substrate specificity the enzyme was given a new number in the Enzyme Nomenclature (EC 3.2.1.155). Apart from Xgl74A, two out of the seven cellulases reported from *M. thermophila* (Cel12A and Cel45A) possess a notable activity against xyloglucan, together with their major activities toward CMC and barley β-glucan (Bukhtojarov *et al.*, 2004).

4.2. The Hemicellulolytic system of *M. thermophila*

As described in **Chapter 2**, hemicellulose polymers have a much more diverse structure than cellulose and consequently several enzymes are needed to completely degrade the polysaccharide into monosaccharides. *M. thermophila's* hemicellulase genes are organized in 8 GH families (3, 10, 11, 30, 43, 51, 62, and 67) (**Figure 4.2**) and nine carbohydrate esterase (CE) families (1, 3, 4, 5, 8, 9, 12, 15, and 16) (**Figure 4.3**). Many of the encoding proteins have been isolated from the WT (wild type) culture supernatant or expressed in heterologous hosts and finally characterized in terms of specific activity and physicochemical properties. The majority of them are predicted to follow the secretion pathway, while modified with *N*- and/or *O*- glucans, comprising a

* Tables are provided at the end of this Chapter.

total amount of 66 enzymes that act synergistically for the degradation of hemicellulose.
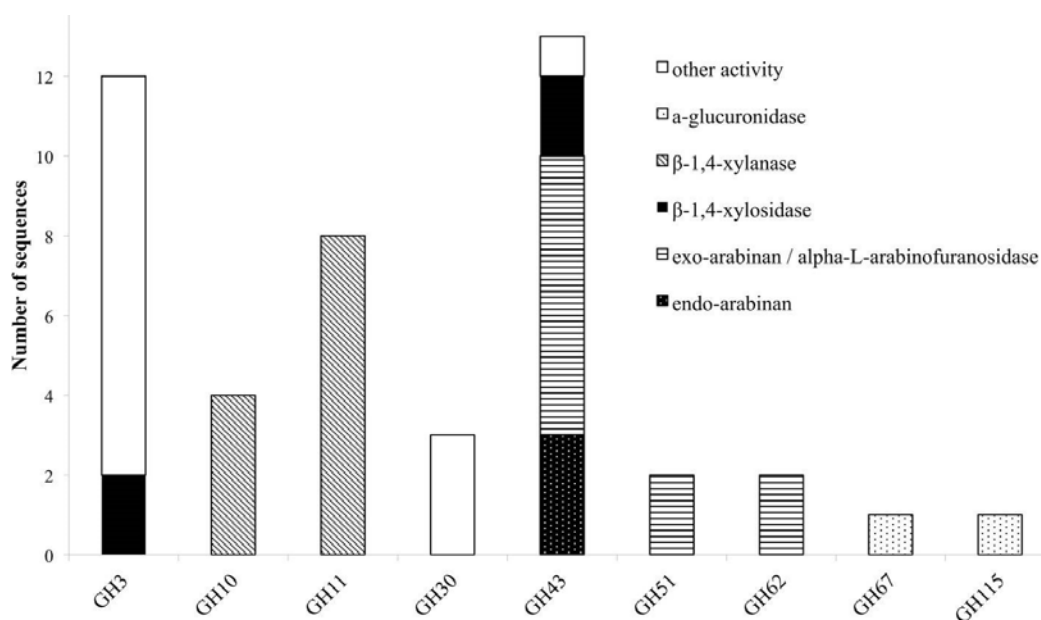
Figure 4.2. Distribution of hemicellulolytic enzymes of *M. thermophila* throughout nine GH families. Other activities refer to β-glycosidase (GH3), xylanase with endo-exo mode of action and xylobiohyrolase (GH30), and galactan 1,3-beta-galactosidase (GH43).

Xylanases / Xylosidases

For the degradation of xylan, the genome of *M. thermophila* encodes totally 12 xylanases with endo- mode of activity, classified to GH 10 and 11 and four xylosidases, classified to GH3 and 43 families (**Table 4.4***). All aforementioned xylan-degrading translated sequences, apart from three, are predicted to exhibit a potential secretion signal. Xylanases possess 1-3 *N*-glycosylation and several *O*-glycosylation sites; whereas more *N*-sites are predicted for xylosidases, though not all of them are glycosylated during post-translational modifications. GH family 30 contains two genes encoding xylanolytic enzymes with endo-exo activity and one sequence for a

* Tables are provided at the end of this Chapter.

characterized xylobiohydrolase, releasing xylobiose units from the substrate (Emalfarb *et al.*, 2012).

Ten xylanases have been purified and characterized from multienzyme preparations of *M. thermophila* modified strains (Ustinov *et al.*, 2008; van Gool *et al.*, 2013). Four of them, belonging to GH10 family (**Table 4.5***), are the products of two genes, either with the presence of a family 1 CBM or displaying only the catalytic domain after partial proteolytic digestion (Ustinov *et al.*, 2008). These enzymes, thought classified to the same family, can hydrolyze different types of decorated xylans. They differ in degradation of high and low substituted substrates and the substitution pattern seems to be an important factor influencing their efficiency (van Gool *et al.*, 2012). Six xylanases, belonging to GH11 family, represent true xylanases, with high specific activities against glucuronoxylans and arabinoxylans. Four of these enzymes exhibit lower thermostability in comparison to GH10 xylanases, in which extended glycosylation has been noticed (Ustinov *et al.*, 2008). One showed a substrate specificity pattern similar to GH10 enzymes and secreted in two forms, with or without CBM (van Gool *et al.*, 2013).
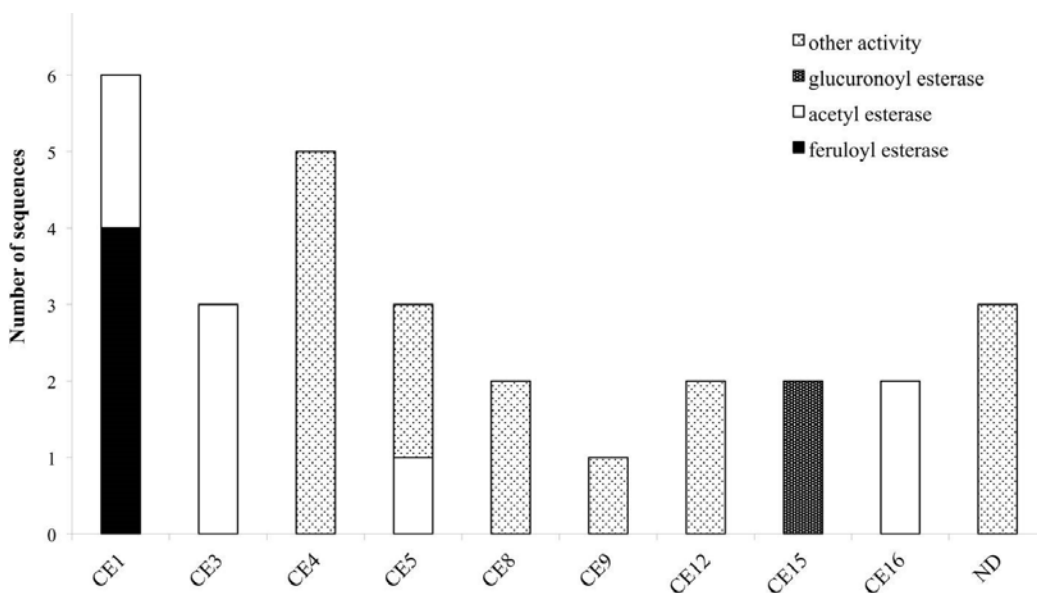
Figure 4.3. Distribution of hemicellulolytic enzymes of *M. thermophila* throughout nine CE families. Family CE4 is comprised of putative proteins with polysaccharide deacetylase activity, CE5 of cutinases and CE8, 12 of pectin esterases. ND (not

* Tables are provided at the end of this Chapter.

determined) refers to sequences encoding putative proteins with unknown activity which are not classified to a specific family.

Esterases

The role of esterases in the breakdown of lignocellulosic material is complex and includes the cleavage of bonds between the main hemicellulose part and many types of side chains. So, upon a closer examination of the genome sequences of *M. thermophila*, there is a wide distribution of enzymatic activities through CE families. These enzymes are classified into nine families and their main activities, among others, include the hydrolysis of feruloyl and acetyl ester bonds.

Feruloyl esterases (FAEs; EC 3.1.1.73) are enzymes responsible for cleaving the ester-link between the polysaccharide main chain of xylans and monomeric or dimeric ferulates. They act synergistically with xylanases to release ferulic acid from cell-wall material and can be divided into four groups, namely A–D. The main difference between groups A and D is their substrate specificity toward synthetic substrates and their capability of liberating diferuloyl bridges (Crepin *et al.*, 2004). One of the first FAEs reported from thermophilic fungi, was produced from *M. thermophila* under solid-state fermentation (SSF) conditions. The esterase activity was isolated and partially characterized for its ability to release ferulic acid from complex substrate, destarched wheat bran (Topakas *et al.*, 2003). Two other FAEs, *St*FaeB, a protein with molecular weight of 66 kDa (homodimers of 33 kDa) (Topakas *et al.*, 2004) and *St*FaeC, 46 kDa (homodimers of 23 kDa) (Topakas *et al.*, 2005), were purified to homogeneity from culture supernatants of *M. thermophila*. *St*FaeB hydrolyzed methyl p-coumarate, methyl caffeate and methyl ferulate and was active on substrates containing ferulic acid ester linked to the C-5 and C-2 linkages of arabinofuranose. *St*FaeC showed maximum catalytic efficiency on 4-hydroxy-3-methoxy cinnamate, a substrate with both hydroxyl and methoxy substituents, indicating that it may be the most promising type of FAE as a biocatalyst for the enzymatic feruloylation of aliphatic alcohols, oligo- and polysaccharides. Properties of characterized FAEs are summarized in **Table 4.6*.** Among the sequences registered to Genome Portal, there are four sequences encoding

* Tables are provided at the end of this Chapter.

proteins with catalytic activity of FAE, all belonging to CE family 1. Two of them (MYCTH_48379, MYCTH_39279) seem to be identical with characterized FAEs secreted from *M. thermophila C1* strain (Kühnel *et al.*, 2012). One sequence (JDI ID: 96478) has been heterologously expressed in *P. pastoris* and encoded a 39 kDa protein (fae1A; *Mt*Fae1a), which showed high activity toward methyl caffeate and p-coumarate and a strong preference for the hydrolysis of n-butyl and iso-butyl ferulate (Topakas *et al.*, 2012). In addition, *Mt*Fae1 esterase releases ferulic acid from destarched wheat bran only by the synergistic action of an endo-xylanase (a maximum of 41% total ferulic acid released after 1 h incubation). MYCTH_2302953 sequence has not yet been characterized; however it still shows 66% identity with a type B FAE from *Neurospora crassa* (CAC05587.1). All proteins encoded by the above sequences appear to be secreted and bring several *N*- and *O*-glycosylation sites, as shown in **Table 4.7***.

About 60–70% of the xylose residues in hardwood xylan are acetylated at the C2 and/or C3 positions (Lindberg *et al.*, 1973). The complete degradation of acetylated xylans by microbes requires the action of *acetyl esterases* (AcEs; EC 3.1.1.72), which cleave acetyl side groups from the heteroxylan backbone, and act in synergy with other hemicellulases (Tenkanen *et al.*, 1996). Eight sequences that encode proteins with AcE activity were detected in the genome of *M. thermophila* and showed identity with characterized enzymes. All of them are secreted, as predicted with SignalP and belong to CE families 1, 3, 5, 16 (**Table 4.7***). Two of them, Axe2 and Axe3, which bare members of CE5 and CE1 families, respectively, were isolated and characterized (Pouvreau *et al.*, 2011a). Annotated genes, encoding the putative enzymes were cloned into the specially designed *M. thermophila C1*-expression host (Verdoes *et al.*, 2010) and over-produced in the culture medium. Axe2 and Axe3 are able to hydrolyze acetyl groups when they are substituted to the O-2 and O-3 positions of acetylated xylo-oligosaccharides and complex insoluble polymeric substrates and had a preference for xylooligosaccharides (Pouvreau *et al.*, 2011a).

Glucuronoyl esterases (GEs) are recently discovered enzymes that are suggested to play an important role in the dissociation of lignin from hemicellulose and cellulose by cleaving the ester bonds between the aromatic alcohols of lignin and the carboxyl groups of 4-O-methyl-D-glucuronic acid residues in glucuronoxylan (Špániková and

* Tables are provided at the end of this Chapter.

Biely, 2006). Sequence alignment studies of these enzymes have revealed a novel conserved amino acid sequence G-C-S-R-X-G that features the characteristic serine residue involved in the mechanism of this esterase family. It has been shown that the mode of action probably involves a nucleophilic serine (Topakas *et al.*, 2010). The genome of *M. thermophila* possesses two genes classified to family CE15 that encode proteins with activity of 4-O-methyl-glucuronoyl esterase. Both putative enzymes are secreted and have potential glycosylation sites. The first GE (*St*EG1), isolated from the culture filtrate of *M. thermophila*, was proved to be a thermophilic enzyme that presents a C-terminal CBM, which was active on substrates containing glucuronic acid methyl ester (Vafiadi *et al.*, 2009). Another CE15 protein molecule, StGE2 was heterologously expressed in yeast *P. pastoris* and was used to prove that nucleophilic serine residue is responsible for catalytic action of GEs, through site-directed mutagenesis studies (Topakas *et al.*, 2010) and crystal structure determination (Charavgi *et al.*, 2013).

Mannan-degrading enzymes

Mannan is a great component of hemicellulose, therefore, as expected the lignocellulolytic toolbox of *M. thermophila* possesses a complete reservoir of genes encoding mannan degrading enzymes. The genome of this fungus encodes three enzymes that putatively catalyze random cleavage of the mannan polysaccharide and belong to GH family 5 and 26. One of these enzymes has been isolated from culture supernatant, characterized and classified as GH5 endo-β-1,4-mannosidase (bMan2, Dotsenko *et al.*, 2012). In addition, there are two genes encoding putative β-mannosidases belonging to GH2 family, while one of them has been characterized in terms of its specificity and physicochemical properties (bMann9, Dotsenko *et al.*, 2012). Two GH27 and one GH26 α-galactosidases boost the efficiency of fungal culture supernatant against hydrolysis of mannan substrate (Emalfarb *et al.*, 2012), while two CE12 family genes encoding proteins with high similarity to known acetyl-mannan esterases have been found.

Arabinohydrolases

L-arabinose is widely present in various hemicellulosic biomass components, such as arabinoxylan, where the main β-D-(1,4)-linked xylopyranosyl backbone is

* Tables are provided at the end of this Chapter.

substituted with arabinose residues. α-L-arabinofuranosidases (AFase; EC 3.2.1.55) are enzymes that release arabinofuranose residues substituted at position O-2 or O-3 of mono or di-substituted xylose residues (Gruppen *et al.*, 1993). Apart from that, AFases act in synergism with other arabinohydrolases, endo-(1,5)-α-L-arabinanases (ABNase; EC 3.2.1.99) for the decomposition of arabinan, a major pectin polysaccharide.

The genome of the *M. thermophila* encodes 14 enzymes that putatively release arabinose or arabinose oligomers from arabinan (Hinz *et al.*, 2009). Throughout CAZy families, arabinohydrolases belong to GH families 43, 51, 54, 62, and 93 (**Figure 4.3***). Eleven sequences contain a secretion signal peptide and produced as extracellular or cell-bounded proteins, while almost all of them exhibit isoelectric point around 4.6–5.6 (**Table 4.8***). Seven of them have been selectively overexpressed homologously in *M. thermophila C1* host and found to release arabinose from wheat arabinoxylan polymers and oligomers (Hinz *et al.*, 2009). *M. thermophila* arabinofuranosidases are selective in releasing arabinose from either single or double substituted xylose residues in arabinoxylans. Eight enzymes, belonging to GH families 43, 51, 62, and 93 with different type of arabinolytic activity have been purified and characterized (Hinz *et al.*, 2009; Kühnel *et al.*, 2011; Pouvreau *et al.*, 2011b) (**Table 4.9***).

Abn7 and Abf3 are GH43 and GH51 arabinases respectively, which were selectively produced in C1 host. Abn7 was found to hydrolyze arabinofuranosyl residues at position O-3 of double substituted xylosyl residues in arabinoxylan-derived oligosaccharides, while Abf3 released arabinose from position O-2 or O-3 of single substituted xyloses. When these enzymes were incubated together, in combination with a GH10 endo-xylanase for the hydrolysis of arabinoxylans, they resulted in a synergistic increase in arabinose release from the substrate (Pouvreau *et al.*, 2011b). In addition, a-L-arabinohydrolases Abn1, Abn2, and Abn4 were overexpressed in C1 and the produced culture supernatant has been shown to produce neutral branched arabino-oligosaccharides from sugar beet arabinan by enzymatic degradation. As found by sugar analysis, neutral arabino-oligosaccharides contained an α-(1,5)-linked backbone of l-arabinosyl residues and carried single substituted α-(1,3)-linked l-arabinosyl residues or consisted of a double substituted α-(1,2,3,5)-linked arabinan structure within the molecule (Westphal *et al.*, 2010). Enzyme Abn4 belongs to GH43 family and

* Tables are provided at the end of this Chapter.

is more active toward branched polymeric arabinan substrate that releases arabinose monomers from single substituted arabinose residues, while Abn1 and Abn2 are active toward linear arabinan (Kühnel *et al.*, 2010). Abn2 is a member of GH93 family that consists of exoarabinases acting on linear arabinan, hydrolyzing the α-1,5-linkages of arabinan polysaccharides presented as side chains of pectin. Their mode of action was studied with Abn2, which binds two arabinose units at the subsites −1 and −2 and releases arabinose. Three more arabinohydrolases were also overexpressed in C1 strain (Hinz *et al.*, 2009). Abn5 was found to be specifically active toward arabinan, but not arabinoxylan. Arabinofuranosidases Abf1 and Abf2, members of GH62 family released O-2 or O-3 substituted arabinose or linked arabinofuranosyl from mono substituted xylose. GH family 62 arabinofuranosidases are reported to be predominantly active toward arabinoxylan and are, therefore, also called arabinoxylan arabinofuranohydrolases (Beldman *et al.*, 1997). Several of these enzymes contain either a CBM1, like Abf1, or a CBM43 (xylan)-binding domain.

4.3. Auxiliary enzymes

In spite of the cooperative activity exhibited by the cellulolytic and hemicellulolytic enzymes, the impressive hydrolytic ability of various microorganisms in nature cannot be attributed only to this endo–exo mechanism. Apart from the hydrolytic system responsible for carbohydrate degradation, it seems that an oxidative system catalyze lignin depolymerization and oxidation of plant cell wall components, yielding reactive molecules (e.g., H_2O_2). Recent evidence highlights the critical role of alternative enzymatic partners involved in the oxidation of cell wall components. Among these enzymes, outstanding role during hydrolysis exhibit the originally described as cellulases LPMO enzymes, CDH and multicopper enzymes such as laccases. The genome of *M. thermophila* possesses more than 30 genes that encode proteins with such auxiliary activities(**Figure 4.4**). Members of the LPMO family AA9, have been shown to be copper-dependent monooxygenases that enhance cellulose degradation in concert with classical cellulases, as aforementioned before and reviewed by Dimarogona *et al.* (2013). These enzymes catalyze the cleavage of cellulose by an oxidative mechanism provided that reduction equivalents are available. These equivalents either involve low molecular weight reducing agents (e.g., ascorbate) or are

* Tables are provided at the end of this Chapter.

produced by CDH activity (Langston *et al.*, 2011). CDHs are extracellular enzymes produced by various wood-degrading fungi that oxidize soluble cellodextrins, mannodextrins and lactose efficiently to their corresponding lactones by a ping-pong mechanism using a wide spectrum of electron acceptors (Henriksson *et al.*, 2000). Throughout the genome of *M. thermophila*, two genes encoding proteins classified to AA3 and 8 families have been identified (**Figure 4.4**). Both of them are predicted to be secreted in the culture supernatant and have potential glycosylation sites. The translated CDH MYCTH_111388 exhibits a C-terminal CBM and a cDNA clone of this sequence has been isolated and biochemically characterized by screening an expression library of *M. thermophila* (Subramaniam *et al.*, 1999). Canevascini *et al.* (1991) purified a monomeric (91 kDa) and a dimeric (192 kDa) form of CDH that differed not only in molecular weight, but amino acid composition and carbohydrate content. Both forms oxidized cellobiose in the presence of cytochrome c or dichlorophenol–indophenol.
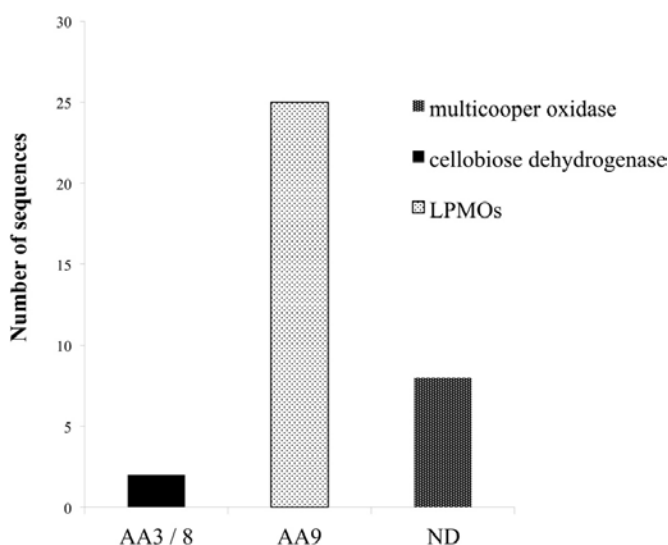
Figure 4.4. Distribution of enzymes of *M. thermophila* with auxiliary activities, classified to AA3/8, AA9 families and multicopper oxidases. *M. thermophila* distinguishes itself from other cellulolytic fungi, exhibiting an impressing number of LPMOs accessory enzymes belonging to AA9 family (previously described as GH61).

* Tables are provided at the end of this Chapter.

Laccases (EC 1.10.3.1) are multicopper enzymes that catalyze the oxidation of a variety of phenolic compounds, with concomitant reduction of O_2 to H_2O. These polyphenol oxidases are produced by most ligninolytic basidiomycetes (Baldrian, 2006) and can degrade lignin and other recalcitrant compounds in the presence of redox mediators (Ruiz-Dueñas and Martínez, 2009). The genome of the *M. thermophila* encodes eight putative enzymes with multicopper oxidase activity. Four of them have been annotated and one (MYCTH_51627) matches the lcc1 gene product encoding an extracellular laccase (Berka *et al.*, 1997). Four sequences are predicted to possess a secretion signal, while one appears to remain membrane-bound. Lcc1 gene has been isolated from fungi's genome, heterologously expressed in *A. oryzae* and the produced 85 kDa enzyme (MtL) was characterized as a thermostable low oxidation potential laccase with high reactivity in aqueous medium at room temperature and neutral pH. MtL was tested for its capacity to catalyze enzymatic oxidation of several phenolic and polyphenolic compounds (ferulic acid, gallic acid, caffeic acid, and catechin) (Mustafa *et al.*, 2005). *M. thermophila* laccases have been reported to oxidize lignin surface, by increasing the amount of radicals during thermomechanical pulp fiber material bleaching (Grönqvist *et al.*, 2003) and promote oxidative polymerization of Kraft lignin from back liquor, which is the main by-product of pulp and paper industry (Gouveia *et al.*, 2013).

4.4. Lignocellulosic potential—statistics

M. thermophila is a powerful lignocellulolytic organism, which secretes a complex system of carbohydrate hydrolases for the breakdown of cellulose and hemicellulose, as well as oxidoreductases embedded in lignin degradation. Genome analysis in this review revealed 30 genes encoding cellulases classified to 10 GH families, 66 genes encoding hemicellulases classified to 10 GHs, 9 CEs and 35 genes encoding auxiliary enzymes. The latter include CDHs (AA3/AA8 family), LPMOs (AA9 family) and multicopper oxidases (laccases or laccase-like enzymes). Out of the total consortium of *M. thermophila* sequences encoding proteins with putative lignocellulosic activity, 80.2% are predicted to have a secretion signal peptide. Almost

* Tables are provided at the end of this Chapter.

76% of cellulases, hemicellulases and 88% of the accessory redox enzymes are targeted to secretion pathway, while only a very small amount remain inside the cell or represent membrane cell—bound macromolecules. Only 15.8% of the secreted enzymes in this review are predicted to possess a CBM and the majority of them comprise of auxiliary enzyme activities. The theoretical average molecular weight of secreted enzymes is 41.36 ± 15.9, varying between 10 and 97 kDa. The majority of secreted enzymes have molecular weight varying between 20 and 50 kDa, whereas β-xylosidases and β-glycosidases (GH3 family), and arabinofuranosidases (GH43 and GH51) appear to be high molecular weight proteins (**Figure 4.5**). The theoretical average isoelectric point of secretory enzymes is calculated 5.27 ± 0.8, at a range 4.34–7.9. In vivo expression and study of these enzymes would give different results, as the proteins are glycosylated, so size and pI value tend to moderate.
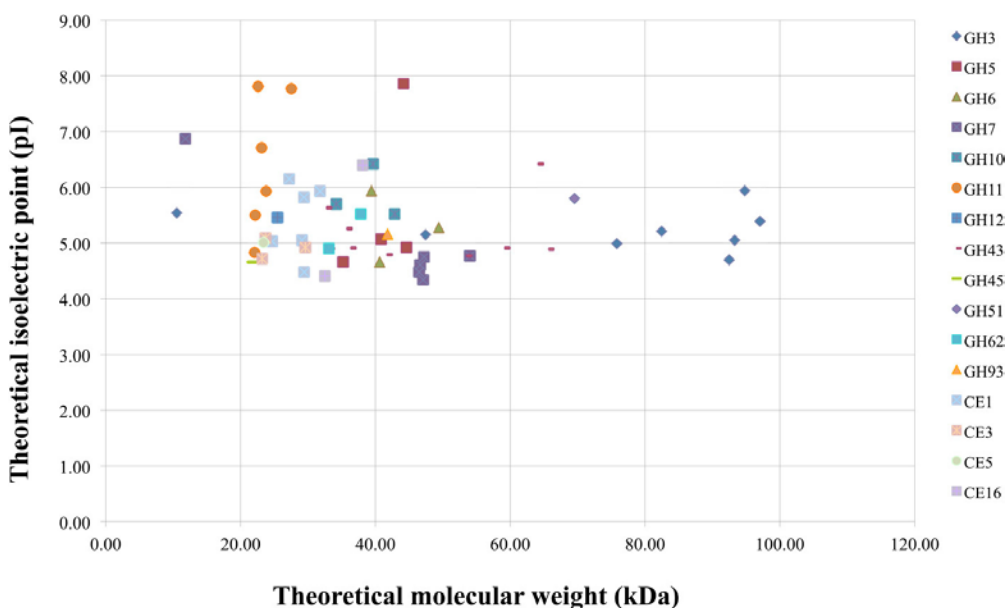
Figure 4.5. Theoretical molecular weight of secreted enzymes of *M. thermophila* classified in several GHs and CEs families, plotted against theoretical pI. The average molecular weight was calculated at 51.05 ± 16.2 kDa (range between 21 and 97 kDa) for cellulolytic enzymes, 35.5 ± 19.5 kDa (range between 22 and 89 kDa) for hemicellulases (GHs/CEs), and 28.51 ± 4.1 kDa (range between 23 and 39 kDa) for the fraction of esterases.

* Tables are provided at the end of this Chapter.

Protein glycosylation

A total proportion of 92.8% of secreted proteins have either *N-* or *O-* putative glycosylation sites. These proteins are often glycosylated due to the existence of many *Asn-Xaa-Ser/Thr* sequons, which are known to be a prerequisite for *N-*glycosylation post-translational modifications. The molecules of many GHs and accessory enzymes have a modular structure consisting of a catalytic module, flexible peptide linker, and CBM. Flexible linker peptides, which are rich in *Ser* and *Thr* residues, are typically *O-* glycosylated (Gilkes *et al.*, 1991). The *N-*glycosylation seems to be restricted to the catalytic modules, and it is usually absent in other parts of enzyme molecules. Various *N-*linked glycan structures have been found in different enzymes from *M. thermophila*, belonging to different enzyme classes and protein families (Gusakov *et al.*, 2008). It has been noticed that glycosylation follows a heterogeneity pattern, meaning that in some molecules, the same *Asn* residue was modified with oligosaccharides having different structure, while not all of the potential glycosylation sites were found to be occupied. The most frequently met *N-*linked glycan was $(Man)_3(GlcNAc)_2$, a pentasaccharide which represents a well-known conserved core structure that forms mammalian-type high-mannose and hybrid/complex glycans in glycoproteins from different organisms (Dwek *et al.*, 1993). Both types of glycosylation occur in 65% of secreted cellulases, 62.1% of secreted hemicellulases, while only *O-*glycosylation patterns appear in most of accessory enzymes. The presence of *N-*linked glycans is common for catalytic domain of the enzymes, while *O-*glycosylation usually occurs in linker region. Even though predicted to, non-secreted enzymes are not modified in vivo with glycans, since this procedure has been noticed as a post-translational modification in proteins targeted to the secretory pathway of the cell (Blom *et al.*, 2004).

4.5. Conclusions

Rapid depolymerization of lignocellulosic material is a distinguishing feature of thermophilic fungi, such as *M. thermophila*, which was isolated from soil and self-heating masses of composted vegetable matter (Domsch *et al.*, 1993). However, the precise biochemical mechanisms and underlying genetics of this procedure are not

* Tables are provided at the end of this Chapter.

completely understood. Systematic examination of the *M. thermophila* genome revealed a unique enzymatic system comprising of an unusual repertoire of auxiliary enzymes, especially those classified to AA9 family, and provided insights into its extraordinary capacity for protein secretion. The current review constitutes, to the best of our knowledge, the first genomic analysis of the lignocellulolytic system of *M. thermophila*. The genomic data, along with the observed enzymatic activity of several isolated and characterized enzymes suggest that this fungus possesses a complete set of enzymes, including 30 cellulases, 66 hemicellulases, and 35 proteins with auxiliary auxiliary enzymes, covering the most of the recognized CAZy families. From its cellulases to its oxido-reductases and multicopper enzymes, *M. thermophila* gene complement represents several avenues for further research and its diverse array of enzymatic capabilities will contribute to the study of lignocellulose degradation and the subsequent ethanol biofuel production.

References

Baldrian P. 2006. Fungal laccases—occurrence and properties. *FEMS Microbiol. Rev.* 30, 215–242.

Berka R. M., Grigoriev I. V., Otillar R., Salamov A., Grimwood J., Reid I., *et al.* 2011. Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris. Nat. Biotechnol.* 29, 922–927.

Berka R. M., Schneider P., Golightly E. J., Brown S. H., Madden M., Brown K. M., *et al.* 1997. Characterization of the gene encoding an extracellular laccase of *Myceliophthora thermophila* and analysis of the recombinant enzyme expressed in *Aspergillus oryzae. Appl. Environ. Microbiol.* 63, 3151–3157.

Bhat K. M., Maheshwari R. 1987. *Sporotrichum thermophile* growth, cellulose degradation and cellulase activity. *Appl. Environ. Microbiol.* 53, 2175–2182.

Blom N., Sicheritz-Ponten T., Gupta R., Gammeltoft S., Brunak S. 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4, 1633–1649.

Bukhtojarov F. E., Ustinov B. B., Salanovich T. N., Antonov A. I., Gusakov A. V., Okunev O. N., *et al.* 2004. Cellulase complex of the fungus *Chrysosporium lucknowense*: isolation and characterization of endoglucanases and cellobiohydrolases. *Biochemistry (Mosc).* 69, 542–551.

Canevascini G., Borer P., Dreyer J.L. 1991. Cellobiose dehydrogenases of *Sporotrichum (Chrysosporium) thermophile. Eur. J. Biochem.* 198, 43–52.

* Tables are provided at the end of this Chapter.

Charavgi M. D., Dimarogona M., Topakas E., Christakopoulos P., Chrysina E. D. 2013. The structure of a novel glucuronoyl esterase from *Myceliophthora thermophila* gives new insights into its role as a potential biocatalyst. *Acta Crystallogr. D Biol. Crystallogr.* 69, 63–73.

Crepin V.F., Faulds C.B., Connerton I.F. 2004. Functional recognition of newclasses of feruloyl esterase. *Appl. Microbiol. Biotechnol.* 63, 647–652.

Dimarogona M., Topakas E., Christakopoulos P. 2013. Recalcitrant polysaccharide degradation by novel oxidative biocatalysts. *Appl. Microbiol. Biotechnol.* 97, 8455–8465.

Domsch K. H., Gams W., Anderson T. H. 1993. Compendium of Soil Fungi. (Reprint of 1980 Ed. Verlag) (New York, NY: Academic Press, 780–783.

Dotsenko G. S., Semenova M. V., Sinitsyna O. A., Hinz S. W., Wery J., Zorov I. N., *et al.* 2012. Cloning, purification, and characterization of galactomannan-degrading enzymes from Myceliophthora thermophila. Biochemistry (Mosc). 77, 1303–1311.

Dwek R. A., Edge C. J., Harvey D. J., Wormald M. R., Parekh R. B. 1993. Analysis of glycoprotein-associated oligosaccharides. *Annu. Rev. Biochem.* 62, 65–100.

Emalfarb M., Hinz S., Joosten V., Koetsier M., Visser J., Visser J., *et al.* 2012. Novel fungal enzymes. US Patent No. US20120030838 A1.

Gilkes N. R., Henrissat B., Kilburn D. G., Miller R. C., Warren R. A. (1991). Domains in microbial beta 1,4-glycanases: sequence conservation, function and enzyme families. *J. Microbiol. Rev.* 55, 303–315

Gouveia S., Fernández-Costas C., Sanromán M. A., Moldes D. 2013. Polymerisation of Kraft lignin from black liquors by laccase from *Myceliophthora thermophila*: effect of operational conditions and black liquor origin. *Bioresour. Technol.* 131, 288–294.

Grönqvist S., Buchert J., Rantanen K., Viikari L., Suurnäkki A. 2003. Activity of laccase on unbleached and bleached thermomechanical pulp. *Enzyme Microb. Technol.* 32, 439–445.

Guillén D., Sánchez S., Rodríguez-Sanoja R. 2009. Carbohydrate-binding domains: multiplicity of biological roles. *Appl. Microbiol. Biotechnol.* 85, 1241–1249.

Gusakov A. V., Antonov A. I., Ustinov B. B. 2008. N-Glycosylation in *Chrysosporium lucknowense* enzymes. *Carbohydr. Res.* 343, 48–55.

Gusakov A. V., Punt P. J., Verdoes J. C., van der Meij J., Sinitsyn A. P., Vlasenko E., *et al.* 2011. Fungal Enzymes. Jupiter, FL: Dyadic International (USA), Inc; US Patent No. 7,923,236.

Gusakov A. V., Salanovich T. N., Antonov A. I., Ustinov B. B., Okunev O. N., Burlingame R., *et al.* 2007. Design of highly efficient cellulose mixtures for enzymatic hydrolysis of cellulose. Biotechnol. Bioeng. 97, 1028–1038 .

Gusakov A. V., Sinitsyna A. P., Salanovicha T. N., Bukhtojarova F. E., Markova A. V., Ustinova B. B., *et al.* 2005. Purification, cloning and characterisation of two forms of thermostable and highly active cellobiohydrolase I (Cel7A) produced by the industrial strain of *Chrysosporium lucknowense. Enz. Microb. Technol.* 36, 57–69.

* Tables are provided at the end of this Chapter.

Henriksson G., Johansson G., Pettersson G. 2000. A critical review of cellobiose dehydrogenases. *J. Biotechnol.* 78, 93–113.

Henrissat B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 280, 309–316.

Hinz S. W. A., Pouvreau L., Joosten R., Bartels J., Jonathan M. C., Wery J., *et al.* 2009. Hemicellulase production in *Chrysosporium lucknowense C1*. *J. Cer. Sci.* 50, 318–323.

Karnaouri A., Topakas E., Paschos T., Taouki I., Christakopoulos P. 2013. Cloning, expression and characterization of an ethanol tolerant GH3 β-glucosidase from *Myceliophthora thermophila*. *PeerJ* 1, e46.

Karnaouri A. C., Topakas E., Christakopoulos P. 2014. Cloning, expression, and characterization of a thermostable GH7 endoglucanase from Myceliophthora thermophila capable of high-consistency enzymatic liquefaction. Appl. Microbiol. Biotechnol. 98, 231–242.

Kühnel S., Hinz S. W., Pouvreau L., Wery J., Schols H. A., Gruppen H. 2010. *Chrysosporium lucknowense* arabinohydrolases effectively degrade sugar beet arabinan. *Bioresour. Technol.* 101, 8300–8307.

Kühnel S., Pouvreau L., Appeldoorn M. M., Hinz S. W., Schols H. A., Gruppen H. 2012. The ferulic acid esterases of *Chrysosporium lucknowense C1*: purification, characterization and their potential application in biorefinery. *Enzyme Microb. Technol.* 50, 77–85.

Kühnel S., Westphal Y., Hinz S. W., Schols H. A., Gruppen H. 2011. Mode of action of *Chrysosporium lucknowense C1* α-l-arabinohydrolases. *Bioresour. Technol.* 102, 1636–1643.

Lindberg B., Rossel K. G., Svensson S. 1973. Position of the O-acetyl groups in birch xylan. *Sven. Papperstidn.* 76, 30–32.

Morgenstern I., Powlowski J., Ishmael N., Darmond C., Marqueteau S., Moisan M. C., *et al.* 2012. A molecular phylogeny of thermophilic fungi. *Fungal Biol.* 116, 489–502.

Moukouli M., Topakas E., Christakopoulos P. 2010. Optimized expression of a GH11 xylanase and a type C feruloyl esterase from *Fusarium oxysporum* in *Pichia pastoris* and study of their synergistic action, in Final Workshop, COST Action 928, Control and exploitation of enzymes for added-value products, Poster Presentation (Naples).

Mustafa R., Muniglia L., Rovel B., Girardin M. 2005. Phenolic colourants obtained by enzymatic synthesis using a fungal laccase in a hydro-organic biphasic system. *Food Res. Int.* 38, 995–1000.

Pouvreau L., Jonathan M. C., Kabel M. A., Hinz S. W., Gruppen H., Schols H. A. 2011a. Characterization and mode of action of two acetyl xylan esterases from *Chrysosporium lucknowense C1* active towards acetylated xylans. *Enzyme Microb. Technol.* 49, 312–320.

Pouvreau L., Joosten R., Hinz S. W., Gruppen H., Schols H. A. 2011b. *Chrysosporium lucknowense C1* arabinofuranosidases are selective in releasing arabinose from either single or double substituted xylose residues in arabinoxylans. *Enzyme Microb. Technol.* 48, 397–403.

Punta M., Coggill P.C., Eberhardt R.Y., Mistry J., Tate J., Boursnell C., *et al.* 2012. The Pfam protein families' database. Nucleic Acids Res. 40, D290–D301.

* Tables are provided at the end of this Chapter.

Ruiz-Dueñas F.J. and Martínez A.T. 2009. Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microb. Biotechnol.* 2, 164–177.

Špániková S., Biely P. 2006. Glucuronoyl esterase–novel carbohydrate esterase produced by *Schizophyllum* commune. *FEBS Lett.* 580, 4597–4601.

Subramaniam S.S., Nagalla S.R., Renganathan V. 1999. Cloning and characterization of a thermostable cellobiose dehydrogenase from *Sporotrichum thermophile. Arch. Biochem. Biophys.* 365, 223–230.

Szijártó N., Horan E., Zhang J., Puranen T., Siika-Aho M., Viikari L. 2011. Thermostable endoglucanases in the liquefaction of hydrothermally pretreated wheat straw. *Biotechnol. Biofuels* 4, 2.

Tambor J.H., Ren H., Ushinsky S., Zheng Y., Riemens A., St-Francois C., *et al.* 2012. Recombinant expression, activity screening and functional characterization identifies three novel endo-1,4-β-glucanases that efficiently hydrolyse cellulosic substrates. *Appl. Microbiol. Biotechnol.* 93, 203–14.

Tenkanen M., Siika-aho M., Hausalo T., Puls J., Viikari L. 1996. Synergism between xylanolytic enzymes of *Trichoderma reesei* in the degradation of acetyl-4-O-methylglucoronoxylan, in *Biotechnology in Pulp and Paper Industly Advances in Applied and, Fundamental Research*, eds Messner K., Srebotnik E., editors. (Vienna: WUA Universitatsverlag;), 503–508.

Topakas E., Christakopoulos P., Faulds C.B. 2005. Comparison of mesophilic and thermophilic feruloyl esterases: characterization of their substrate specificity for methyl phenylalkanoates. *J. Biotechnol.* 115, 355–366.

Topakas E., Kalogeris E., Kekos D., Macris B. J., Christakopoulos P. 2003. Production and partial characterisation of feruloyl esterase by *Sporotrichum thermophile* in solid-state fermentation. *Proc. Biochem.* 38, 1539–1543.

Topakas E., Moukouli M., Dimarogona M., Christakopoulos P. (2012). Expression, characterization and structural modelling of a feruloyl esterase from the thermophilic fungus *Myceliophthora thermophila. Appl. Microbiol. Biotechnol.* 94, 399–41.

Topakas E., Moukouli M., Dimarogona M., Vafiadi C., Christakopoulos P. 2010. Functional expression of a thermophilic glucuronyl esterase from *Sporotrichum thermophile:* identification of the nucleophilic serine. *Appl. Microbiol. Biotechnol.* 87, 1765–1772.

Topakas E., Stamatis H., Biely P., Christakopoulos P. 2004. Purification and characterization of a type B feruloyl esterase (StFAE-A) from the thermophilic fungus *Sporotrichum thermophile. Appl. Microbiol. Biotechnol.* 63, 686–690.

Topakas E., Vafiadi C., Christakopoulos P. 2007. Microbial production, characterization and applications of feruloyl esterases. *Proc. Biochem.* 42, 497–509.

Ustinov B.B., Gusakov A.V., Antonov A.I., Sinitsyn A.P. (2008). Comparison of properties and mode of action of six secreted xylanases from *Chrysosporium lucknowense. Enzyme Microb. Technol.* 43, 56–65.

* Tables are provided at the end of this Chapter.

Vafiadi C., Topakas E., Biely P., Christakopoulos P. 2009. Purification, characterization and mass spectrometric sequencing of a thermophilic glucuronoyl esterase from *Sporotrichum thermophile*. *FEMS Microbiol. Lett.* 296, 178–184.

van Gool M.P., van Muiswinkel G.C., Hinz S.W., Schols H.A., Sinitsyn A.P., Gruppen H. 2012. Two GH10 endo-xylanases from *Myceliophthora thermophila C1* with and without cellulose binding module act differently towards soluble and insoluble xylans. *Bioresour. Technol.* 119, 123–132.

van Gool M.P., van Muiswinkel G.C., Hinz S.W., Schols H.A., Sinitsyn A.P., Gruppen H. 2013. Two novel GH11 endo-xylanases from *Myceliophthora thermophila* C1 act differently toward soluble and insoluble xylans. *Enzyme Microb. Technol.* 53, 25–32.

Verdoes J.C., Punt P.J., Burlingame R.P., Pynnonen C.M., Olson P.T., Wery J. 2010. New fungal production system. Patent No. WO2010107303 A2.

Visser H., Joosten V., Punt P. J., Gusakov A. V., Olson P. T., Joosten R., *et al.* 2011. Development of a mature fungal technology and production platform for industrial enzymes based on a *Myceliophthora thermophila* isolate, previously known as *Chrysosporium lucknowense C1*. *Ind. Biotechnol.* 7, 214–223.

Westphal Y., Kühnel S., de Waard P., Hinz S.W., Schols H.A., Voragen A.G., *et al.* 2010. Branched arabino-oligosaccharides isolated from sugar beet arabinan. *Carbohydr. Res.* 345, 1180–1189.

Zhang J., Pakarinen A., Viikari L. 2013. Synergy between cellulases and pectinases in the hydrolysis of hemp. *Bioresour. Technol.* 29, 302–307.

* Tables are provided at the end of this Chapter.

Table 4.1. Number of predicted CAZymes encoded in the genome of *M. thermophila*. **GH**s, Glycoside hydrolases; **CE**s, carbohydrate esterases; and **PL**s, polysaccharide lyases are included, covering the most of the recognized families.

Specific activity		CAZy module(s)	No id. seq.
Cellulases	endoglucanases	GH 5, 7, 12, 45	8
	cellobiohydrolases	GH 6, 7	7
	β-glucosidases	GH 1,3	8
Xylanases	xylanases	GH 10, 11	12
	xylosidases	GH 3, 43	4
Arabinases	endoarabinases	GH 43	3
	exo-arabinases / arabinofuranosidases	GH 43, 51, 62	11
Mannanases	endomannanases	GH 5, 26	3
	mannosidases	GH 2	2
Pectinases	polygalacturonases	GH28	2
	rhamnosidases	GH78	1
	pectin lyases	PL1, PL3, PL4, PL20	8
	pectin esterases	CE 8, 12	4
Esterases	feruloyl esterases	CE 1	4
	acetyl esterases	CE 3, 5, 16	8
	acetylmannanesterases	CE 12	2
	glycuronoyl esterases	CE 15	2

* Tables are provided at the end of this Chapter.

Table 4.2. Number of predicted sequences encoding enzymes with cellulolytic activity (EGs,CBHs,andBGLs).

EGs

Protein ID	InterPro ID	InterPro description	CAZy module	GenBank ID	Location	Secretion signal	Exons	pI/MW (kDa)	Lenght	N-gly sites	O-gly sites	CBM 1	BLAST parameters
MYCTH_52068	IPR017853	glycoside hydrolase, catalytic core	GH5	AEO58455.1	4:188041-189524	18aa	7	4.66 / 35.2	319aa	3	2	-	
MYCTH_86753	IPR017853	glycoside hydrolase, catalytic core	GH5	AEO53769.1	1:2823610-2825549	16aa	3	5.07 / 40.85	389aa	3	17	+	65% identity with endoglucanase GH5 from *Thermoascus aurantiacus* [PDB ID: 1GZJ] & 79% with β-1,4-endoglucanase from *Penicillium brasilianum* [GenBank: ACB06750]
MYCTH_94336	IPR017853	glycoside hydrolase, catalytic core	GH5	AEO57401.1	3:908597-909880	21aa	1	4.92 / 44.6	406aa	2	4	-	61% identity with cellulase family protein from *Ophiostoma piceae* UAMH 11346 [GenBank: EPE03278.1]
MYCTH_43356	IPR017853	glycoside hydrolase, catalytic core	GH5	AEO53175.1	1:580471-581733	26aa	1	7.86 / 44.18	394aa	1	2	-	71% identity with EG from endoglucanase *Verticillium dahliae* [GenBank: EGY21718.1]
MYCTH_116157	IPR017853	glycoside hydrolase, catalytic core	GH7	AEO59361.1	4:4081701-4083768	20aa	2	4.75 / 47.20	436aa	2	2	-	68% identity with endo-1,4-beta-glucanase from *Humicola grisea* [UniProtKB/Swiss-Prot:Q12622.1]
MYCTH_111372	IPR017853	glycoside hydrolase, catalytic core	GH7	AEO58196.1	3:4135959-4137568	22aa	1	4.61 / 46.632	442aa	2	16	+	64% identity with endoglucanase I from *Hypocrea orientalis* [GenBank: AFD50194.1]
MYCTH_109444	IPR008985	concanavalin A-like lectin/glucanases superfamily	GH12	AEO60532.1	6:212893-214302	15aa	3	5.46 / 25. 48	247aa	-	-	-	56% identity with Cel12A from *Humicola grisea* [PDB ID: 1OLR] & 63% with endoglucanase from *Fusarium oxysporum* [GenBank: ENH72093.1]
MYCTH_76901	IPR014733	barwin-like endoglucanase	GH45	AEO54078.1	1:4001048-4002222	18aa	4	4.66 / 21.8	207aa	-	3	-	78% identity with endoglucanase from *Humicola insolens* [PDB ID: 1HD5] & 82% with endoglucanase from *Humicola grisea* [GenBank: BAA74957.1]

CBHs													
Protein_ID	InterPro_ID	InterPro_description	CAZy module	GenBank ID	Location	Secretion signal	Exons	pI/MW (kDa)	Lenght	N-gly positions	O-gly positions	CBM 1	BLAST parameters
MYCTH_51545	IPR016288	1,4-beta cellobiohydrolase	GH6	AEO59280.1	4:3726723-3728113	17aa	3	4.66 / 40.64	378aa	-	4	-	49% identity with CBHII from *Trichoderma reesei* [GenBank: ADC83999.1] & 51% with CBHII from *Trichoderma reesei* [PDB ID: 3CBH]
MYCTH_66729	IPR016288	1,4-beta cellobiohydrolase	GH6	AEO55787.1	2:46305-48489	17aa	4	5.28 / 49.41	465aa	1	35	+	79% identity with CBH II from *Humicola insolens* [PDB ID: 1BVW] & 64% with CBHII from *Trichoderma viride* [GenBank: AAQ76094.1]
MYCTH_2303045	IPR016288	1,4-beta cellobiohydrolase	GH6	AEO57190.1	2:5389544-5391132	18aa	2	5.94 / 39.41	363aa	4	5	-	56% identity with endoglucanase Cel6b from *Humicola insolens* [PDB ID: 1DYS] & 39% with Cellobiohydrolase II (Cel6a) from *Humicola insolens* [1BVW]
MYCTH_109566	IPR008985	concanavalin A-like lectin / glucanases superfamily	GH7	AEO55544.1	1:9753507-9755507	17aa	2	4.77 / 54	509aa	1	20	+	61% identity with CBH I from *Humicola grisea* [GenBank: BAA09785.1] & 67% with cellobiohydrolase Cel7d (Cbh 58) *Phanerochaete chrysosporium* [PDB: 1GPI]
MYCTH_42937	IPR008985	concanavalin A-like lectin/ glucanases superfamily	GH7	AEO53522.1	1:1922814-1924693	20aa	5	4.34 / 47.08	436aa	2	2	-	66% identity with β- 1,4-cellobiohydrolase from *Phanerochaete chrysosporium* [UniProtKB: P13860.1]
MYCTH_97137	IPR008985	concanavalin A-like lectin / glucanases superfamily	GH7	AEO61262.1	7:24661-26448	20aa	6	4.48 / 46.5	430aa			-	79% identity with β- 1,4-beta-cellobiosidase from *Melanocarpus albomyces* [GenBank: CAD56667.1]
MYCTH_95095 fragment	IPR008985	concanavalin A-like lectin / glucanases superfamily	GH7	AEO58824.1	4:1967106-1967560	19aa	2	6.87 / 11.82	109aa	1	4	C-term, ND	77% identity with β- 1,4-beta-cellobiosidase from *Acremonium thermophilum* [GenBank: CAM98446.1]

	BGLs												
Protein_ID	InterPro_ID	InterPro_description	CAZy module	GenBank ID	Location	Secretion signal	Exons	pI/ MW (kDa)	Lenght	N-gly positions	O-gly positions	CBM 1	BLAST parameters
MYCTH_115968	IPR 017853	glycoside hydrolase, catalytic core	GH1	AEO57459.1	3:1092854-1095039	-	2	5.53 / 54.1	476aa	1	-	-	92% identity with β-glucosidase from *Humicola grisea* [GenBank: BAA74958.1]
MYCTH_38200	IPR 017853	glycoside hydrolase, catalytic core	GH3	AEO61246.1	6:4096305-4099311	-	2	5.54 / 10.53	968aa	7	3	-	78% identity with β-glucosidase from *Chaetomium thermophilum var. thermophilum* [GenBank: EGS22574.1]
MYCTH_2059579	IPR 017853 / IPR 026891	glycoside hydrolase / fibronectin type III-like domain	GH3	AEO56238.1	2:1843191-1846185	23aa	4	5.94 / 94.8	880aa	6	5	-	57% identity with β-1,4-glucosidase from *Marssonina brunnea* [GenBank: EKD11918.1]
MYCTH_66804	IPR 017853 / IPR 026891	glycoside hydrolase/ fibronectin type III-like domain	GH3	AEO58343.1	3:4861135-4863642	17aa	2	4.99 / 75.83	716aa	2	2	-	72% identity with β-glucosidase from *Trichoderma reesei* [PRF: 227874]
MYCTH_80304	IPR 017853 / IPR 026891	glycoside hydrolase / fibronectin type III-like domain	GH3	AEO58175.1	3:3949561-3952737	19aa	4	5.05 / 93.3	851aa	11	-	-	77% identity with β-glucosidase from *Chaetomium thermophilum* [GenBank: ABR 57325.2]
MYCTH_62925	IPR 017853	glycoside hydrolase, catalytic core	GH3	AEO53892.1	1:3330354-3333832	-	7	5.39 / 97.05	884aa	3	4	-	76% identity with β-glucosidase from *Fusarium oxysporum* [GenBank: EMT 71863.1]
MYCTH_2302509	IPR 026891	fibronectin type III-like domain	GH3	AEO56946.1	2:4612453-4614052	-	1	5.15 / 47.75	440aa	6	2	-	60% identity with β-glucosidase from *Grosmannia clavigera* [GenBank: EFX 01027.1]
MYCTH_58882	IPR 017853 / IPR 026891	glycoside hydrolase/ fibronectin type III-like domain	GH3	AEO60477.1	6:51326-53769	16aa	2	5.21 / 82.46	761aa	5	-		
MYCTH_2129052 (fragment)	IPR 001764	glycoside hydrolase, family 3, N terminal	GH3	AEO59952.1	5:2414101-2414414	-	2	4.70 / 92.48	83aa	1	-		46% identity with β-glucosidase from *Parastagonospora avenae* [GenBank: CAB82861.1]

Table 4.3. Description of the characterized cellulolytic enzymes either isolated from the culture broth of a *M. thermophila C1* mutant strain or expressed in an heterologous host.

Enzyme	Type of action	GH	MW-monomer (kDa)	pH_{opt}	T_{opt} (°C)	p*I*	gene	source	Reference
StCel5A	endoglucanase	5	46	6	70	ND		cDNA library-EST analysis, expessed in *A. niger*	Tambor *et al.*, 2012
EG51	endoglucanase	5	51	4.7	70	4.8		isolated from the culture broth of a C1 mutant strain	Bukhtojarov *et al.*, 2004/Gusakov *et al.*, 2005
MtEG7a	endoglucanase	7	65	5	60	multiple bands in 3.8–4.5	*eg7a*	expressed in *P. pastoris*	Karnaouri *et al.*, 2014
EG60	endoglucanase	7	60	4.7	60	3.7		isolated from the culture broth of a C1 mutant strain	Bukhtojarov *et al.*, 2004
EG28	endoglucanase	12	28	5.35	60	5.7		isolated from the culture broth of a C1 mutant strain	Bukhtojarov *et al.*, 2004
EG25	endoglucanase	45	25	5.5	65	4		isolated from the culture broth of a C1 mutant strain	Bukhtojarov *et al.*, 2004/Gusakov *et al.*, 2005
CBH Ia	1,4-beta cellobiohydrolase	7	65	5	ND	4.5		isolated from the culture broth of a C1 mutant strain	Bukhtojarov *et al.*, 2004/Gusakov et al., 2005
CBH IIa	1,4-beta cellobiohydrolase	6	43	5.4	65	4.2		isolated from the culture broth of a C1 mutant strain	Bukhtojarov *et al.*, 2004/Gusakov *et al.*, 2005
CBH Ib	1,4-beta cellobiohydrolase	7	60	ND	ND	ND		isolated from the culture broth of a C1 mutant strain	Gusakov *et al*., 2007
CBH IIb	1,4-beta cellobiohydrolase	6	70	ND	ND	5.6		isolated from the culture broth of a C1 mutant strain	Gusakov *et al*., 2007
MtBgl3a	beta-glucosidase	3	90	5	70	4.0	*bgl3a*	expressed in *P. pastoris*	Karnaouri *et al*., 2013
Bxl5	beta-glucosidase	3	120 ± 5	4.6	75	5.2	*bxl5*	homologously expressed in C1	Dotsenko *et al.*, 2012

Table 4.4. Number of predicted sequences encoding enzymes with hemicellulolytic activity (β-1,4-xylanases and β-xylosidases)

XYLs													
Protein ID	InterPro ID	InterPro description	CAZy module	GenBank ID	Location	Secretion signal	Exons	pI/MW (kDa)	Lenght	N-gly sites	O-gly sites	Conserved domains	BLAST parameters
MYCTH_2139438	IPR013781	glycoside hydrolase, catalytic domain	GH10	AEO58598.1	4:852357-853828	19aa	3	6.42 / 39.65	356aa	1	-		84% identity with XynA from *Humicola insolens* [GenBank: AGG68962.1]
MYCTH_52904	IPR013781	glycoside hydrolase, catalytic domain	GH10	AEO59320.1	4:3932271-3933431	16aa	3	5.70 / 34.20	311aa	1	1		77% identity with endo-1,4-beta-xylanase protein from *Chaetomium thermophilum* [GenBank: EGS20178.1]
MYCTH_2125938 (fragment)	IPR013781	glycoside hydrolase, catalytic domain	GH10	AEO56947.1	2:4614086-4614433	ND	1	10.11 / 13.53	115aa	1	1		63% identity with XynA from *Humicola insolens* [GenBank: AGG68962.1]
MYCTH_112050	IPR013781	glycoside hydrolase, catalytic domain	GH10	AEO60457.1	5:4306110-4307709	17aa	2	5.52 / 42.86	396aa	-	22	CBM1	66% identity with endo-1,4-beta-xylanase A from *Gaeumannomyces graminis* [GenBank: EJT78185.1]
MYCTH_100068	IPR008985	concanavalin A-like lectin/glucanases superfamily	GH11	AEO57157.1	2:5265193-5266157	19aa	2	7.77 / 27.53	259aa	1	1	CBM1	74% identity with endo-beta1,4-xylanase from *Chaetomium gracile* [GenBank: BAA08650.1]
MYCTH_89603	IPR008985	concanavalin A-like lectin/glucanases superfamily	GH11	AEO55365.1	1:9130476-9131260	20aa	2	6.71 / 23.12	208aa	1	1		87% identity with endo-beta-1,4-xylanase from *Chaetomium sp.* CQ31 [GenBank: ADW78258.1]
MYCTH_2121801	IPR008985	concanavalin A-like lectin/glucanases superfamily	GH11	AEO62054.1	7:3382576-3383357	18aa	2	7.81 / 22.59	200aa	1	2		86% identity with endo-1,4-beta-xylanase protein from *Chaetomium thermophilum* [GenBank: EGS23409.1]
MYCTH_99786	IPR008985	concanavalin A-like lectin/glucanases superfamily	GH11	AEO54512.1	1:5821400-5822224	21aa	2	5.50 / 22.17	205aa	1	2		56% identity from xylanase II from *Hypocrea orientalis* [GenBank: AFD50199.1]
MYCTH_2309574 (fragment)	IPR008985	concanavalin A-like lectin/glucanases superfamily	GH11	AEO60385.1	5:4039395-4039724	ND	1	9.19 / 7.10	66aa	1	-		63% identity with xyn11C from *Chaetomium thermophilum* [GenBank: CAD48751.1]

MYCTH_2309575 (fragment)	IPR008985	concanavalin A-like lectin/glucanases superfamily	GH11	AEO60386.1	5:4039732-4040297	-	1	5.49 / 16.34	146aa	2	4		77% identity with endoxylanase 11C from *Chaetomium thermophilum* [GenBank: CAD48751.1]
MYCTH_56237	IPR008985	concanavalin A-like lectin/glucanases superfamily	GH11	AEO59539.1	5:215065-215872	18aa	2	4.83 / 22.08	205aa	1	2		72% identity with xylanase from *Scytalidium thermophilum* [GenBank: BAD07040.1]
MYCTH_49824	IPR008985	concanavalin A-like lectin/glucanases superfamily	GH11	AEO58284.1	3:4545212-4546113	16aa	2	5.93 / 23.79	214aa	3	1		79% identity with endo-1,4-beta-xylanase from *Chaetomium cupreum* [GenBank: ABI48363.1]

BXYLs

Protein_ID	InterPro_ID	InterPro_description	CAZy module	GenBank ID	Location	Secretion signal	Exons	pI/MW (kDa)	Lenght	N-gly positions	O-gly positions	Conserved domains	BLAST parameters
MYCTH_104628	IPR017853	glycoside hydrolase, catalytic core	GH3	AEO60531.1	6:209038-211675	-	2	4.69 / 89.69	835aa	6	1	BglX	55% identity with 1,4-beta-xylosidase from *Fusarium fujikuroi* IMI 58289 [GenBank: CCT61496.1]
MYCTH_50705	IPR017853 / IPR026891	glycoside hydrolase, catalytic core / fibronectin type III-like domain	GH3	AEO58346.1	3:4871063-4873395	21aa	2	5.57 / 79.51	739aa	6	2		53% identity with 1,4-beta-xylosidase from *Neurospora crassa* [GenBank: CAB91343.2]
MYCTH_80104	IPR008985	concanavalin A-like lectin/glucanases superfamily	GH43	AEO58399.1	3:5027172-5028785	-	1	5.86 / 61.23	537aa	1	-	XynB	80% identity with beta-xylosidase from *Fusarium oxysporum* [GenBank: EMT73385.1]
MYCTH_2072383	IPR008985	concanavalin A-like lectin/glucanases superfamily	GH43	AEO61672.1	7:1863035-1864780	19aa	2	4.77 / 53.52	494aa	4	3	XynB	45% identity with beta-xylosidase from *Micromonospora sp.* ATCC39149 [NCBI Ref: WP_007073468.1]

Table 4.5. Description of the characterized β‑1,4–xylanases isolated from the culture broth of a *M. thermophila* C1 mutant strain.

XYLs

Enzyme	CAZy module	MW-monomer (kDa)	pH$_{opt}$	T$_{opt}$ (ºC)	pI	gene	Reference
Xyn10A	GH10	42 / 31*	5.5 - 7.0	65 - 70	7.8 / 8.9	*xyl1*	Ustinov *et al.*, 2008; van Gool *et al.*, 2012
Xyn10B	GH10	57 / 46*	5.5 - 7.0	80 - 85	4.4 / 4.3	*xyl3*	Ustinov *et al.*, 2008; van Gool *et al.*, 2012
Xyn10C	GH11	40	5.0	80	4.8	*xyl4*	Ustinov *et al.*, 2008
Xyn11A	GH11	24	6.5	70	7.9	*xyl2*	Ustinov *et al.*, 2008
Xyn11B	GH11	23	6.0 - 6.5	65 - 70	8.4	*xyl6*	Ustinov *et al.*, 2008
Xyn11C	GH11	22	4.5	65	6.7	*xyl5*	Ustinov *et al.*, 2008
Xyl7	GH11	22 / 30*	5.5 - 6.5	50 - 60	7.3 / 7.6	*xyl7*	van Gool *et al.*, 2013
Xyl8	GH11	22	5.5 - 6.0	50 - 65	6.2	*xyl8*	van Gool *et al.*, 2013

*isolated in two different forms, with (high molecular weight enzyme) or without CBM (low molecular weight enzyme).

Table 4.6. Number of predicted sequences encoding enzymes with hemicellulolytic activity (endoarabinases and arabinofuranosidases).

ABFs +													
Protein_ID	InterPro_ID	InterPro description	CAZy module	GenBank ID	Location	Secretion signal	Exons	pI/MW (kDa)	Lenght	N-gly positions	O-gly positions	Conserved domains	BLAST parameters
MYCTH_39555	IPR006710	glycoside hydrolase family 43	GH43	AEO61077.1	6:3204367-3205350	-	1	4.87 / 37.12	327aa	-	1		82% identity with a-N-arabinofuranosidase / alpha-L-arabinofuranosidase from *Fusarium fujikuroi* [GenBank: CCT69715.1]
MYCTH_2303298	IPR006710	glycoside hydrolase family 43	GH43	AEO57303.1	3:514704-516772	23aa	1	4.91 / 59.16	535aa	1	1	XynB	55% identity with arabinase from *Auricularia delicata* [GenBank: EJD41599.1]
MYCTH_2305738	IPR006710	glycoside hydrolase family 43	GH43	AEO58423.1	4:27829-29545	19aa	4	4.90 / 33.18	301aa	1	2		52% identity with a-N-arabinofuranosidase 2 from *Fusarium oxysporum* [GenBank: EMT68952.1]
MYCTH_103032	IPR016840	glycoside hydrolase family 43, endo-1,5-alpha-L-arabinosidase	GH43	AEO58422.1	4:25889-27048	20aa	3	5.63 / 32.74	301aa	1	4		60% identity with arabinan endo-1,5-alpha-L-arabinosidase A from *Aspergillus kawachii* [GenBank: GAA90221.1]
MYCTH_2300677	IPR006710	glycoside hydrolase family 43	GH43	AEO56123.1	2:1449574-1451195	22aa	1	4.79 / 41.72	410aa	5	35		38% identity with endo-arabinase from *Colletotrichum gloeosporioides* [GenBank: ELA34066.1]
MYCTH_2064169	IPR006710	glycoside hydrolase family 43	GH43	AEO58916.1	4:2252824-2255361	16aa	11	6.42 / 64.17	573aa	7	5	XynB	65% identity with arabinofuranosidase from *Coprinopsis cinerea* [NCBI Ref: XP_001836293.2] & 52% with beta-
MYCTH_2301869	IPR006710	glycoside hydrolase family 43	GH43	AEO56692.1	2:3685597-3687704	18aa	1	4.89 / 65.71	613αα	3	1	Xyl1	55% identity with Xylosidase/arabinosidase (arabinofuranoside) from *Fusarium oxysporum* [GenBank: ENH67211.1]
MYCTH_2127683	IPR007934	a-L-arabinofuranosidase B	GH43	AEO58631.1	4:1109767-1110916	-	3	4.91 / 36.35	339aa	1	3	CBM42	65% with a-N-arabinofuranosidase from *Chaetomium sp.* [GenBank: AFU88757.1]
MYCTH_2306666	IPR023296	glycosyl hydrolase, five-bladed-beta-propellor domain	GH43	AEO58919.1	4:2260967-2262690	25aa	4	5.26 / 35.76	329aa	1	1		72% identity with arabinosidase (arabinofuranosidase) from *Glomerella graminicola* [GenBank: EFQ32395.1]

MYCTH_	IPR	Description	GH	Accession	Location			pI / MW	Length			CBM	Homology
MYCTH_42071	IPR010720	a-L-arabinofuranosidase, C-terminal	GH51	AEO53569.1	1:2109152-2111615	18aa	6	5.80 / 69.54	636aa	6	2		54% identity with a-L-arabinofuranosidase from *Penicillium chrysogenum* [GenBank: BAH70480.1]
MYCTH_83019	IPR010720	a-L-arabinofuranosidase, C-terminal	GH51	AEO58452.1	4:178906-181205	-	5	5.81 / 57.67	512aa	2	2		75% identity with a-L-arabinofuranosidase from *Glomerella graminicola* [GenBank: EFQ27215.1]
MYCTH_98003	IPR005193	glycoside hydrolase, family 62, arabinosidase	GH62	AEO60934.1	6:2448019-2449251	22aa	2	5.52 / 37.84	353aa	-	3	CBM1	73% identity with a-L-arabinofuranosidase axhA-2 from *Colletotrichum higginsianum* [GenBank: CCF42219.1]
MYCTH_55982	IPR005193	glycoside hydrolase, family 62, arabinosidase	GH62	AEO59813.1	5:1711636-1712601	19aa	1	4.90 / 33.1	302aa	1	4		77% identity with a-L-arabinofuranosidase from *Streptomyces ghanaensis* [NCBI Ref: WP_004979944.1]
MYCTH_104827	IPR011040	Sialidases	GH93	AEO55492.1	1:9532077-9534361	17aa	2	5.16 / 41.8	378aa	3			56% with exo-arabinanase from *Penicillium chrysogenum* [NCBI Ref: XP_002562032.1]

Table 4.7. Description of the characterized endoarabinases and arabinofuranosidases isolated from the culture broth of a *M. thermophila* C1 mutant strain.

ABs

Enzyme	CAZy module	Mode of action	MW-monomer (kDa)	pH$_{opt}$	T$_{opt}$ (^{O}C)	Uniprot / SwissProt ID	gene	source	Reference
Abf1	GH62	Arabinofuranosidase / releases O-2 or O-3 arabinose from mono-substituted xylose					*abf1*	selectively produced in C1-host	Hinz *et al.*, 2009
Abf2	GH62	Arabinofuranosidase / releases O-2 or O-3 linked arabinofuranosyl residues from mono-substituted xylose					*abf2*	selectively produced in C1-host	Hinz *et al.*, 2009
Abf3	GH51	Arabinofuranosidase / releases arabinose from the non-reducing end of reduced arabinose oligomers	70	5	40	HQ324254	*abf3*	selectively produced in C1-host	Pouvreau *et al.*, 2011b
Abn7	GH43	Arabinofuranosidase / releases O-3 linked arabinofuranosyl residues from di-substituted xylose	70	5	40	HQ324255	*abn7*	selectively produced in C1-host	Pouvreau *et al.*, 2011b
Abn1	GH43	Endoarabinase	36	5.5	60	HQ324251	*abn1*	overexpressed in fermentation supernatant	Kühnel *et al.*, 2011
Abn2	GH93	Exoarabinase / arabinobiose from the non-reducing end of reduced arabinose oligomers.	40	4	50		*abn2*	overexpressed in fermentation supernatant	Kühnel *et al.*, 2011
Abn4	GH43	Arabinofuranosidase / releases arabinose from the non-reducing end of reduced arabinose oligomers	33	5.5	60	HQ324253	*abn4*	overexpressed in fermentation supernatant	Kühnel *et al.*, 2011
Abn5		Arabinofuranosidase					*abn5*	selectively produced in C1-host	Hinz *et al.*, 2009

Table 4.8. Description of *M. thermophila* characterized esterases (FAEs, AcEs, and GEs).

FAEs

Enzyme	type	MW-monomer (kDa)	pH$_{opt}$	T$_{opt}$ (^{O}C)	pI	gene	source	Reference
StFaeB	B	33*	6	55-60	3,5		isolated from culture supernatant	Topakas *et al*., 2004
StFaeC	C	23*	6	55	<3,5		isolated from culture supernatant	Topakas *et al*., 2005b
FAE	ND	27	8	60	5		isolated from culture supernatant	Topakas *et al*., 2003
FaeA1	A	29	6.5	45	≈ 5.5		overexpressed in fermentation supernatant	Kühnel *et al.*, 2012
FaeA2	A	36	7.5	40	≈ 5.2		overexpressed in fermentation supernatant	Kühnel *et al*., 2012
FaeB2	B	33	7.5	45	≈ 6.0		overexpressed in fermentation supernatant	Kühnel *et al.*, 2012
MtFae1a	B	39	7	50	ND	*fae1a*	expressed in *P. pastoris*	Topakas *et al*., 2012

* dimeric structures are detected, after comparing the results of SDS-PAGE and native electrophoresis

AcEs

Enzyme	CE	MW-monomer (kDa)	pH$_{opt}$	T$_{opt}$ (^{O}C)	pI	gene	source	Reference
MtAxe3	1	33.6	7	40	ND	*axe3*	overexpressed in fermentation supernatant	Pouvreau *et al.*, 2011a
MtAxe2	5	23.6	7	40	ND	*axe2*	overexpressed in fermentation supernatant	Pouvreau *et al.*, 2011a

GEs

Enzyme	CE	MW-monomer (kDa)	pH$_{opt}$	T$_{opt}$ (^{O}C)	pI	gene	source	Reference
StGE2	15	43	7	55	ND	*ge2*	expressed in *P. pastoris*	Topakas *et al.*, 2010
StEG1	15	58	6	60	ND		isolated from culture supernatant	Vafiadi *et al.*, 2009

CHAPTER 5

Cloning, expression and characterization of EG7 endoglucanase from *M. thermophila* (PAPER II)

Endoglucanases [E.C. 3.2.1.4] play a key role in cellulose degradation and act at the initial stages of hydrolytic conversion of the cellulosic components into fermentable sugars. These enzymes catalyze the initial attack on the polymer by hydrolyzing β-1,4 glucosidic bonds within amorphous regions of cellulose chains (Xiros *et al.* 2012). Enzyme thermostability is essential during the saccharification process, which converts lignocellulosic biomass to reducing sugars, because steam is usually used as a pretreatment step in order to make the biomass more suitable for enzymatic hydrolysis. The thermostable enzymes can be used directly after the heating step, thereby decreasing the processing time, saving energy and improving fermentation yields. Considerable emphasis has also been placed on identifying alkalistable enzymes that are active on an alkaline pH range used in many biotechnological applications, such as modifying the structure of cellulose fibrils, as additives to laundry detergents, as well as in cotton treatment process (Anish *et al.* 2007). Together with the stability of the enzymes used in the conversion of lignocellulosic biomass to reducing sugars and the subsequent ethanol production, the ability of liquefaction is very important; when aiming at ethanol concentrations above 4% (w/w) in the broth (which is considered as a minimum prerequisite for a feasible large-scale distillation technology) the initial dry matter (DM) loading has to be over 15% for most technical substrates (Fan *et al.* 2003). The *high gravity processes* (HG, initial DM concentrations of lignocellulose above 12% (w/w)) have gained much attention, as they result in high final ethanol concentrations and therefore at a significantly decreased cost of the distillation step. However, these high-solid contents create an environment in which practically no free water exists in the pretreated material, resulting in a difficulty to handle the slurry. The poor mass-transfer conditions can be improved by means of a partial enzymatic hydrolysis stage (*liquefaction*), where the structured, porous and water-absorbing high solid substrate is converted to a more flowable fluid (Szijártó *et al.* 2011a). This change in viscosity is
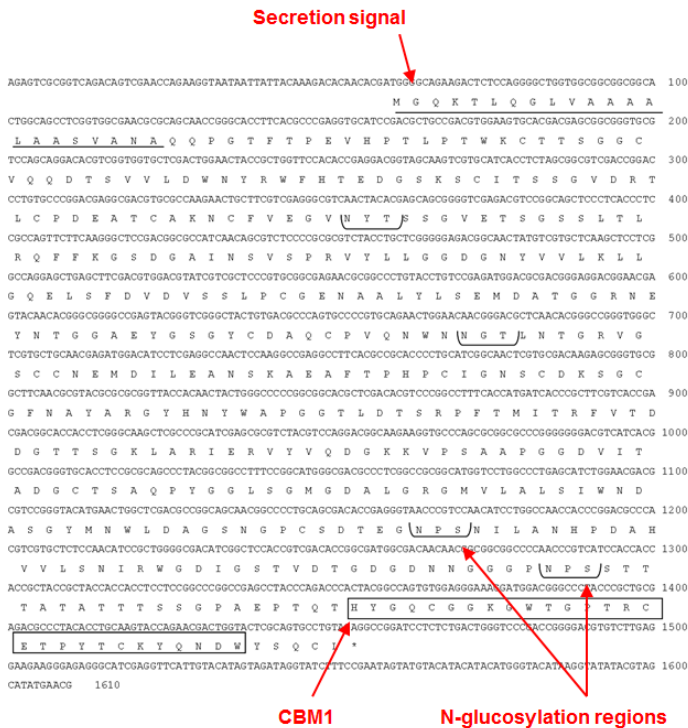
probably due to the gradual reduction of the average chain length of cellulose polysaccharides by endo-acting enzymes, such as endoglucanases.

The thermophilic fungi *Myceliophthora thermophila* ATCC 42464, as described in **Chapter 4**, synthesizes a complete set of endoglucanases. In this chapter, the successful cloning of the complete genomic DNA sequence of an endoglucanase gene belonging to GH7 family, the heterologous expression in methylotrophic yeast *P. pastoris* and the characterization of the recombinant enzyme are described. Although many endoglucanases were characterized from fungi previously, the properties of the recombinant *M. thermophila Mt*EG7a described here are distinct from them in terms of catalytic efficiency and thermostability. Apart from that, viscosity measurements showed the liquefaction efficiency of the enzyme on high solid lignocellulose substrates.

5.1. Identification and cloning of MtEG7a

From genome analysis, the translation of *eg7a* open reading frame (ORF) (Model ID 111372) from the *M. thermophila* genome database shows significant primary sequence identity with characterized endoglucanases which have been classified to family GH7 on CAZy database (http://www.cazy.org/; Cantarel *et al.* 2009). The putative endoglucanase showed high sequence identity (64%) with endoglucanase I from *Hypocrea orientalis* [GenBank: AFD50194.1]. The hypothetical protein of 111372 was selected as a candidate endoglucanase and the corresponding gene, which was provisionally named *eg7a*, has no introns leading us to the direct cloning and expression of the encoding enzyme named *Mt*EG7a (**Table 5.1**).

The ORF of eg7a encodes a protein of 464 amino acids including a secretion signal peptide of 22 amino acids (MGQKTLQGLVAAAALAASVANA) based upon the prediction using SignalP v4.0, which is a web-based program (http://www.cbs.dtu.dk/services/SignalP/). The predicted mass and isoelectric point (pI) of the mature protein was 46629 Da and pH 4.61, respectively, by calculations using the ProtParam tool of ExPASY (http://web.expasy.org/protparam/).

Secretion signal

```
AGAGTCGCGGTCAGACAGTCGAACCAGAAGGTAATAATTATTACAAAGACACAACACGATGGGGCAGAAGACTCTCCAGGGGCTGGTGGCGGCGGCGGCA  100
                                                             M  G  Q  K  T  L  Q  G  L  V  A  A  A  A
CTGGCAGCCTCGGTGGCGAACGCGCAGCAACCGGGCACCTTCACGCCCGAGGTGCATCCGACGCTGCCGACGTGGAAGTGCACGACGAGCGGCGGGTGCG  200
 L  A  A  S  V  A  N  A  Q  Q  P  G  T  F  T  P  E  V  H  P  T  L  P  T  W  K  C  T  T  S  G  G  C
TCCAGCAGGACACGTCGGTGGTGCTCGACTGGAACTACCGCTGGTTCCACACCGAGGACGGTAGCAAGTCGTGCATCACCTCTAGCGGCGTCGACCGGAC  300
 V  Q  Q  D  T  S  V  V  L  D  W  N  Y  R  W  F  H  T  E  D  G  S  K  S  C  I  T  S  S  G  V  D  R  T
CCTGTGCCCGGACGAGGCGACGTGCGCCAAGAACTGCTTCGTCGAGGGCGTCAACTACACGAGCAGCGGGGTCGAGACGTCCGGCAGCTCCCTCACCCTC  400
 L  C  P  D  E  A  T  C  A  K  N  C  F  V  E  G  V  N  Y  T  S  S  G  V  E  T  S  G  S  S  L  T  L
CGCCAGTTCTTCAAGGGCTCCGACGGCGCCATCAACAGCGTCTCCCCGCGCGTCTACCTGCTCGGGGGAGACGGCAACTATGTCGTGCTCAAGCTCCTCG  500
 R  Q  F  F  K  G  S  D  G  A  I  N  S  V  S  P  R  V  Y  L  L  G  G  D  G  N  Y  V  V  L  K  L  L
GCCAGGAGCTGAGCTTCGACGTGGACGTATCGTCGCTCCCGTGCGGCGAGAACGCGGCCCTGTACCTGTCCGAGATGGACGCGACGGGAGGACGGAACGA  600
 G  Q  E  L  S  F  D  V  D  V  S  S  L  P  C  G  E  N  A  A  L  Y  L  S  E  M  D  A  T  G  G  R  N  E
GTACAACACGGGCGGGGCCGAGTACGGGTCGGGCTACTGTGACGCCCAGTGCCCGGTGCAGAACTGGAACAACGGGACGCTCAACACGGGCCGGGTGGGC  700
 Y  N  T  G  G  A  E  Y  G  S  G  Y  C  D  A  Q  C  P  V  Q  N  W  N  N  G  T  L  N  T  G  R  V  G
TCGTGCTGCAACGAGATGGACATCCTCGAGGCCAACTCCAAGGCCGAGGCCTTCACGCCGCACCCCTGCATCGGCAACTCGTGCGACAAGAGCGGGTGCG  800
 S  C  C  N  E  M  D  I  L  E  A  N  S  K  A  E  A  F  T  P  H  P  C  I  G  N  S  C  D  K  S  G  C
GCTTCAACGCGTACGCGCGCGGTTACCACAACTACTGGGCCCCCGGCGGCACGCTCGACACGTCCCGGCCTTTCACCATGATCACCCGCTTCGTCACCGA  900
 G  F  N  A  Y  A  R  G  Y  H  N  Y  W  A  P  G  G  T  L  D  T  S  R  P  F  T  M  I  T  R  F  V  T  D
CGACGGCACCACCTCGGGCAAGCTCGCCCGCATCGAGCGCGTCTACGTCCAGGACGGCAAGAAGGTGCCCAGCGCGGCGCCCGGGGGGGACGTCATCACG  1000
 D  G  T  T  S  G  K  L  A  R  I  E  R  V  Y  V  Q  D  G  K  K  V  P  S  A  A  P  G  G  D  V  I  T
GCCGACGGGTGCACCTCCGCGCAGCCCTACGGCGGCCTTTCCGGCATGGGCGACGCCCTCGGCCGCGGCATGGTCCTGGCCCTGAGCATCTGGAACGACG  1100
 A  D  G  C  T  S  A  Q  P  Y  G  G  L  S  G  M  G  D  A  L  G  R  G  M  V  L  A  L  S  I  W  N  D
CGTCCGGGTACATGAACTGGCTCGACGCCGGCAGCAACGGCCCCTGCAGCGACACCGAGGGTAACCCGTCCAACATCCTGGCCAACCACCCGGACGCCCA  1200
 A  S  G  Y  M  N  W  L  D  A  G  S  N  G  P  C  S  D  T  E  G  N  P  S  N  I  L  A  N  H  P  D  A  H
CGTCGTGCTCTCCAACATCCGCTGGGGCGACATCGGCTCCACCGTCGACACCGGCGATGGCGACAACAACGGCGGCGGCCCCAACCCGTCATCCACCACC  1300
 V  V  L  S  N  I  R  W  G  D  I  G  S  T  V  D  T  G  D  G  D  N  N  G  G  G  P  N  P  S  S  T  T
ACCGCTACCGCTACCACCACCTCCTCCGGCCCGGCCGAGCCTACCCAGACCCACTACGGCCAGTGTGGAGGGAAATGATGGACGGGCCCACCCGCTGCG  1400
 T  A  T  A  T  T  T  S  S  G  P  A  E  P  T  Q  T  H  Y  G  Q  C  G  G  K  Q  W  T  G  T  R  C
AGACGCCCTACACCTGCAAGTACCAGAACGACTGGTACTCGCAGTGCCTGTAAGGCCGGATCCTCTCTGACTGGGTCCCGACCGGGGACGTGTCTTGAG  1500
 E  T  P  Y  T  C  K  Y  Q  N  D  W  Y  S  Q  C  L  *
GAAGAAGGGAGAGGGCATCGAGGTTCATTGTACATAGTAGATAGGTATCATTCCGAATAGTATGTACATACATACATGGGTACATAGGTATATACGTAG  1600
CATATGAACG    1610
```

CBM1   N-glucosylation regions

Genome Portal ID	66950	JGI
Chromosome	3:4135959-4137568	
Family	Glycoside hydrolase7	CAZY
Domains	CBM_1 [Pfam: PR00734, InterProScan]	Pfam
Gene (translation)	1395 bp	
Gene (trancription) [3'UTP, 5'UTP]	1610 bp	JGI
Protein	464 aa	
Exons	1	
Secretion signal	MGQKTLQGLVAAAALAASVANA (22 aa)	SignalP 3.0
Theoretical predicted MW	49.45 kDa	ProtParam
theoretical pI	4.76	
Glucosylation sites N-Glyc	4	NetNGlyc 1.0

SignalP (http://www.cbs.dtu.dk/services/SignalP/) [Bendtsen el al., 2004]
ProtParam (http://expasy.org/tools/protparam/) [Wilkins et al., 1999]
NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/) [Gupta et al.,2004]

Table 5.1. Properties of *Mt*EG7 obtained from genome analysis.

The gene coding for the hypothetical protein *Mt*EG7a (Model ID 111372; chromosome_3:4135959-4137568) was PCR amplified from genomic DNA using primers designed accordingly to the available gene sequence (http://genome.jgi-psf.org/, DOE Joint Genome Institute) including the *ClaI* and *XbaI* restriction enzyme sites at their respective 5'-ends. The two specific primers were designed as follows:

	primer sequence	Tm (⁰C)	% GC
StEGa766950F (30 bp)	5'- GCA TCG ATG CAG CAA CCG GGC ACC TTC ACG - 3' ClaI	60	63.3
StEG7a66950R (29 bp)	5'- CGT CTA GAA GGC ACT GCG AGT ACC AGT CG - 3' XbaI	58	58.6

A high fidelity KOD Hot Start DNA polymerase producing blunt ends was used for the DNA amplification which was carried out with 30 cycles of denaturation (20 s at 95∘C), annealing (10 s at 58∘C), and extension (26 s at 70∘C), followed by 1 min of further extension at 70∘C. In order to determine the DNA sequence, the PCR product was cloned into the PCR-Blunt vector according to the method described by the Zero Blunt PCR Cloning Kit (Invitrogen, USA). The PCR Blunt vector carrying the *eg7a* gene was digested with *ClaI* and *XbaI*, and the produced fragment was gel-purified before cloning into the pPICZαC vector resulting in the recombinant pPICZαC/*eg7a* which was amplified in *E. coli* TOP10F′, and the transformants were selected by scoring for zeocin resistance (25 μg/ml). The recombinant vector pPICZαC/*eg7a* was confirmed by restriction analysis (Figure 4.1). Then, it was linearized with restriction enzyme *SacI* to allow gene replacement at the AOX1 locus and was used to transform *P. pastoris* X33.

5.2. Transformation of P. pastoris and screening of recombinant transformants

P. pastoris host strain X33 and pPICZαC (Invitrogen, USA) were used for protein expression. P. *pastoris* was routinely grown in shaking flasks at 30∘C according to the instructions in the EasySelect *Pichia* Expression Kit (Invitrogen, USA). High-level expression transformants were screened from the YPDS plates containing zeocin

at a final concentration of 100 μg/mL. To screen the *P. pastoris* transformants for endoglucanase expression, 50 colonies were plated out on methanol medium (MM) (1.34% (w/v) yeast nitrogen base, 4x10-5% (w/v) biotin and 0.5% (v/v) methanol top agar containing 1% (w/v) CMC (carboxymethyl-cellulose) at a density of 1 colony/cm². After incubation at 30°C for 24 h, the plates were overlaid with 4 ml of 0.1% Congo red in phosphate–citrate 100 mM pH 5.0 buffer solution. After incubation for 15 min, the plates were washed with 1 M NaCl to reveal clear zones against a red background that were developed by the hydrolysis of CMC (**Figure 4.2**).
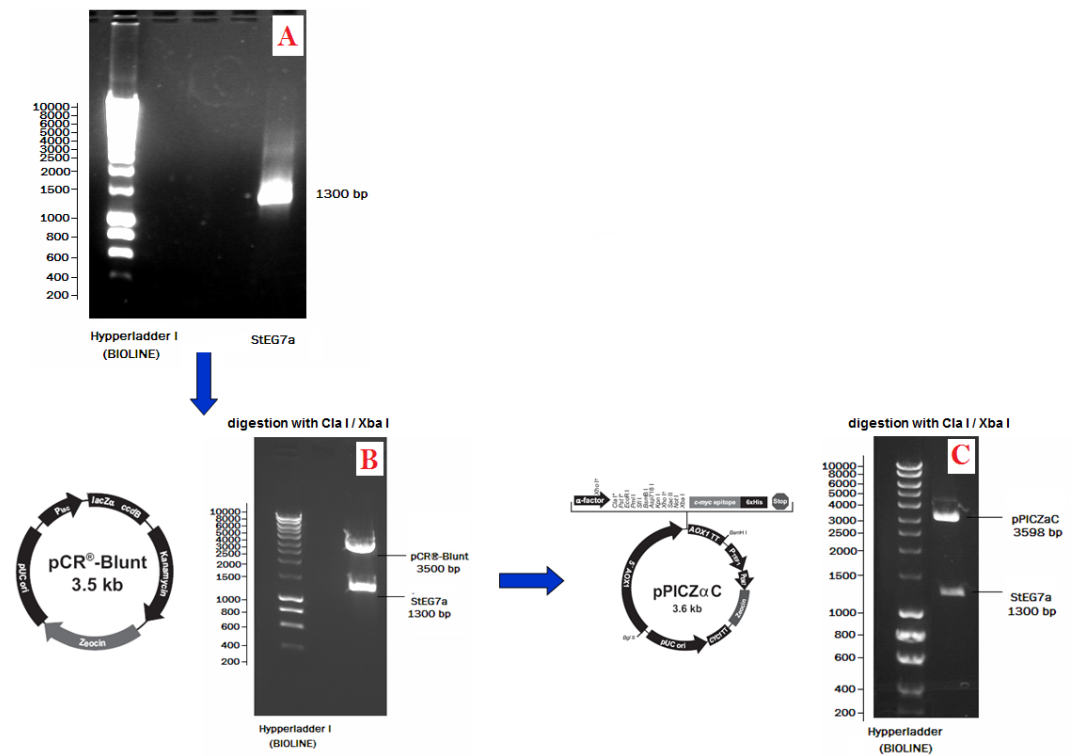
Figure 5.1. Amplification of eg7 gene through PCR (A), cloning of the PCR product into the PCR® Blunt vector (Invitrogen, USA), amplification in *E. coli* TOP10 cells and digestion with *ClaI* and *XbaI* (B). The produced fragment was gel-purified before cloning into the pPICZαC vector resulting in the recombinant pPICZαC/*eg7a* (C) which was amplified in *E. coli* TOP10F cells.
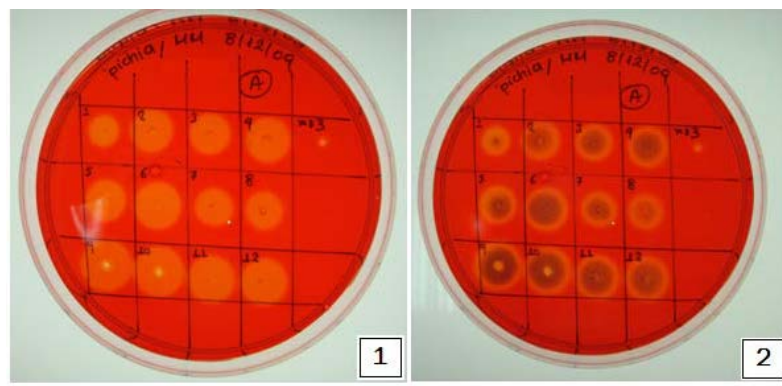
Figure 5.2. The pPICZαC/*eg7a* transformants were selected by their ability to grow on agar plates containing zeocin and their ability to produce surrounding clear zones on MM agar plates containing CMC after Congo red treatment, which are indicative of endoglucanase expression.

5.3. Production and purification of recombinant MtEG7a – Enzyme assay

The production of the recombinant *P. pastoris* harboring *eg7a* gene was grown and harvested, as previously described (Topakas *et al.* 2012). The cultures were kept in a shaking incubator at 30°C for 6 days (200 rpm) with the addition of 0.75 ml methanol once a day to maintain induction (0.5% v/v). Ten colonies zeocin resistant were screened for protein expression and secretion under methanol induction. All transformants produced a major secreted protein product of ca. 65 kDa upon examination of culture supernatants by SDS-PAGE, whereas no protein could be detected with the vector control.

To confirm the production of endoglucanase activity by the transformants, all ten independent clones were assayed using CMC. The *endoglucanase activity* was determined by incubating the enzyme with 1% CMC, for 15 min in 0.1 M citrate-phosphate buffer pH 5.0. The concentration of reducing ends was determined using the dinitrosalicylic acid reagent (DNS) (Miller 1959). Glucose was used for the standard curve. One unit (U) of activity was defined as the amount of enzyme which released 1 μmol of glucose equivalents per min under assay conditions. Recombinant protein was determined by measuring A_{280nm} using molar absorptivity of 83070 M^{-1} cm^{-1} (Stoscheck

1990). The clone showing the highest activity was retained for further study. Endoglucanase activity could be first detected in the medium 24 h after inoculation and peaked at *144 h* with a titer of *468 U/ml* (**Figure 5.3**).
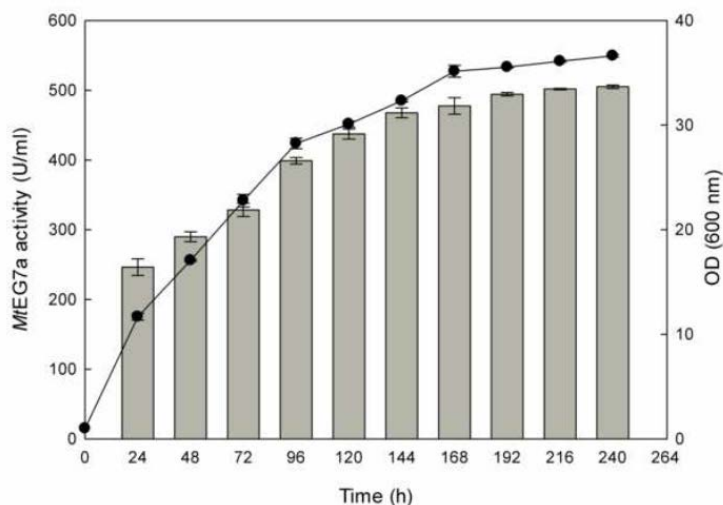
Figure 5.3. Time course of *Mt*EG7a endoglucanase activity (▭) and biomass (●) production of the recombinant *P. pastoris* harbouring the *eg7a* gene. The endoglucanase was expressed in culture broth by induction with 0.5% methanol and measured with CMC as substrate.

For the purification of the recombinant endoglucanase, 800 ml of culture broth were centrifuged and concentrated 30-fold using an Amicon ultrafiltration apparatus (Amicon chamber 8400 with membrane Diaflo PM-10, exclusion size 10 kDa), (Millipore, Billerica, USA). The concentrate was dialyzed overnight at 4 °C against a 20 mM Tris-HCl buffer containing 300 mM NaCl (pH 8.0) and loaded onto a immobilized metal-ion affinity chromatography (IMAC) column (Talon, Clontech; 1.0 cm i.d., 15 cm length) equilibrated with the same buffer. The column was first washed with 300 ml buffer, then a linear gradient from 0 to 100 mM imidazole in 20 mM Tris-HCl buffer containing 300 mM NaCl (60 ml, pH 8.0) was applied at a flow rate of 2 ml/min. Fractions (2 ml) containing endoglucanase activity were concentrated and the homogeneity of the purified recombinant *Mt*EG7a was examined on a SDS-PAGE, which appeared as a single band. The molecular weight was estimated to be ca. 65 kDa (**Figure 5.4a**), which appears to be higher than the predicted value using the

ProtParam tool of ExPASY (49455 Da) considering the presence of the *myc* epitope and the polyhistidine tag which contribute 2.8 kDa to the size of *Mt*EG7a. The nominal mass discrepancy observed for *Mt*EG7a might be explained by the existence of *Asn-Xaa-Ser/Thr* sequons, which are known to be a prerequisite for *N*-glycosylation post-translational modifications. Indeed, 2 *N*-glycosylation sites were predicted by using the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) to occur at *Asn* residues, while another 2 Asn-Pro-Ser sequons were found in which Pro residue is known to preclude *N*-linked glycosylation in most cases by rendering *Asn* inaccessible. Moreover, several potential *O*-glycosylation sites were expected along the flexible linker peptide rich in *Ser* and *Thr* residues between the catalytic and non-catalytic CBM of protein, which is often *O*-glycosylated due to the use of a yeast expression system (Conde *et al.* 2004). Indeed, 13 potential *O*-glycosylation sites (4 *Ser* and 9 *Thr*) were identified, as predicted by using the NetOGlyc 3.1 server (http://www.cbs.dtu.dk/services/NetOGlyc/).
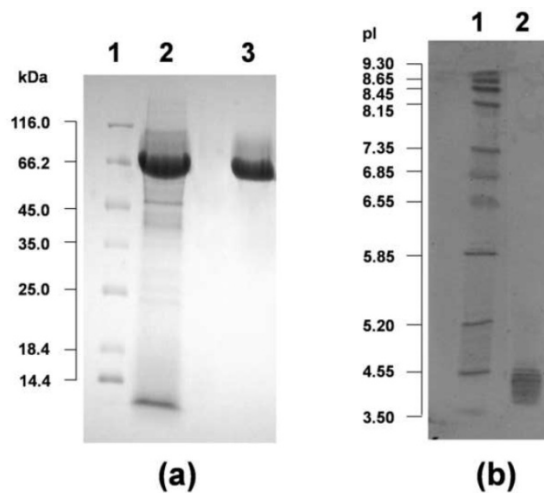
Figure 5.4 SDS-PAGE (**a**) and IEF (**b**) of *Mt*EG7a. (**a**) *Lanes: 1*, LMW standard protein markers; 2, *P. pastoris* culture broth; 3, purified *Mt*EG7a. (**b**) *Lanes: 1*, standard protein markers with pI range 3.5-9.3; 2, purified *Mt*EG7a.

For the determination of isoelectric point (pI), isoelectric focusing (IEF) was performed with the Phastsystem using PhastGel IEF (Amersham Biosciences AB) using broad-range IEF markers (pH 3-9) from Pharmacia. The calculated pI value of

the translated recombinant *Mt*EG7a was found 4.76, which is very close to the experimentally determined value range found from the IEF in the pH range of 3–9 (multiple bands in the range of 3.8-4.5; **Figure 5.4b**).

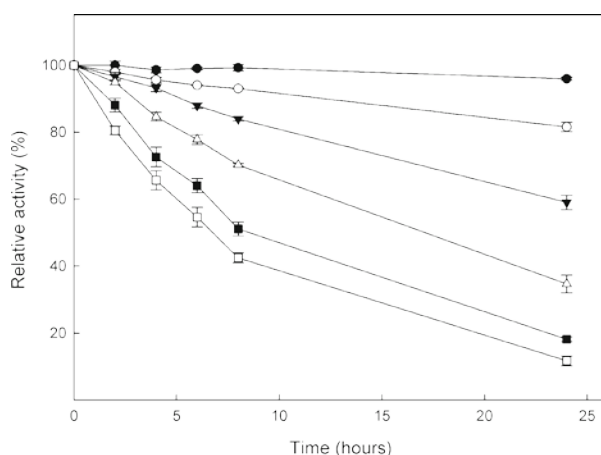*5.4. Enzyme characterization (I) – Temperature and pH optimal activity / stability*

The optimal temperature was determined using the standard assay procedure at temperatures ranging from 30 to 90°C in 0.1 M citrate-phosphate buffer pH 5.0. Temperature stability was determined by measuring the residual activity under the standard assay procedure, after incubation of 0.28 mg of purified *Mt*EG7a at various temperatures for different amount of time. The optimum temperature activity was observed at 60 °C, losing rapidly its activity for temperatures over 65°C (**Figure 5.5b**). The endoglucanase remained fairly stable up to 40 °C after preincubation for 8 hours in 100 mM phosphate-citrate buffer (pH 5.0) at different temperatures (**Figure 5.6**). However, at higher temperatures the enzyme lost part of its initial activity, retaining more than 40% at temperatures up to 80°C for 8 hours of incubation. *Mt*EG7a exhibits half-lives of 9.96 h and 6.5 h at 70°C and 80°C, respectively.
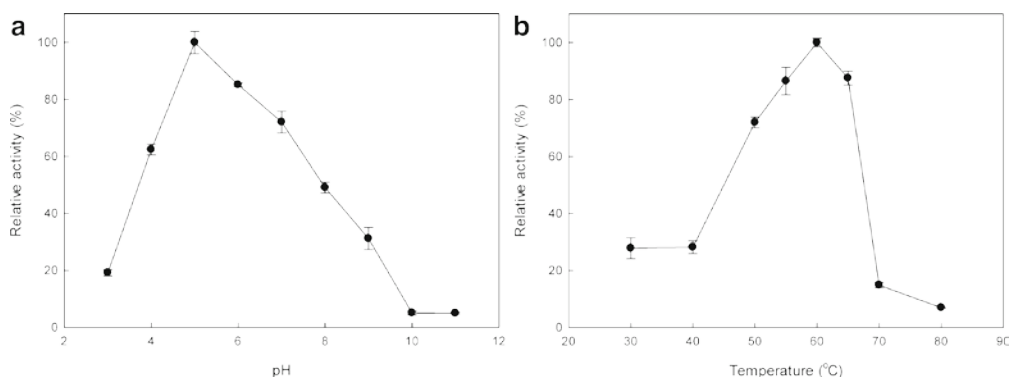
Figure 5.5 Effect of pH (**a**) and temperature (**b**) on the activity of *Mt*EG7a. The experiment was carried out in duplicates.

The optimal pH was determined by the standard assay at 50°C over the pH range 3.0-11.0 using either 0.1 M citrate-phosphate buffer pH 3.0-7.0, 0.1 M Tris-HCl pH 7-9 or 0.1 M glycine-NaOH buffer pH 9-11. The stability at different pH was determined after incubating the enzyme in the above buffers at 4°C for 24 h and then

measuring the activity remaining using the standard assay. The enzyme presented the highest activity levels at pH 5.0 and the >80% of the peak activity was displayed at pH 6, while the activity drops rapidly for pH less than 4 or higher than 7 (**Figure 5.5a**). The enzyme was found remarkably stable in the pH range 3-11 after 24 h retaining its initial activity (data not shown). The activity of *Mt*EG7a at different pH values and temperature conditions reveals that this enzyme corresponds to an acidic thermophilic endoglucanase.

Figure 5.6 Effect of temperature (● 30°C, ○ 40°C, ▼ 50°C, Δ 60°C, ■ 70°C and □ 80°C) on the stability of *Mt*EG7a. The experiment was carried out in duplicates.

5.5. Enzyme characterization (II) – Adsorption on microcrystalline cellulose

The translated sequence of eg7a gene, feature a recognizable carbohydrate binding module (CBM), which is quite common not only to other known endoglucanases but also to cellulolytic enzymes generally in literature. Through this domain, which belongs to the CBM family 1 of the CAZy database, the enzyme adsorbs on the surface of the substrate at the initial step of hydrolysis. It contains conserved cysteines which form disulfide bridges, five aromatic amino acid residues (Tyr433, Trp441, Tyr451, Tyr455 and Trp459), and two glutamines (Gln435 and Gln456) which are essential for the adsorption capacity (Linder *et al.* 1995).

Adsorption of the *Mt*EG7a on cellulose was estimated using 10 mg/ml Avicel as substrate in 0.1 M citrate-phosphate buffer pH 5.0 (Medve *et al.* 1998). Enzyme solutions with known concentrations of protein (5.7 − 113.6 µg/ml) were incubated for 1 hour at 4°C under continuous mixing then filtered and the residual activity was estimated by standard assay. The amount of unbound protein was also determined by measuring A_{280nm}. This adsorption can often be described by the Langmuir isotherm model: $[A] = [A_{max}] \times K_{ad} \times [E] / (1 + K_{ad} \times [E])$ where $[A]$ represents the adsorbed enzyme, $[A_{max}]$ the maximum adsorption amount, K_{ad} the adsorption equilibrium constant, and $[E]$ the initial enzyme concentration in the reaction (Nimz 1974).

The enzyme adsorption on Avicel cellulose is influenced by the concentration of the protein relative to the substrate. At initial enzyme concentration 5.7 µg/mg substrate, the amount of bounded enzyme was 92 %, at 4 °C for 1 hour. The parameters in the Langmuir isotherm were optimized using non-linear regression. The maximum adsorbed amount (W_{max}) was found 110.4 ± 1.8 µg/mg substrate. At the point where the adsorbed amount of *Mt*EG7a was half of the maximum amount, the equilibrium concentration ($1/K$) was 128.7 ± 0.02 mg/ml. With the assumptions that the substrate binding sites are equivalent and the reaction mix was homogeneous, the Langmuir isotherm was found to describe the adsorption well.

5.6. Enzyme characterization (III) − Substrate specificity

To investigate *substrate specificity* of *Mt*EG7a, multiple substrates, such as isolated polysaccharides (lichenan, barley β-glucan, laminarin, wheat arabinoxylan, HEC, HPMC, Avicel) or xylans (oat spelt, beechwood and birchwood xylan) were selected. Enzyme activity was determined after incubation in 0.1 M citrate-phosphate buffer (pH 5.0) containing 1.0% of each substrate at 50°C for 15 min. The amount of reducing sugars produced was estimated using the DNS method, as described above. The activity against filter paper was determined, as previously described (Wood and Bhat 1988). Enzyme activity on p-NPG and p-NPX was tested in reaction mixtures (1 ml) containing 0.145 mg *Mt*EG7a and 10 mM p-NPG or p-NPX in 0.1 M citratephosphate buffer pH 5.0. The mixtures were incubated for 10 min at 50°C. The amount of p-nitrophenol released was measured at A_{410nm} after addition of 0.2 ml 1 M

Na$_2$CO$_3$. For determining the kinetic constants (K$_m$, k$_{cat}$ and k$_{cat}$/K$_m$ ratio) of *Mt*EG7a for the hydrolysis of CMC, the enzyme was incubated at different concentrations of the polysaccharide ranging from 0.10 − 20 % (w/v) in 0.1 M citrate-phosphate buffer pH 5.0 at 50°C.

The enzyme showed relatively high activity toward barley β-glucan (298 U/mg) and 1% low viscosity CMC (177 U/mg) and lower activity toward lichenan (26 U/mg), wheat arabinoxylan (5 U/mg) and other xylans suggesting that its primal enzymatic activity is to hydrolyze the β-1,4 linkages of the substrates (**Table 5.2**). The enzyme exhibited very low activity on Avicel cellulose, while showed no activity toward laminarin, arabic gum, and *p*-NPG. Since β-glucan is a linear polysaccharide formed from glucose residues bound by β-1,4 and β-1,3 glucoside bonds, this enzyme could be classified as β-1,3-1,4-glucanase. However, *Mt*EG7a completely lacked activity against laminarin (β-1,3-glucan), thus it is obvious that the enzyme is strictly specific in hydrolyzing β-1,4-bonds.

Substrate	Specific activity (U/mg protein)
Barley β glucan 1%	298 ± 9
CMC 4% low viscosity	230 ± 4
CMC 1% low viscosity	177 ± 6
CMC 1% high viscosity	106 ± 8
Lichenan 1%	26 ± 0
Oat spelt xylan 1%	10 ± 0.2
Wheat arabinoxylan 1%	5 ± 0.2
Beechwood xylan 1%	5 ± 0
Birchwood xylan 1%	4 ± 0.1
Filter paper	5 ± 0.3
HEC 1%	2 ± 0
Avicel	0.24 ± 0

Table 5.2 Activity of purified *Mt*EG7a on polysaccharide substrates. The enzyme was not active on the following substrates: *p*-NPG, *p*-NPX, HPMC, laminarin and arabic gum.

For studying the hydrolysis products of cellooligosaccharides under the action of the purified enzyme, 1 ng of *Mt*EG7a and 30 μM cellotriose (G3), cellotetraose (G4) or cellopentaose (G5), respectively, in 4 ml citrate-phosphate buffer (pH 5.0) was incubated at 50°C for 100 min. During hydrolysis, samples of 250 μl were taken and inactivated by boiling for 15 min prior to analysis using high-performance anion exchange chromatography (HPAEC), by a CarboPac PA-1 (4 x 250 mm, Dionex) column with a pulsed amperometric detector equipped with a gold electrode. For the analysis, NaOH (66 mM), NaOAc (500 mM) in 66 mM NaOH and 500 mM NaOH as eluent A, B and C respectively, at a flow rate of 1 ml/min. The column was pre equilibrated for 20 min in 100% A. Following sample injection, a gradient run was performed as follows: 0–5 min, isocratic step (100% A), 5.1– 30 min 0- 20% B, 30.1– 35 min 100% B, 35.1 − 40 min 100%C. The column was stabilized for 10 min between separate injections (100% A). For the identification and quantification of hydrolysis products D-glucose and G2–G5 cello-oligosaccharides were used as carbohydrate standards. Assuming that the condition $[E_o] \ll [S_o] \ll K_m$ is satisfied (where $[E_o]$ and $[S_o]$ represent the concentrations of the enzyme and substrate, respectively), the enzymatic hydrolysis of cello-oligosaccharides was regarded as a first-order reaction. Since the integrated rate equation for the first-order kinetics can be written as $k*t = \ln([S_o]/[S_t])$, where $k = (k_{cat}/K_m)[enzyme]$, whereas $[S_o]$ and $[S_t]$ represent substrate concentration prior to the start of the reaction and at a specified time during the reaction, respectively (Matsui *et al.* 1991), the estimation of the catalytic efficiency of *Mt*EG7a against cello-oligosaccharides was made using the above equation.

Analysis of the hydrolysis patterns of cellooligosaccharides G3, G4 and G5 were showed that the enzyme displaying no detectable activity against cellobiose (G2). The hydrolysis of G3 produced G2 and glucose, the hydrolysis of G4 produced only G2, whereas when the enzyme catalyzed the hydrolysis of G5, the products were G2 and G3 which was subsequently cleaved to glucose and G2. The catalytic efficiency ($k$cat/Km) of *Mt*EG7a against G3, G4 and G5 was 4.1 x 103, 5.6 x 104 and 6.8 x 104 min^{-1} M^{-1}, respectively. Kinetics parameters were also calculated for the hydrolysis of CMC at concentrations between 0.10 − 20 % (w/v), where the endoglucanase exhibited a catalytic efficiency (k_{cat}/K_m) of 18.82 ml mg^{-1} min^{-1} ($V_{max} = 622.5 \pm 86.4$ U min^{-1} mg^{-1} of protein and $K_m = 24 \pm 0.5$ mg CMC ml^{-1}).

5.7. Viscosity measurements

The change in viscosity of CMC solution was determined using an Ostwald viscometer and incubating a 5-ml reaction mixture containing 4% CMC (low viscosity) and the enzyme (0.28 µg) in 0.1 M citrate-phosphate buffer, pH 5.0 at 50°C. The time of outflow was measured at different time intervals up to 2 hours. Simultaneously, another set of reactions was carried out under identical conditions for different time intervals and the amount of reducing sugar released was estimated by DNS method. The specific viscosity (n_{sp}) was calculated by the formula $n_{sp} = (t-t_o)/t_o$ where to and t represent the flow time of the buffer and reaction mixture respectively. This was plotted against the amount of released sugars and showed that even a small amount of *Mt*EG7a caused rapid decrease in the relative viscosity of the CMC solution. When 0.28 µg of purified enzyme was added to 5 ml of CMC 4% (w/v) solution, almost 50 % loss of viscosity was observed at the initial stage of the reaction, with simultaneous release of small sugar amounts (**Figure 5.7**), which suggested that it is an endo-acting endoglucanase catalyzing randomly the cleavage of polymeric cellulose (Christakopoulos *et al.* 1995).
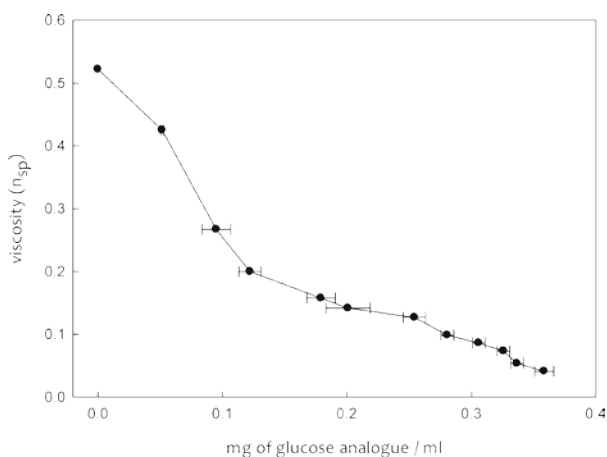
Figure 5.7 Plot of viscosity decrease (n_{sp}) versus the release of reducing sugars for the hydrolysis of CMC by the *Mt*EG7a endoglucanase.

The capability of *Mt*EG7a to reduce the viscosity of high-solid biomass suspensions using an oscillation viscometric measurement method was investigated.

The substrate used for the hydrolysis was hydrothermally pretreated wheat straw (*Triticum aestivum* L.; PWS), obtained from Inbicon A/S (Fredericia, Denmark). Wheat straw residence time in the hydrothermal reactor averaged 12 min with the reactor temperature maintained at 190°C by injection of steam (Thomsen *et al.* 2006). The solid fiber fraction was analyzed according to the procedure of Xiros *et al.* 2009 and was found to contain 50.2% glucan, 3.8% hemicellulose, 25.5% acid insoluble lignin and 2.8% starch, based on DM (w/w). The hydrolysis step of PWS catalyzed by purified *Mt*EG7a was carried out in a liquefaction reactor based on fall mixing consisted of a 6 cm wide and 25 cm in diameter chamber (**Figure 5.8**). A horizontal rotating shaft mounted with three paddlers in each chamber was used for mixing/agitation. A 0.37 kW motor was used as drive and the rotation speed was set at 7 rpm. The direction of rotation was programmed to shift twice a minute between clockwise and anti-clockwise. Mixing and falling down of material results to better saccharification at shorter time duration. An oil-filled heating jacket on the outside enables the control of the temperature up to 90°C.

The experiments were performed at 45 - 60°C, using 5 mg/g DM of *Mt*EG7a that is contained in *P. pastoris* concentrated culture supernatant adjusted at pH=5.0, resulting in 800 g of 18% (w/w) DM. Microbial contaminations were prevented by the addition of 0.02% (w/v) sodium azide. For the determination of the viscosity, aliquots of the liquefacted PWS were taken in different time intervals and apparent viscosities of slurries were measured with an Anton Paar Physica MCR rheometer (Anton Paar Gmbh, Austria), using a parallel plate system with roughened plates. This plate system is reported to be well suited for measuring nearly all of the rheological properties in biomass slurry consisted of pretreated corn stover (Stickel *et al.* 2009). Oscillatory measurements were taken at 25°C, at angular velocities (ω) of the spindle between $10 - 100$ rad/s and normal force of the parallel plates was set at 0 N. Aliquots of the reaction mixture were transferred to pre-weighed Falcon tubes, boiled for 10 min in order to inactivate the enzyme, diluted approximately 10-times and centrifuged. Clear supernatants of the diluted samples were analyzed for reducing sugars by the DNS method (Miller 1959) using D-glucose as standard.

Figure 5.8 The laboratory scale liquefaction reactor based on fall mixing that was employed for the hydrolysis and the reduction of viscosity of PWS (18% w/w, DM) by the recombinant *Mt*EG7a.
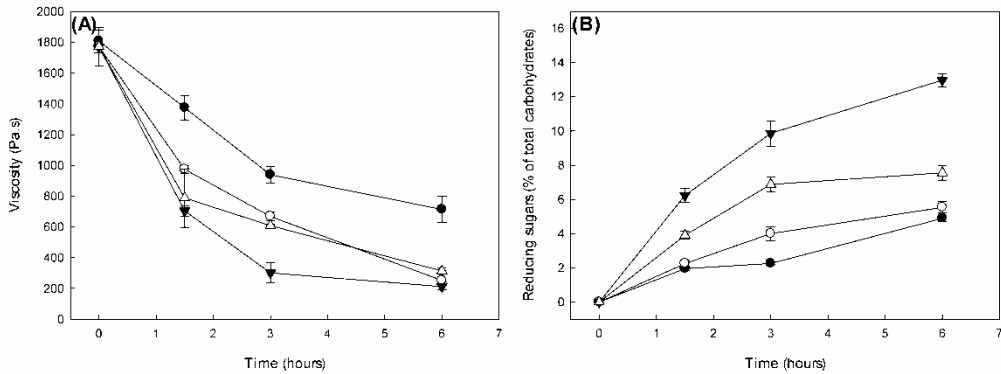

Figure 5.9 Effect of the reaction temperature (● 45°C, ○ 50°C, ▼ 55°C and Δ 60°C) of *Mt*EG7a (5 mg/g DM) on the viscosity (**a**) and release of reducing sugars (**b**), on the hydrolysis of PWS (18% (w/w) DM) during 6 hours of incubation in a laboratory scale liquefaction reactor based on fall mixing at 800 g reaction load.

In all reactions the apparent viscosity decreased with increasing shear rates typically found for non-Newtonian liquids (Stickel *et al.* 2009). The apparent initial viscosity of the mixture (t = 0) was found very high (1860 Pa·s), and, as expected, the viscosity decreased with reaction time, owing to the loss of the lignocellulose structure and water-binding capacity in cellulose degradation. The decrease in viscosity was found to be most pronounced within the first 1.5 hour of reaction (**Figure 5.9a**). Since the viscosity response to angular velocity ω was proportional in all reactions at angular

velocities of less than 60 rad/s (data not shown), the reactions could be compared at a single ω and for the specific velocity of 50 rad/s, at which all measurements at different time points for the different reactions were within the linear range. *Mt*EG7a endoglucanase was found to catalyze the liquefaction of PWS more efficiently at 55°C (**Figure 4.9a**). At this temperature, the viscosity decreased drastically from 1773 to 302 Pa·s within the initial 3 hours of hydrolysis and then became relatively stable for the rest of the liquefaction reaction. Moreover, the saccharification levels obtained at 55°C were far above those obtained at 50°C and 60°C (**Figure 5.9b**).

5.8. Scanning electron microscopy

Scanning electron microscopy (SEM) was performed to image the surface morphology of the initial and the resulting dried fiber fraction from wheat straw after 6 hours of enzymatic hydrolysis at 55 °C. The SEM images were done using a FEI Quanta 200 (FEI Company, Eindhoven, The Netherlands), operated at an accelerating voltage of 25–30 kV. All samples were coated with gold in an Emitech K550X sputter coater. The straw initially exhibited rigid and highly ordered fibrils with smooth and contiguous surface (**Figure 5.10a**), while the fibers of hydrolyzed samples were distorted (**Figure 5.10b**). When the biomass was treated with *Mt*EG7a clefts appeared on the surface of PWS, which seems to be slightly corroded and became heterogeneous and loose.
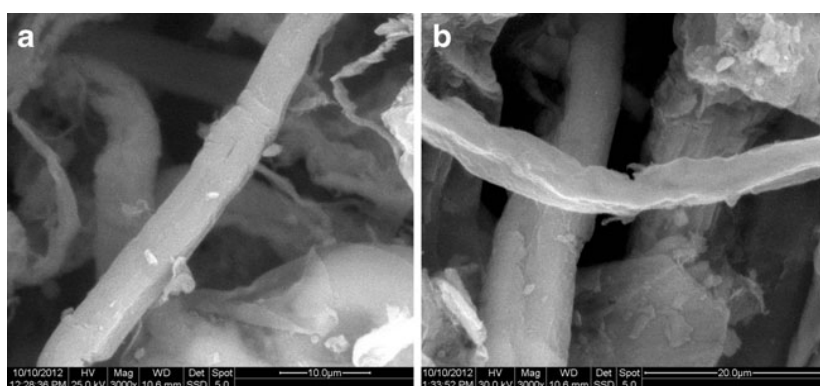
Figure 5.10 SEM images (×3,000 of magnitude) of PWS before (**a**) and after 6 h of hydrolysis (**b**) catalyzed by *Mt*EG7a in a laboratory scale liquefaction reactor based on fall mixing.

5.9. Conclusions - Discussion

The gene encoding *Mt*EG7a from *M. thermophila* was expressed in the heterologous host *P. pastoris*. After induction with methanol, the hydrolytic activity towards CMC and the accumulation of the recombinant enzyme in the culture broth increased significantly up to 468 U/ml after 6-day culture. This expression yield is 20–30-fold higher than levels observed at *Volvariella volvacea* endoglucanase using the *P. pastoris* system ($15.7 - 24.1$ U/ml; Ding *et al.* 2002). Total extracellular enzyme activity in the yeast expression system was about 80 times higher compared with that reported previously for a *E. coli* construct developed to produce recombinant *Trichoderma reseei* Cel7B endoglucanase (5.78 U/ml; Nakazawa *et al.* 2008) and *Erwinia chrysanthemi* endoglucanase (7.8 U/ml; Zhou *et al.* 1999) for use in industrial applications.

The *P. pastoris* expression system has been used successfully to study recombinant endoglucanases from various fungi species, like for example from *Aspergilli* species *Aspergillus kawachii* (AkCel61) (Koseki *et al.* 2008) and *Aspergillus niger* (EGI) (Shumiao *et al.* 2010). The methylotrophic yeast *P. pastoris* is a favorite system for expressing heterologous proteins due to its many advantages, such as protein processing, protein folding and post-translational modification (Cereghino and Cregg 2004).

The calculated molecular mass for *Mt*EG7a and the pI value (46629 Da and pH 4.61, respectively) are similar to other fungal endoglucanases isolated from *Gloeophyllum sepiarium* (45 kDa, pI 3.8) (Mansfield *et al.* 1998) and *Aspergillus aculeatus* (45 kDa, pI 4.3) (Naika *et al.* 2007). However, the higher molecular weight found by SDS-PAGE (ca. 65 kDa) might be explained by *N*- and *O*-glycosylation post-translational modifications, particularly along the flexible linker peptide rich in Ser and Thr residues between the catalytic and non-catalytic CBM of protein. The same observation was indicated in the heterologous expression of a GH61 cellulase from the same thermophilic fungus (*St*Cel61a) in *P. pastoris*, where 23 potential *O*-glycosylation sites were predicted on the *Ser-Thr* linker region (Dimarogona *et al.* 2012). The glycosylation patterns of the recombinant endoglucanase produced by *P. pastoris* may shift the pI (multiple bands in the range of pH 3.8-4.5), as a result of the charged carbohydrate groups added to the molecule during post-translational modification.

These glycosylation patterns have been showed to vary considerably in the length and the type of oligosaccharides. Even in the same glycosylation Golgi machinery of one cell, it is possible for two proteins to be differently glycosylated and has been proposed that this heterogeneity is related to specific sequences recognized by the various glycosylation systems. It has been observed that some proteins heterologously expressed in *P. pastoris* vary considerably in terms of the number of the mannose units added to the same polysaccharide core (Daly and Hearn 2005).

The purified recombinant *Mt*EG7a showed the highest activity toward barley β-glucan (298 U/mg), while considerable activity was found on oat spelt xylan (10 U/mg), wheat arabinoxylan (5.3 U/mg) and other xylans underpinning its primal enzymatic activity on the hydrolysis of the β-1,4 linkages. The enzyme was more active against barley β-glucan than CMC, maybe due to the reason that CMC is highly substituted with methoxy side chains, which may interfere with the enzyme activity (Karlsson *et al.* 2002). Other endoglucanases exhibiting significant xylanolytic activity have been reported previously from *A. niger* (EG1C) (Shumiao *et al.* 2010) and *Fomitopsis pinicola* (Yoon *et al.* 2008). It is worth mentioning that EGI of *T. reesei* is the only known cellulase of *T. reesei*, which displays significant xylanase activity, with a specific activity of 7.9 U/mg on glucuronoxylan and 8.3 U/mg arabinoxylan (Markov *et al.* 2006). It has been previously reported that activity on xylans is a common feature shared only by endoglucanases of GH7 family. This can be ascribed to the structural homology between the GH7 EGs and xylanases. GH7 EGs and xylanases might originate from a diverged ancestral gene, and its remnants might enable Cel7 EGs active on xylan (Vlasenko *et al.* 2010).

The optimum temperature activity of the recombinant enzyme was observed at 60°C while a considerable amount of activity was retained up to 80°C exhibiting more than 40% of its initial activity. Thermophilic enzymes would have advantages in stability during the course of harsh process conditions, and increased catalytic rates at higher temperatures. Other endoglucanases with similar properties (pH=5.0 and T=60°C) have been isolated from different fungal species, such as *F. pinicola* (Yoon *et al.* 2008) or *Bispora* sp. MEY-1 (Luo *et al.* 2010).

The comparison of k_{cat}/K_m values of *Mt*EG7a against the oligosaccharides showed preference against G5 following a G5>G4>G3 pattern. Similar hydrolysis patterns have been observed in the endoglucanases isolated from *Gloeophyllum sepiarium* and *Gloeophyllum trabeum* (Mansfield *et al.* 1998), but differ in terms of hydrolysis of G3. The recombinant *Mt*EG7a exhibited lower *Km* than an endoglucanase isolated from *Bispora* sp. MEY-1 (V_{max}= 3460 U min⁻¹ mg⁻¹ of protein and K_m = 287 mg CMC ml⁻¹; Luo *et al.* 2010).

M. thermophila has been previously characterized as a powerful source of endoglucanases. A very thermostable endoglucanase with properties similar to *Mt*EG7a was isolated from the culture filtrates of *M. thermophila* (Klyosov *et al.* 1988). The molecular weight of the enzyme was 52 kDa, the pI was found 4.7 while the pH optimum was determined between 5.0-6.0. The enzyme exhibited high thermostability after incubation at different temperatures, at 0.1 M sodium acetate buffer (170 hours at 65∘C), which is higher than that of the recombinant *Mt*EG7a. Another group (Bukhtojarov *et al.* 2004) investigated the properties of individual cellulases from the multienzyme complex produced by a mutant strain of the fungus *C. lucknowense* C1, later reclassified as a *M. thermophila* isolate (Visser *et al.* 2011). Among endoglucanases, the highest saccharification activity was displayed by EG60 (60 kDa) and EG51 (51 kDa). The enzymes exhibited pI 3.6 and 5.0 respectively. It has been shown later that the EG51 and EG60 represent the GH5 and GH7 endoglucanases from *M. thermophila*, respectively (Gusakov *et al.* 2011). The latter enzyme first described by Bukhtojarov *et al.* (2004) shows highest activity at 60∘C, pH 5.0 and lacks a CBM domain, which implies that EG60 is different from the *Mt*EG7a described in the present work. Another endoglucanase (*St*Cel5A) from *S. thermophile* (synonym of *M. thermophila*) was discovered by Tambor *et al.* (2012). *St*Cel5A displays a typical GH5 domain, exhibiting optimal activity at pH 6.0 and 70∘C. The enzyme retained greater than 50% of its activity following 2 hours of incubation at 55∘C, diluted in 10 mM citrate buffer pH 4.5.

Only a few experiments on hydrolysis and rheological examination of biomass slurries are reported under high DM conditions, which involve either a monoenzyme treatment under low enzyme loading conditions and a short liquefaction period (Szijártó *et al.* 2011a, 2011b) or a multienzyme treatment under normal enzyme loading conditions and a typical liquefaction period (Rosgaard *et al.* 2007). Under this view, in

the present investigation we evaluated *Mt*EG7a monoenzyme for its ability to liquefy high-consistency lignocellulosic biomass under conditions that mimics the real behavior of that process. *Mt*EG7a was proved to liquefy the biomass efficiently and rapidly, similar to EGI/Cel7B endoglucanase of *T. reesei* (Szijártó *et al.* 2011b). *Mt*EG7a endoglucanase catalyzed the liquefaction of PWS more efficiently at 55°C by decreasing significantly the viscosity of the slurry within the first 3 hours of incubation. The cleavage of the long-chain cellulose fibers during liquefaction process occurs at the initial stages of liquefaction process by the endoglucanases, as reported by Szijártó *et al.* (2011b) and Lu *et al.* (2008), who also observed dramatic rheological changes in the suspensions during the first hours of the hydrolysis. In addition, hydrolysis resulted in the release of small amount of reducing sugars, indicating low saccharification rates and the ability of *Mt*EG7a to alter the viscoelastic properties of wheat straw, possibly by separating the microfibrils and degrading part of the chains. It seems that at the beginning of HG liquefaction experiments, lignocellulosic biomass slurry is insoluble in the solvent and the viscosity of the reaction is high. Along with the process, more and more macromolecules are dissolved by the solvent, resulting in change of the rheological properties and decline in mechanical strength of the material. At elevated temperatures, the reaction rates of thermostable endoglucanases are higher, potentially reducing processing times. The high thermostability and extended life-time of this enzyme can make the liquefaction stage possible at high temperatures.

In the present study for the first time, to our knowledge, a monoenzyme such as the recombinant thermophilic GH7 endoglucanase from *M. thermophila* has been evaluated for its ability to liquefy high-consistency lignocellulosic biomass under conditions that mimics the real behavior of that process. Although many endoglucanases have been reported to date, the high catalytic efficiency and thermostability of *Mt*EG7a, renders it a good candidate for industrial applications, including the saccharification of lignocellulosic materials at high DM content for high ethanol production at the subsequent fermentation.

References

Anish R, Rahman MS, Rao M (2007) Application of cellulases from an alkalothermophilic *Thermomonospora sp.* in biopolishing of denims. *Biotechnol Bioeng* 96:48–56.

Berka RM, Grigoriev IV, Otillar R, *et al.*, (2011) Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nat Biotechnol* 29:922-927.

Bukhtojarov FE, Ustinov BB, Salanovich TN, Antonov AI, Gusakov AV, Okunev ON, Sinitsyn AP (2004) Cellulase complex of the fungus *Chrysosporium lucknowense*: isolation and characterization of endoglucanases and cellobiohydrolases. *Biochemistry (Moscow)* 69:542–551.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* 37:D233-238.

Cereghino JL, Cregg JM (2004) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* 24:45-66.

Christakopoulos P, Kekos D, Macris BJ, Claeyssens M, Bhat MK (1995) Purification and mode of action of a low molecular mass endo- l,4-D-glucanase from *Fusarium oxysporum*. J *Biotechnol* 39:85–93.

Conde R, Cueva R, Pablo G, Polaina J, Larriba G (2004) A search for hyperglycosylation signals in yeast glycoproteins. *J Biol Chem* 279:43789–43798.

Daly R, Hearn M (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit* 18:119–138.

Dimarogona M, Topakas E, Olsson L, Christakopoulos P (2012) Lignin boosts the cellulose performance of a GH-61 enzyme from *Sporotrichum thermophile*. *Bioresour Technol* 110:480-487.

Ding S, Gea W, Buswell JA (2002) Secretion, purification and characterisation of a recombinant *Volvariella volvacea* endoglucanase expressed in the yeast *Pichia pastoris*. Enz Microbial Technology 31:621–626.

Fan Z, South C, Lyford K, Munsie J, van Walsum P, Lynd L (2003) Conversion of paper sludge to ethanol in a semicontinuous solids-fed reactor. *Bioprocess Biosyst Eng* 26:93–101.

Gusakov AV, Punt PJ, Verdoes JC, van der Meij J, Sinitsyn AP, Vlasenko E, Hinz SW, Gosink M, Jiang Z (2011) Fungal enzymes. US Patent No.7,923,236.

Higgins DR, Busser K, Comiskey J, Whittier PS, Purcell TJ, Hoeffler JP (1998) Small vectors for expression based on dominant drug resistance with direct multicopy selection. In: Higgins, D.R., Cregg, J.M., (eds) Methods in molecular biology: *Pichia* protocols, Humana, Totowa, pp 28-41.

Karlsson J, Momcilovic D, Wittgren B, Schülein M, Tjerneld F, Brinkmalm G (2002) Enzymatic degradation of carboxymethyl cellulose hydrolyzed by the endoglucanases Cel5A, Cel7B, and Cel45A from *Humicola insolens* and Cel7B, Cel12A, and Cel45A core from *Trichoderma reesei*. *Biopolymers* 63: 32-40.

Klyosov AA, Ermolova OV, Chernoglazov VM (1988) A thermostable endo-1,4-/β glucanase from *Myceliophthora thermophila*. *Biotechnol Lett*, 10:351-354.

Koseki T, Mese Y, Fushinobu S, Masaki K, Fujii T, Ito K, Shiono Y, Murayama T, Iefuji H (2008) Biochemical characterization of a glycoside hydrolase family 61 endoglucanase from *Aspergillus kawachii*. *Appl Microbiol Biot* 77:1279-1285.

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

Lu Y, Wang Y, Xu G, Chu J, Zhuang Y, Zhang S (2008) Influence of high solid concentration on enzymatic hydrolysis and fermentation of steam-exploded corn stover biomass. *Appl Biochem Biotechnol* 160:360–369.

Luo H, Yang J, Yang P, Li J, Huang H, Shi P, Bai Y, Wang Y, Fan Y, Yao B (2010) Gene cloning and expression of a new acidic family 7 endo-beta-1,3-1,4-glucanase from the acidophilic fungus *Bispora* sp. MEY-1. *Appl Microbiol Biotechnol* 85:1015-1023.

Maheshwari R, Bharadwaj G, Bhat MK (2000) Thermophilic fungi: their physiology and enzymes. *Microbiol Mol Biol* Rev 64:461–488.

Mansfield SD, Saddler JN, Gübitz GM (1998) Characterization of endoglucanases from the brown rot fungi *Gloeophyllum sepiarium* and *Gloeophyllum trabeum*. *Enzyme Microb Technol* 23:133–140.

Markov AV, Gusakov AV, Dzedziulia EI, Ustinov BB, Antonov AA, Okunev ON, Bekkarevich AO, Sinitsyn AP (2006) Properties of Hemicellulases of the Enzyme Complex From *Trichoderma longibrachiatum*. *Appl Biochem Microbiol* 42:573-583.

Matsui I, Ishikawa K, Matsui E, Miyairi S, Fukui S, Honda K (1991) Subsite structure of *Saccharomycopsis* alpha-amylase secreted from *Saccharomyces cerevisiae*. *J Biochem (Tokyo)* 109:566–569.

Medve J, Karlsson J, Lee D, Tjerneld F (1998) Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from *Trichoderma reesei*:adsorption, sugar production pattern, and synergism of the enzymes. *Biotechnol Bioeng* 59:621-34.

Miller GL (1959) Use of dinitrosalicyclic acid reagent for determination of reducing sugars. *Anal Chem* 31:426–428.

Naika GS,Kaul P, Prakash V (2007) Purification and characterization of a new endoglucanase from *Aspergillus aculeatus*. *J Agric Food Chem* 55:7566-7572.

Nakazawa H, Okada K, Kobayashi R, Kubota T, Onodera T, Ochiai N, Omata N, Ogasawara W, Okada H, Morikawa Y (2008) Characterization of the catalytic domains of *Trichoderma reesei* endoglucanase I, II, and III, expressed in *Escherichia coli*. *Appl Microbiol Biotechnol* 81:681-889.

Nimz H (1974) Beech lignin—proposal of a constitutional scheme. *Angew Chem* 13:313–321

Rosgaard L, Andric P, Dam-Johansen K, Pedersen S, Meyer AS (2007) Effects of substrate loading on enzymatic hydrolysis and viscosity of pretreated barley straw. *Appl Biochem Biotechnol* 143:27-40.

Shumiao Z, Huang J, Zhang C, Deng L, Hu N, Liang Y (2010) High-level expression of an *Aspergillus niger* endo-beta-1,4-glucanase in *Pichia pastoris* through gene codon optimization and synthesis. *J Microbiol Biotechn* 20:467-73.

Stickel JJ, Knutsen JS, Liberatore MW, Luu W, Bousfield DW, Klingenberg DJ, Scott CT, Root TW, Ehrhardt MR, Monz TO (2009) Rheology measurements of biomass slurry: an inter-laboratory study. *Rheol Acta* 48:1005–1015.

Stoscheck CM (1990) Quantification of protein. Methods Enzymol 182:50-68 Szijártó N, Horan E, Zhang J, Puranen T, Siika-Aho M, Viikari L (2011a) Thermostable endoglucanases in the liquefaction of hydrothermally pretreated wheat straw. *Biotechnol Biofuels* 26:2.

Szijártó N, Siika-aho M, Sontag-Strohm T, Viikari L (2011b) Liquefaction of hydrothermally pretreated wheat straw at high-solids content by purified *Trichoderma* enzymes. *Bioresour Techn* 102:1968–1974.

Tambor JH, Ren H, Ushinsky S, Zheng Y, Riemens A, St-Francois C, Tsang A, Powlowski J, Storms R (2012) Recombinant expression, activity screening and functional characterization identifies three novel endo-1,4-β-glucanases that efficiently hydrolyse cellulosic substrates. *Appl Microbiol Biotechnol* 93:203-14.

Thomsen MH, Thygesen A, Jørgensen H, Larsen J, Christensen BH, Thomsen AB (2006) Preliminary results on optimisation of pilot scale pretreatment of wheat straw used in coproduction of bioethanol and electricity. *Appl Biochem Biotechnol* 129–132:448–460.

Topakas E, Moukouli M, Dimarogona M, Christakopoulos P (2012) Expression, characterization and structural modelling of a feruloyl esterase from the thermophilic fungus *Myceliophthora thermophila*. *Appl Microbiol Biotechnol* 94:399-411.

Visser H, Joosten V, Punt PJ, Gusakov AV, Olson PT, Joosten R, Bartels J, Visser J, Sinitsyn AP, Emalfarb MA, Verdoes JC, Wery J (2011) Development of a mature fungal technology and production platform for industrial enzymes based on a *Myceliophthora thermophila* isolate, previously known as *Chrysosporium lucknowense* C1. *Ind Biotechnol* 7:214-223.

Vlasenko E, Schólein M, Cherry J, Xu F (2010) Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases. *Bioresour Technol* 101:2405-11.

Wood TM, Bhat MK (1988) Methods for measuring cellulase activities. *Method Enzymol* 160:87–112.

Xiros C, Katapodis P, Christakopoulos P (2009) Evaluation of *Fusarium oxysporum* cellulolytic system for an efficient hydrolysis of hydrothermally treated wheat straw. *Bioresour Technol* 100:5362-5365.

Xiros C, Topakas E, Christakopoulos P (2012) Hydrolysis and fermentation for cellulosic ethanol production. *Wiley Interdiscip Rev Energ Env,* 2:633–654.

Yeoman CJ, Han Y, Dodd D, Schroeder CM, Mackie RI, Cann IK (2010) Thermostable enzymes as biocatalysts in the biofuel industry. *Adv Appl Microbiol* 70:1-55.

Yoon JJ, Cha CJ, Kim YS, Kim W (2008) Degradation of cellulose by the major endoglucanase produced from the brown-rot fungus *Fomitopsis pinicola*. Biotechnol Lett 30:1373-1378.

Zhou S, Yomano LP, Saleh AZ, Davis FC, Aldrich HC, Ingram LO (1999) Enhancement of expression and apparent secretion of *Erwinia chrysanthemi* endoglucanase (encoded by celZ) in *Escherichia coli* B. *Appl Environ Microbiol* 65:2439-2445.

CHAPTER 6

Cloning, expression and characterization of BGL3 β-glucosidase from *M. thermophila* (PAPER III)

Beta-glucosidases (EC 3.2.1.21) are enzymes that participate in the final step of cellulose degradation and belong to GH families 1, 3, 5, 9, 30 and 116. They act on soluble cello-oligosaccharides produced by the action of β-1,4-endoglucanases (EC 3.2.1.4) and cellobiohydrolases (EC 3.2.1.91), including cellobiose, towards the release of D-glucose, β-Glucosidases hydrolyze soluble cellodextrins and cellobiose to D-glucose and thus relieve the system from end product inhibition (Himmel *et al.*, 2007). As shown in various studies (Bezerra & Dias, 2005; Gruno *et al.*, 2004; Zhao *et al.*, 2004) the cellobiose released during the hydrolysis of lignocellulose for ethanol production has a high inhibitory effect on cellulases (Sorensen *et al.*, 2013). Thus the need of instant cellobiose removal is crucial for the efficiency of the process. Due to the fact that the most common cellulolytic commercial complex (Celluclast 1.5L) lacks β-glucosidase activity, efforts have been made for the production of an enzyme supplement that will be rich in this enzymatic activity (Zhang *et al.*, 2010; Gurgu *et al.*, 2011; Pitson *et al.*, 1997). The presence of sufficient β-glucosidase activity in the enzyme mixture is shown to increase the hydrolysis performance by more than 20% reaching even 40% increase in the total ethanol production of the processes (Xin *et al.*, 1993; Han & Chen, 2008). Here lies the need of β-glucosidase activity in the lignocellulosic multi-enzyme systems.

Recently, β-glucosidases have become the focus of many applied studies because they are essential part not only in the cellulose breakdown but also in the synthesis of oligomers and other complex molecules, such as alkyl-glucosides. The synthetic behavior is a result of the transglycosylation activity that has occurred by the presence of a stronger nucleophile compared to water, such as methanol. Therefore, these enzymes can be used for synthesizing a variety of glycoconjugates, such as alkyl glucosides, aminoglycosides and special disaccharide fragments of phytoalexin-elecitor oligosaccharides, which are involved in plant and other microbial defence mechanisms (Bhatia *et al.*, 2002). The enzymatic production of alkyl glycosides, which are

surfactants with good biodegradability and low toxicity, is attractive forming stereochemically well-defined products. To date, much of the effort in the enzymatic synthesis of alkyl glucosides has been placed into GH1 enzymes (*Hansson & Adlercreutz, 2001*; Goyal *et al.*, 2001), however, there is still a great need to find better glycosidases to compete traditional chemical synthesis (Goyal *et al.*, 2001; Turner *et al.*, 2006).

The thermophilic fungi *Myceliophthora thermophila* (synonym *Sporotrichum thermophile*) ATCC 42464, as described in **Chapter 4**, is an exceptionally powerful cellulolytic organism which possesses a great variety of genes encoding proteins with β-glucosidase activity, that are necessary for the breakdown of cellulose. This paper describes, for the first time, the successful cloning of the complete genomic DNA sequence of *M. thermophila* β-glucosidase gene belonging to GH3 family, and its heterologous expression in methylotrophic yeast *P. pastoris*. This enzyme exhibiting high specific activity towards hydrolysis of glucosidic substrates is stimulated by alcohols and has been shown to be an efficient biocatalyst in alkyl glucoside synthesis at increased cellobiose concentrations due to its transglycosylation activity. Moreover, it exhibits a combined tolerance to ethanol, pH and temperature, characteristics that render it a good candidate to be used in SSF processes for biofuel production and reflect potential commercial significance of the enzyme.

6.1. Identification and cloning of MtBgl3a

The translation of *bgl3a* open reading frame (ORF) (Model ID 66804) from the *M. thermophila* genome database shows significant primary sequence identity with known β-glucosidases which have been classified to family GH3 on CAZy database (http://www.cazy.org/; *Cantarel et al.*, 2009). The putative β-glucosidase shows high sequence identity with β-glucosidases identified in *Trichoderma* species, such as BGL1 (70%) from *Trichoderma viride* (*Liu et al.*, 2004) and BGL1 (71%) from *Hypocrea jecorina* (anamorph *T. reesei*) (*Mach, 1993*). The hypothetical protein of 66804 was selected as a candidate β-glucosidase and the corresponding gene, which was provisionally named *bgl3a*, was cloned and used to transform *P. pastoris* X33; the encoded enzyme named *Mt*Bgl3a was expressed and finally characterized.

The ORF of *bgl3a* encodes a protein of 733 amino acids including a secretion signal peptide of 17 amino acids MTLQAFALLAAAALVRG based upon the prediction using SignalP v4.0, which is a web-based program (http://www.cbs.dtu.dk/services/SignalP/). The predicted mass and isoelectric point (pI) of the mature protein was 79819 Da and pH 5.05, respectively, by calculations using the ProtParam tool of ExPASY (http://web. expasy.org/protparam/).

Genome Portal ID	66804
Chromosome	3:4861135-4863642
Family	Glycoside hydrolase 3
Domains	CBM_1 [Pfam: PR00734, InterProScan]
Gene (translation)	2202 bp
Gene (trancription) [3'UTP, 5'UTP]	2508 bp
Protein	733 aa
Exons	2
Secretion signal	MTLQAFALLAAAALVRG (17 aa)
Theoretical predicted MW	79.82 kDa
theoretical pI	5.05
Glucosylation sites N-Glyc	2

SignalP (http://www.cbs.dtu.dk/services/SignalP/) [Bendtsen el al., 2004]
ProtParam (http://expasy.org/tools/protparam/) [Wilkins et al., 1999]
NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/) [Gupta et al.,2004]

Table 6.1. Properties of *MtBgl3a* obtained from genome analysis.

The gene coding for the hypothetical protein *Mt*Bgl3a (Model ID 66804; chromosome 3:4861135-4863642) was PCR amplified from genomic DNA using primers **EF/ER** (**Table 6.2**) designed accordingly to the available gene sequence(http://genome.jgi-psf.org/, DOE Joint Genome Institute, (*Berka et al., 2011*) including the *ClaI* and *XbaI* restriction enzyme sites at their respective 5'-ends. A high fidelity KOD Hot Start® DNA polymerase producing blunt ends was used for the DNA amplification, which was carried out with 30 cycles of denaturation (20 s at 95°C), annealing (10 s at 60°C), and extension (50 s at 70°C), followed by 1 min of further extension at 70°C. In order to determine the DNA sequence, the PCR product was cloned into the PCR Blunt® vector according to the method described by the Zero Blunt® PCR Cloning Kit.

Intron removal was achieved using the molecular technique of overlap extension polymerase chain reaction (OEPCR) (*Topakas et al., 2012*) using the polymerase KOD Hot Start® (Novagen, USA). Two complementary DNA primers per intron, two external primers (**EF/EeR**, **EeF/ER**, **Table 6.2**) and the appropriate PCR amplification process were used to generate two DNA fragments harbouring overlapping ends. The recombinant plasmid pCRBlunt/*bgl3a*, at an appropriate dilution, was used as template DNA and the PCR conditions for each reaction are given as the following: 95°C for 2 min, ensued by 30 cycles of 95°C for 20 s, 60°C for 10 s and 70°C for 16 s (fragment 407 bp) or 20 s (fragment 1795 bp) respectively, with a final extension step at 70°C for 1 min. The two PCR products were combined together in a subsequent hybridization reaction. The generated "*fusion*" fragment was amplified further by overlapping PCR through the utilization of the two external primers, EF end ER, with an initial denaturation step at 95°C for 2 min, followed by 45 cycles at 95°C for 20 s, 60°C for 10 s, 70°C for 25 s and a final extension step at 70°C for 1 min. An extended annealing was performed (25 min) in order to improve base-pairing between the complementary ends of each fragments that have to be fused. The produced *bgl3a* DNA was digested with the enzymes *ClaI* and *XbaI* and the DNA fragment gel-purified before cloning into the pPICZαC vector, resulting in the recombinant pPICZαC/*bgl3a* which was amplified in *E. coli* TOP10F', and the transformants were selected by scoring for Zeocin™ resistance (25 μg/ml). The recombinant vector pPICZαC/*bgl3a* was confirmed by restriction analysis and DNA sequencing and finally transformed into *P. pastoris*. The recombinant plasmid pPICZαC/*bgl3a* was linearized with *SacI*, and then transformation of *P. pastoris* and cultivation in shaken flasks were performed according to the EasySelect™ *Pichia* Expression Kit.

Primer	Primer sequences
EF	5'-GCAT|**CGAT**GGAGACCCCGACCAAGGTCC-3' *ClaI*
ER	5'-CGT|**CTAGA**ACTCGGAAGCTCCCCGTCAG-3' *XbaI*
EeR	5'-GTGAAGTTGGTGTATGACAGCCCAAACCCAAACTCGTACC-3'
EeF	5'-CATACACCAACTTCACGTACGCC-3'

**Table 6.1**. PCR primers for amplification and intron splicing.
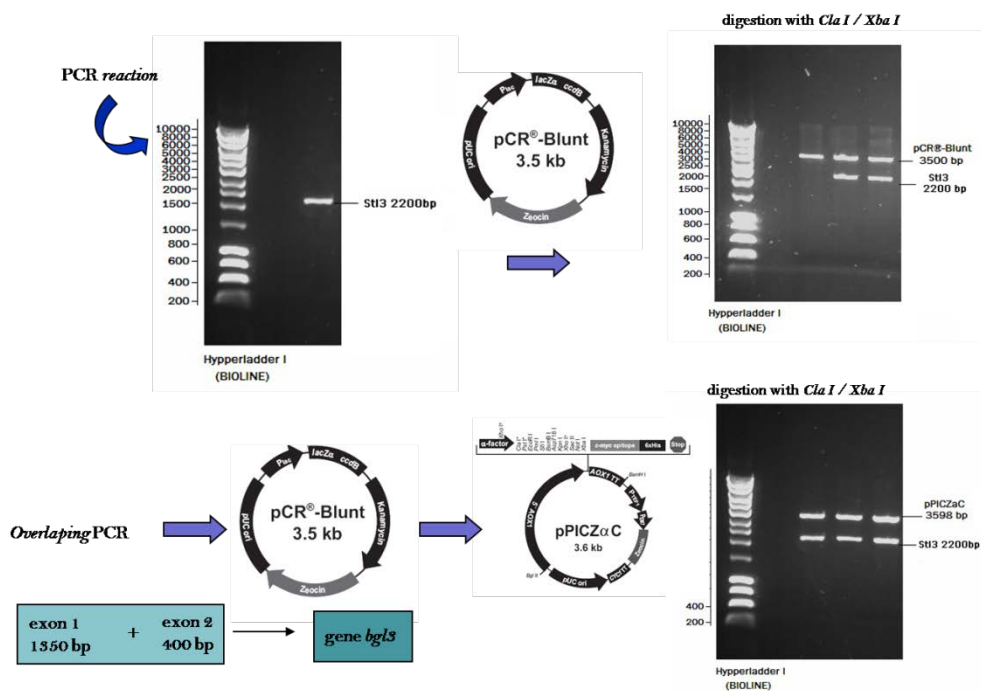
Figure 5.1. Amplification of *bgl3* gene through PCR, cloning of the PCR product into the PCR® Blunt vector (Invitrogen, USA), amplification in *E. coli* TOP10' cells and digestion with *ClaI* and XbaI, subsequent amplification of two exons, final overlapping PCR and cloning into the PCR® Blunt vector as previously. The produced fragment was gel-purified before cloning into the pPICZαC vector resulting in the recombinant pPICZαC/*bgl3* (C) which was amplified in *E. coli* TOP10F' cells.

6.2. Transformation of P. pastoris and screening of recombinant transformants

High-level expression transformants were screened from the YPDS plates containing Zeocin™ at a final concentration of 100 μg/ml. The presence of the *bgl3a* gene in the transformants was confirmed by PCR using yeast genomic DNA as template and gene specific primers (**EF** and **ER**; **Table 6.1**). For the cloning of the β-glucosidase gene from *M. thermophila*, *Escherichia coli* One Shot®Top10 (Invitrogen, USA) and Zero Blunt® PCR Cloning Kit (Invitrogen, USA) were used as the host-vector system. *P. pastoris* host strain X-33 and pPICZαC (Invitrogen, USA) were used for protein expression. The WT strain of *M. thermophila* ATCC 42464 was maintained on 1.5% malt-peptone-agar slants at 4°C. *P. pastoris* was routinely grown in shaking

flasks at 30°C according to the instructions in the EasySelect™ *Pichia* Expression Kit (Invitrogen, USA). Genomic DNA was prepared and isolated as previously described (*Topakas et al., 2012*). An *E. coli/P. pastoris* vector, pPICZαC, was used to achieve secreted expression of *Mt*bgl3a. pPICZαC contains the tightly regulated *AOX1* promoter and the *Saccharomyces cerevisiae* α-factor secretion signal located immediately upstream of the multiple cloning site (*Higgins et al., 1998*).

To screen the *P. pastoris* transformants for β-glucosidase expression, 50 colonies were plated out on MM (1.34% (w/v) yeast nitrogen base, 4×10^{-5}% (w/v) biotin and 0.5% (v/v) MeOH top agar at a density of 1 colony/cm². After incubation at 30 $^{\circ}$C for 24 h, the plates were overlaid with 4 ml of 1% agarose containing 10 mM MeUmbGlc and incubated at room temperature for up to 5 min. The plates were inspected regularly under UV light for fluorescent haloes surrounding recombinant colonies, which is indicative of MeUmbGlc hydrolysis (**Figure 6.2**).
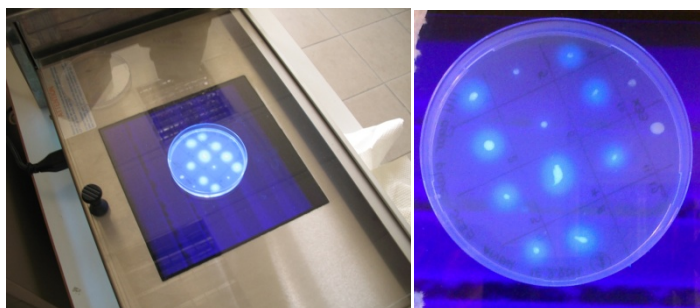
Figure 6.2. The pPICZαC/*bgl3* transformants were selected by their ability to hydrolyse MeUmbGlc substrate, as indicated by the fluorescent haloes surrounding the recombinant colonies.

For the quantification of β-glucosidase activity found in the fluorescent positive *P. pastoris* colonies, the different transformants were cultivated in BMGY medium for 18–24 h, at 30 $^{\circ}$C in a shaker (200 rpm) and then inoculated into the production medium BMMY reaching $OD_{600} = 1$. The extracellular secreted protein was tested for β-glucosidase activity after 24 h of incubation at 30 $^{\circ}$C and 200 rpm. The best recombinant *P. pastoris* harbouring *bgl3a* gene was grown and harvested, as previously described (Topakas *et al., 2012*). The cultures were kept in a shaking incubator at 30 $^{\circ}$C

for 6 days (200 rpm) with the addition of 0.75 ml methanol once a day to maintain induction (0.5% v/v).

6.3. Production and purification of recombinant MtBgl3a – Enzyme assay

After selection of *P. pastoris* transformants by their ability to produce fluorescent haloes under UV light when covered with agar containing MeUmbGlc substrate, ten colonies ZeocinTM resistant were screened for protein expression and secretion under methanol induction. All transformants produced a major secreted protein product of ca. 90 kDa upon examination of culture supernatants by SDS-PAGE, whereas no protein could be detected with the vector control (data not shown). The *β-glucosidase activity* was determined by incubating the enzyme with p-β-NPG. The enzymatic reaction mixtures (1 ml) containing 50 μl of enzyme solution and 1 mM p-β-NPG (final concentration) in 0.1Mcitrate-phosphate buffer pH 5.0 were incubated for 10 min at $50^{o}C$. The amount of p-nitrophenol (p-NP) released was measured at A_{410}, after addition of 0.2 ml 1 M Na_2CO_3 to the reaction mixtures, using a standard curved prepared under the same conditions. β-Glucosidase activity could be first detected in the medium 24 h after inoculation and peaked at 192 h with a titer of 41 U/ml (**Figure 6.3**).
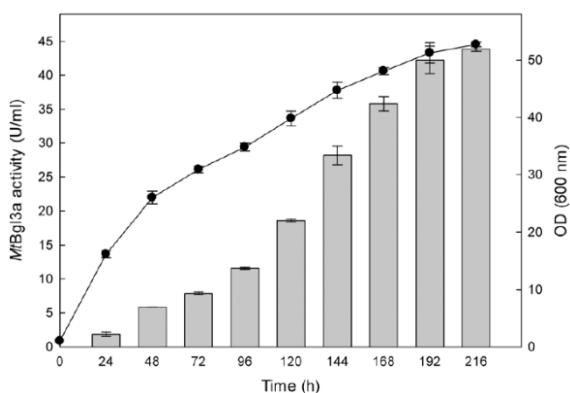

Figure 6.3. Time course of *MtBgl3*a β-glucosidase activity (▭) and biomass (●) production of the recombinant *P. pastoris* harbouring the *blg3a* gene. The β-glucosidase was expressed in culture broth by induction with 0.5% methanol and measured with p-β-NPG as substrate.

For the purification of the recombinant β-glucosidase, 800 ml of culture broth was centrifuged and concentrated 30-fold using an Amicon ultrafiltration apparatus (Amicon chamber 8400 with membrane Diaflo PM-30, exclusion size 30 kDa), (Millipore, Billerica, USA). The concentrate was dialyzed overnight at 4 ^{o}C against a 20 mM Tris-HCl buffer containing 300 mM NaCl (pH 8.0) and loaded onto a immobilized metal-ion affinity chromatography (IMAC) column (Talon, Clontech; 1.0 cm i.d., 15 cm length) equilibrated with the same buffer. The column was first washed with 300 ml buffer, then a linear gradient from 0 to 100 mM imidazole in 20 mM Tris-HCl buffer containing 300 mM NaCl (60 ml, pH 8.0) was applied at a flow rate of 2 ml/min. Fractions (2 ml) containing β-glucosidase activity were concentrated and the homogeneity was checked by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) using 10% acrylamide separating gels, which appeared as a single band. The molecular weight was estimated to be ca. 90 kDa (**Figure 5.6**), which appears to be higher than the predicted value using the ProtParam tool of ExPASY (79819 Da) considering the presence of the myc epitope and the polyhistidine tag which contribute 2.8 kDa to the size of *Mt*Bgl3a. The nominal mass discrepancy observed for *Mt*Bgl3a might be explained by the existence of Asn-Xaa-Ser/Thr sequons and Ser-Thr residues, which are known to be a prerequisite for *N*- and *O*-glycosylation post-translational modifications respectively. Indeed, 2 *N*-glycosylation and 95 potential *O*-glycosylation sites (50 Ser and 45 Thr) were predicted by using the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) and the NetOGlyc 3.1 server (http://www.cbs.dtu.dk/services/NetOGlyc/). The same observation was indicated in the heterologous expression of a feruloyl esterase from the same thermophilic fungus (*St*FaeB) in *P. pastoris*, where 3 potential *N*-glycosylation sites were predicted (Topakas *et al.*, 2012).

For the determination of isoelectric point (pI), isoelectric focusing (IEF) was performed with the Phastsystem using PhastGel IEF (Amersham Biosciences AB) using broad-range IEF markers (pH 3–9) from Pharmacia. Both gels were stained with Coomassie brilliant blue G-250. The calculated pI value of the translated recombinant *Mt*Bgl3a was found 4.0, which is close to the experimentally determined value range found from the IEF in the pH range of 3–9 (multiple bands in the range of 3.8–4.5; **Figure 5.6**). Glycosylation patterns may shift the pI as a result of the charged

carbohydrate groups added to the molecule during post-translational modification. *P. pastoris* glycosylation patterns have been showed to vary considerably in the length and the type of oligosaccharides, thus resulting in diverse structural heterogeneity of the protein population. Within the same cell, two different molecules of the same protein may have different oligosaccharides; even they have been exposed to the same enzymes and glycosylation machinery. It has been observed that some proteins heterologously expressed in *P. pastoris* vary considerably in terms of the number of the mannose units added to the same polysaccharide core (Daly and Hearn, 2005). Both molecular mass and pI values of the isolated recombinant *Mt*Bgl3a are similar to other fungal _-glucosidases, such as β-glucosidase from *Penicillium brasilianum* (92.9 kDa, pI 3.9) (Krogh *et al.*, 2010).
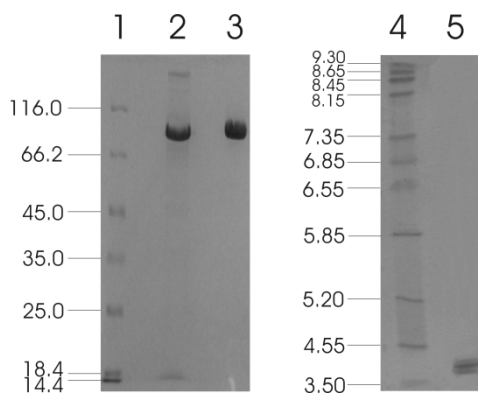

Figure 6.6. SDS-PAGE (A) and IEF (B) of *Mt*Bgl3a. (A) *Lanes: 1*, LMW standard protein markers; *2*, *P. pastoris* culture broth; *3*, purified *Mt*Bgl3a. (B) *Lanes: 1*, standard protein markers with pI range 3.5–9.3; *2*, purified *Mt*Bgl3a.

6.4. Enzyme characterization – Specificity, Temperature and pH optimal activity / stability

The optimal temperature was determined using the standard assay procedure at temperatures ranging from 30 to 80°C in 0.1 M citrate-phosphate buffer pH 5.0. Temperature stability was determined by measuring the residual activity under the standard assay procedure, after incubation of purified *Mt*bgl3a at various temperatures for different amount of time in 50mMMOPS buffer (pH D 6:5) in the presence of 1 mg/ml BSA. The optimal pH was determined by the standard assay at 50°C over the pH range 3.0–11.0 using either 0.1Mcitrate-phosphate buffer pH 3.0–7.0, 0.1 M Tris-

HCl pH 7–9 or 0.1 M glycine-NaOH buffer pH 9–11. The stability at different pH was determined after incubating the enzyme in the above buffers at 4 °C for 4 h and then measuring the activity remaining using the standard assay.
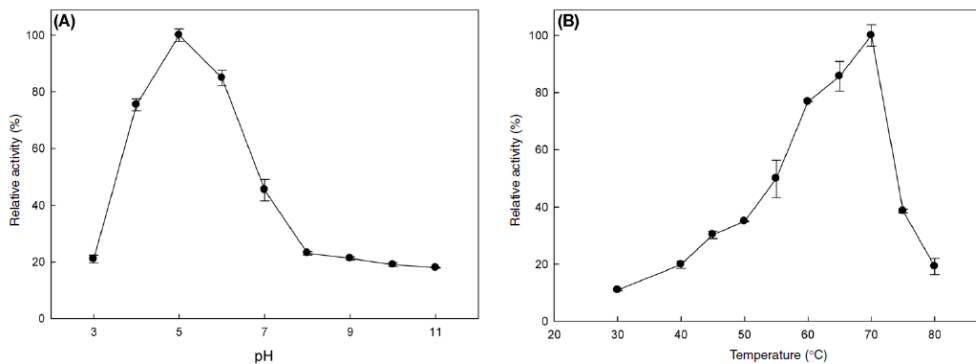
Figure 6.7. Effect of pH (A) and temperature (B) on the activity of *Mt*Bgl3a.
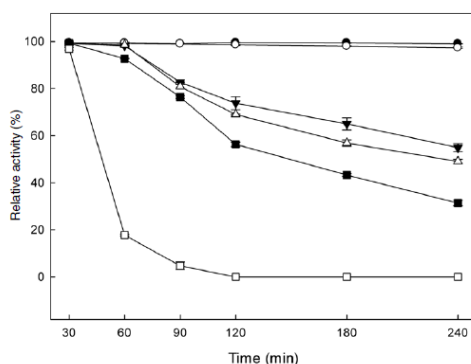
Figure 6.8. The thermal stability of *Mt*Bgl3a preincubated at different temperatures (pH 6.5, 30–65°C) in the absence of substrate and assayed for residual activity on *p*-β-NPG under standard assay conditions. Incubation temperatures: 30 °C (●), 40 °C (○), 50 °C (▼), 55 °C (Δ), 60 °C (■), 65 °C (□) and 70 °C (+). The experiment was carried out in triplicates.

The enzyme presented the highest activity levels at pH 5.0 and the >80% of the peak activity was displayed at pH 6, while the activity drops rapidly for pH less than 4 or higher than 7 (**Figure 6.7**). The enzyme was found remarkably stable in the pH range 3–7 after 24 h of incubation, retaining almost 90% of the initial activity. The specific activity of the *Mt*Bgl3a at different temperatures under standard assay

conditions was measured and the enzyme exhibited its optimal activity at 70°C (**Figure 6.7**). The β-glucosidase was fairly stable up to 60°C for 60 min and retained 56.3% of its activity after 2 h preincubation at the same temperature (**Figure 6.8**). *Mt*Bgl3a exhibits half lives of 274 min, 214 min and 143 min at 50°C, 55 ^{o}C and 60 ^{o}C, respectively. Other β-glucosidases showing similar values for optimal temperature and pH of enzymatic activity have been isolated from different fungal species, such as *A. fumigatus* (pH 6, 60°C; Liu *et al.*, 2012), *P. brasilianum* (pH 4.8, 70 ^{o}C; Krog*h et al.*, 2010) and *T. koningii* (pH 5, 50 ^{o}C; Lin *et al.*, 2010).

The *substrate specificity* of *Mt*bgl3a against p-α-NPG, p-β-NPGal, p-α-NPGal was tested with reaction mixtures containing 0.064 mg *Mt*bgl3a and 5 mM of each substrate under the standard assay conditions. Enzyme activity on p-NPCell was tested at the same conditions, using 1 mM substrate. Enzyme activity on multiple polysaccharide substrates (lichenan, barley β-glucan, laminarin, Avicel) or birchwood xylan was also investigated. Enzyme activity was determined after incubation in 0.1 M citrate-phosphate buffer (pH 5.0) containing 1.0% of each substrate at 50 ^{o}C for 15 min. The amount of reducing sugars released was estimated using the dinitrosalicylic acid reagent (DNS) (Miller, 1959), using D-glucose for the standard curve. The activity on cellobiose was estimated by assaying the amount of released D-glucose using GOD– POD method (Lin *et al.*, 1999). One unit of activity was defined as the amount of enzyme which released 1 μmol of D-glucose equivalents or p-nitrophenol (p-NP) per min under assay conditions. The protein was determined by the absorbance at 280 nm using molar extinction coefficient of 115655M^{-1}cm^{-1} (*Stoscheck, 1990*).

β-Glucosidase was preferentially active against *p*-β-NPG when compared to cellobiose. The highest activity was observed with 1 mM *p*-β-NPG (97.7 U/mg) followed by laminarin (52.0 U/mg), 10 mM cellobiose (30.7 U/mg) and lichenan (20.6 U/mg) (**Table 6.2**). The purified enzyme had lower activity on *p*-NPCell and barley-β-glucan, whereas no detectable activity towards *p*-α-NPGal, *p*-α-NPG, CMC, xylans (wheat and birchwood) and Avicel was observed. Although polymers are not usually substrates for β-glucosidases, *Mt*Bgl3a can hydrolyze long glucans, such as laminarin and lichenan and in this respect resembles an exoglucanase. Several β-glucosidases have been found to hydrolyze laminarin and/or lichenan (Mamma *et al.*, 2004; Riou *et*

al., 1998). According to their specific activity β-glucosidases are classified into three major groups including aryl β-glucosidases with a strong affinity for aryl β-glucosides, cellobiases, which only hydrolyze oligosaccharides (including cellobiose) and β-glucosidases that are active with both type of substrates (Enari and Niku-Paavola, 1987). The above results indicate that β-glucosidase *Mt*Bgl3a was active against both aryl β-glucosides and cellobiose therefore it can be concluded that belong to the last group.

Substrate	Specific activity (U / mg protein)
p-β-NPG (1 mM)	97.7 ± 1.07
p-β-NPGal (5 mM)	traces
p-NPCell 1mM	15.9 ± 0.07
cellobiose 10mM	30.7 ± 0.97
laminarin 0.5%	52.0 ± 2.20
lichenan 1%	20.6 ± 0.04
barley β-glucan 1%	12.4 ± 0.18

Table 6.2. Activity of purified *Mt*Bgl3a on polysaccharide substrates. Activity was measured, as described in "Methods" section. Activity not detected for substrates *p*-α-NPG, *p*-α-NPGal, wheat arabinoxylan, CMC, Avicel, filter paper and birchwood xylan. The experiment was carried out in triplicates.

The effects of various metal ions or other substances at 10 mM on *Mt*bgl3a activity were determined by preincubating the enzyme with the individual compounds in 100 mM citrate-phosphate buffer pH 5.0 at 4 °C for 40 min. Activities were then measured at 50 °C, under standard assay conditions, in the presence of the metal ions or chemical agents. The activity assayed in the absence of metal ions or agents was recorded as 100%. The enzyme was slightly inhibited by Ca^{2+}, Co^{2+}, EDTA and SDS while, it was activated by Mn^{2+}, Mg^{2+}, Zn^{2+} and Cu^{2+} (**Table 6.3**). β-Glucosidases of microbial origin are usually affected by EDTA and the cations Mg^{2+}, Ca^{2+}, Co^{2+}, Mn^{2+}, Cu^{2+}, and Zn^{2+} at 1.0 mM (Venturi *et al.* 2002; Peralta *et al.* 1997). It has been demonstrated that some fungal β-glucosidases are activated by several cations, including Ca^{2+}, Mg^{2+}, Co^{2+}, and/ or Mn^{2+} (Riou *et al.* 1998; Karnchanatat *et al.* 2007).

Metal ion or reagent	Residual activity (%)
KCl	103 %
ZnSO$_4$	149 %
CuSO$_4$	208 %
MgSO$_4$	110 %
MnSO$_4$	167 %
CoCl$_2$	85 %
CaCl$_2$	94 %
EDTA	90 %
urea	95 %
SDS	91 %

Table 6.3. Effect of metal ions and other chemical reagents (10 mM each) on β-glucosidase *Mt*Bgl3a. The experiment was carried out in triplicates.

6.5. Determination of MtBgl3a kinetic parameters – inhibition studies

The values of the Michaelis constant (K_m) and the maximum velocity (V_{max}) for Mtbgl3a were determined by incubating the enzyme in 100 mM citrate-phosphate buffer pH 5.0 at 40 ^{o}C with p-β-NPG and cellobiose at concentrations ranging from 0.1 to 10 mM. The inhibition of *Mt*bgl3a by D-glucose and xylose was determined by assay the enzymatic activity on p-β-NPG in the presence of different inhibitor concentrations. Data were fitted to the Michaelis–Menten equation to generate estimates of values for K_m, V_{max} and K_i, using GraFit data analysis software that also gives an estimate of the standard error of each parameter (Leatherbarrow 1998). The recombinant β-glucosidase showed higher specificity for the hydrolysis of *p*-β-NPG compared to cellobiose, exhibiting *K*m values of 0.39 ± 0.12 mM and 2.64 ± 0.30 mM, respectively. However, both substrates were hydrolysed by the enzyme exhibiting similar velocities of 47.9 ± 3.8 and 49.4 ± 2.4 µmol/min/mg for *p*-β-NPG and cellobiose, respectively. A broad range of *K*m values for *p*-β-NPG and cellobiose has been reported for β-glucosidases produced from different fungal sources, such as *Aspergillus oryzae* (*Km* =0.29 mM for *p*-β-NPG; Langston *et al.*, 2006), *Aspergillus niger* (*Km* =0.57 mM for *p*-β-NPG and 0.88 mM for cellobiose; Chauve *et al.* 2010) and *Trichoderma reesei* (*Km* =0.38 mM for *p*-β-NPG and 1.36 mM for cellobiose; Chauve *et al.* 2010).

The effect of different amounts of *D-glucose* (0-25 mM), *D-xylose* (0-10 mM) and *D-gluconic acid* (0-10 mM) on the hydrolysis of *p*-β-NPG by β-glucosidase was investigated (**Table 6.4**). The enzyme was competitively inhibited by D-glucose and D-xylose with *Ki* values of 282 µM and 30 µM, respectively. β-Glucosidase is frequently a rate-limiting factor during enzymatic hydrolysis of cellulose and is very sensitive to D-glucose inhibition (Mamma *et al.* 2004; Riou *et al.* 1998; Karnchanatat *et al.* 2007). Most of the microbial β-glucosidases, reported to date, are competitively inhibited by D-glucose and exhibit *Ki* values ranging from as low as 0.2 mM to no more than 100 mM (Pitson *et al.*1997; Karnchanatat *et al.* 2007; Mamma, *et al.*, 2004; Seidle *et al.*, 2004; Harhangi *et al.* 2002; Lin, Pillay & Singh 1999; Ferreira Filho 1996; Parry *et al.* 2001). However, several fungal β-glucosidases show high glucose tolerance with *Ki* values of more than 100 mM (Riou *et al.* 1998).

Inhibitor (mM)	*Km* (mM)	*Vmax* (µmol/min.mg protein)	*Ki* (µM)
Glucose	7.262 ± 1.2	100.656 ± 38	282 ± 100
Xylose	9.017 ± 3.5	389.6 ± 72	30 ± 5
glucono-δ-lactone /gluconic acid	1.84 ± 0.5	205.4 ± 49	22 ± 2

Table 6.4. Kinetic parameters of inhibition of *p*-β-NPG hydrolysis by 0-25 mM glucose, 0-10 mM xylose and 0-10 mM glucono-δ-lactone/gluconic acid. The *Ki* values given are the averages of separate experiments on four different substrate concentrations performed in duplicate.

D-Gluconic acid, as a transition state analogue, is by far the most potent inhibitor for the microbial β-glucosidases (Harhangi *et al.* 2002). Cellulolytic fungi have been proposed to own an oxidoreductive cellulolytic system composed by different enzymes, such as GH-61 proteins that co-exist with the well-studied fungal cellulases resulting in efficient lignocellulose conversion (Dimarogona *et al.* 2012; Langston *et al.*, 2011). These enzymes act through a mechanism that involves an hydrolytic and an oxidative step, thus generating two new chain ends on the crystalline surface, one normal non-

reducing and an "oxidized reducing end", i.e an aldonic acid (Forsberg *et al.* 2011; Langston *et al.* 2011). Thus the use of β-glucosidases with tolerance for these oxidized derivatives, such as D-gluconic acid seems to be a promising approach. Like other fungal β-glucosidases, the *Mt*Bgl3a was competitively inhibited by D-gluconic acid (Pitson *et al.*, 1997; Parry *et al.* 2001; Riou *et al.* 1998) with a Ki value of 22 μM in this study. Ki values reported for D-gluconic acid are in the range of 3–30 μM (Harhangi *et al.* 2002; Pitson *et al.* 1997; Parry *et al.* 2001) with exception of the β-glucosidase from *A. oryzae*, which exhibited a Ki value of 12.5 mM (Riou *et al.* 1998).

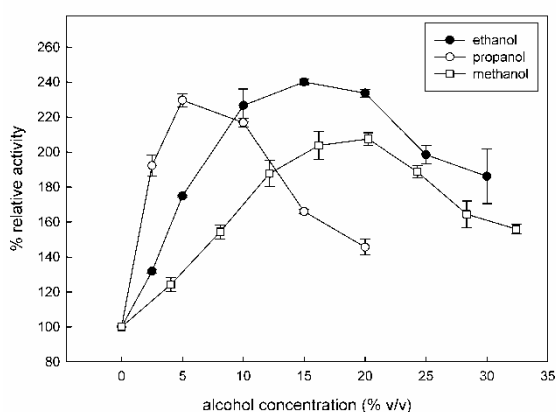### *6.6. Effect of alcohols and transglycosylation activity*

β-Glucosidases catalyse the transglycosylation reaction in an aqueous solution, in the presence of a second nucleophile stronger than water, such as methanol or ethanol (Tsitsimpikou *et al.* 1997).The *effect of alcohols* (ethanol, methanol and propanol) as strong nucleophile reagents on the hydrolysis of *p*-β-NPG was studied. Reaction mixtures containing 1mM *p*-β-NPG in 100 mM citrate phosphate buffer, pH 5.0, with varying concentrations of short chain alcohols were incubated at 50 $^{\circ}$C and the activity was measured under standard assay conditions. The stability at various concentrations of ethanol up to 50% (v/v) at 30 $^{\circ}$C was determined, after incubating the enzyme in 100 mM citrate-phosphate buffer pH 5.0 for 6 hours and then measuring the residual activity using the standard assay.

In the presence of these alcohols, an increase in enzyme activity was observed. Analysis of the reaction products revealed that the optimum *methanol* concentration was 20% (v/v) (**Figure 6.9).** At concentrations higher than 20 % (v/v), the activation was decreased probably due to the denaturation effect of methanol, as many proteins break down in response to alcohol exposure. Furthermore, ethanol and propanol stimulated the activity of β-glucosidase to concentrations up to 15% (v/v) and 5% (v/v), respectively. These results indicate that the presence of short chain alcohols have a positive influence on the hydrolytic activity of β-glucosidase. It has been reported that the change in polarity of the medium induced by alcohols could stabilize enzyme conformation (Mateo & Stefano 2011). Activation by short chain alcohols has been

earlier observed for β-glucosidase from *Thermoascus aurantiacus* (Parry *et al.* 2001), *A. oryzae* (Riou *et al.* 1998), *Fusarium oxysporum* (Christakopoulos *et al.*1994).

*Mt*Bgl3a remained stable after 6 hours of incubation at different *ethanol* concentrations up to 50% (v/v) at 30 °C. This result is important at the SSF process for industrial ethanol production since the enzyme and ethanol coexist in the reactor. Various researches have shown that ethanol reduces the enzyme activities of cellulaces (Wu & Lee 1997; Jorgensen *et al.* 2007) thus the presence of an ethanol tolerant β-glucosidase appears to be critical for the efficiency of the process, as it will be able to remove cellobiose even at the late fermentation process eliminating its inhibitory effect.

Figure 6.9. Effect of increasing concentrations of alcohols, such as methanol (□), ethanol (●) and propanol (○) on the activity of *Mt*Bgl3a. The experiment was carried out in triplicates.

Transglycosylation activity was examined using cellobiose as a donor and methanol as an acceptor. A 2 ml incubation mixture contained 20% (v/v) of methanol, 0.02% NaN$_3$ and different concentrations of cellobiose (1, 2, 4, 5 and 6% w/v) in 0.1 M citrate-phosphate buffer, pH 5.0 and 64 μg mg of enzyme was used. The reaction mixtures were incubated at 50 °C and samples were withdrawn at different time intervals for 5 hours. Methyl-D-glucoside synthesis was monitored by employing an HPLC system (Shimadzu LC-20AD) equipped with a refractive index detector (Shimadzu RID 10A) and a Macherey-Nagel CC 250 x 4.6 mm Nucleosil 100-5 NH2 column. The mobile phase was acetonitrile:water (87:13 v/v) and the sugars were

eluted at a flow rate of 1 ml/min, as described previously (Tsitsimpikou *et al.* 1997). The products were quantified based on peak areas using standard methyl-β-D-glucopyranoside, D-glucose and cellobiose.
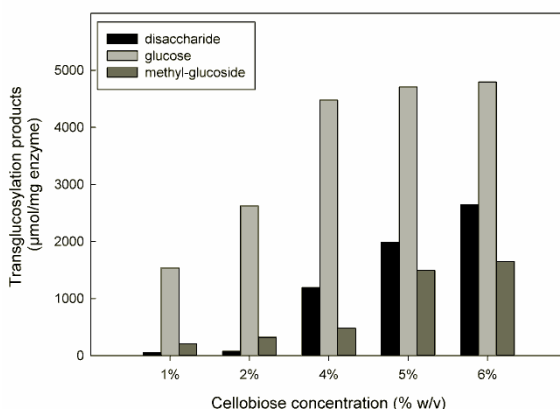
Figure 6.10. Effect of different amounts of cellobiose in transglycosylation reaction with methanol activity catalyzed by *Mt*Bgl3a.

The *inductive effect of the methyl group* and the transglycosylation activity of *Mt*Bgl3a were further investigated by studying the effect of increasing cellobiose concentration on the composition of the product mixture (**Figure 6.10**). Methanol was used at the optimum concentration (20% v/v) found previously for methyl-D-glucoside synthesis. Analysis of the reaction products by HPLC revealed that methyl-D-glucoside synthesis increased sharply at cellobiose concentrations above 4% (w/v) in the reaction. The Km values of cellobiose for transglycosylation (methanolysis) were calculated to be 91 mM. Similar affinity towards cellobiose for transglycosylation (methanolysis) has been reported for β-glucosidase from *F. oxysporum* (Km=138.2 mM) (Tsitsimpikou, 1997). Methanol has a competitive role as a nucleophilic glycosyl acceptor, as glycosylation proceeds through a nucleophilic attack to the stronger nucleophilic character of methanol compared to that of water (Drueckhammer *et al.* 1991). Glycosidase-catalyzed transglycosylation is a promising alternative to classical chemical glycosylation methods, with numerous applications not only in the Food and Cosmetics but also in Pharmaceutical industry for the production of bioactive compounds (Pal *et al.* 2010). The transferase activity was also studied in β-glucosidase

from *P. thermophila* (Yang *et al.*, 2008), *Thermotoga neapolitana* (Park *et al.*, 2005), thermophilic fungus *Melanocarpus sp.* (Kaur *et al.*, 2007).

6.7. Conclusions

Currently, heterologous expression is the main tool for the production of industrial enzymes, with *P. pastoris* being one of the favorite expression hosts for the heterologous expression of eukaryotic biocatalysts. In this study, the gene encoding *Mt*Bgl3a from *M. thermophila* was functionally expressed and secreted by the heterologous host *P. pastoris*. β-Glucosidases are of key importance, as they are needed to supplement the cellobiohydrolase and endoglucanase activities for ensuring final glucose release and at the same time decreasing the accumulation of cellobiose and shorter cellooligmers, which are known as product inhibitors for the cellobiohydrolases. The low inhibition rate by glucose and ethanol renders this enzyme a good candidate for use in many biotechnological processes, including cellulose degradation, where combined stability is appreciated. In addition, the ability of the enzyme to catalyze transglycosylation reactions seems to be very promising for the synthesis of glycoside-containing compounds and bioactive products with potential application in both Food and Pharmaceutical industries.

References

Berka RM, Grigoriev IV, Otillar R, Salamov A, Grimwood J. *et al.* 2011. Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nature Biotechnology* 29:922-927.

Bezerra RMF, Dias AA. 2005. Enzymatic kinetic of cellulose hydrolysis inhibition by ethanol and cellobiose. *Applied Biochemistry and Biotechnlogy* 126: 49-59.

Bhat MK. 2000 Cellulases and related enzymes in biotechnology. *Biotechnology Advances* 18: 355-383.

Bhatia Y, Mishra S, Bisaria VS. 2002. Microbial b-glucosidases: cloning, properties and applications. *Critical Reviews in Biotechnology* 22: 375–407.

Cantarel BL Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37: D233-238.

Cereghino JL, Cregg JM. 2004. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiology Reviews* 24:45-66.

Chauve M, Mathis H, Huc D, Casanave D, Monot F, Lopes Ferreira N. 2010. Comparative kinetic analysis of two fungal beta-glucosidases. *Biotechnology for Biofuels* 3 :3.

Cheng G, Varanasi P, Li C, Liu H, Melnichenko YB, Simmons BA, Kent MS, Singh S. 2011. Transition of cellulose crystalline structure and surface morphology of biomass as a function of ionic liquid pretreatment and its relation to enzymatic hydrolysis. *Biomacromolecules* 12: 933-941.

Christakopoulos P, Goodenough P, Kekos D, Macris BJ, Claeyssen M, Bhat MK. 1994. Purification and characterization of an extracellular β-glucosidase with transglycosylation and exo-glucosidase activities from *Fusarium oxysporum*. *European Journal of Biochemistry* 224: 379-385.

Daly R, Hearn M. 2005. Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *Journal of Molecular Recognition* 18: 119–138.

Dimarogona M, Topakas E, Olsson L, Christakopoulos P. 2012. Lignin boosts the cellulase performance of a GH-61 enzyme from *Sporotrichum thermophile*. *Bioresource Technology* 110: 480–487.

Drueckhammer D, Hennen W, Pederson R, Barbas C, Gautheron C, Krach T, Wong CH. 1991. Enzyme catalysis in synthetic carbohydrate chemistry. *Synthesis* 7: 499-525.

Enari TM, Niku-Paavola ML. 1987. Enzymatic hydrolysis of cellulose: is the current theory of the mechanisms of hydrolysis valid? *Critical reviews in biotechnology* 5: 67-87.

Ferreira Filho EX. 1996. Purification and characterization of a β-glucosidase from solid-state cultures of *Humicola grisea* var. *thermoidea*. *Canadian Journal of Microbiology* 42: 1–5.

Forsberg Z, Vaaje-Kolstad G, Westereng B, Bunæs AC, Stenstrøm Y, MacKenzie A, Sørlie M, Horn SJ, Eijsink VG. 2011. Cleavage of cellulose by a CBM33 protein. *Protein Science* 20:1479-83.

Goyal K, Selvakumar P & Hayashi K. 2001. Characterization of a thermostable β-glucosidase (BglB) from *Thermotoga maritima* showing transglycosylation activity. *Journal of Molecular Catalysis B: Enzymatic* 15: 45–53.

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

Gurgu L, Lafraya Á, Polaina J, Marín-Navarro J. 2011. Fermentation of cellobiose to ethanol by industrial *Saccharomyces* strains carrying the β- glucosidase gene (BGL1) from *Saccharomycopsis fibuligera*. *Bioresource Technology* 102: 5229-5236.

Han Y, Chen H. 2008. Characterization of β-glucosidase from corn stover and its application in simultaneous saccharification and fermentation. *Bioresource Technology*, 99: 6081-6087.

Hansson T & Adlercreutz P. 2001. Enhanced transglycosylation/hydrolysis ratio of mutants of *Pyrococcus furiosus* β-glucosidase: effects of donor concentration, water content, and temperature on activity and selectivity in hexanol. *Biotechnoogy and Bioengineering* 75: 656–665.

Harhangi HR, Steenbakkers PJ, Akhmanova A, Jetten MS, van der Drift C, Op den Camp HJ. 2002. A highly expressed family 1 β-glucosidase with transglycosylation capacity from the anaerobic fungus *Piromyces* sp. E2. *Biochimica et Biophysica Acta* 1574: 293–303.

Higgins DR, Busser K, Comiskey J, Whittier PS, Purcell TJ, Hoeffler JP. 1998. Small vectors for expression based on dominant drug resistance with direct multicopy selection. *In: Higgins, D.R., Cregg, J.M., (eds) Methods in molecular biology: Pichia protocols*, Humana, Totowa, pp 28-41.

Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. 2007. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science* 315: 804-807.

Jorgensen H, Vibe-Pedersen J, Larsen J, Felby C. 2007. Liquefaction of lignocellulose at high-solids concentrations. *Biotechnology and Bioengineering* 96: 862-870.

Karnchanatat A, Petsom A, Sangvanich P, Piaphukiew J, Whalley AJ, Reynolds CD, Sihanonth P. 2007. Purification and biochemical characterization of an extracellular beta-glucosidase from the wood-decaying fungus *Daldinia eschscholzii* (Ehrenb.: Fr.) Rehm. *FEMS Microbiology Letters* 270: 162–170.

Kaur J, Chadha B, Kumar B, SAINI, Harvinder S. 2007. Purification and characterization of two endoglucanases from *Melanocarpus* sp. MTCC 3922. *Bioresource Technology*, 98: 74-81.

Krogh KB, Harris PV, Olsen CL, Johansen KS, Hojer-Pedersen J, Borjesson J, Olsson L. 2010. Characterization and kinetic analysis of a thermostable GH3 β-glucosidase from *Penicillium brasilianum*. *Applied microbiology and Biotechnology* 86: 143-54.

Langston J, Sheehy N, Xu F. 2006. Substrate specificity of *Aspergillus oryzae* family 3 β-glucosidase. *Biochimica et Biophysica Acta*, 1764: 972–978.

Langston JA, Shaghasi T, Abbate E, Xu F, Vlasenko E, Sweeney MD. 2011. Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Applied and Environmental Microbiology* 77: 7007–7015.

Leatherbarrow RJ. 1998. GraFit version 4.0. Erithacus Software Ltd., Staines, UK

Lin J, Pillay B, Singh S. 1999. Purification and biochemical characterization of β-glucosidase from a thermophilic fungus, *Thermomyces lanuginosus*-SSBP. *Biotechnology and Applied Biochemistry* 30:81–87.

Lin Y, Chen G, Ling M, Liang Z. 2010. A method of purification, identification and characterization of β-glucosidase from *Trichoderma koningii* AS3.2774. *Journal of Microbiological Methods* 83: 74–81.

Liu BD, Yang Q, Zhou Q, Song JZ, Chen DF, Liu H. 2004. Cloning and expression of the endo-beta-glucanase III cDNA gene from *Trichoderma viride* AS3.3711. *Huan Jing Ke Xue* 25:127-32.

Liu D, Zhang R, Yang X, Zhang Z, Song S, Miao Y, Shen Q. 2012. Characterization of a thermostable β-glucosidase from *Aspergillus fumigatus* Z5, and its functional expression in *Pichia pastoris* X33. *Microbial Cell Factories* 17:11-25.

Mach RL. 1993. Klonierung und Charakterisierung einiger Gene des Kohlenstoffmetabolismus von *Trichoderma reesei*. Thesis, Mikrobielle Biochemie, Inst.

Mamma D, Hatzinikolaou DG, Christakopoulos P. 2004. Biochemical and catalytic properties of two intracellular β-glucosidases from the fungus *Penicillium decumbens* active on flavonoid glucosides. *Journal of Molecular Catalysis B: Enzymatic* 27: 183–190.

Mansfield SD, Mooney C, Saddler JN. 1999. Substrate and enzyme characteristics that limit cellulose hydrolysis. *Biotechnology Progress* 15: 804-816.

Mateo JJ, Di Stefano R. 1997. Description of the β-glucosidase activity of wine yeasts. *Food Microbiology* 14: 583–591.

Miller GL. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugars. *Analytical Chemistry* 31: 426–428.

Pal S, Banik SP, Ghorai S, Chowdhury S, Khowala S. 2010. Purification and characterization of a thermostable intra-cellular β-glucosidase with transglycosylation properties from filamentous fungus *Termitomyces clypeatus. Bioresource Technology* 101: 2412–2420.

Park TH, Choi KW, Park CS, Lee SB, Kang HY, Shon KJ, Park JS, Cha J. 2005. Substrate specificity and transglycosylation catalyzed by a thermostable beta-glucosidase from marine hyperthermophile *Thermotoga neapolitana. Applied Microbiology and Biotechnology* 69: 411-22.

Parry NJ, Beever DE, Owen E, Vandenberghe I, Van Beeumen J, Bhat MK. 2001. Biochemical characterization and mechanism of action of a thermostable β-glucosidase purified from *Thermoascus aurantiacus. The Biochemical Journal* 353: 117–127.

Peralta RM, Kadowaki MK, Terenzi HF, Jorge JA 1997. A highly thermostable β-glucosidase activity from the thermophilic fungus *Humicola grisea* var. *thermoidea*: purification and biochemical characterization. *FEMS Microbiology Letters* 146: 291–295.

Pitson SM, Seviour RJ, McDougall BM. 1997. Purification and characterization of an extracellular P-glucosidase from the filamentous fungus *Acremonium persicinum* and its probable role in β-glucan degradation. *Enzyme and Microbial Technology* 21:182-190.

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ Jr, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T. 2006. The path forward for biofuels and biomaterials. *Science* 311: 484-489.

Riou C, Salmon JM, Vallier MJ, Günata Z, Barre P. 1998. Purification, characterization, and substrate specificity of a novel highly glucose-tolerant beta-glucosidase from *Aspergillus oryzae. Applied and Environmental Microbiology* 64: 3607–3614.

Seidle HF, Marten I, Shoseyov O, Huber RE. 2004. Physical and kinetic properties of the family 3 β-glucosidase from *Aspergillus niger* which is important for cellulose breakdown. *The Protein Journal* 23: 11–23.

Sorensen A, Lubeck M, Lubeck PS, Ahring BK 2013. Fungal beta-glucosidases: A bottleneck in industrial use of linocellulosic materials. *Biomolecules* 3: 612-631.

Stoscheck CM. 1990. Quantification of protein. *Methods in Enzymology* 182: 50-68.

Topakas E, Moukouli M, Dimarogona M, Christakopoulos P. 2012. Expression, characterization and structural modelling of a feruloyl esterase from the thermophilic fungus *Myceliophthora thermophila. Applied Microbiology and Biotechnology* 94: 399-411.

Tsitsimpikou C, Christakopoulos P, Makropoulou M, Kekos D, Macris BJ and Kolisis FN. 1997. Role of methanol on the catalytic behavior of b-glucosidase from *Fusarium oxysporum Biotechnology Letters* 19: 31–33.

Turner C, Turner P, Jacobson G, Waldebäck M, Sjöberg P, Nordberg Karlsson E & Markides K. 2006. Subcritical water extraction and β-glucosidase catalyzed hydrolysis of quercetin in onion waste. *Green Chemistry* 8:949–959.

Venturi LL, Polizeli Mde L, Terenzi HF, Furriel Rdos P, Jorge JA. 2002. Extracellular beta-D-glucosidase from *Chaetomium thermophilum* var. *coprophilum*: production, purification and some biochemical properties. *Journal of Basic Microbiology* 42: 55–66.

Wu Z, Lee YY. 1997. Inhibition of the enzymatic hydrolysis of cellulose by ethanol. *Biotechnology Letters* 19:977–979.

Xin, Z, Yinbo Q, Peiji G. 1993 Acceleration of ethanol production from paper mill waste fiber by supplementation with β-glucosidase. *Enzyme and Microbial Technology* 15: 62-65.

Yan Q, Hua C, Yang S, Li Y, Jiang Z. 2012. High level expression of extracellular secretion of a β-glucosidase gene (PtBglu3) from *Paecilomyces thermophila* in *Pichia pastoris*. *Protein Expression and Purification* 84: 64–72.

Yang S, Qiaojuan Y, Jiang Z, Fan G, Wang L. Biochemical characterization of a novel thermostable beta-1,3-1,4-glucanase (lichenase) from *Paecilomyces thermophila*. 2008. *Journal of Argricultural and Food Chemistry* 56:5345-51

Zhang M, Su R, Qi W, He Z. 2010. Enhanced enzymatic hydrolysis of lignocellulose by optimizing enzyme complexes. *Applied Biochemistry and Biotechnology* 160: 1407-1414.

Zhao Y, Wu B, Yan B, Gao P. 2004 Mechanism of cellobiose inhibition in cellulose hydrolysis by cellobiohydrolase. *Science in China, Series C: Life Sciences* 47: 8-24.

CHAPTER 7

Lignocellulolytic enzymes from *M. thermophila*

In this Chapter, the cloning, heterologous expression and characterization of one GH5 endoglucanase and two cellobiohydrolases (GH6 and GH7), all from *M. thermophila*, are described. The genes encoding the above enzymes were isolated from the fungal genomic DNA, then cloned and amplified in *E. coli* strains and finally heterologously expressed in *P. pastoris*. The recombinant proteins were secreted to the culture medium, purified to their homogeneity and characterized.

The productivity of *P. pastoris* in shake flasks is typically low and is improved greatly by fermentor culturing. The first reason is that only in the controlled environment of a fermentor is it possible to grow the organism to high cell densities. The second reason is that the level of transcription initiated from the AOX1 promoter is greater in *P. pastoris* cells fed with methanol at growth-limiting rates in fermentor culture than in cells grown in excess of methanol (Chirovolu *et al.*, 1997). Therefore improving the fermentation methodology is important for *P. pastoris* based processes. These improvements include substrate feeding strategies, oxygen supplementation to allow higher cell densities while avoiding oxygen limitation, and mixed-substrate feeding strategies, as described in detail in **Chapter 2**. In this thesis, two basic strategies were followed for the production of one endoglucanase belonging to the glycoside hydrolase family 5, one cellobiohydrolase belonging to the GH6 and one lytic-polysaccharide monooxygenase belonging to the recently reclassified AA9 family (previously reported as GH61). These strategies include control of proteolysis through low temperature and addition of amino acid rich supplements to the culture medium, as well as smooth transition of cell culture from glycerol to methanol feed phase. A common procedure was followed for all three enzymes and resulted in the successful production of the recombinant proteins in the culture medium.

Apart from the great variety of cellulases, *M. thermophila* is capable of synthesizing a complete set of enzymes with xylanase activity, as described in **Chapter 4**. In this chapter, the partial purification of enzymes with xylanolytic activity is

described. Totally three different fractions were isolated from the culture medium of *M. thermophila* either grown on wheat straw or on corn cob. All these fractions exhibited relatively high xylanase specific activity, in absence of cellulolytic activity and thus, they were used for enrichment of the cellulases cocktail in hydrolysis experiments described in **Chapter 8**.

✓ Subjects described in this Chapter:

7.1. Cloning, expression and characterization of Endoglucanase MtEG5

7.2. Cloning, expression and characterization of Cellobiohydrolase MtCBH6

7.3 Cloning, expression and characterization of Cellobiohydrolase MtCBH7

7.4. Production of MtGH61 in fermentor

7.5. Production and purification of enzymes with xylanase activity

7.1. Cloning, expression and characterization of Endoglucanase *Mt*EG5

7.1.1. Identification and cloning of MtEG5

From genome analysis, as described in **Chapter 4**, the translation of *eg5* open reading frame (ORF) (Model ID 86753) from the *M. thermophila* genome database shows significant primary sequence identity with characterized endoglucanases which have been classified to family GH5 on CAZy database (http://www.cazy.org/; Cantarel *et al.* 2009). The putative endoglucanase showed high sequence identity (65%) with structure identified endoglucanase GH5 from *Thermoascus aurantiacus* [PDB ID: 1GZJ] and 79% with endoglucanase from *Penicillium brasilianum* [GenBank: ACB06750]. The hypothetical protein of 86753 was selected as a candidate endoglucanase and the corresponding gene, which was provisionally named *eg5*, was cloned and used to transform *P. pastoris* X33; the encoded enzyme named *MtEG5* was expressed and finally characterized (**Table 7.1**). The ORF of *eg5* encodes a protein of 389 amino acids including a secretion signal peptide of 17 amino acids (MKSSILASVFATGAVA) based upon the prediction using SignalP v4.0, which is a web-based program (http://www.cbs.dtu.dk/services/SignalP/). The predicted mass and isoelectric point (pI) of the mature protein was 40.85 kDa and pH 5.07, respectively, by calculations using the ProtParam tool of ExPASY (http://web.expasy.org/protparam/).

Genome Portal ID	86753
Chromosome	1: 2823610 - 2825549
Family	Glycoside hydrolase 5
Domains	CBM_1, [Pfam: PR00734, InterProScan]
Gene (translation)	1170 bp
Gene (trancription) [3'UTP, 5'UTP]	1940 bp
Protein	389 aa
Exons	3
Secretion signal	MKSSILASVFATGAVA (17 aa)
Theoretical predicted MW	40.85 kDa
theoretical pI	5.07
Glucosylation sites N-Glyc	3
Glucosylation sites O-Glyc	17

Table 7.1. Properties of MtEG5 obtained from genome analysis.

For the cloning of the endoglucanase gene from *M. thermophila*, *Escherichia coli* One Shot® Top10 (Invitrogen, USA) and Zero Blunt® PCR Cloning Kit (Invitrogen, USA) were used as the host-vector system. *P. pastoris* host strain X33 and pPICZαC (Invitrogen, USA) were used for protein expression. The WT strain of *M. thermophila* ATCC 42464 was maintained on 1.5% malt-peptone-agar slants at 4 ^{o}C. *P. pastoris* was routinely grown in shaking flasks at 30 ^{o}C according to the instructions in the EasySelect™ *Pichia Expression Kit* (Invitrogen, USA). Genomic DNA was prepared and isolated as previously described (Topakas *et al.*, 2012).

An *E. coli/P. pastoris* vector, pPICZαC, was used to achieve secreted expression of *Mt*EG5. pPICZαC contains the tightly regulated *AOX1* promoter and the *Saccharomyces cerevisiae* α-factor secretion signal located immediately upstream of the multiple cloning site (Higgins *et al.*, 1998). The gene coding for the hypothetical protein *Mt*EG5 (Model ID 86753, chromosome 1:2823610-2825549) was PCR amplified from genomic DNA using primers **EF/ER** (**Table 7.2**) designed accordingly to the available gene sequence (http://genome.jgi-psf.org/, DOE Joint Genome Institute, (Berka *et al.*, 2011) including the *ClaI* and *XbaI* restriction enzyme sites at their respective 5'-ends. A high fidelity KOD Hot Start® DNA polymerase producing blunt ends was used for the DNA amplification, which was carried out with 30 cycles of denaturation (20 s at 95 ^{o}C), annealing (10s at 56 ^{o}C), and extension (25s at 70 ^{o}C), followed by 1 min of further extension at 70 ^{o}C. In order to determine the DNA sequence, the PCR product was cloned into the pCRBlunt® vector according to the method described by the Zero Blunt® PCR Cloning Kit.

Intron 1 and *intron 2* removal was achieved using the molecular technique of overlap extension polymerase chain reaction (OEPCR) (Topakas *et al.*, 2012) using the polymerase KOD Hot Start® (Novagen, USA). Two complementary DNA primers per intron, two external primers (**EF/Ee1R**, **Ee2F/Ee2R**, **Ee3F/ER**, **Table 7.2**) and the appropriate PCR amplification process were used to generate two DNA fragments harbouring overlapping ends. The recombinant plasmid pCRBlunt/*eg5*, at an appropriate dilution, was used as template DNA and the PCR conditions for each reaction are given as the following: 95 ^{o}C for 2 min, ensued by 30 cycles of 95 ^{o}C for 20 s, annealing for 10 s and extension step, with a final extension step at 70 ^{o}C for 1 min. Annealing and extension conditions for each fragment are described in **Table 7.2**..
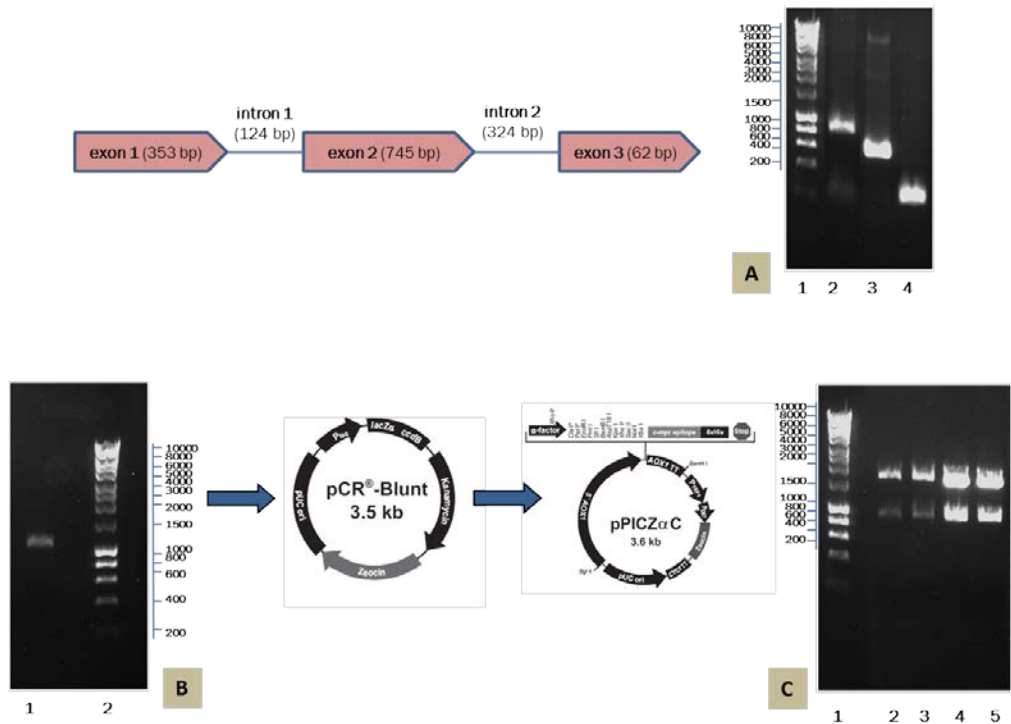
Figure 7.1. Amplification of *eg5* gene through PCR. *Intron 1* and *2* removal (A) and final OE-PCR (B), cloning of the PCR product into the PCR® Blunt vector/amplification in *E. coli* TOP10 cells and final cloning to pPICZαC vector/amplification in *E. coli* TOP10F' cells. (**A**) *Lane 1*: Hyperladder Marker (Bioline), *Lanes 2-4*: exons 1-3, (**B**) *Lane 1*: final OE-PCR product, *Lane 2*: Hyperladder Marker (Bioline), (**C**) *Lane 1*: Hyperladder Marker (Bioline), *Lanes 2-5*: Digestion of pPICZαC/*eg5* with *ClaI/XbaI*.

The two PCR products were combined together in a subsequent hybridization reaction. The generated "*fusion*" fragment was amplified further by overlapping PCR through the utilization of the two external primers, **EF** end **ER**, with an initial denaturation step at 95 °C for 2 min, followed by 45 cycles at 95 °C for 20 s, 56 °C for 10 s, 59 °C for 26 s and a final extension step at 70 °C for 1 min. An extended annealing was performed in order to improve base-pairing between the complementary ends of each fragments that have to be fused. The produced *eg5* DNA was digested with the enzymes *ClaI* and *XbaI* and the DNA fragment gel-purified before cloning into the pPICZαC vector, resulting in the recombinant pPICZαC/*eg5* which was amplified in *E. coli* TOP10F', and the transformants were selected by scoring for Zeocin™ resistance

(25 μg/ml). The recombinant vector pPICZαC/*eg5* was confirmed by restriction analysis and DNA sequencing and finally transformed into *P. pastoris*.

	primer sequence	Tm (ºC)	% GC
StEG5a86753F (32 bp)	5' - GCA TCG ATG CAA AGT GGT CCG TGG CAG CAA TG - 3' *ClaI*	72.1	56.3
StEG5a86753R (32 bp)	5' - CGT CTA GAG GCA AGT ACT TCT TCA AGA TCG AG - 3' *XbaI*	68.2	46.9

|	primer sequence (OE-PCR)	Tm (ºC)	% GC
StEG5a86753e1R (37 bp)	5' - CCA TCA TTG ATG AGC GTC TGA ATC GCC GAA GTC GAC G - 3'	85.9	
StEG5a86753e2F (24 bp)	5' - ACG CTC ATC AAT GAT GGA TAC AAC - 3'	65.8	
StEG5a86753e2R (39 bp)	5' - GTG CCC GAA GGA GGC TCG AAC GAG TAC ATG TAG TCG CCC - 3'	86.1	
StEG5a86753e3F (19 bp)	5' - AGC CTC CTT CGG GCA CCG G - 3'	74.9	

PCR	Primers	Fragment targeted	Conditions
#1	EF – ER	1570 bp	annealing: 56ºC/10s extension: 70ºC/25s, 30 cycles
#2	EF – Ee1R	315 bp	annealing: 56ºC/10s extension: 70ºC/4s, 30 cycles
#3	Ee2F – Ee2R	745 bp	annealing: 56ºC/10s extension: 56ºC/12s, 30 cycles
#4	Ee3F – ER	1600 bp	annealing: 56ºC/10s extension: 70ºC/1s, 30 cycles
#5	EF – ER	1122 bp	annealing: 56ºC/10s extension: 59ºC/26s, 45 cycles

Table 7.2. Primers and conditions used for the amplification of *eg5* gene through PCR (**#1**), removal of introns (**#2-4**) and final overlapping PCR (**#5**). Colored sequences represent the restriction sites of *ClaI* and *XbaI* enzymes (red, purple), as well as the complementary DNA fragments that allowed the hybridization reaction and the amplification of the generated "*fusion*" fragment by overlapping PCR (blue: annealing, green: overhang).

Protein expression of the recombinant enzyme was first evaluated in small scale cultures, in shake flasks. For the production of endoglucanase, one single *P. pastoris* colony harboring *eg5* gene was cultivated in BMGY medium for 18-24 hours, at 30°C in a shaker (200 rpm) and then inoculated into the production medium BMMY reaching $OD_{600}=1$. The extracellular secreted protein was tested for endoglucanase activity against β-glucan 1%w/v in 100mM phosphate-citrate buffer pH=5.0, after 24 hours of incubation at 30 °C and 200 rpm. The clone exhibiting the highest activity was chosen for the production of the recombinant enzyme and further characterization studies. The cultures were kept in a shaking incubator at 30ºC for 6 days (200 rpm) with the addition of 0.75 ml methanol once a day to maintain induction (0.5% v/v).
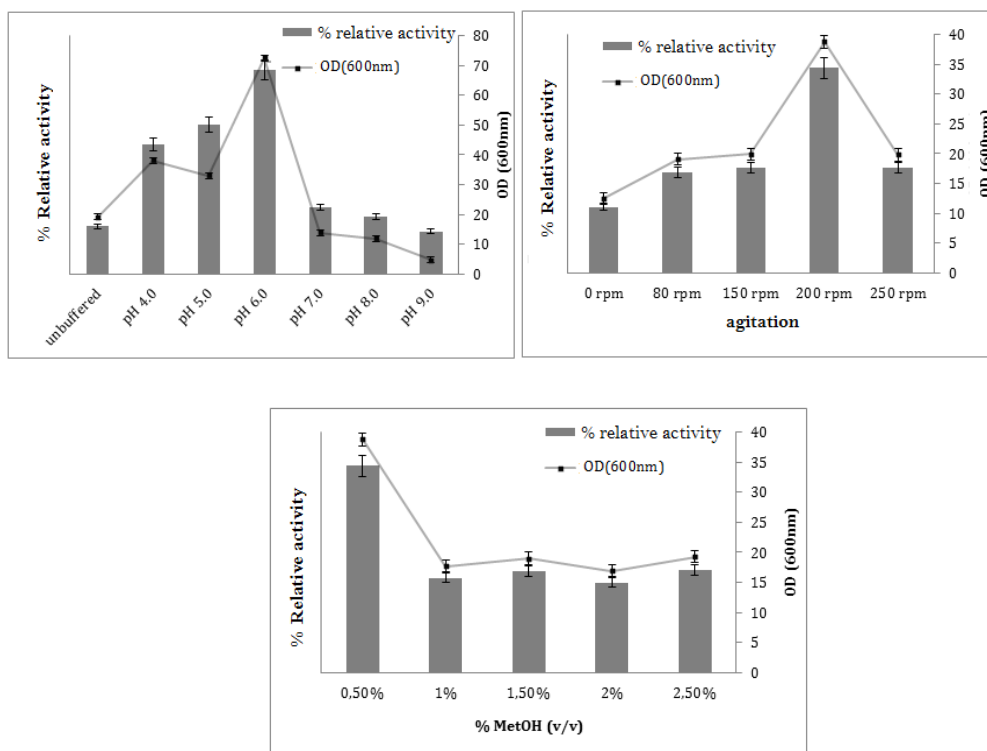
Figure 7.2. Studies of initial pH, methanol concentration and agitation influence at the MtEG5 production in shake flasks. All experiments were conducted in 50 mL BMMY medium, in 250 mL Erlenmeyer flasks.

To achieve higher enzyme production, the influence of initial pH, methanol concentration and agitation was evaluated (**Figure 7.2**). According to these studies, the enzyme reaches the highest levels of expression at shakes flasks when cultured at pH 6.0, 200 rpm agitation and 0.5% (v/v) methanol addition. Under optimum culture conditions, production of *Mt*EG5 reached a final yield of *377 U/mL* of culture supernatant in shake flasks containing BMMY culture medium.

Cultivation of recombinant strain expressing the endoglucanase in *high cell-density fermentation* was performed in the basal salts medium (BSM), as described in the *Pichia* fermentation guidelines provided by Invitrogen (Invitrogen, *Pichia* Fermentation Process Guidelines). The basal salts medium consists of mineral salts and glycerol as the sole carbon source at the initial phase of cultivation. In addition the BSM is supplemented by a trace element solution referred to as PTM_1 trace salts which also includes biotin (Wegner, 1983). The PTM_1 trace salt solution is also included in the glycerol- and methanol feeds during glycerol and methanol fed-batch phases. The only nitrogen source is ammonium hydroxide which was added as the pH was regulated. Cultivation started at $28^{o}C$, aeration was set at 4 vvm and agitation at 800 rpm.


Figure 7.3. Fermentation of *Pichia pastoris* recombinant strains was carried out in a 3 lt glass autoclavable Applikon bioreactor, equipped with an *ez*-Control system (Applikon Biotechnology B.V., Netherlands). Glycerol and methanol feeding was performed with *Alitea* XV pump with *MasterFlex* 96400 L/S 14 silicon tube (glycerol) and *PharmaMed* 0.51mm (methanol).

After 24 hours of batch fermentation in glycerol medium (30 g/lt initial concentration), the dry cell mass of the culture reached about 17.80 g/lt. Analysis of the culture supernatant, after growth in glycerol medium, showed absence of recombinant enzyme (based on β-glucan 0.1%w/v activity assay and SDS-PAGE). The end of glycerol batch was indicated by a sharp increase in the dissolved oxygen (DO) tension. This stage was followed by a 5-hour step of fed-batch glycerol one; during this step 50% w/v glycerol, with PTM_1 salts was fed at an initial flow-rate of 12 mL/h/lt of culture medium and was reduced gradually until it was fully consumed. At the same time, temperature was reduced from to $28^{o}C$ to $25^{o}C$ and finally to $23^{o}C$ and 2 mL of methanol were added manually in small aliquots with syringe. Total consumption of glycerol was again indicated by a spike in the DO. At the end of this stage the dry cell mass of the cells reached the amount of 40.1 g/lt.

At the onset of methanol fed-batch phase, casamino acids solution was added at a final concentration of 3 g/lt and then, a feed of 100% MeOH, with PTM_1 was initiated at a flow rate of 1.9 mL/h/lt. The methanol consumption rate was monitored indirectly by stopping the feed and checking the "*lag phase*", while increasing the methanol feed rate manually. After 8h, feed rate was adjusted to a maximum of 5,46 mL/h/lt and maintained for ~20h, causing extracellular expression of the recombinant enzyme into the supernatant.. Then, the temperature was decreased to $21^{o}C$ and pure oxygen supply was set to maintain dissolved oxygen levels between 60-30 %. Induction time lasted 153h in total and approximately 700 mL of methanol were consumed. The level of enzyme expression increased with fermentation time and maximum level obtained was 6.3 U/mL (β-glucan activity for varying time points shown at **Figure 7.4**). As methanol was used as carbon source, there was an increase in cell-density during the fed batch phase (**Figure 7.4**). At the end of the fermentation, the dry weight of cells reached 98,63 g/lt of culture medium. The amount of extracellular protein produced reached 0,98 g.

For the purification of the recombinant enzyme, the culture broth was centrifuged and concentrated 30-fold using a LabScale Tangential flow filtration system (TFF with membrane Pellicon XL Ultrafiltration Module Biomax, exclusion size 10 kDa; Millipore, Billerica, USA). The concentrate was dialyzed overnight at $4^{o}C$ against a 20 mM Tris-HCl buffer containing 300 mM NaCl (pH 8.0) and *Mt*EG5 was

rapidly purified by single-step immobilized metal ion affinity chromatography (IMAC), with a cobalt charged resin on an ÄKTA Prime Plus system, using 0-100 mM imidazole gradient, at a flow rate of 2 ml/min (**Figure 7.5**). Fractions (2 ml) containing endoglucanase activity were concentrated and the homogeneity was checked by SDS-PAGE (**Figure 7.6**). *Mt*EG5 was further polished using S300 Gel Filtration chromatography column to remove trace endoglucanase contaminants. Removal of background impurities from the fermentation broth resulted in 564 mg of pure *Mt*EG5 per lt of culture supernatant. The molecular weight of the enzyme was estimated to be ca. 75 kDa (**Figure 7.6**), which appears to be higher than the predicted value using the ProtParam tool of ExPASY (40.85 kDa) considering the presence of the *myc* epitope and the polyhistidine tag which contribute 2.8 kDa to the size of *Mt*EG5a. The nominal mass discrepancy observed might be explained by the existence of *Asn-Xaa-Ser/Thr* sequons and *Ser-Thr* residues, which are known to be a prerequisite for *N-* and *O-*glycosylation post-translational modifications respectively. Indeed, 3 *N-*glycosylation and 17 potential *O-*glycosylation sites were predicted by using the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) and the NetOGlyc 3.1 server (http://www.cbs.dtu.dk/services/NetOGlyc/).
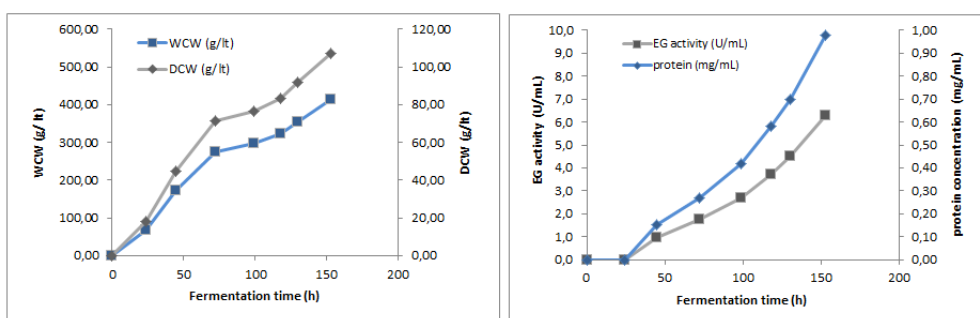
Figure 7.4. (**A**) Cell mass concentration during *Mt*EG5 fermentation. Dry cell weight reached 40.1 g/lt after glycerol fed-batch phase and 110 g/lt at the end of the cultivation. (**B**) Protein concentration and endoglucanase activity detected in the culture medium during fermentation. Protein concentration was determined using the BCA protein assay microplate procedure (Pierce Chemical Co., Rockford, IL), according to the manufacturer's recommendations. Specific activity was tested against β-glucan 0.1% (w/v), pH 5.0, 60°C in 100mM phosphate-citrate buffer
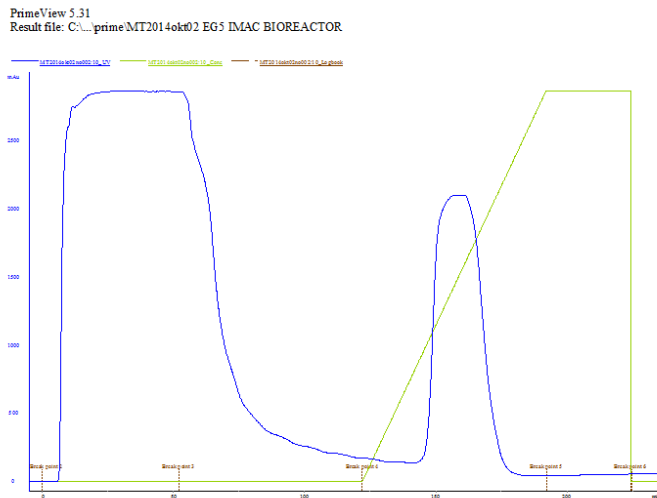
Figure 7.5. Immobilized metal ion affinity chromatography for the purification of the recombinant *Mt*EG5. The left peak corresponds to proteins that were not bound to the resin and the right one to the *His-tagged* endoglucanase that was eluted using a 0-100 mM gradient imidazole. Elution started at 36 mM imidazole, total protein eluted was 602 mg.
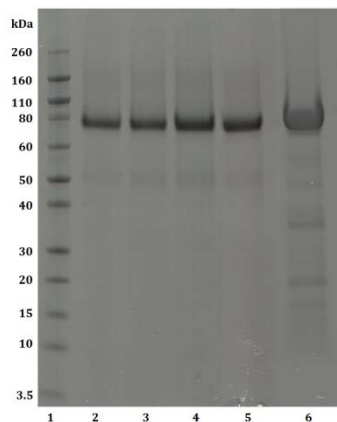

Figure 7.6. SDS – PAGE of *Mt*EG5 during fermentation. *Lane 1*: Novex® sharp pre-stained protein marker, *lanes 2-5*: samples from fermentation culture broth at 99, 118, 130 and 153 h of hydrolysis, *lane 6*: culture broth with *Mt*EG5 after ultrafiltration.

*Mt*EG5 activity was determined on β-glucan 0.1% w/v for 15 min, at 60°C in 100 mM citrate–phosphate buffer pH 5.0. The concentration of reducing ends was determined using the dinitrosalicylic acid reagent (DNS; Miller 1959). Glucose was used for the standard curve. One unit (U) of activity was defined as the amount of

enzyme which released 1 μmol of glucose equivalents per minute under assay conditions. Protein concentration was determined using the BCA protein assay microplate procedure (Pierce Chemical Co., Rockford, IL), according to the manufacturer's recommendations, using bovine serum albumin as standard (Smith *et al.*, 1985).

7.1.3. Characterization of purified MtEG5

The optimal temperature was determined using β-glucan 0.1% (w/v) as a substrate, as described above, at temperatures ranging from 30 to 90°C in 0.1 M citrate-phosphate buffer pH 5.0. Temperature stability was determined by measuring the residual activity under the same assay procedure, after incubation of 0.23 mg of purified *Mt*EG7a at various temperatures for different amount of time. The optimum temperature activity was observed at 70 °C, losing rapidly its activity for temperatures over 80°C. The endoglucanase remained fairly stable up to 50 °C after preincubation for 8 hours in 100 mM phosphate-citrate buffer (pH 5.0) at different temperatures (**Figure 7.7**). *Mt*EG5a exhibits half-lives of 26.9 h and 6.02 h at 55°C and 60°C, respectively.
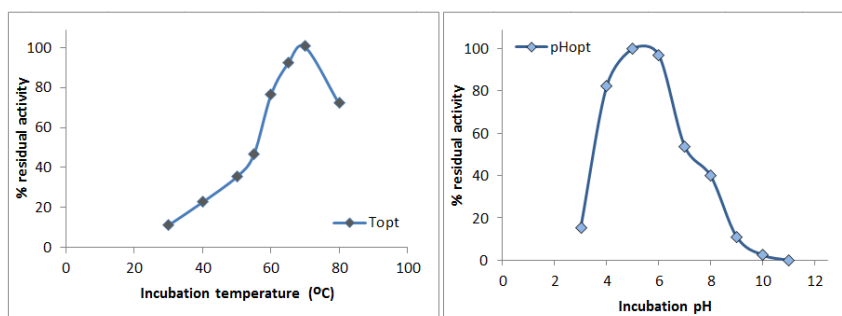
MW	pH$_{opt}$	T$_{opt}$ (°C)	Specific activity (U/mg)		T stability
75	5-6	70	β-glucan 0.1% w/v	9,53	t$_{1/2}$ =26.93h at 55°C
			CMC 1% w/v	0,12	t$_{1/2}$ =6.02h at 60°C
			Avicel 5% w/v	2,23	

Figure 7.7. Properties of purified *Mt*EG5 endoglucanase.

The optimal temperature was determined using the β-glucan assay procedure at temperature ranging from 30 to 80°C in 100 mM citrate–phosphate buffer pH 5.0. Temperature stability was determined by measuring the residual activity under the

standard assay procedure, after incubation of 0.43 mg of purified *Mt*EG5 at various temperatures for different amount of time. The optimal pH was determined by the standard assay at 60°C over the pH range 3.0–11.0 using either 0.1 M citrate–phosphate buffer pH 3.0–7.0, 0.1 M Tris-HCl pH 7–9 or 0.1 M glycine–NaOH buffer pH 9–11. The stability at different pH was determined after incubating the enzyme in the above buffers at 4°C for 24 h and then measuring the activity remaining using the β-glucan assay. The enzyme presented the highest activity levels at pH 5.0 – 6.0, while the activity drops rapidly for pH less than 4 or higher than 7. The enzyme was found remarkably stable in the pH range 3-11 after 24 h retaining its initial activity. Properties of *Mt*EG5 are described in **Figure 7.7**.

7.1.4 Conclusions - Discussion

*Mt*EG5a is an endoglucanase of glycoside hydrolases family 5, consisting of an N-terminal carbohydrate-binding module (CBM1) and a catalytic domain (CD). It has specific activity both on microcrystalline cellulose (Avicel) and β-glucan and exhibits properties that render it a suitable candidate for use in biotechnological applications, such as high temperature stability in 55°C. Its relatively higher activity on microcrystalline cellulose in comparison with CMC is indicative of its processivity properties. *Processivity* is thought to be a critical strategy for improving the catalytic efficiency for hydrolysis of crystalline substrates and is mainly attributed to cellobiohydrolases (Kurasin and Väljamäe, 2011), which are the major components of most cellulolytic systems and are responsible for the degradation of crystalline cellulose (Teeri, 1997). EGs are typically non processive enzymes that are expressed in smaller amounts than CBHs and assist CBHs by randomly attacking internal sites in the cellulose chain, thereby generating new chain ends. However, there have been reported EGs with processive activity, which cleave cellulose internally and also release soluble oligosaccharides before detaching from the polysaccharide. These enzymes belong almost exclusively to the GH9 family, but processive EGs belonging to the GH5 family have been found to be produced by the brown rot basidiomycete *Gloeophyllum trabeum* (Cohen *et al.*, 2005) and *Volvariella volvacea* (Zheng and Ding, 2013). It has been also suggested that there is a strong link between enzyme processivity and adsorption-desorption properties attributed to CBM1 domain (Zheng and Ding, 2013) which is also present in *Mt*EG5 endoglucanase.

7.2. Cloning, expression and characterization of Cellobiohydrolase *Mt*CBH6

7.2.1. Identification and cloning of MtCBH6

From genome analysis, as described in **Chapter 4**, the translation of *cbh6* open reading frame (ORF) (Model ID 66729) from the *M. thermophila* genome database shows significant primary sequence identity with characterized cellobiohydrolases acting on the non-reducing end of the carbohydrate molecules, which have been classified to family GH6 on CAZy database (http://www.cazy.org/; Cantarel *et al.* 2009). The putative cellobiohydrolase showed high sequence identity (79%) with structure identified CBHII from *Humicola insolens* [PDB ID: 1BVW] and 64% with CBHII from *Trichoderma viride* [GenBank: AAQ76094.1]. The hypothetical protein of 66729 was selected as a candidate cellobiohydrolase and the corresponding gene, which was provisionally named *cbh6*, was cloned and used to transform *P. pastoris* X33; the encoded enzyme named *Mt*CBH6 was expressed and finally characterized (**Table 7.3**). The ORF of *cbh6* encodes a protein of 465 amino acids including a secretion signal peptide of 17 amino acids (MAKKLFITAALAAAVLA) based upon the prediction using SignalP v4.0 (http://www.cbs.dtu.dk/services/SignalP/). The predicted mass and isoelectric point (pI) of the mature protein was 49.41 kDa and pH 5.28, respectively, by calculations using the ProtParam tool of ExPASY (http://web.expasy.org/protparam/).

For the cloning of the cellobiohydrolase gene from *M. thermophila*, *Escherichia coli* One Shot® Top10 (Invitrogen, USA) and Zero Blunt® PCR Cloning Kit (Invitrogen, USA) were used as the host-vector system. *P. pastoris* host strain X33 and pPICZαC (Invitrogen, USA) were used for protein expression. The WT strain of *M. thermophila* ATCC 42464 was maintained on 1.5% malt-peptone-agar slants at 4 °C. *P. pastoris* was routinely grown in shaking flasks at 30 °C according to the instructions in the EasySelect™ *Pichia Expression Kit* (Invitrogen, USA). Genomic DNA was prepared and isolated as previously described (Topakas *et al.*, 2012).

An *E. coli/P. pastoris* vector, pPICZαC, was used to achieve secreted expression of *Mt*CBH6. pPICZαC contains the tightly regulated *AOX1* promoter and the *Saccharomyces cerevisiae* α-factor secretion signal located immediately upstream of the

multiple cloning site (Higgins *et al.*, 1998). The gene coding for the hypothetical protein *Mt*EG5 (Model ID 66729, chromosome 2:46305-48489) was PCR amplified from genomic DNA using primers **EF/ER** (**Table 7.4**) designed accordingly to the available gene sequence (http://genome.jgi-psf.org/, DOE Joint Genome Institute, (Berka *et al.*, 2011). The **EF** primer did not include restriction site and was used only for fishing gene from *M. thermophila*'s genome, while the **ER** primer included the *XbaI* restriction enzyme site at 5'-end. A high fidelity KOD Hot Start® DNA polymerase producing blunt ends was used for the DNA amplification, which was carried out with 30 cycles of denaturation (20 s at 95 °C), annealing (10s at 58°C), and extension (32s at 70 °C), followed by 1 min of further extension at 70°C. In order to determine the DNA sequence, the PCR product, containing exons 2-4 and introns was cloned into the pCRBlunt® vector according to the method described by the Zero Blunt® PCR Cloning Kit.

Genome Portal ID	66729
Chromosome	2:46305-48489
Family	Glycoside hydrolase 6
Domains	CBM_1, [Pfam: PR00734, InterProScan]
Gene (translation)	1449 bp
Gene (trancription) [3'UTP, 5'UTP]	1832 bp
Protein	465 aa
Exons	4
Secretion signal	MAKKLFITAALAAAVLA (17 aa)
Theoretical predicted MW	49.41 kDa
theoretical pI	5.28
Glucosylation sites N-Glyc	1
Glucosylation sites O-Glyc	35

Table 7.3. Properties of *Mt*CBH6 obtained from genome analysis.

Introns removal was achieved using the molecular technique of overlap extension polymerase chain reaction (OEPCR) (Topakas *et al.*, 2012) using the polymerase KOD Hot Start® (Novagen, USA). Two complementary DNA primers per intron, two external primers (**Ee2F/Ee2R**, **Ee3F/Ee3R**, **Ee4F/ER**, **Table 7.4**) and the appropriate PCR amplification process were used to generate two DNA fragments harbouring overlapping ends. The primer **Ee2F** included the sequence of *exon 1*, as well

as the *ClaI* restriction enzyme site at 5'–end and was used for the synthesis of the *N*-terminal part of the protein in order to avoid overlapping PCR (**Figure 7.8**). The recombinant plasmid pCRBlunt/*cbh6*, at an appropriate dilution, was used as template DNA and the PCR conditions for each reaction are given as the following: 95 °C for 2 min, ensued by 30 cycles of 95 °C for 20 s, annealing for 10 s and extension step, with a final extension step at 70 °C for 1 min. Annealing and extension conditions for each fragment are described in **Table 7.4**.
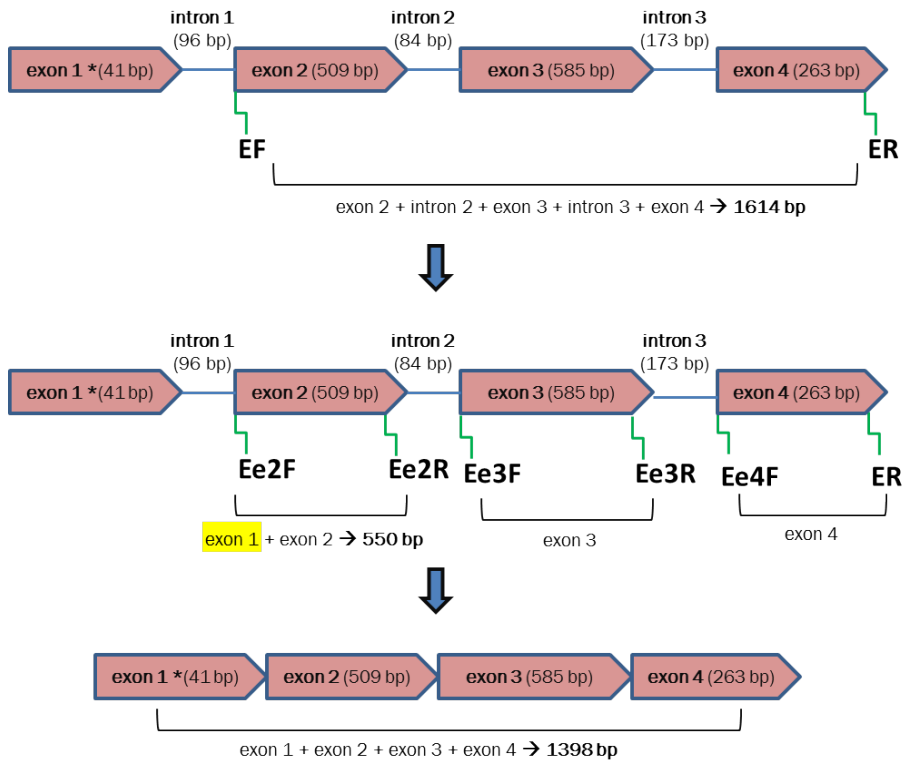
* native signal peptide not included

Figure 7.8. Amplification of *cbh6* gene through PCR. *Intron 2* and *3* removal were done with complementary DNA primers, while *exon 1* was added as part of the **Ee2F** primer. The final OE-PCR resulted in a DNA sequence 1398 bp that is able to encode the *Mt*CBH6 protein. Primers **Ee2F** and **ER** included the *ClaI* and *XbaI* restriction sites respectively at their 5' ends.

	primer sequence	Tm (ºC)	% GC
StCBH6a66729F (21 bp)	5' - GAC TCA ATG CGG CGG TAA CGG - 3'	72.5	62
StCBH6a66729R (29 bp)	5' - CGT CTA GAA AGG GCG GGT TGG CGT TGG TG - 3' *XbaI*	81.7	62

	primer sequence (OE-PCR)	Tm (ºC)	% GC
StCBH6a66729e2F (69 bp)	*ClaI* 5' - GCA TCG ATG GCC CCC GTC ATT GAG GAG CGC CAG AAC TGC GGC GCT GTG TGG ACT CAA TGC GGC GGT AAC - 3'	100	64
StCBH6a66729e2R (38 bp)	5' - GTA GAC GAC GAG TTG GGC AGC ATA GGG AGG ATT GGC AC - 3'	83.9	58
StCBH6a66729e3F (21 bp)	5' - CCC AAC TCG TCG TCT ACG ACC - 3'	67.9	62
StCBH6a66729e3R (40 bp)	5' - CAC CCC ACT GTT GTT GGC CGG TAG GTT GTT TGC CGT TGC G - 3'	89.6	60
StCBH6a66729e4F (22 bp)	5' - GCC AAC AAC AGT GGG GTG ACT G - 3'	71	59

PCR	Primers	Fragment targeted	Conditions
#1	EF - ER	1614 bp	annealing: 58ºC, 10s extension: 70ºC/ 32s, 30 cycles
#2	Ee2F – Ee2R	550 bp	annealing: 55ºC, 10s extension: 68ºC/ 12s, 30 cycles
#3	Ee3F – Ee3R	585 bp	annealing: 55ºC, 10s extension: 68ºC/ 15s, 30 cycles
#4	Ee4F - ER	263 bp	annealing: 57ºC, 10s extension: 70ºC/ 7s, 30 cycles
#5	Ee2F - ER	1398bp	annealing: 54.5ºC, 10s extension: 70ºC/ 26s, 45 cycles

Table 7.4. Primers and conditions used for the amplification of *cbh6* gene through PCR (**#1**), removal of introns (**#2-4**) and final overlapping PCR (**#5**). Colored sequences represent the restriction sites of *ClaI* and *XbaI* enzymes (red, purple), as well as the complementary DNA fragments that allowed the hybridization reaction and the amplification of the generated "*fusion*" fragment by overlapping PCR (blue: annealing, green: overhang).

The three PCR products were combined together in a subsequent hybridization reaction. The generated "*fusion*" fragment was amplified further by overlapping PCR through the utilization of the two external primers, **EF** end **ER**, with an initial denaturation step at 95 ºC for 2 min, followed by 45 cycles at 95 ºC for 20 s, 54.5 ºC for 10 s, 70 ºC for 26 s and a final extension step at 70ºC for 1 min. An extended annealing was performed in order to improve base-pairing between the complementary ends of

each fragments that have to be fused. The produced *cbh6* DNA was digested with the enzymes *ClaI* and *XbaI* and the DNA fragment gel-purified before cloning into the pPICZαC vector, resulting in the recombinant pPICZαC/*cbh6* which was amplified in *E. coli* TOP10F′, and the transformants were selected by scoring for Zeocin™ resistance (25 μg/ml). The recombinant vector pPICZαC/*cbh6* was confirmed by restriction analysis and DNA sequencing and finally transformed into *P. pastoris*.

7.2.2. Expression in high-cell density cultures and purification of *Mt*CBH6

Protein expression of the recombinant enzyme was first evaluated in small scale cultures, in shake flasks. For the production of cellobiohydrolase, one single *P. pastoris* colony harboring *cbh6* gene was cultivated in BMGY medium for 18-24 hours, at 30°C in a shaker (200 rpm) and then inoculated into the production medium BMMY reaching $OD_{600}=1$. The extracellular secreted protein was tested for cellobiohydrolase activity against Avicel 5% (w/v) in 100 mM phosphate-citrate buffer pH=5.0, after 24 hours of incubation at 30°C and 200 rpm. The clone exhibiting the highest activity was chosen for the production of the recombinant enzyme and further characterization studies. The cultures were kept in a shaking incubator at 30°C for 6 days (200 rpm) with the addition of 0.75 ml methanol once a day to maintain induction (0.5% v/v).

Cultivation of recombinant strain expressing the cellobiohydrolase in *high cell-density fermentation* was performed in the basal salts medium (BSM) supplemented with PTM_1 trace salts, as described in the *Pichia* fermentation guidelines provided by Invitrogen (Invitrogen, *Pichia* Fermentation Process Guidelines) and in **§ 7.1.2.** Batch growth on glycerol was used in the first step in order to provide biomass (70,5 g/l wet cell biomass), while product formation was prevented due to repression of AOX1. The second stage was fed-batch growth on glycerol, followed by a transition phase when glycerol was fed together with small amounts of methanol, causing a slight increase of the enzyme alcohol oxidase due to the derepression of the AOX1 (Jahic *et al.*, 2002). The third stage started with fed-batch growth on methanol, in which the AOX1 was strongly induced, and was maintained for the rest of the cultivation, in order to control the oxygen demand at the high cell density.

Analysis of the culture supernatant, after growth in glycerol medium, showed absence of recombinant enzyme (based on Avicel 5%w/v activity assay and SDS-

PAGE). The end of glycerol batch was indicated by a sharp increase in the dissolved oxygen (DO) tension. This stage was followed by a 5-hour step of fed-batch glycerol one; during this step 50% w/v glycerol, with PTM_1 salts was fed at an initial flow-rate of 12 mL/h/lt of culture medium and was reduced gradually until it was fully consumed. At the same time, temperature was reduced from to $28^{\circ}C$ to $25^{\circ}C$ and finally to $23^{\circ}C$ and 2 mL of methanol were added manually in small aliquots with syringe. Total consumption of glycerol was again indicated by a spike in the DO. At the end of this stage the dry cell mass of the cells reached the amount of 44.93 g/lt. The level of enzyme expression increased with fermentation time and maximum level obtained was 0.65 U/mL (activity against Avicel for varying time points shown at **Figure 7.9**). As methanol was used as carbon source, there was an increase in cell-density during the fed batch phase (**Figure 7.9**). At the end of the fermentation, the dry weight of cells reached 98,63 g/lt of culture medium. The amount of extracellular protein produced reached 0,98 g.
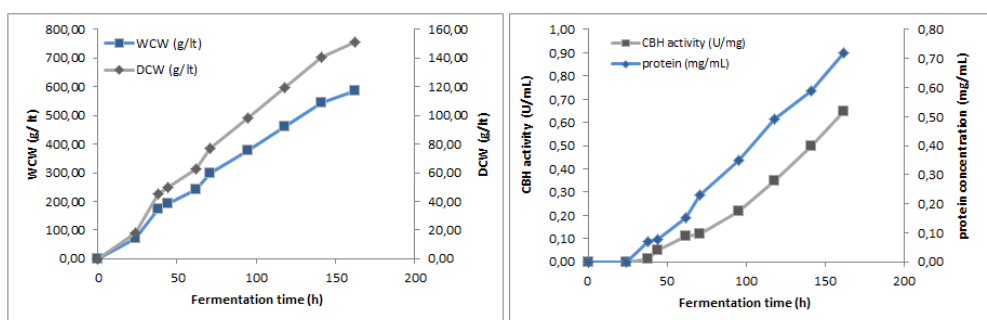

Figure 7.9. (**A**) Cell mass concentration during *Mt*CBH6 fermentation. Dry cell weight reached 44.93 g/lt after glycerol fed-batch phase and 151 g/lt at the end of the cultivation. (**B**) Protein concentration and cellobiohydrolase activity detected in the culture medium during fermentation. Protein concentration was determined using the BCA protein assay microplate procedure (Pierce Chemical Co., Rockford, IL), according to the manufacturer's recommendations. Specific activity was tested against Avicel 5% (w/v), pH 5.0, $60^{\circ}C$ in 100mM phosphate-citrate buffer.

For the purification of the recombinant enzyme, the culture broth was centrifuged and concentrated 30-fold using a LabScale Tangential flow filtration system (TFF with membrane Pellicon XL Ultrafiltration Module Biomax, exclusion

size 10 kDa; Millipore, Billerica, USA). The concentrate was dialyzed overnight at 4ºC against a 20 mM Tris-HCl buffer containing 300 mM NaCl (pH 8.0) and *Mt*EG5 was rapidly purified by single-step immobilized metal ion affinity chromatography (IMAC), with a cobalt charged resin on an ÄKTA Prime Plus system, using 0-100 mM imidazole gradient, at a flow rate of 2 ml/min (**Figure 7.11**). Fractions (2 ml) containing cellobiohydrolase activity were concentrated and the homogeneity was checked by SDS-PAGE (**Figure 7.11**). *Mt*CBH6 was further polished using S300 Gel Filtration chromatography column to remove trace cellobiohydrolase contaminants. Removal of background impurities from the fermentation broth resulted in 564 mg of pure *Mt*CBH6 per lt of culture supernatant. The molecular weight was estimated to be ca. 75 kDa (**Figure 7.10**), which appears to be higher than the predicted value using the ProtParam tool of ExPASY (49.41 kDa) considering the presence of the myc epitope and the polyhistidine tag which contribute 2.8 kDa to the size of *Mt*CBH6. The nominal mass discrepancy observed might be explained by the existence of *Asn-Xaa-Ser/Thr* sequons and *Ser-Thr* residues, which are known to be a prerequisite for *N*- and *O*-glycosylation post-translational modifications respectively. Indeed, 1 *N*-glycosylation and 35 potential *O*-glycosylation sites were predicted by using the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) and the NetOGlyc 3.1 server (http://www.cbs.dtu.dk/services/NetOGlyc/).
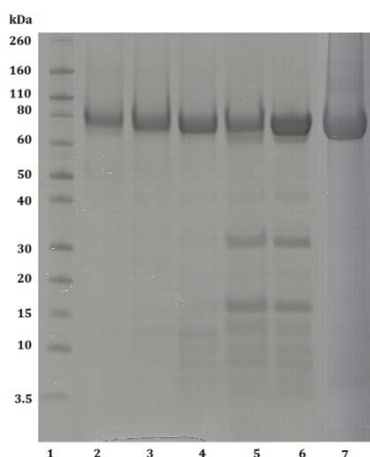
Figure 7.10. SDS – PAGE of *Mt*CBH6 during fermentation. *Lane 1*: Novex® sharp pre-stained protein marker, *lanes 2-6*: samples from fermentation culture broth at 71, 95, 118, 141 and 162 h of hydrolysis, *lane 7:* purified *Mt*CBH6 after IMAC.
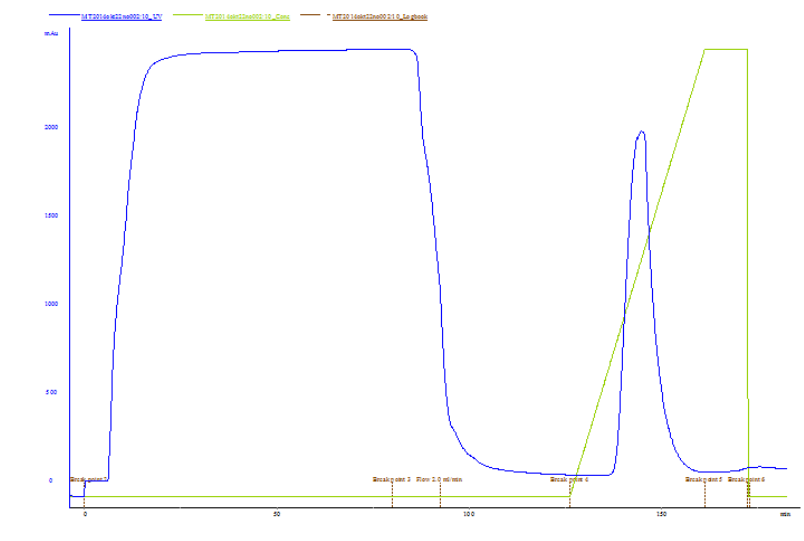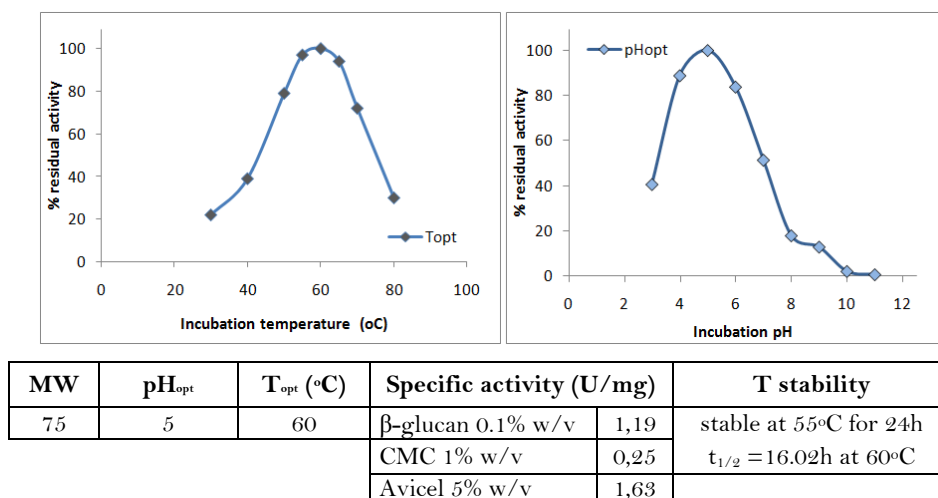
Figure 7.11. Immobilized metal ion affinity chromatography for the purification of the recombinant *Mt*CBH6. The left peak corresponds to proteins that were not bound to the resin and the right one to the *His-tagged* cellobiohydrolase that was eluted using a 0-100 mM gradient imidazole. Elution started at 31 mM imidazole, total protein eluted was 348 mg.

7.2.3. Characterization of purified MtCBH6

The optimal temperature was determined using Avicel 5% (w/v) as a substrate, as described above, at temperatures ranging from 30 to 90°C in 0.1 M citrate-phosphate buffer pH 5.0. Temperature stability was determined by measuring the residual activity under the same assay procedure, after incubation of 0.48 mg of purified *Mt*CBH6 at various temperatures for different amount of time. The optimum temperature activity was observed at 60 °C, losing rapidly its activity for temperatures over 70 °C. The cellobiohydrolase remained fairly stable up to 55 °C after preincubation for 24 hours in 100 mM phosphate-citrate buffer (pH 5.0) at different temperatures (**Figure 7.12**) and exhibited half-life of 16.02 h at 60°C. The optimal temperature was determined using the Avicel assay procedure at temperature ranging from 30 to 80°C in 100 mM citrate–phosphate buffer pH 5.0. Temperature stability

was determined by measuring the residual activity under the standard assay procedure, after incubation of 0.56 mg of purified *Mt*CBH6 at various temperatures for different amount of time. The optimal pH was determined by the standard assay at 60°C over the pH range 3.0–11.0 using either 0.1 M citrate–phosphate buffer pH 3.0–7.0, 0.1 M Tris-HCl pH 7–9 or 0.1 M glycine–NaOH buffer pH 9–11. The stability at different pH was determined after incubating the enzyme in the above buffers at 4°C for 24 h and then measuring the activity remaining using the Avicel assay. The enzyme presented the highest activity levels at pH 5.0, while the activity drops rapidly for pH less than 4 or higher than 6. The enzyme was found remarkably stable in the pH range 3-11 after 24 h retaining its initial activity. Properties of *Mt*CBH6 are described in **Figure 7.12**.

MW	pH$_{opt}$	T$_{opt}$ (°C)	Specific activity (U/mg)		T stability
75	5	60	β-glucan 0.1% w/v	1,19	stable at 55°C for 24h
			CMC 1% w/v	0,25	t$_{1/2}$ =16.02h at 60°C
			Avicel 5% w/v	1,63	

Figure 7.12. Properties of purified *Mt*CBH6 cellobiohydrolase.

7.2.4 Conclusions - Discussion

Cellobiohydrolases are chacterized as CBH I (GH7 family) and CBH II (GH6 family) and act on cellulose molecules from reducing and non-reducing ends, respectively, as described in **Chapter 2**. These enzymes can achieve complete, although slow, solubilization of cellulose crystals even without help of endoglucanases (Teeri, 1997). In the presence of endoglucanases, the rate of hydrolysis of crystalline cellulose by CBHs increases drastically because of an endo–exo synergy between two classes of the enzymes (Wood and McCrae, 1976, Henrissat *et al.*, 1985). An exo–exo synergism between two types of cellobiohydrolases has been reported too (Medve *et al.*, 1994). The enzymes from each family typically share a high degree of identity in their amino

acid sequence and display a common fold. *Family 6* enzymes perform catalysis with inversion of anomeric configuration of the substrate, whereas *family 7* enzymes operate with retention of configuration (Schulein, 2000). Fungal cellobiohydrolases usually have molecular structure typical for most cellulases, i.e. they have a core catalytic module and a cellulose-binding module (CBM) connected by a flexible peptide linker (Van Tilbeurgh *et al.*, 1986; Gilkes *et al.*, 1991) and are often glycosylated, involving both *O*-linked and N-linked carbohydrate chains (Maras *et al.*, 1997; Hui *et al.*, 2001). *Processivity* is a feature common to many cellobiohydrolases and is thought to be a critical strategy for improving the catalytic efficiency for hydrolysis of crystalline substrates (Teeri, 1997). Structure analyses have revealed that CBHs such as CBHI and CBHII from *Trichoderma reesei* have enclosed active site tunnels for substrate binding and catalysis (Kurasin *et al.*, 2011). A single glucan chain enters the tunnel from one end, and disaccharides are cleaved off at the catalytic center during its passage.

The cellulolytic system of *M. thermophila* consists of a repertoire of enzymes with cellobiohydrolase (CBH) activity. Totally, four CBHs have been isolated from crude supernatant and studied, as described in **Chapter 4**. MtCBH6 is the product of MYCTH_66729 gene that represents an enzyme of GH6 family, which is attached to polysaccharide substrate through a CBM and in the same study, was cloned, heterologously expressed in *Pichia pastoris* and partially characterized. The relatively high thermostability (stable for 24h at 55°C) that the enzyme exhibited is indicative of the potential biotechnological use of the enzyme and can be attributed partially to the mannose units added by the post-translational system of *P. pastoris* that tends to hyperglycosylate the secreted proteins, as described in **Chapter 2.**

7.3. Cloning, expression and characterization of Cellobiohydrolase *Mt*CBH7

7.3.1. Identification and cloning of MtCBH7

From genome analysis, as described in **Chapter 4**, the translation of *cbh6* open reading frame (ORF) (Model ID 109566) from the *M. thermophila* genome database shows significant primary sequence identity with characterized cellobiohydrolases acting on the reducing end of the carbohydrate molecules, which have been classified to family GH7 on CAZy database (http://www.cazy.org/; Cantarel *et al.* 2009). The putative cellobiohydrolase showed high sequence identity (67%) with structure identified Cel7d (Cbh58) *Phanerochaete chrysosporium* [PDB: 1GPI] and 61% identity with CBH I from *Humicola grisea* [GenBank: BAA09785.1]. The hypothetical protein of 109566 was selected as a candidate cellobiohydrolase and the corresponding gene, which was provisionally named *cbh7*, was cloned and used to transform *P. pastoris* X33; the encoded enzyme named *MtCBH7* was expressed and finally characterized (**Table 7.5**). The ORF of *cbh7* encodes a protein of 509 amino acids including a secretion signal peptide of 17 amino acids (MYAKFATLAALVAGAAA) based upon the prediction using SignalP v4.0 (http://www.cbs.dtu.dk/services/SignalP/). The predicted mass and isoelectric point (pI) of the mature protein was 54 kDa and pH 4.77, respectively, by calculations using the ProtParam tool of ExPASY (http://web.expasy.org/protparam/).

Genome Portal ID	109566
Chromosome	1:9753507-9755507
Family	Glycoside hydrolase 7
Domains	CBM_1, [Pfam: PR00734, InterProScan]
Gene (translation)	1581 bp
Gene (trancription) [3'UTP, 5'UTP]	1934 bp
Protein	509 aa
Exons	2
Secretion signal	MYAKFATLAALVAGAAA (17 aa)
Theoretical predicted MW	54 kDa
theoretical pI	4.77
Glucosylation sites N-Glyc	1
Glucosylation sites O-Glyc	20

Table 7.5. Properties of *Mt*CBH7 obtained from genome analysis.

For the cloning of the cellobiohydrolase gene from *M. thermophila, Escherichia coli* One Shot® Top10 (Invitrogen, USA) and Zero Blunt® PCR Cloning Kit (Invitrogen, USA) were used as the host-vector system. *P. pastoris* host strain X-33 and pPICZαC (Invitrogen, USA) were used for protein expression. The WT strain of *M. thermophila* ATCC 42464 was maintained on 1.5% malt-peptone-agar slants at 4 °C. *P. pastoris* was routinely grown in shaking flasks at 30 °C according to the instructions in the EasySelect™ *Pichia Expression Kit* (Invitrogen, USA). Genomic DNA was prepared and isolated as previously described (Topakas *et al.*, 2012).

	primer sequence	Tm (°C)	% GC
StCBH7109566F (30 bp)	*ClaI* 5' - GC ATCG AT G CAG AAC GCC TGC ACT CTG ACC - 3'	72.2	60
StCBH109566R (32 bp)	5' - CG T CTA GA A GGC ACT GCG AGT ACC AGT CAT TC -3' *XbaI*	70.8	56.3

	primer sequence (OE-PCR)	Tm (°C)	% GC
StCBH109566e2F (21 bp)	5' - TGT TCC AGC TCC TCG GCA ACG - 3'	73.8	61.9
StCBH109566e1R (38 bp)	5' - CGA GGA GCT GGA ACA TCT GGT ACT GGG TGT CGC TCT CC - 3'	83.6	57.9

| PCR	Primers	Fragment targeted	Conditions
#1	EF – ER	2001 bp	annealing: 60°C/10s extension: 70°C/33s, 30 cycles
#2	EF – Ee1R	409 bp	annealing: 60°C/10s extension: 70°C/4s, 30 cycles
#3	Ee2F – ER	1172 bp	annealing: 60°C/10s extension: 56°C/12s, 30 cycles
#5	EF – ER	1581 bp	annealing: 59°C/10s extension: 59°C/25s, 45 cycles

Table 7.6. Primers and conditions used for the amplification of *cbh7* gene through PCR (**#1**), removal of introns (**#2-3**) and final overlapping PCR (**#5**). Colored sequences represent the restriction sites of *ClaI* and *XbaI* enzymes (red, purple).

An *E. coli/ P. pastoris* vector, pPICZαC, was used to achieve secreted expression of *Mt*CBH7. pPICZαC contains the tightly regulated *AOX1* promoter and the *Saccharomyces cerevisiae* α-factor secretion signal located immediately upstream of the

multiple cloning site (Higgins *et al.*, 1998). The gene coding for the hypothetical protein *Mt*CBH7 (Model ID 109566, chromosome 1:9753507-9755507) was PCR amplified from genomic DNA using primers **EF/ER** (**Table 7.6**) designed accordingly to the available gene sequence (http://genome.jgi-psf.org/, DOE Joint Genome Institute, (Berka *et al.*, 2011) including the *ClaI* and *XbaI* restriction enzyme sites at their respective 5'-ends. A high fidelity KOD Hot Start® DNA polymerase producing blunt ends was used for the DNA amplification, which was carried out with 30 cycles of denaturation (20 s at 95°C), annealing (10 s at 60°C), and extension (33 s at 70°C), followed by 1 min of further extension at 70 °C. In order to determine the DNA sequence, the PCR product was cloned into the pCRBlunt® vector according to the method described by the Zero Blunt® PCR Cloning Kit.

Intron removal was achieved using the molecular technique of overlap extension polymerase chain reaction (OEPCR) (Topakas *et al.*, 2012) using the polymerase KOD Hot Start® (Novagen, USA). Two complementary DNA primers per intron, two external primers (**EF/Ee1R, Ee2F/ER**, **Table 7.6**) and the appropriate PCR amplification process were used to generate two DNA fragments harbouring overlapping ends. The recombinant plasmid pCRBlunt/*cbh7*, at an appropriate dilution, was used as template DNA and the PCR conditions for each reaction are given as the following: 95 °C for 2 min, ensued by 30 cycles of 95 °C for 20 s, 60 °C for 10 s and 70 °C for 4 s (fragment 409 bp) or 12 s (fragment 1172 bp) respectively, with a final extension step at 70 °C for 1 min. The two PCR products were combined together in a subsequent hybridization reaction. The generated "fusion" fragment was amplified further by overlapping PCR through the utilization of the two external primers, EF end ER, with an initial denaturation step at 95 °C for 2 min, followed by 45 cycles at 95 °C for 20 s, 59 °C for 10 s, 70 °C for 25 s and a final extension step at 70 °C for 1 min. An extended annealing was performed (25 min) in order to improve base-pairing between the complementary ends of each fragments that have to be fused. The produced *cbh7* DNA was digested with the enzymes *ClaI* and *XbaI* and the DNA fragment gel-purified before cloning into the pPICZαC vector, resulting in the recombinant pPICZαC/*cbh7* which was amplified in *E. coli* TOP10F0, and the transformants were selected by scoring for Zeocin™ resistance (25 μg/ml). The recombinant vector pPICZαC/*cbh7* was confirmed by restriction analysis and DNA

sequencing and finally transformed into *P. pastoris*. The recombinant plasmid pPICZαC/*cbh7* was linearized with *PmeI* and then transformation of *P. pastoris* and cultivation in shaken flasks were performed according to the EasySelect™ *Pichia* Expression Kit. High-level expression transformants were screened from the YPDS plates containing Zeocin™ at a final concentration of 100 μg/ml. The presence of the *cbh7* gene in the transformants was confirmed by PCR using yeast genomic DNA as template and gene specific primers (**EF and ER; Table 7.6**).
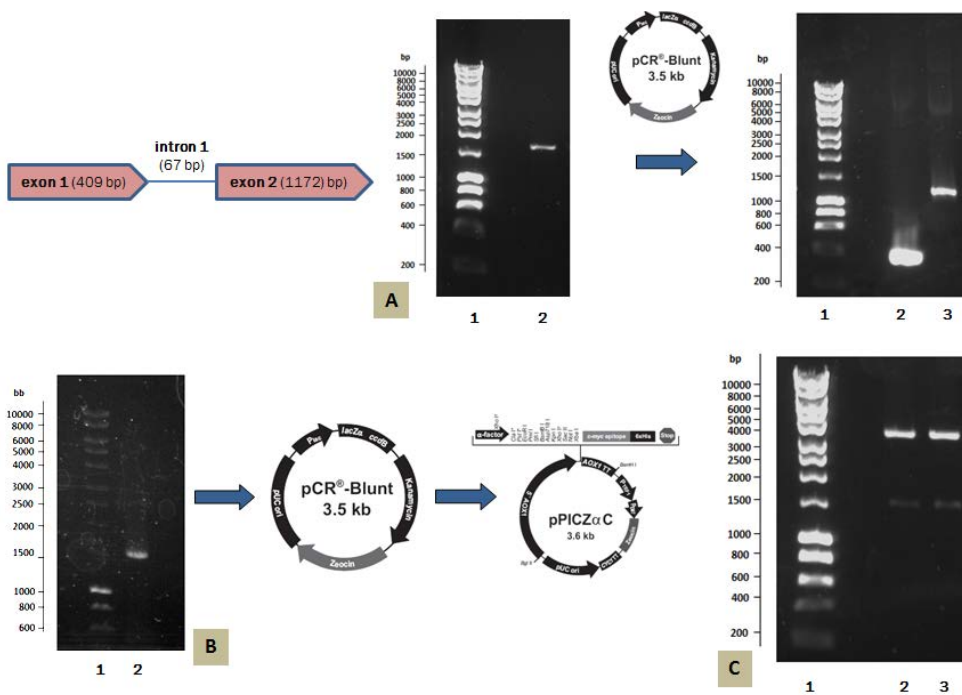
Figure 7.12. Amplification of *cbh7* gene through PCR. *Intron 1* and *2* removal (A) and final OE-PCR (B), cloning of the PCR product into the PCR® Blunt vector/amplification in *E. coli* TOP10 cells and final cloning to pPICZαC vector/amplification in *E. coli* TOP10F' cells. (**A**) *Lane 1*: Hyperladder Marker (Bioline), *Lane 2*: gene *cbh7 (*left) and *Lanes 2-3*: exons 1,2 (right), (**B**) *Lane 1*: final OE-PCR product, *Lane 2*: Hyperladder Marker (Bioline), (**C**) *Lane 1*: Hyperladder Marker (Bioline), *Lanes 2-3*: Digestion of pPICZαC/*cbh7* with *ClaI/XbaI*.

Protein expression of the recombinant enzyme was first evaluated in small scale cultures, in shake flasks. For the production of cellobiohydrolase, one single *P. pastoris* colony harboring *cbh7* gene was cultivated in BMGY medium for 18-24 hours, at 30°C in a shaker (200 rpm) and then inoculated into the production medium BMMY reaching $OD_{600}=1$. The extracellular secreted protein was tested for celobiohydrolase activity against Avicel 5%w/v n 100mM phosphate-citrate buffer pH=5.0, after 24 hours of incubation at 30°C and 200 rpm. The clone exhibiting the highest activity was chosen for the production of the recombinant enzyme and further characterization studies. The cultures were kept in a shaking incubator at 30ºC for 6 days (200 rpm) with the addition of 0.75 ml methanol once a day to maintain induction (0.5% v/v). After examination of cellobiohydrolase activity, no efficient yield of recombinant protein has been achieved. The major factor causing this problem was primarily the *proteolytic degradation* of enzymes produced, which hampered the yield. Proteolysis led to low recombinant protein levels and active products that were smaller than the full-length protein. Degraded proteins ran as a "smear" after SDS-PAGE (**Figure 7.13**). As a result, in spite of high protein amounts measured at the culture medium, only a small proportion was biologically active.

In order to achieve eliminate proteolysis and achieve higher production levels of homogenous and stable enzyme, several strategies were followed and a series of different parameters were evaluated, such as the reduction of incubation temperature, the influence of initial pH, ammonium sulfate and methanol concentration and agitation. Of all the above parameters, it was found that when the initial concentration of ammonium sulfate in the culture medium was 2-fold higher than the one usually used (10% w/v, as suggested by EasySelect™ *Pichia* Expression Kit protocol), the protein appeared full length sized and homogenous (**Figure 7.13**).

Addition of ammonium ions has also been recommended, in the form of ammonium sulfate. Tsujikawa and colleagues (1996), observed 10 fold reduction of proteolysis by supplementing medium with ammonium ions. Proteolysis is found to amplify over the induction period when the amount of viable cells in the culture is reduced. *Salting-out* is attractive for separation and purification of proteins in both laboratory and industrial

scales. Examples include the recovery of proteins, i.e. diagnostic enzymes, insulin, human growth hormone, interferon, and food proteins, by salting-out with $(NH_4)_2SO_4$ or Na_2SO_4 (Shih *et al.* 1992, McNay *et al.* 2001). However, there are some data on literature that report recombinant protein production with salting-out effect to protect against proteolytic degradation have been published. Ammonium sulfate had a salting-out effect on extracellular proteins shown by the higher protein purity on SDS-PAGE gels. Kobayashi and Nakamura, 2003 proposed that lower proteolysis can be attributed to the fact that the solubility of the proteases decreases with the addition of $(NH_4)_2SO_4$, so total proteolytic activity in the supernatant liquid decreases by sulfate conjugation. It was found that the addition of 20 g $(NH_4)_2SO_4$ l^{-1} to a culture medium was effective for not only increasing the amount of glucoamylase produced but also for maintaining glucoamylase activity at a high level. País-Chanfrau *et al.*, 2004 reported an increased expression of recombinant mini-proinsulin in *Pichia pastoris* in bioreactors, achieved by, among others, periodical addition of ammonium sulfate.
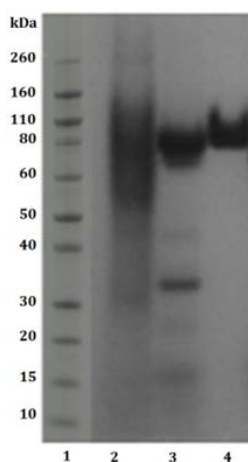
Figure 7.13. SDS-PAGE of cellobiohydrolase MtCBH7. *Lane 1*: Novex® sharp pre-stained protein marker, *Lane 2*: samples from culture medium, where BMMY was used as substrate, with addition of 10% w/v ammonium sulfate. Signs of proteolysis are dominant, as protein exhibit lower molecular weight, run as a "smear" and smaller size molecules appear. *Lanes 3-4:* samples of *Mt*CBH7 from culture medium with addition of 20% w/v ammonium sulfate, after ultrafitration *(3)* and after purification with IMAC *(4)*. The enzyme appears full length sized and homogenous.
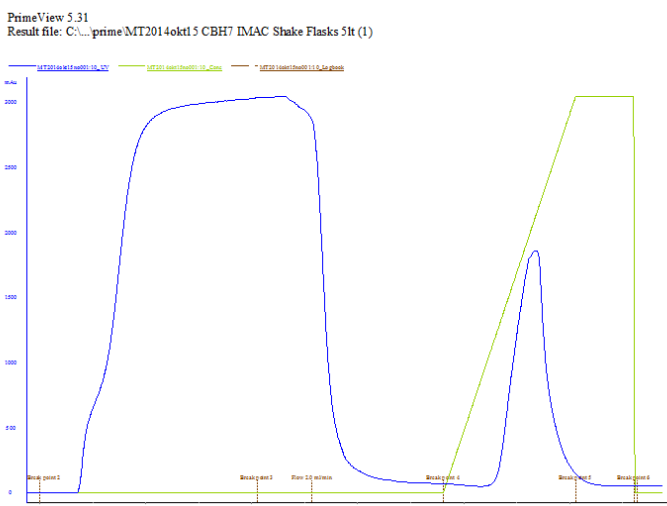
Figure 7.14. Immobilized metal ion affinity chromatography for the purification of the recombinant *Mt*CBH7. The left peak corresponds to proteins that were not bound to the resin and the right one to the *His-tagged* cellobiohydrolase that was eluted using a 0-100 mM gradient imidazole. Elution started at 38 mM imidazole, total protein eluted was 197.8 mg from 5 lt of initial culture.

For the purification of the recombinant enzyme, the culture broth was centrifuged and concentrated 30-fold using a LabScale Tangential flow filtration system (TFF with membrane Pellicon XL Ultrafiltration Module Biomax, exclusion size 10 kDa; Millipore, Billerica, USA). The concentrate was dialyzed overnight at 4°C against a 20 mM Tris-HCl buffer containing 300 mM NaCl (pH 8.0) and *Mt*EG5 was rapidly purified by single-step immobilized metal ion affinity chromatography (IMAC), with a cobalt charged resin on an ÄKTA Prime Plus system, using 0-100 mM imidazole gradient, at a flow rate of 2 ml/min (**Figure 7.14**). Fractions (2 ml) containing cellobiohydrolase activity were concentrated and the homogeneity was checked by SDS-PAGE (**Figure 7.13**). Removal of background impurities from the fermentation broth resulted in 39.5 mg of pure *Mt*CBH7 per lt of culture supernatant. The molecular weight was estimated to be ca. 78 kDa (**Figure 7.14**), which appears to be higher than the predicted value using the ProtParam tool of ExPASY (54 kDa) considering the presence of the myc epitope and the polyhistidine tag which contribute

2.8 kDa to the size of *Mt*CBH6. The nominal mass discrepancy observed might be explained by the existence of *Asn-Xaa-Ser/Thr* sequons and *Ser-Thr* residues, which are known to be a prerequisite for *N-* and *O-*glycosylation post-translational modifications respectively. Indeed, 1 *N*-glycosylation and 20 potential *O*-glycosylation sites were predicted by using the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) and the NetOGlyc 3.1 server (http://www.cbs.dtu.dk/services/NetOGlyc/).

7.3.3. Characterization of purified MtCBH7

The optimal temperature was determined using Avicel 5% (w/v) as a substrate, as described above, at temperatures ranging from 30 to 90°C in 0.1 M citrate-phosphate buffer pH 5.0. Temperature stability was determined by measuring the residual activity under the same assay procedure, after incubation of 0.52 mg of purified *Mt*CBH7 at various temperatures for different amount of time. The optimum temperature activity was observed at 60 °C, losing rapidly its activity for temperatures over 70°C. The cellobiohydrolase remained fairly stable up to 50 °C after preincubation for 24 hours in 100 mM phosphate-citrate buffer (pH 5.0) at different temperatures (**Figure 7.15**) and exhibited half-life of 18.1 h at 55°C and 9.41 h at 60°C, respectively. The optimal temperature was determined using the Avicel assay procedure at temperature ranging from 30 to 80°C in 100 mM citrate–phosphate buffer pH 5.0. Temperature stability was determined by measuring the residual activity under the standard assay procedure, after incubation of 0.55 mg of purified *Mt*CBH7 at various temperatures for different amount of time. The optimal pH was determined by the standard assay at 60°C over the pH range 3.0–11.0 using either 0.1 M citrate–phosphate buffer pH 3.0–7.0, 0.1 M Tris-HCl pH 7–9 or 0.1 M glycine–NaOH buffer pH 9–11. The stability at different pH was determined after incubating the enzyme in the above buffers at 4°C for 24 h and then measuring the activity remaining using the Avicel assay. The enzyme presented the highest activity levels at pH 5.0, while the activity drops rapidly for pH less than 4 or higher than 6. The enzyme was found remarkably stable in the pH range 3-11 after 24 h retaining its initial activity. Properties of *Mt*CBH7 are described in **Figure 7.15**.
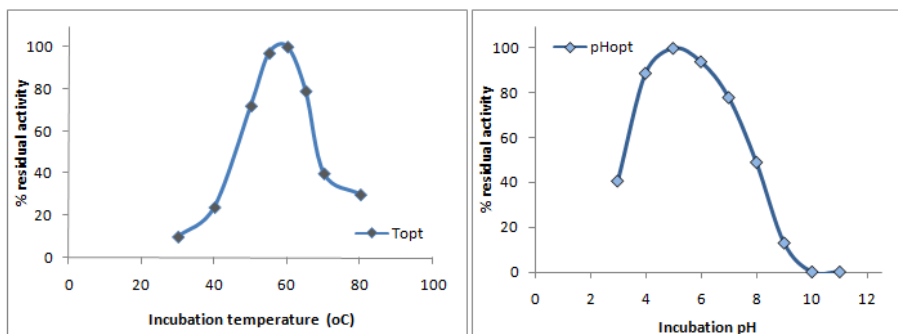
MW	pH$_{opt}$	T$_{opt}$ (°C)	Specific activity (U/mg)		T stability
78	5	60	pNP-β-cellobioside (5mM)	0,13	stable at 50°C for 24h
			pNP-β-lactopyranoside (5 mM)	2,25	t$_{1/2}$ =18.1h at 55°C t$_{1/2}$ =9.41h at 60°C
			CMC 1% w/v	0,2	
			Avicel 5% w/v	2,8	

Figure 7.15. Properties of purified *Mt*CBH7 cellobiohydrolase.

7.3.4. Conclusions - Discussion

Proteolytic degradation has been a perpetual problem when yeasts are employed to express recombinant proteins in *P. pastoris* (Gimenez *et al.*, 2000). Yeast vacuoles contain various proteases whose levels vary according to the nutritional conditions (van den Hazel *et al.*, 1996). Though several strategies have been followed to limit proteolytic degradation of the recombinant protein, no in-depth analysis on the conditions that promote proteolysis or the nature of the proteases acting on the desired protein is exactly known. In this study, it is reported the successful expression of cellobiohydrolase *Mt*CBH7 in *P. pastoris*, under high osmotic pressure and increased salinity conditions. 2-fold increase of the ammonium sulfate in the culture medium resulted in the production of full length product. *Mt*CBH7 is a process exo-acting, cellobiose-releasing enzyme of GH7 family, acting on the reducing end of the cellulosic substrate. As other retaining carbohydrases, family 7 cellobiohydrolases may catalyse transglycosylation (Gusakov *et al.*, 1991). *Mt*CBH7 is stable at 50°C after 24h incubation and exhibits activity on Avicel and pNP-substituted substrates, as described above. Though CBHs have a key role to the conversion of biomass to fermentable sugars, they are refered to be notoriously slow and susceptible to inhibition. Xylan and xylan-fragments have been suggested repeatedly as one cause of the reduced activity of

CBHs, together with product inhibition by cellobiose (Selig *et al*, 2008; Bauman *et al*., 2011). The addition of enzymes with xylanolytic activity will increase the yield of soluble sugars and provide remedies for xylan inhibition of CBHs' activity.

7.4. Production of *Mt*GH61 in fermentor

The recently discovered family of AA9 lytic polysaccharide monooxygenases (LPMOs) includes metallo-enzymes enzymes that have been shown to enhance the hydrolytic potential of a cellulase mixture during the enzymatic hydrolysis of lignocellulosic substrates (Hu *et al*., 2013). Unlike the canonical cellulase enzymes which have been shown to cleave cellulose by a hydrolytic mechanism involving the conserved carboxylic acid residues within either channel or cleft shape substrate loading sites, AA9 are thought to cleave cellulose chain by an oxidative mechanism at the protein's planar active site which contains a divalent metal ion (Aachman *et al*., 2012; Li *et al*., 2012; Quinlan *et al*., 2011). By using electron paramagnetic resonance spectroscopy and single-crystal X-ray diffraction, the active site of GH61 is revealed to contain a *type II copper* and, uniquely, a *methylated histidine* in the copper's coordination sphere, thus providing an innovative paradigm in bioinorganic enzymatic catalysis (Quinlan *et al*., 2011). Although a final model mechanism of action has not yet been found, some common features of these enzymes include their need of a reductant cofactor that works as an external electron donor to enhance their activity and the oxidation at the C1 carbon in the glucose ring structure as the most represented (Forsberg *et al*., 2011; Quinlan *et al*., 2011). Cannella *et al*., 2012 showed that under commercially relevant conditions, around 4.1% of the glycosidic bonds in cellulose were oxidatively cleaved by presumably GH61 enzymes, which provides new entry sites for the hydrolytic enzymes and increase the access to the substrate.

*Mt*GH61 represents an enzyme belonging to the AA9 family which has been functionally expressed in *Pichia pastoris* under the transcriptional control of the alcohol oxidase (AOX1) promoter and characterized (Dimarogona *et al*., 2012). In this study, the enzyme was produced in fermentor and subsequently used in hydrolysis experiments described in **Chapter 8**. Cultivation of recombinant strain expressing the *Mt*GH61 in *high cell-density fermentation* was performed in the basal salts medium (BSM), supplemented by trace element solution PTM$_1$, as described in the *Pichia*

fermentation guidelines provided by Invitrogen (Invitrogen, *Pichia* Fermentation Process Guidelines). The PTM_1 trace salt solution was also included in the glycerol- and methanol feeds during glycerol and methanol fed-batch phases. The only nitrogen source was ammonium hydroxide which was added as the pH was regulated. Cultivation started at 28°C, aeration was set at 4 vvm and agitation at 800 rpm.

After 24 hours of batch fermentation in glycerol medium (30 g/lt initial concentration), the end of glycerol batch was indicated by a sharp increase in the dissolved oxygen (DO) tension. This stage was followed by a 5-hour step of fed-batch glycerol one; during this step 50% w/v glycerol, with PTM_1 salts was fed at an initial flow-rate of 12 mL/h/lt of culture medium and was reduced gradually until it was fully consumed. At the same time, temperature was reduced from to 28°C to 25°C and finally to 23°C and 2 mL of methanol were added manually in small aliquots with syringe. Total consumption of glycerol was again indicated by a spike in the DO. At the onset of methanol fed-batch phase, casamino acids solution was added at a final concentration of 3 g/lt and then, a feed of 100% MeOH, with PTM_1 was initiated at a flow rate of 1.9 mL/h/lt. The methanol consumption rate was monitored indirectly by stopping the feed and checking the "*lag phase*", while increasing the methanol feed rate manually. After 8h, feed rate was adjusted to a maximum of 5,46 mL/h/lt and maintained for ~20h, causing extracellular expression of the recombinant enzyme into the supernatant.. Then, the temperature was decreased to 21°C and pure oxygen supply was set to maintain dissolved oxygen levels between 60-30 %. Induction time lasted 160 in total and approximately 700 mL of methanol were consumed. At the end of the fermentation, the dry weight of cells reached 117,5 g/lt of culture medium. The amount of extracellular protein produced reached 1,27 g.

For the purification of *Mt*GH61, the culture broth was centrifuged and concentrated 30-fold using a LabScale Tangential flow filtration system (TFF with membrane Pellicon XL Ultrafiltration Module Biomax, exclusion size 10 kDa; Millipore, Billerica, USA). The concentrate was dialyzed overnight at 4°C against a 20 mM Tris-HCl buffer containing 300 mM NaCl (pH 8.0) and then the enzyme was rapidly purified by single-step immobilized metal ion affinity chromatography (IMAC), with a cobalt charged resin on an ÄKTA Prime Plus system, using 0-100 mM imidazole gradient, at a flow rate of 2 ml/min (**Figure 7.16**). The eluent, containing

*Mt*GH61 with an estimated molecular weight of ~58 kDa (**Figure 7.17**), was concentrated and later was used in hydrolysis experiments.
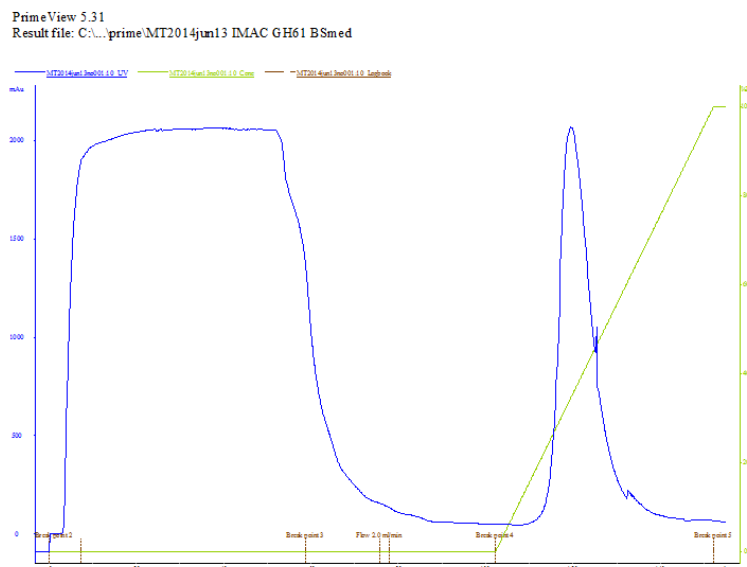
Figure 7.16. IMAC chromatography for the purification of the recombinant *Mt*GH61. The left peak corresponds to proteins that were not bound to the resin and the right one to the *His-tagged* enzyme that was eluted using a 0-100 mM gradient imidazole. Elution started at 26 mM imidazole, total protein eluted was 715,8 mg.
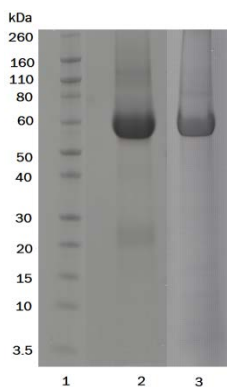

Figure 7.17. SDS – PAGE of *Mt*GH61 samples after fermentation. *Lane 1*: Novex® sharp pre-stained protein marker, *lane 2*: sample from fermentation culture broth after ultrafiltration, *lane 3*: purified protein after IMAC.

7.5. Production and purification of enzymes with xylanase activity

7.5.1. Purification of Xylanases from M. thermophila grown on corn cob

 M. thermophila produces high levels of xylanases, when grown on a cheap carbon source such as corn cob. In this study, partially purified crude enzyme mixture from *M. thermophila* ATCC 34628 strain, was kindly offered from Assoc. Prof. Katapodis P., University of Ioannina, Department of Biological Applications and Technologies, Greece and was used as a source of proteins exhibiting xylanase activity. This mixture had been originally purified from submerged cultures supernatant containing corb cod as carbon source (Katapodis *et al.*, 2003; **Figure 7.18**). Xylan represents more than 60% of the polysaccharides existing in the cell wall of corn cob, thus inducing the expression of xylanases in fungus' secretome. The sample, exhibiting initial xylanase specific activity of 67.31 U/mg against birchwood xylan, was loaded onto a S-300 16/60 Sephacryl gel filtration column and fractions were eluted at a 0.4 mL/min flow rate using 100mM phosphate-citrate buffer pH=5.0. Four major fractions were eluted and tested for xylanase and cellulose activity against birchwood xylan 1%w/v and carboxy-methyl-cellulose 1%w/v respectively, in 100mM phosphate-citrate buffer pH=5.0 and 15min incubation at 50°C (**Figure 7.19**). One unit of enzyme activity was defined as the amount of enzyme liberating 1 μmol of reducing sugars measured as xylose / glucose equivalent per min. *Fractions III* and *IV*, that showed relatively high xylanase specific activity (**Table 7.7**), but no cellulose activity corresponded to ~ 38 kDa and 45 kDa protein band; enzymes were concentrated to a final volume of 2 mL and used in the hydrolysis experiments of natural substrates described in **Chapter 8**.

Figure 7.18. SDS-PAGE gel. *Lane 1:* Partially purified crude enzyme mixture from *M. thermophila* ATCC 34628 strain grown on corn cob. *Lane 2:* Novex® sharp pre-stained protein marker.
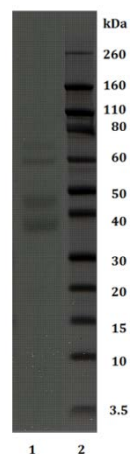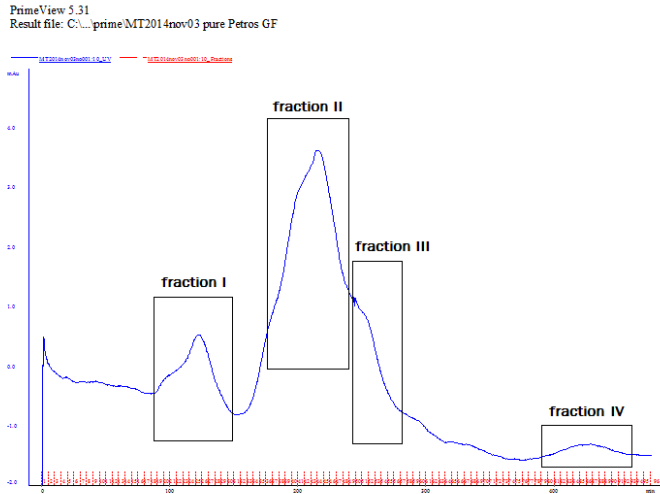
Figure 7.18. Gel filtration on S-300 16/60 Sephacryl column of the partially purified crude enzyme mixture resulted in the elution of four fractions. *Fractions III* and *IV* showed relatively high xylanase specific activity, but no cellulose activity.

7.5.2. Purification of Xylanases from M. thermophila grown on wheat straw

The mineral medium used for the production of xylanolytic enzymes contained per liter:

|	
KH_2PO_4	3.0
K_2HPO_4	2.0
$MgSO_4 \cdot 7H_2O$	0.5
$CaCl_2 \cdot 2H_2O$	0.1
$FeSO_4 \cdot 7H_2O$	0.005
$MnSO_4 \cdot 4H_2O$	0.0016
$ZnSO_4 \cdot 7H_2O$	0.0014
$CoCl_2 \cdot 2H_2O$	0.0002

A 5 mL mycelia and spore suspension of *M. thermophila* ATCC 42464 from a 5-day old culture, grown on PDA Petri dish at 45°C, was inoculated to two 1lt Erlenmeyer flasks each containing 300mL of the above mineral medium supplemented with wheat straw (30g/lt) as carbon source and $(NH_4)_2HPO_4$ (7.5 g/lt) as nitrogen source. *M. thermophila*

ATCC 42464 strain has been found to reach the highest yield of xylanase activity in culture medium when cultured with wheat straw as carbon source (Katapodis *et al.*, 2006). The flasks were incubated at 46°C in an orbital shaker operating at 200 rpm for 2 days. After gentle centrifuge and resuspension in 50 mL sterile ultrapure water, 200mL of the preculture were used to inoculate a 3 lt glass autoclavable Applikon bioreactor, equipped with an *ez*-Control system (Applikon Biotechnology B.V., Netherlands), containing 1 lt mineral medium having the above composition and wheat straw (30g/lt). The fungus was cultured for 5 days at 46°C and pH 5.0. Agitation speed was set at 800 rpm, enabling the oxygen concentration to remain above 60% saturation during the cultivation process.

The fermentation broth was centrifuged at 10,000g for 30 min, at 4°C and then filtrated consecutively through filters with 0.45 μm and 0.2 μm pore diameter. The clear supernatant was assayed for xylanase against 1%w/v birchwood xylan in 100mM phosphate-citrate buffer, pH 5.0 and 15 min incubation at 50°C and concentrated 30-fold using a LabScale Tangential flow filtration system (TFF with membrane Pellicon XL Ultrafiltration Module Biomax, exclusion size 10 kDa; Millipore, Billerica, USA). The retentate, containing endo-xylanase activity, was used for further purification of the enzyme. It was dialyzed overnight at 4 ^{o}C against a 20 mM Tris-HCl buffer (pH 8.0) and further concentrated to 5 mL with *Vivaspin 2*, 10 kDa MWCO PES (GE Healthcare).

Anion exchange chromatography. After equilibration with 20 mM Tris-HCl buffer (pH 8.0), crude enzyme preparation was applied on a HiPrep Q HP 16/10 anion exchange column (GE Healthcare). The column was washed with 80 mL of the equilibrating buffer, following by a linear gradient of 0 - 1 M NaCl in 150 mL of the same buffer. Some indicative eluted fractions (# 36, 40, 44, and 50) as well as the flow-through (FT) fraction that has been separately kept were tested for xylanase activity against birchwood xylan, as described above. The FT and #36 fractions exhibited detectable xylanase activity, so the corresponding fractions (see Picture XX) were concentrated to 2 mL and tested for xylanase and cellulase activity against birchwood xylan and carboxymethyl-cellulase respectively. Protein concentration was determined by the BCA method (Pierce Chemical Co., Rockford, IL), using bovine serum albumin as a standard. Specific activity was calculated (Table XX).
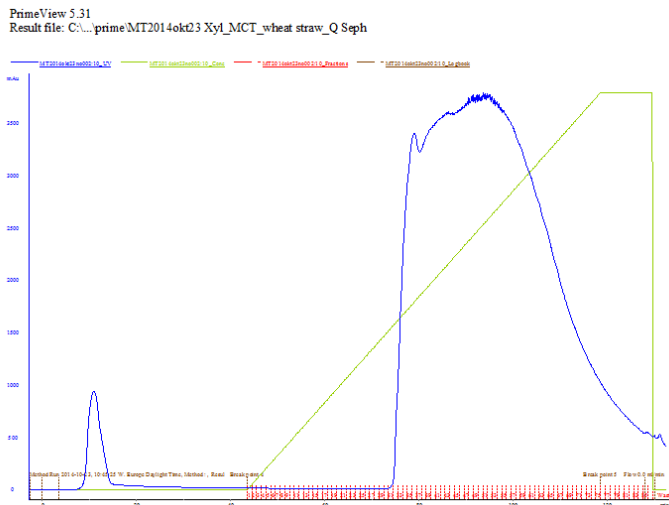
Figure 7.19. Anion exchange chromatography with HiPrep Q HP 16/10 anion exchange column resulted. The fraction that was eluted before gradient of NaCl was started (flow through) contained proteins with xylanolytic activity.
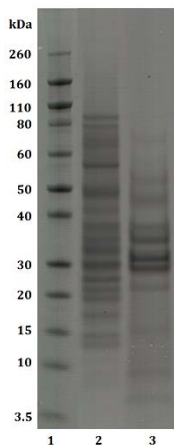

Figure 7.20. SDS-PAGE of samples after anion exchange chromatography, *Lane 1*: Novex® sharp pre-stained protein marker, *Lane 2*: protein fraction that was eluted with linear gradient of 0 – 1 M NaCl, *Lane 3*: protein fraction that was eluted with 20 mM Tris-HCl buffer, before gradient of NaCl was started (flow through). This fraction contains proteins with isoelectric point above 8.0, which exhibited relatively high xylanolytic, but very low cellulolytic activity.

The FT fraction, containing proteins with isoelectric point below 8.0, was applied on a S-300 16/60 Sephacryl gel filtration column pre-equilibrated with 100mM prosphate-citrate buffer and fractions were eluted at a 0.4 mL/min flow rate using 100mM phosphate-citrate buffer pH=5.0. Three fractions were eluted. This purification step yielded one endoxylanase active fraction (*Fraction II*), consisted of two proteins with molecular weight ~30-35 kDa, as checked by SDS-PAGE. This fraction showed relatively high xylanase specific activity, but no cellulose activity. It was concentrated to a final volume of 2 mL and used in the hydrolysis experiments of natural substrates described in Chapter 7.
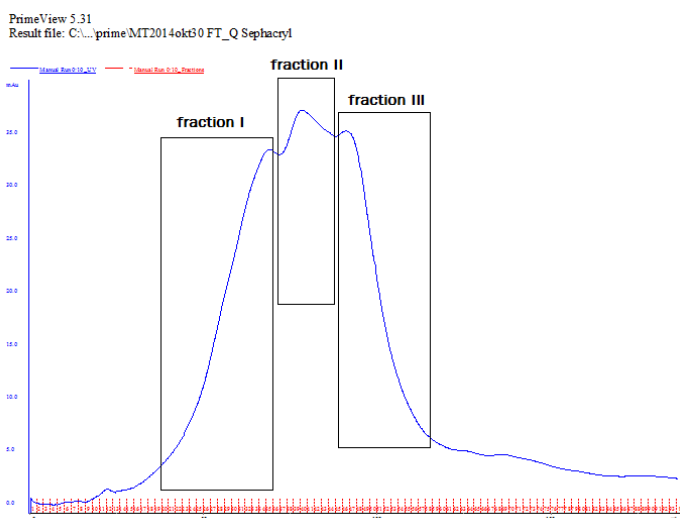
Figure 7.22. Gel filtration of *FT fraction* resulted in three fractions.

Figure 7.23. SDS-PAGE of xylanase fragments after gel filtration column. *Lane 1*: Novex® sharp pre-stained protein marker, *Lanes 1-3*: Fractions I-III. *Fraction II* exhibited relatively high xylanolytic activity and was used in hydrolysis experiments.
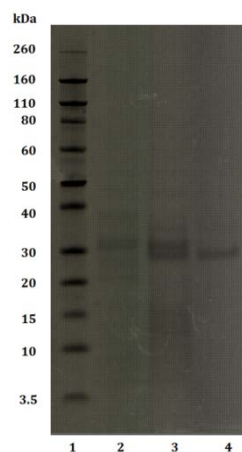
		PROTEIN		XYLANASE ACTIVITY		CELLULASE ACTIVITY	
	Volume (mL)	C (mg/mL)	TOTAL mg	Units/mL	Units/mg	Units/mL	Units/mg
Xyl_MCT_Corn Cob							
crude enzyme	13	2,36	30,73	70,11	**29,66**	1,62	**0,69**
pure enzyme	14	0,16	2,23	10,73	**67,31**	0,58	**3,64**
GF FRACTION III	4	0,03	0,14	0,57	**4,10**	0,03	**0,85**
GF FRACTION IV	2	0,02	0,04	0,14	**3,75**	0,00	**0,00**
Xyl_MCT_Wheat Straw							
Crude	10	35,51	355,06	142,57	**0,40**	267,59	**2,09**
Q-Seph Flow through	4,7	1,86	8,75	135,63	**15,50**	3,25	**0,37**
Q-Seph FRACTION I	7,5	4,03	30,20	46,51	**1,54**	43,17	**1,43**
GF FRACTION I	6,5	0,21	1,35	0,88	**0,66**	0,3035	**1,47**
GF FRACTION II	3	0,35	1,06	5,98	**5,64**	0,1425	**0,40**
GF FRACTION III	10	0,08	0,77	0,67	**0,87**	0,188	**2,44**

Table 7.7. Specific activities of fractions purified from *M. thermophila* culture medium grown on two different carbon sources. Xylanase and cellulose activity were tested against birchwood xylan 1%w/v and carboxy-methyl-cellulose 1%w/v respectively, in 100mM phosphate-citrate buffer pH=5.0 and 15min incubation at 50°C. Colored fractions were used in hydrolysis experiments.

Xylanase mixture that was used at hydrolysis experiments described in **Chapter 8** was consisted of 87% Fraction II, 11% Fraction III and 2% Fraction IV. Had a total activity of 14.82 Units/mg when tested against birchwood xylan 1% w/v and remained stable after 8 hours of incubation at 50°C.

References

Aachmann, F.L., Sørlie, M., Skjåk-Bræk, G., Eijsink, V.G., Vaaje-Kolstad, G. 2012. NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. *Proc Natl Acad Sci USA.* 109, 18779-84.

Baumann, M.J., Borch, K., Westh, P., 2011. Xylan oligosaccharides and cellobiohydrolase I (*Tr*Cel7A) interaction and effect on activity. *Biotechnol Biofuels* 4, 45.

Berka, R.M., Grigoriev, I.V., Otillar, R., Salamov, A., Grimwood, J. *et al.* 2011. Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris. Nature Biotechnology* 29:922-927.

Cannella, D., Hsief, C.C., Felby, C., Jorgensen, H. 2012. Production and effects of aldonic acids during enzymatic hydrolysis of lignocelluloses at high dry matter content. *Biotechnol Biof* 5, 26.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B. 2009. The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37: D233-238.

Chiruvolu, V., Cregg, J.M., Meagher, M.M. 1997. Recombinant protein production in an alcohol oxidase-defective strain of *Pichia pastoris* in fedbatch fermentations. *Enz Microb Technol*, 21, 277-283.

Cohen, R., Suzuki, M.R.,, Hammel, K.E. 2005. Processive endoglucanase active in crystalline cellulose hydrolysis by the brown rot basidiomycete *Gloeophyllum trabeum. Appl. Environ. Microbiol.* 71:2412–2417.

Dimarogona, M., Topakas, E., Olsson, L., Christakopoulos, P. 2012. Lignin boosts the cellulase performance of a GH61 enzyme from *Sporotrichum thermophile. Bioresource Techn* 110, 480-487.

Forsberg, Z., Vaaje-Kolstad, G., Westereng, B., Bunaes, A.C., Stenstrom, Y., MacKenzie, A., Sorlie, M., Horn, S.J., Eijsink, V.J.H. 2011. Cleavage of cellulose by CBM33 protein. *Protein Sci.* 20, 1479–1483.

Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C., Warren, R.A.J. 1991. Domains in microbial _-1,4–glycanases—sequence conservation, function, and enzyme families. *Microbiol Rev 55:303–*15.

Gimenez, J.A., Monkovic, D.D., Dekleva, M.L. 2000 Identification and monitoring of protease activity in recombinant *Saccharomyces cerevisiae. Biotechnol Bioeng.*67, 245-51.

Gusakov, A.V., Protas, O.V., Chernoglazov, V.M., Sinitsyn, A.P., Kovalysheva, G.V., Shpanchenko, O.V. *et al.* 1991. Transglycosylation activity of cellobiohydrolase I from *Trichoderma longibrachiatum* on synthetic and natural substrates. *Biochim Biophys Acta* 1073:481–5.

Henrissat, B., Driguez, H., Viet, C., Schulein, M. 1985. Synergism of cellulose from *Trichoderma reesei* in degradation of cellulose. *Biotechnology* 3:722–6.

Higgins, D.R., Busser, K., Comiskey, J., Whittier, P.S., Purcell, T.J., Hoeffler, J.P. 1998. Small vectors for expression based on dominant drug resistance with direct multicopy selection. *In: Higgins, D.R., Cregg, J.M., (eds) Methods in molecular biology: Pichia protocols*, Humana, Totowa, pp 28-41.

Hu, J., Arantes, V., Pribowo, A., Gourlay, K., Saddler, J. 2014. Substrate factors that influence the synergistic interaction of AA9 and cellulases during the enzymatic hydrolysis of biomass.

Hui, J.P.M., Lanthier, P., White, T.C., McHugh, S.G., Yaguchi, M., Roy, R. *et al.* 2001. Characterization of cellobiohydrolase I (Cel7A) glycoforms from extracts of *Trichoderma reesei* using capillary isoelectric focusing and electrospray mass spectrometry. *J Chromatogr B* 752:349–68.

Jahic, M., Rotticci-Mulder, J.C., Martinelle, M., Hult, K., Enforset, S-O. 2002. Modeling of growth and energy metabolism of *Pichia pastoris* producing a fusion protein. *Bioprocess Biosyst Eng* 24, 385-393.

Katapodis, P., Vrsanská, M., Kekos, D., Nerinckx, W., Biely, P., Claeyssens, M., Macris, B.J., Christakopoulos, P. 2003. Biochemical and catalytic properties of an endoxylanase purified from the culture filtrate of Sporotrichum thermophile. *Carbohydr Res.* 338:1881-90.

Kobayashi, F. and Nakamura, Y. 2003 Efficient production by *Escherichia coli* of recombinant protein using salting-out effect protecting against proteolytic degradation. *Biotechnol. Lett.* 25: 779–782.

Kurasin, M. and Väljamäe, P. 2011. Processivity of cellobiohydrolases is limited by the substrate. *J. Biol. Chem.* 286:169 −177

Li, X., Beeson, W.T., Phillips, C.M., Marletta, M.A., Cate, J.H. 2012. Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure*. 20, 1051-61.

Maras, M., De Bruyn, A., Schraml, J., Herdewijn, P., Claeyssens, M., Fiers, W., *et al.* 1997. Structural characterization of *N*-linked oligosaccharides from cellobiohydrolase I secreted by the filamentous fungus *Trichoderma reesei* RUTC 30. *Eur J Biochem* 245:617−25.

McNay, J.L., O'Connell, J.P., Fernandez, E.J. 2001. Protein unfolding during reversed-phase chromatography: II. Role of salt type and ionic strength. *Biotechnol Bioeng.* 76, 233-40.

Medve, J., Stahlberg, J., Tjerneld, F. 1994. Adsorption and synergism of cellobiohydrolase I and II of *Trichoderma reesei* during hydrolysis of microcrystalline cellulose. *Biotechnol Bioeng* 44:1064−73.

Miller, G.L. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugars. *Analytical Chemistry* 31: 426−428.

País-Chanfrau, J.M., García, Y., Licor, L., Besada, V., Castellanos-Serra, L., Cabello, C.I., Hernández, L., Mansur, M., Plana, L., Hidalgo, A., Támbara, Y., del C Abrahantes-Pérez, M., del Toro, Y., Valdés, J., Martínez, E. 2004. Improving the expression of mini-proinsulin in *Pichia pastoris*, *Biotechnol Lett.* 26, 1269-72.

Quinlan, R.J., Sweeney, M.D., Lo Leggio, L., Otten, H., Poulsen, J.C., Johansen, K.S., Krogh, K.B., Jørgensen, C.I., Tovborg, M., Anthonsen, A., Tryfona, T., Walter, C.P., Dupree, P., Xu, F., Davies, G.J., Walton, P.H. 2011. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci USA.* 108, 15079-84.

Schulein, M. 2000. Protein engineering of cellulases. *Biochim Biophys Acta* 1543:239−52.

Selig, M.J., Knoshaug, E.P., Adney, W.S., Himmel, M.E., Decker, S.R. 2008. Synergistic enhancement of cellobiohydrolase performance on pretreated corn stover by addition of xylanase and esterase activities. *Bioresour Technol* 99, 4997-5005.

Shih, Y.C., Prausnitz, J.M., Blanch, H.W. 1992. Some characteristics of protein precipitation by salts. *Biotechnol Bioeng.* 40, 1155-64.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., Klenk, D.C. 1985. Measurement of protein using bicinchoninic acid. *Anal Biochem.* 150, 76-85.

Teeri TT. 1997. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends Biotechnol.* 15:160 −167.

Topakas, E., Moukouli, M., Dimarogona, M., Christakopoulo,s P. 2012. Expression, characterization and structural modelling of a feruloyl esterase from the thermophilic fungus *Myceliophthora thermophila. Applied Microbiology and Biotechnology* 94: 399-411.

Tsujikawa, M., Okabayashi, K., Morita, M., Tanabe, T. 1996. Secretion of a variant of human single-chain urokinase-type plasminogen activator without an N-glycosylation site in the methylotrophic yeast, *Pichia pastoris* and characterization of the secreted product. *Yeast.* 12, 541-53.

Van Den Hazel, H.B., Kielland-Brandt, M.C., Winther, J.R.1996. Review: biosynthesis and function of yeast vacuolar proteases. *Yeast* 12, 1-16.

Van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R., Pettersson G. 1986. Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. Separation of the functional domains. FEBS Lett 204:223–7.

Wegner, E.H., 1983. Biochemical conversions by yeast fermentation at high cell densities. United States Patent 4414329.

Wood, T.M. and McCrae, S.I. 1979. Synergism between enzymes involved in the solubilization of native cellulose. *Adv Chem Ser* 1979;181:181–209.

Zheng, F. and Ding, S. 2013. Processivity and Enzymatic Mode of a glycoside hydrolase family 5 endoglucanase from *Volvariella volvacea*. *Appl. Environ. Microbiol.* 79(3):989-996.

CHAPTER 8

Optimization of *tailor-made* enzyme cocktail for deconstruction of agricultural and forest residues

In this chapter, four core enzymes, in the presence of other four "accessory" enzymes, all encoded by *M. thermophila*'s genes and isolated using various purification and heterologous expression strategies, were tested against natural substrates to determine optimal combinations at 20 mg/g glucan protein loading, using a suitable design of experiments methodology. Synergistic interactions among different enzymes were then determined through various mixture optimization experiments. Optimal combinations were predicted from suitable statistical models which were able to further increase hydrolysis yields, suggesting that tailor-made enzyme mixtures targeted towards a particular feed stock can help maximize hydrolysis yields.

Even though the main components of all lignocellulosic feedstocks include cellulose, hemicellulose, as well as the protective lignin matrix, there are some differences in structure that may influence the degradability of the materials. Agricultural residues, such as wheat straw, have the advantage that in most cases they are easier to degrade in comparison with forest residues. This is due to the lower lignin content, as well as the dimension of the straw being relatively small, which results in a material that is more easily accessible for the microbial enzymes. Also, there is usually more difficult to hydrolyze softwoods than hardwoods because of the higher lignin content of softwoods than that of hardwoods. In hardwood and agricultural plants, xylan is the dominant hemicellulosic structure, whereas for softwoods, it is glucomannan, leading to the hypothesis that various types of biomass require a minimal set of enzymes that has to be tailor-made, i.e. more xylanases for hardwoods or more mannanases for softwoods.

In order to understand better the role of the individual enzymes and their synergistic interactions, in the present study, the hydrolysis of wheat straw, two types of softwood (pine and spruce) and one type of hardwood (birch) by a six component mixture at different stages was analyzed. All the substrates had been hydrothermally

pretreated (**Table 8.1**). In order to ensure an efficient enzymatic hydrolysis of cellulose, *pretreatment* (also called *first-stage hydrolysis*) is needed to break down the shield formed by lignin and hemicellulose, resulting in the disruption of the crystalline structure and the reduction of the degree of polymerization of cellulose (Xiros *et al.,* 2013). Different pretreatment technologies have varying effects on product yield and subsequent process steps (Wyman *et al.,* 2005).

Hydrothermal pretreatment with sulphuric acid

Spruce, tough treated	212°C for 4-8 minutes and pH 1,6-1,8
Pine	210-215°C for 5 minutes and pH 1,5-1,7
Birch	190°C for 4-6 minutes and pH 1,8-2,0

Hydrothermal pretreatment

Wheat straw	190°C for 12 minutes

Table 8.1 Pretreatment specifications of the materials used in hydrolysis experiments.

Spruce, pine and birch were purchased from SEKAB E-Technology AB (Örnsköldsvik, Sweden), while wheat straw from *Triticum aestivum* L. (PWS) which was purchased from Inbicon A/S, (Fredericia, Denmark). PWS was hydrothermally pretreated; residence time in the hydrothermal reactor averaged 12 min with the reactor temperature maintained at 190 °C by injection of steam (Thomsen *et al.* 2006). The solid fiber fraction was analyzed according to the procedure of Xiros *et al.* (2009) and was found to contain 50.2 % glucan, 3.8 % hemicellulose, 25.5 % acid insoluble lignin, and 2.8 % starch, based on DM (*w/w*). . Each of forest materials was hydrothermally pretreated with sulphuric acid and was received as pretreated slurries of low pH. Specifications are given **Table 8.1**.

An experimental mixture plan was set up for the four major cellulases *Mt*CBH7, *Mt*CBH6, *Mt*EG5 and*Mt*EG7, whereas "accessory" enzymes (mannanase, *Mt*CH61, xylanases) and β-glucosidase were added at specific loadings. In order to optimize the mixture of cellulases so as to achieve increased saccharification, an experimental design was employed. More specifically the software Design Expert® 7.0.0 (Stat-Ease inc.) was employed where the design *D-optimal* was used in order to generate 20 experimental conditions (**Table 8.2**) where the enzymes varied at specified levels (**Table 8.3**). In all the experimental combinations the summary of the enzymes was set to be equal to 1. The same software was used in order to evaluate the results and determine the most appropriate model that would be used to fit the experimental data. The two models applied during this work were either the *quadratic* (Eq. 1) or the *special cubic* (Eq. 2):

$$y = \sum_{j=1}^{4} b_j x_j + \sum_{1j<k4} b_{jk} x_j x_k \qquad \text{Eq. 1}$$

$$y = \sum_{j=1}^{4} b_j x_j + \sum_{1j<k4} b_{jk} x_j x_k + \sum_{1j<k<m4} b_{jkm} x_j x_k x_m \qquad \text{Eq. 2}$$

where *y* is the response (either total reducing sugars, TRS or glucose, Glc, mg/mL), *b* are the coefficients that were estimated by the model and *x* are the variables of the model. Optimization of the mixture was also performed by the same software, where the option to maximize either TRS or glucose was set. At the same moment the concentration of the enzymes were set to be in the limits that they were chosen to vary (**Table 8.5**). The efficiency of the model was evaluated by calculating the *p*-value and R^2.

Enzymes were prepared as described in **Chapters 4–7**. With the exception of xylanases, all individual enzymes used in these experiments were produced in *Pichia pastoris* from native *M. thermophila* genes. *Mt*EG5, *Mt*CBH6 and *Mt*GH61 were expressed in BSM medium in fermentors, while *Mt*EG7, *Mt*BGL3, *Mt*CBH7 in BMMY medium in shake flasks. Mannanase *Mt*Man26a was generously offered from Assist. Prof. E. Topakas, Biotechnology laboratory, School of Chemical Engineering, NTUA,

Greece. Xylanases were purified from *M. thermophila* crude, after induction with wheat straw, as described in **Chapter 7**.

# React	Enzyme proportions			
	*Mt*EG5	*Mt*EG7	*Mt*CBH6	*Mt*CBH7
1	0,06	0,07	0,65	0,21
2	0,20	0,05	0,53	0,22
3	0,01	0,40	0,20	0,39
4	0,12	0,33	0,24	0,31
5	0,20	0,40	0,20	0,20
6	0,01	0,05	0,47	0,47
7	0,01	0,08	0,21	0,70
8	0,09	0,19	0,52	0,20
9	0,03	0,19	0,35	0,44
10	0,01	0,40	0,20	0,39
11	0,01	0,24	0,21	0,54
12	0,01	0,05	0,47	0,47
13	0,01	0,33	0,46	0,20
14	0,20	0,40	0,20	0,20
15	0,20	0,21	0,20	0,38
16	0,13	0,05	0,20	0,62
17	0,01	0,33	0,46	0,20
18	0,06	0,07	0,65	0,21
19	0,11	0,24	0,37	0,28
20	0,20	0,05	0,37	0,38

Table 8.2. Experimental combinations used for the hydrolysis tests.
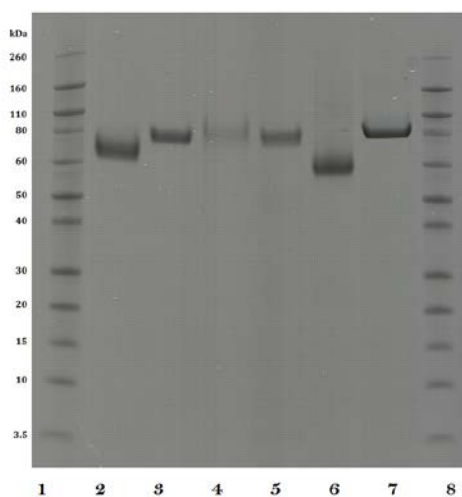
Figure 8.1. SDS – PAGE of purified enzymes produced and used in hydrolysis experiments. *Lanes 1, 8*: Novex® sharp pre-stained protein marker, *2: Mt*EG7a, *3: Mt*EG5, *4: Mt*CBH7, *5: Mt*CBH6, *6: Mt*GH61, *7: Mt*BGL3a.

Protein concentrations were determined using the bicinchoninic acid (BCA) protein assay microplate procedure (Pierce Chemical Co., Rockford, IL), according to the manufacturer's recommendations, using bovine serum albumin as standard (Smith *et al.*, 1985). Enzyme activities were assayed as described in previous Chapters. The supernatants were filtrated, concentrated using a tangential flow filtration system with a 10 kDa-cutoff membrane (Pellicon XL Ultrafiltration Module Biomax 10 kDa, Millipore), buffer exchanged in dialysis tubing membrane with ten volumes of 100mM phosphate – citrate buffer, pH 5.0, and then concentrated further to 20mL. Concentrated desalted enzymes were aliquoted into 1.5 mL tubes and stored at 4 $^{\circ}$C, with the addition of sodium azide was added (0.02% w/v). Final stock enzyme concentrations ranged from 0.2 to1 mg/mL. Purity of the final protein preparations was determined by 12.5% SDS-PAGE electrophoresis (**Figure 8.1**).

The relative abundances of each of the four major enzymes were varied using an experimental design. The borders of the experimental domain were carefully chosen. Special attention was paid to avoid a too large domain as this may impact the reliability of predictions within the domain. On the other hand, it should not be too small and contain the optimum, since extrapolation outside the domain borders is impossible. The lower and upper limits of each component were, therefore, determined following rational considerations (Billard *et al.*, 2012):

✓ *Cellobiohydrolases* are known to be important for cellulose hydrolysis (Teeri, 1997) and the sum of *Mt*CBH7 and *Mt*CBH6 should constitute the majority of the enzyme cocktail (>50%). In addition, literature data showed that higher CBH7/CBH6 ratios are more beneficial for hydrolysis of steam-exploded wheat straw than lower ones (Rosgaard *et al.*, 2007). Their relative abundances were thus varied from 13:4 to 2:7. *Mt*CBH7 is produced as a major protein of fungi's secretome (20–25% of the total extracellular protein) and adsorbed strongly on microcrystalline cellulose, so it is arguably the most important single enzyme involved in cellulose deconstruction, which however has limited activity on un-pretreated lignocellulosic material, due to limited accessible sites of action. Moreover, it has been shown that there is a significant synergism between *Mt*CBH7 and *Mt*CBH6 enzymes during substrate hydrolysis (Gusakov *et al.*, 2007).

✓ The presence of *endoglucanases* is necessary and *Mt*EG7 and *Mt*EG5 should each at least make up 1% of the mixture. A higher upper limit was chosen for *Mt*EG7, as this enzyme was shown to major importance for optimized initial conversion rates and final yield for steam-exploded wheat straw according to the literature (Billard *et al.*, 2012). Moreover, as described in **Chapter 5**, *Mt*EG7 has been proved to liquefy efficiently high-consistency lignocellulosic biomass by decreasing significantly the viscosity of the slurry in the first stage of reaction, underlining the crucial role of this enzyme for hydrolysis.

	Variable in model	Lower limit (%)	Upper limit (%)
*Mt*EG5	χ_1	1	20
*Mt*EG7	χ_2	5	40
*Mt*CBH6	χ_3	20	65
*Mt*CBH7	χ_4	2	70

Table 8.3. The respective borders for all variables used the experimental design.

Hydrolysis of phosphoric acid swollen cellulose (PASC) was performed using only the four core cellulolytic enzymes and β-glucosidase. In, case of lignocellulosic feedstocks, the addition of accessory enzymes was based on the structure and type of each material. In all experiments, *Mt*GH61 consisted 4% of the enzyme mixture, while gallic acid was added at 10mM concentration, as electron donor. The enzymes with xylanase activity consisted 3% of the enzyme mixture for birch (hardwood) and 2% for wheat straw and softwoods. In case of softwoods, *Mt*Man26a was used at 3%. Enzymatic hydrolysis was performed in safe lock 2 mL volume microtubes. The surfactant octylphenol (ethyleneglycol)$_{9.6}$ ether (Triton X-100) was added at all reactions with natural substrates at concentration 1 mg/mL, which is equivalent to a surfactant addition of 4% of the substrate dry matter. Enhancement of cellulose hydrolysis by adding surfactants to the hydrolysis mixture has been reported by several authors (Eriksson *et al.*, 2002; Castanon and Wilke, 1981; Helle *et al.*, 1993; Ooshima *et al.*, 1986; Park *et al.*, 1992). Reactions were performed at 50°C, with 2.5% initial dry matter content when natural substrates were used and 0.25% in case of pure substrate (PASC). The enzymes were loaded at *1 mg/g substrate* for PASC and *20 mg/g substrate* for natural materials. In all reactions, β-glucosidase was added in excess in

order to prevent inhibition caused by cellobiose produced by the combined action of exo- and endo-1,4-β-glucanases. After 12h of incubation, more enzyme was added, so as to ensure the effective release of glucose. All assays were replicated once, sampled twice and assayed twice for total reducing sugars (TRS) and glucose (Glc) at 48 h of hydrolysis, for a total of four replicates of each mixture each time for each variable (TRS and Glc). Total reducing sugars were measured with dinitrosalicylic acid colorimetric method (Miller, 1958) and Glc with GO assay (Sigma). All reactions were performed with 1200 rpm agitation and contained 0.02% w/v sodium azide to prevent microbial contamination.

Figure 8.2. Enzymatic hydrolysis was performed in safe lock 2mL volume microtubes, with 1200 rpm agitation in Thermomixer incubator, which offers simultaneous mixing and temperature control (Eppendorf, Germany).

8.1. Hydrolysis of phosphoric acis swollen cellulose (PASC)

PASC is considered to be a representative of amorphous cellulose. It is usually prepared from Avicel by phosphoric acid treatment (as described by Schulein, 1997) and, opposed to the dry, solid, powder-like composition of Avicel, PASC is relatively viscous, unclear (cloudy) liquid. It has been shown that in the cellulose swelling process, the macromolecular structure of cellulose as a moiety of fibers is maintained, while the physical properties of the sample are changed (Zhang, *et al.*, 2006). In particular, specific surface area and sample volume have been shown to increase, due to

swelling, while DP of PASC have not changed radically compared to the untreated Avicel (Zhang and Lynd, 2005; Zhang, *et al.*, 2006). In this study, the PASC that was used was a generous offer from M. Dimarogona, Swedish University of Agricultural Sciences, Uppsala, Sweden.

	Enzyme proportions						
# React	**EG5**	**EG7**	**CBH6**	**CBH7**	**RS (mg/mL)**	**% Hydrolysis**	**Glucose (mg/mL)**
1	0,06	0,07	0,65	0,21	0,518	18,660	0,491
2	0,20	0,05	0,53	0,22	0,603	21,709	0,553
3	0,01	0,40	0,20	0,39	0,552	19,879	0,531
4	0,12	0,33	0,24	0,31	0,637	22,928	0,559
5	0,20	0,40	0,20	0,20	0,683	24,584	0,597
6	0,01	0,05	0,47	0,47	0,431	15,524	0,384
7	0,01	0,08	0,21	0,70	0,313	11,255	0,301
8	0,09	0,19	0,52	0,20	0,511	18,399	0,376
9	0,03	0,19	0,35	0,44	0,601	21,622	0,391
10	0,01	0,40	0,20	0,39	0,557	20,054	0,421
11	0,01	0,24	0,21	0,54	0,409	14,740	0,372
12	0,01	0,05	0,47	0,47	0,436	15,698	0,398
13	0,01	0,33	0,46	0,20	0,625	22,493	0,554
14	0,20	0,40	0,20	0,20	0,688	24,758	0,579
15	0,20	0,21	0,20	0,38	0,441	15,872	0,437
16	0,13	0,05	0,20	0,62	0,567	20,402	0,516
17	0,01	0,33	0,46	0,20	0,620	22,319	0,506
18	0,06	0,07	0,65	0,21	0,526	18,921	0,500
19	0,11	0,24	0,37	0,28	0,441	15,872	0,409
20	0,20	0,05	0,37	0,38	0,291	10,471	0,294
BLANK					0,017	0,627	0,060
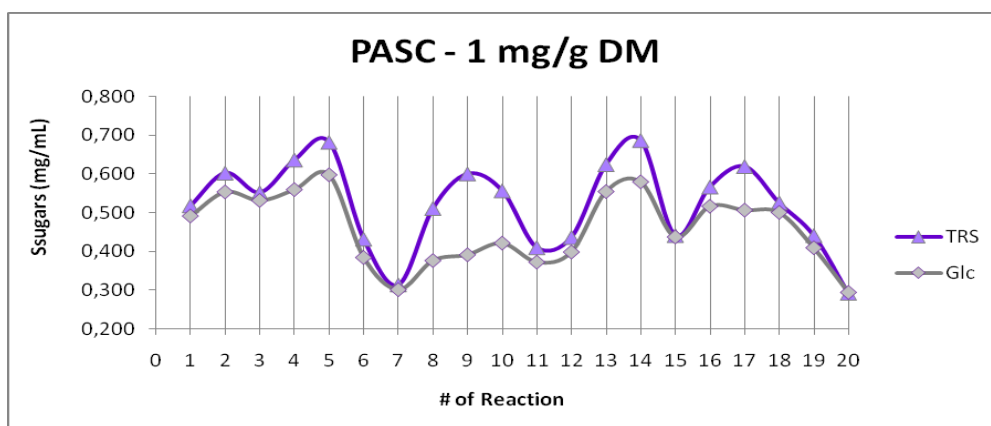
Table 8.4. TRS and Glc yields for each enzyme combination used for the hydrolysis of *PASC*. The marked mixture indicates the reaction that produced the highest amount of sugars.

*PASC, Total Reducing Sugars (TRS):* special cubic, p=0.0002, R^2 =0.9849

TRS =   - 21.66724 * EG5
         + 1.09867 * EG7
         - 0.54947 * CBH6
         - 0.23873 * CBH7
         + 45.79912 * EG5 * EG7
         + 45.36321 * EG5 * CBH6
         + 37.05488 * EG5 * CBH7
         + 0.22492 * EG7 * CBH6
         - 2.87501 * EG7 * CBH7
         + 2.75752 * CBH6 * CBH7
         - 97.02937 * EG5 * EG7 * CBH6
         - 19.25658 * EG5 * EG7 * CBH7
         - 64.09204 * EG5 * CBH6 * CBH7
         +14.71254 * EG7 * CBH6 * CBH7

TRS opt: **EG5** (0.128) **EG7** (0.400) **CBH6** (0.200) **CBH7** (0.272), Yield: **0.754853 mg/mL**

PASC, Glucose (Glc): special cubic, p=0.0116, R^2 = 0.9401

Glc =   - 20.42708 * EG5
         + 0.45905 * EG7
         - 0.50713 * CBH6
         - 0.19593 * CBH7
         + 40.53931 * EG5 * EG7
         + 40.07249 * EG5 * CBH6
         + 30.82599 * EG5 * CBH7
         + 3.85197 * EG7 * CBH6
         + 1.05697 * EG7 * CBH7
         + 2.91880 * CBH6 * CBH7
         - 92.35394 * EG5 * EG7 * CBH6
         + 0.082939 * EG5 * EG7 * CBH7
         - 42.30706 * EG5 * CBH6 * CBH7
         - 9.17932 * EG7 * CBH6 * CBH7

Glc opt: **EG5** (0.118) **EG7** (0.400) **CBH6** (0.200) **CBH7** (0.282), Yield: **0.734118 mg/mL**
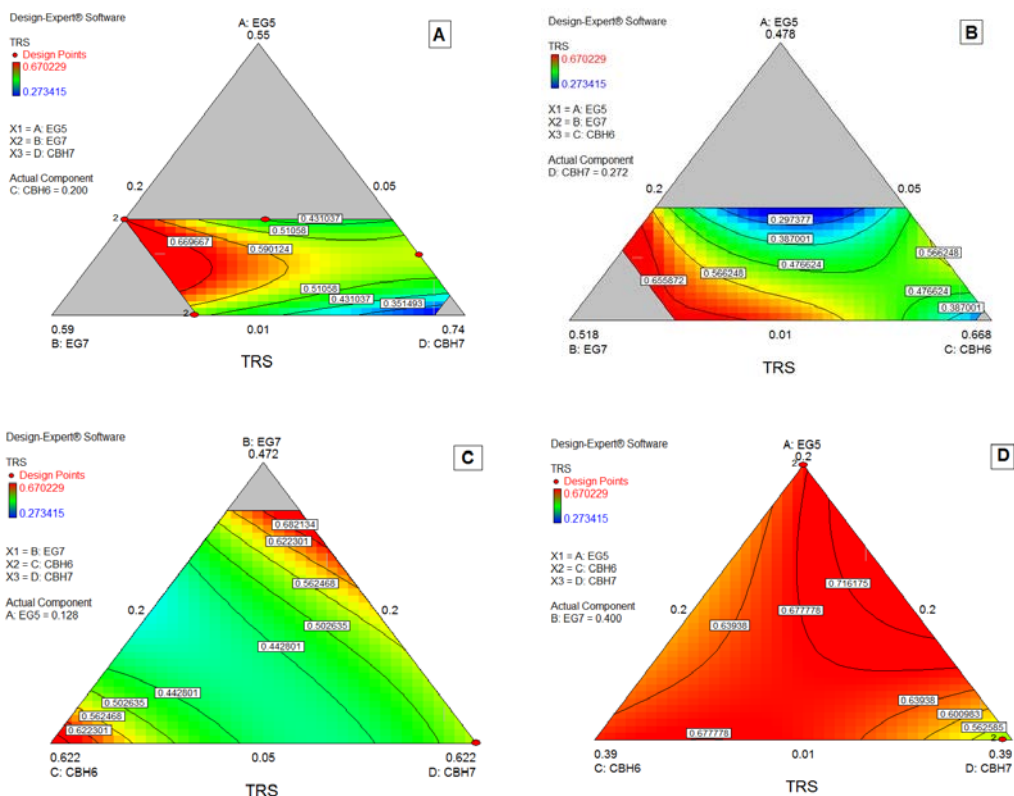
Figure 8.3. Ternary plots showing predicted final Total Reducing Sugars (TRS) yields from PASC hydrolysis, as a function of three out of four "core" enzymes (X1, X2, X3) content. For each plot, the forth enzyme ("actual component") has been fixed to the proportion of the point resulting in the optimal TRS yield, as predicted by the model. As illustrated in (A), a decrease in *Mt*EG7 proportion results in a decrease in hydrolysis yield, even if *Mt*EG5 levels are higher, indicating the key role of GH7 endoglucanase for the reaction. Even though it cannot compensate for *Mt*EG7, *Mt*EG5 is also an important enzyme; as moving towards D point, where *Mt*EG7 and *Mt*CBH7 are in moderate levels and *Mt*EG5 in its lower limit proportion, the hydrolysis rate is very low.
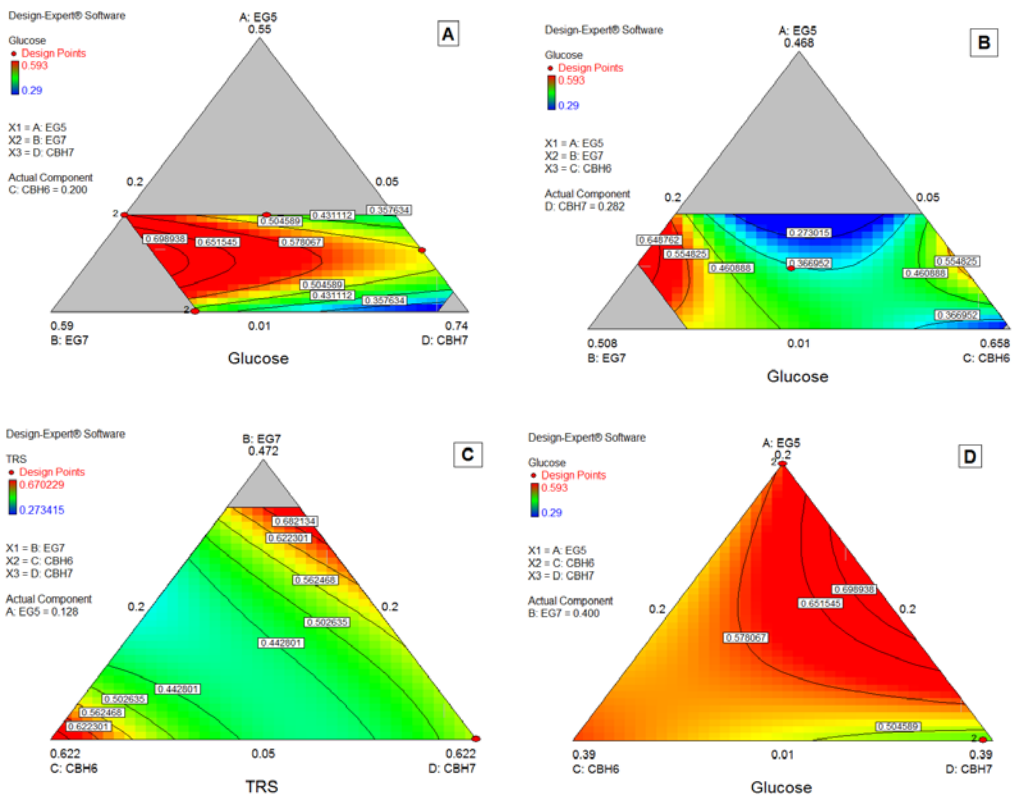
Figure 8.4. Ternary plots showing predicted final glucose (Glc) yields from PASC hydrolysis, as a function of three out of four "core" enzymes content (X1, X2, X3). For each plot, the forth enzyme ("actual component") has been fixed to the proportion of the point resulting in the optimal glucose yield, as predicted by the model.

Core enzyme mixtures were tested for their hydrolysis performance on hydrothermally pretreated wheat straw. All enzyme loadings were based on equivalent bovine serum albumin BCA based measurement. *Mt*GH61 and the xylanase cocktail were added as accessory enzymes, as described above. *Mt*BGL3 was loaded in excess (10% of the total cellulose loading), in order to prevent cellobiose inhibition. The trends for both total reducing sugars (TRS) and glucose (Glc) yields among different enzyme mixtures were dependent on the unique enzyme combinations. % rate of hydrolysis was calculated from the following equation, assuming that the glucose content of wheat straw is 50.2%, according to the analysis of solid fiber fraction by Xiros *et al.* (2009):

$$\% \text{ hydrolysis rate} = \frac{\text{TRS (mg / mL)}}{\text{DM (mg/mL)} * 0.502 * 1.111} * 100$$

Enzyme proportions

# React	EG5	EG7	CBH6	CBH7	RS (mg/mL)	% Hydrolysis	Glucose (mg/mL)
1	0,06	0,07	0,65	0,21	2,967	21,282	2,930
2	0,20	0,05	0,53	0,22	2,793	20,033	2,652
3	0,01	0,40	0,20	0,39	3,229	23,156	3,094
4	0,12	0,33	0,24	0,31	3,142	22,532	3,094
5	0,20	0,40	0,20	0,20	3,635	26,072	3,486
6	0,01	0,05	0,47	0,47	3,084	22,115	3,077
7	0,01	0,08	0,21	0,70	3,432	24,614	3,110
8	0,09	0,19	0,52	0,20	3,113	22,323	3,077
9	0,03	0,19	0,35	0,44	3,316	23,781	3,323
10	0,01	0,40	0,20	0,39	3,258	23,365	3,110
11	0,01	0,24	0,21	0,54	3,606	25,864	3,126
12	0,01	0,05	0,47	0,47	3,113	22,323	3,126
13	0,01	0,33	0,46	0,20	3,345	23,989	2,979
14	0,20	0,40	0,20	0,20	3,693	26,488	3,519
15	0,20	0,21	0,20	0,38	3,809	27,321	3,634
16	0,13	0,05	0,20	0,62	3,490	25,031	3,486
17	0,01	0,33	0,46	0,20	3,287	23,573	2,963
18	0,06	0,07	0,65	0,21	3,025	21,699	2,619
19	0,11	0,24	0,37	0,28	3,200	22,948	3,077
20	0,20	0,05	0,37	0,38	3,055	21,907	2,881
BLANK					0,064	0,458	0,246

Table 8.5. TRS and Glc yields for each enzyme combination used for the hydrolysis of pretreated wheat straw. The marked mixture indicates the reaction that produced the highest amount of sugars.

Wheat Straw
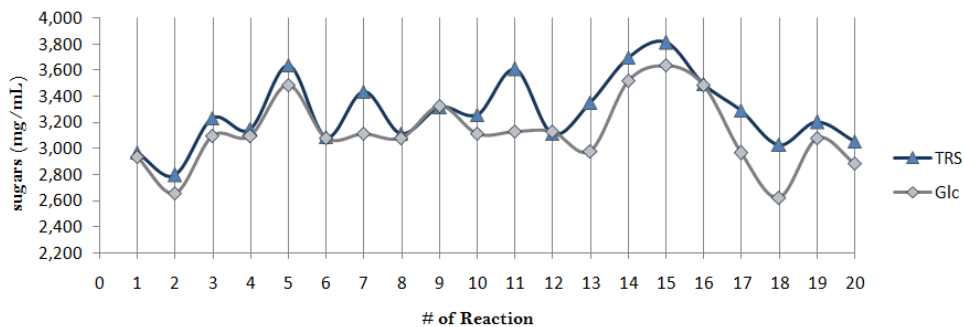
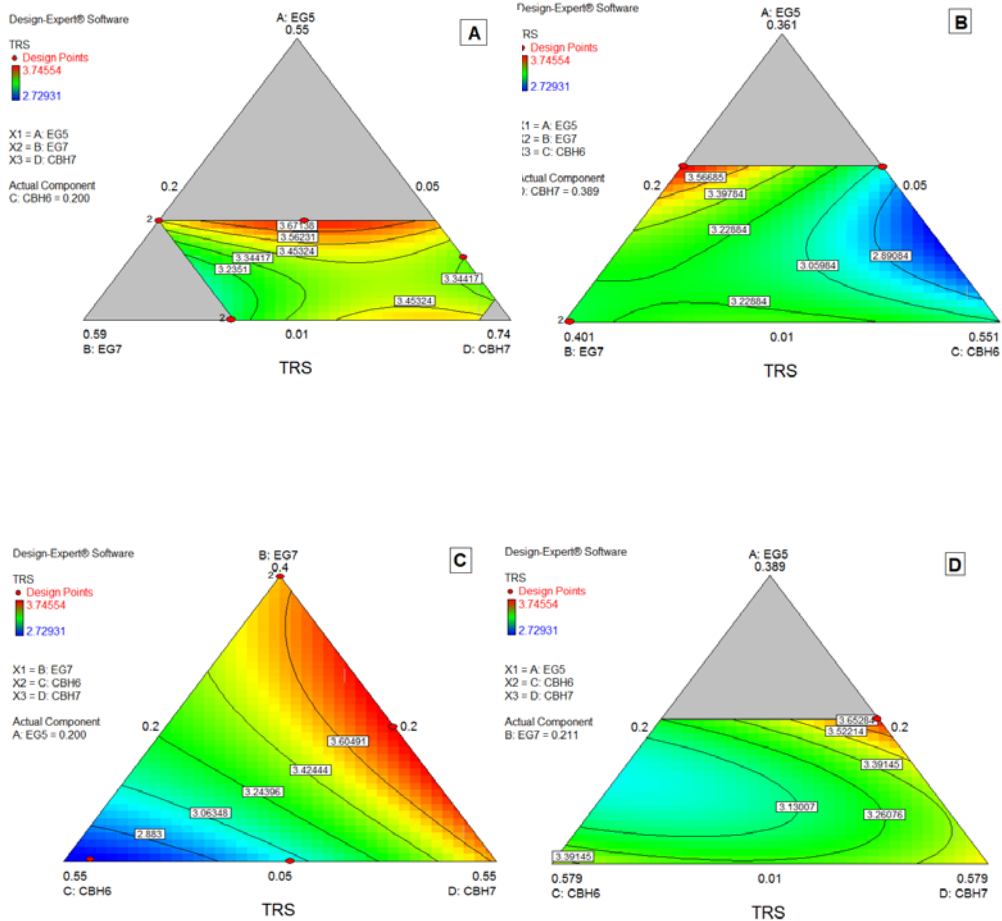*Wheat Straw, Total Reducing Sugars (TRS)*: quadratic model p< 0.0001, R^2 = 0.9548

TRS = + 26.24805 * EG5

    - 0.87730 * EG7

    + 4.34075 * CBH6

    + 4.23657 * CBH7

    - 15.49701 * EG5 * EG7

    - 38.29441 * EG5 * CBH6

    - 24.78159 * EG5 * CBH7

    + 5.40389 * EG7 * CBH6

    + 6.19993 * EG7 * CBH7

    - 5.36302 * CBH6 * CBH7

TRS opt: **EG5** (0.200) **EG7** (0.211) **CBH6** (0.200) **CBH7** (0.389), Yield: **3.7866 mg/mL**

Wheat Straw, Glucose (Glc): quadratic p=0.0053, R^2 = 0.8411

Glc =   + 7.15288 * EG5

    + 0.70601 * EG7

    + 2.69945 * CBH6

    + 2.81969 * CBH7

    + 4.99455 * EG5 * EG7

    - 13.04901 * EG5 * CBH6

    + 0.051320 * EG5 * CBH7

    + 2.90623 * EG7 * CBH6

    + 3.63262 * EG7 * CBH7

    + 0.23688 * CBH6 * CBH7

Glc opt: **EG5** (0.200) **EG7** (0.219) **CBH6** (0.200) **CBH7** (0.381), Yield: **3.34854 mg/mL**

Figure 8.5. Ternary plots showing predicted final Total Reducing Sugars (TRS) yields from wheat straw hydrolysis, as a function of three out of four "core" enzymes (X1, X2, X3) content. For each plot, the forth enzyme ("actual component") has been fixed to the proportion of the point resulting in the optimal TRS yield, as predicted by the model.
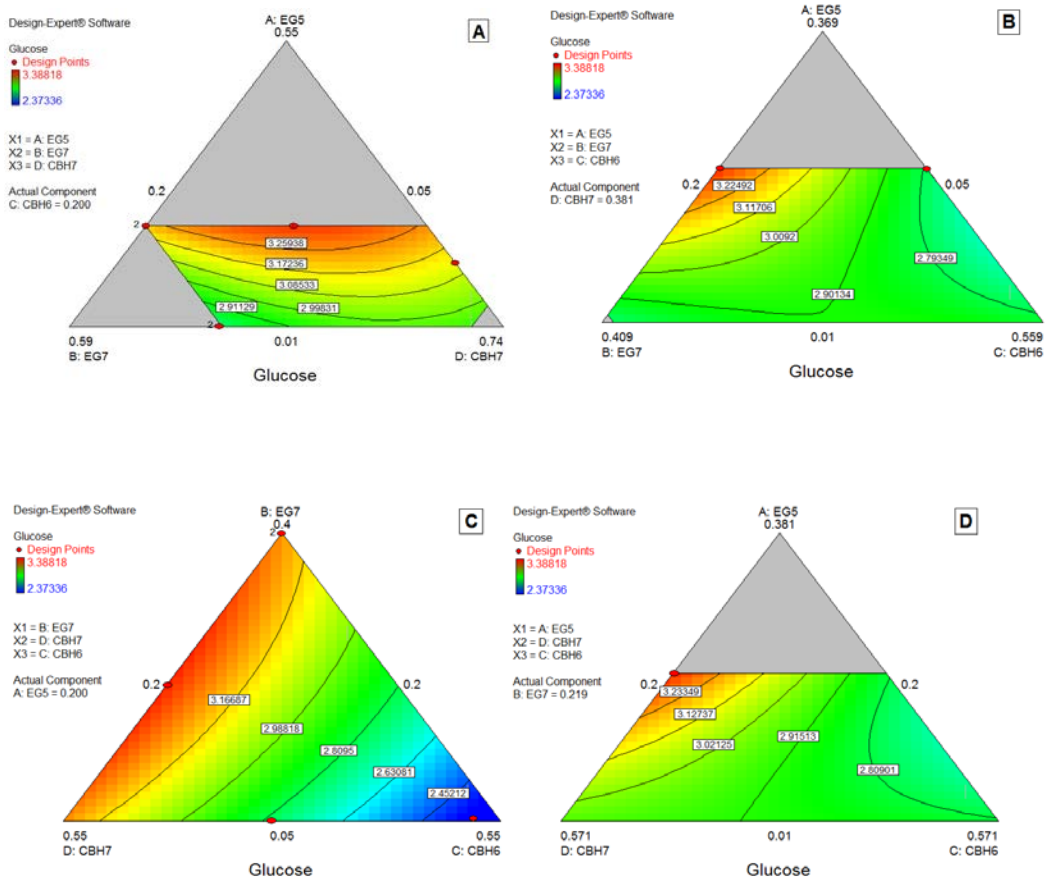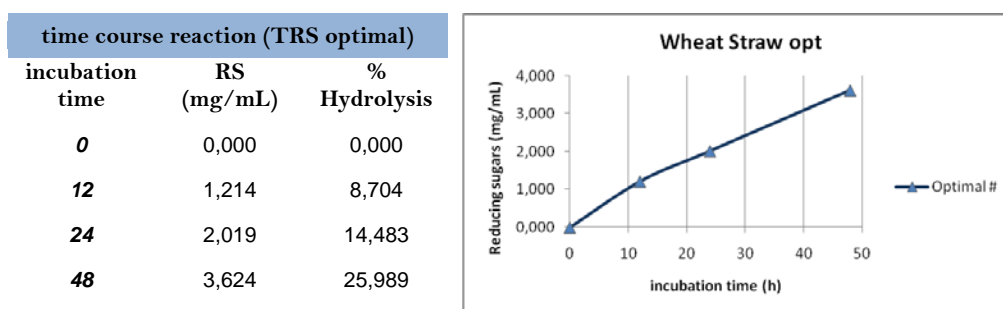
Figure 8.6. Ternary plots showing predicted final glucose (Glc) yields from wheat straw hydrolysis, as a function of three out of four "core" enzymes content (X1, X2, X3). For each plot, the forth enzyme ("actual component") has been fixed to the proportion of the point resulting in the optimal glucose yield, as predicted by the model.

The ternary plot **C** of **Figure 8.3** shows that, when the proportions of *Mt*EG7 are increasing or decreasing over a large range, a high final yield can be conserved. Similar to Billard *et al.*, 2012, that showed that the optimum yield is conserved over a range from about 13% to 23% EG7. Same can be noticed also for *Mt*CBH7. Comparing hydrolysis yields obtained with high and low *Mt*CBH6/*Mt*CBH7 ratio, it seems that

those when *Mt*CBH7 is in higher proportions, the yields are better. As the ratio decreases, the hydrolysis yield also follows the same tension, so *Mt*CBH6 does not compensate for *Mt*CBH7. Even when cellobiohydrolases participate in low proportions (lower limits), there is some hydrolysis that can be attributed to the action of *Mt*EG7. Moving vertically towards lower *Mt*EG7 proportions, yields are maintained, so *Mt*CBH7 can compensate for *Mt*EG7 (at least partially). Similar assumptions may be made for ternary plot **C** of **Figure 8.4** concerning the glucose yield.

Endoglucanases are of major importance for the efficient hydrolysis of pretreated wheat straw, catalyzing the initial attack on the amorphous regions of cellulose chains and the gradual reduction of the average chain length of these polysaccharides. Szijarto *et al.* identified EG2 (Cel5a) as a key component for the liquefaction of pretreated wheat straw, while *Mt*EG7 has also been proved to liquefy efficiently high-consistency lignocellulosic biomass by decreasing significantly the viscosity of the slurry in the first stage of reaction, underlining the crucial role of this enzyme for hydrolysis (Karnaouri *et al.*, 2014). The crucial role of these enzymes is highlighted in the results of present study.

time course reaction (TRS optimal)		
incubation time	RS (mg/mL)	% Hydrolysis
0	0,000	0,000
12	1,214	8,704
24	2,019	14,483
48	3,624	25,989

TRS opt (predicted): **EG5** (0.200) **EG7** (0.211) **CBH6** (0.200) **CBH7** (0.389), Yield: **3.7866 mg/mL**

Figure 8.5. Time-course hydrolysis using the enzyme combination that was predicted to lead to the highest Total Reducing Sugars production. Experimental yield appear to be slightly lower than the theoretically predicted one. The rate of hydrolysis remains almost stable during the first 48h of hydrolysis.

Core enzyme mixtures were tested for their hydrolysis performance on hydrothermally pretreated with sulphuric acid birch. All enzyme loadings were based on equivalent bovine serum albumin BCA based measurement. *Mt*GH61 and the xylanase cocktail were added as accessory enzymes, as described above. Xylanases' concentration was 3%, as the main hemicellulose of birch is xylan. *Mt*BGL3 was loaded in excess (10% of the total cellulose loading), in order to prevent cellobiose inhibition. The trends for both total reducing sugars (TRS) and glucose (Glc) yields among different enzyme mixtures were dependent on the unique enzyme combinations. % rate of hydrolysis was calculated from the following equation, assuming that the glucose content of wheat straw is 41%, according to the literature data (Alén, 2000; Sjöström, 1993; Willför *et al.*, 2005):

$$\% \text{ hydrolysis rate} = \frac{\text{TRS (mg / mL)}}{\text{DM (mg/mL)} * 0.410 * 1.111} * 100$$

	Enzyme proportions						
# React	EG5	EG7	CBH6	CBH7	RS (mg/mL)	% Hydrolysis	Glucose (mg/mL)
1	0,06	0,07	0,65	0,21	0,735	6,456	0,617
2	0,20	0,05	0,53	0,22	0,764	6,711	0,722
3	0,01	0,40	0,20	0,39	0,715	6,277	0,696
4	0,12	0,33	0,24	0,31	0,677	5,946	0,651
5	0,20	0,40	0,20	0,20	0,753	6,609	0,668
6	0,01	0,05	0,47	0,47	0,755	6,634	0,651
7	0,01	0,08	0,21	0,70	0,741	6,507	0,619
8	0,09	0,19	0,52	0,20	0,689	6,048	0,644
9	0,03	0,19	0,35	0,44	0,843	7,399	0,642
10	0,01	0,40	0,20	0,39	0,721	6,328	0,701
11	0,01	0,24	0,21	0,54	0,581	5,104	0,577
12	0,01	0,05	0,47	0,47	0,761	6,685	0,686
13	0,01	0,33	0,46	0,20	0,631	5,538	0,582
14	0,20	0,40	0,20	0,20	0,761	6,685	0,678
15	0,20	0,21	0,20	0,38	0,837	7,348	0,640
16	0,13	0,05	0,20	0,62	0,755	6,634	0,617
17	0,01	0,33	0,46	0,20	0,625	5,487	0,574
18	0,06	0,07	0,65	0,21	0,744	6,532	0,614
19	0,11	0,24	0,37	0,28	0,799	7,017	0,702
20	0,20	0,05	0,37	0,38	0,843	7,399	0,685
BLANK					0,056	0,490	0,061

Table 8.6. TRS and Glc yields for each enzyme combination used for the hydrolysis of pretreated birch. The marked mixture indicates the reaction that produced the highest amount of sugars.

Birch
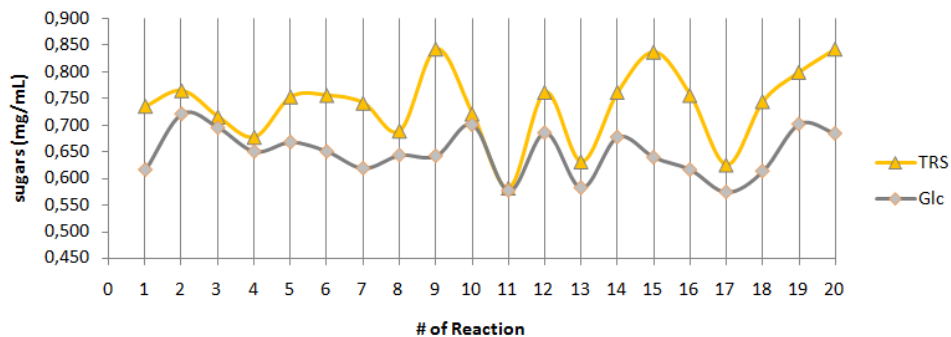
Birch, Total Reducing sugars (TRS): special cubic p< 0.0001, R^2= 0.9981

TRS = + 23.86839 * EG5

+ 7.02440 * EG7

+ 2.52957 * CBH6

+ 1.81127 * CBH7

- 55.96731 * EG5 * EG7

- 37.14665 * EG5 * CBH6

- 26.39382 * EG5 * CBH7

- 22.53481 * EG7 * CBH6

- 20.28144 * EG7 * CBH7

- 6.27797 * CBH6 * CBH7

+ 92.42527 * EG5 * EG7 * CBH6

+ 20.87724 * EG5 * EG7 * CBH7

+ 4.84800 * EG5 * CBH6 * CBH7

+ 68.32745 * EG7 * CBH6 * CBH7

TRS opt: **EG5** (0.200) **EG7** (0.196) **CBH6** (0.356) **CBH7** (0.248), Yield: **1.00049 mg/mL**

Birch, Glucose (Glc): special cubic p=0.0008, R^2= 0.9764

Glc =   + 7.18980 * EG5

+ 3.62601 * EG7

+ 0.79553 * CBH6

+ 0.75054 * CBH7

- 23.13030 * EG5 * EG7

- 8.40950 * EG5 * CBH6

- 7.08057 * EG5 * CBH7

- 8.05122 * EG7 * CBH6

- 7.10649 * EG7 * CBH7

- 0.57697 * CBH6 * CBH7

+ 44.79136 * EG5 * EG7 * CBH6

+ 10.25298 * EG5 * EG7 * CBH7

- 5.08012 * EG5 * CBH6 * CBH7

+ 17.51918 * EG7 * CBH6 * CBH7

<u>Glc opt:</u> **EG5** (0.200) **EG7** (0.209) **CBH6** (0.391) **CBH7** (0.200), Yield: **0.773808 mg/mL**
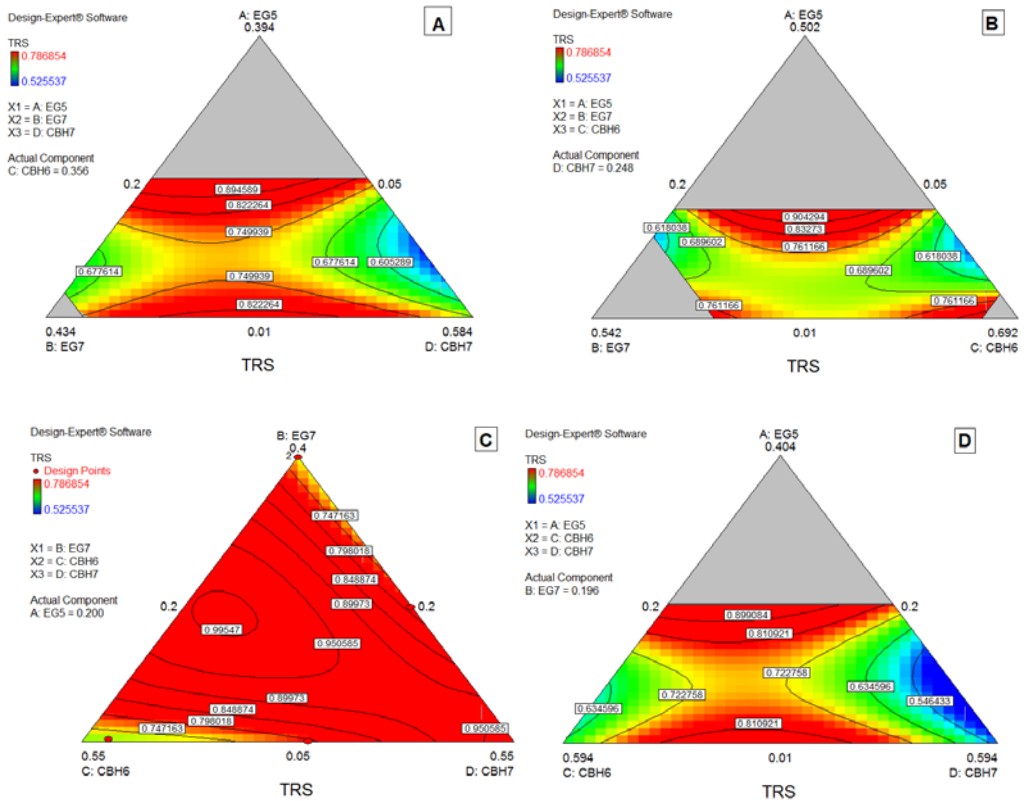

Figure 8.7. Ternary plots showing predicted final Total Reducing Sugars (TRS) yields from birch hydrolysis, as a function of three out of four "core" enzymes (X1, X2, X3) content. For each plot, the forth enzyme ("actual component") has been fixed to the proportion of the point resulting in the optimal TRS yield, as predicted by the model. As illustrated at (C), the location of the optimal domain indicates that *Mt*EG7 and *Mt*CBH6 are needed in moderate levels for the maximum hydrolysis.
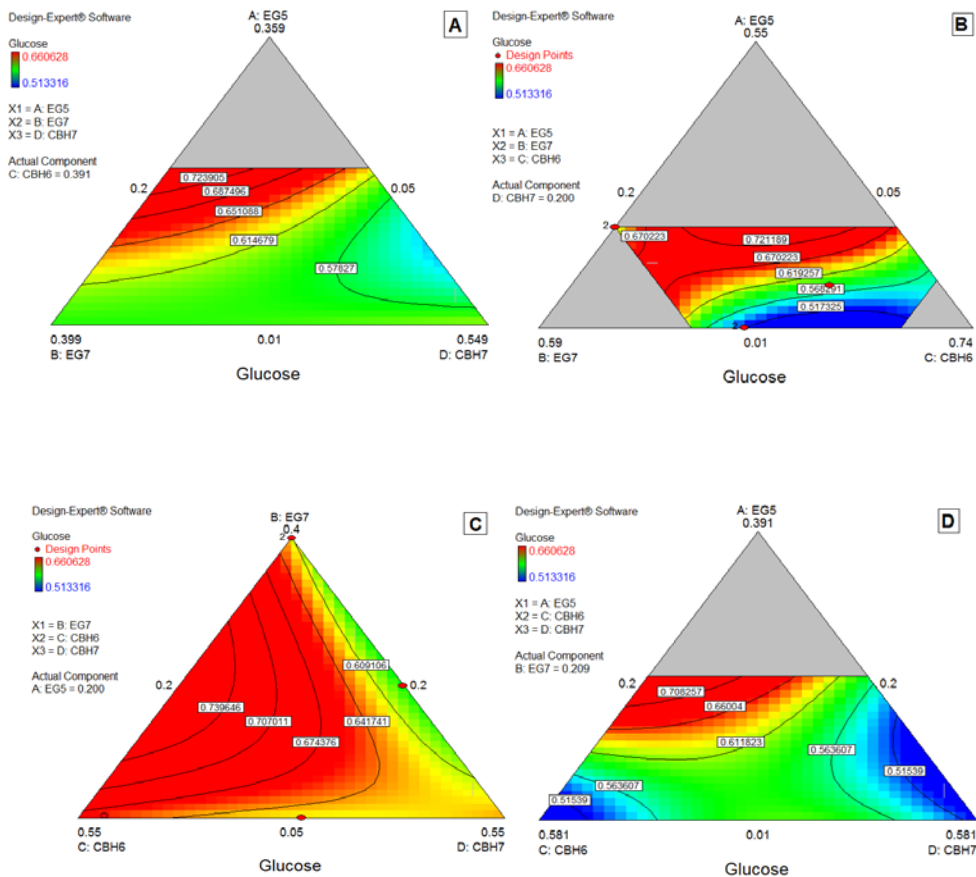
Figure 8.8. Ternary plots showing predicted final glucose (Glc) yields from birch hydrolysis, as a function of three out of four "core" enzymes content (X1, X2, X3). For each plot, the forth enzyme ("actual component") has been fixed to the proportion of the point resulting in the optimal glucose yield, as predicted by the model. The ternary plots **A** and **C** show that, when the proportions of *Mt*EG7 and *Mt*EG5 are maintained in a moderate range can result in efficient hydrolysis, whereas cellobiohydrolases seem to be effective in smaller proportions. Cellobiohydrolases may be inhibited by xylan and xylan-fragments produced during the hydrolysis. Birch is a hardwood with its main hemicellulolytic component is xylan, so the addition of efficient proportion of xylanases and β-xylosidases is of great importance to maintain an optimal yield and eliminate the inhibitory effect that hampers the activity of CBHs.

time course reaction (TRS optimal)		
incubation time	RS (mg/mL)	% Hydrolysis
0	0,000	0,000
12	0,448	3,932
24	0,561	4,926
48	0,838	7,320

Birch opt
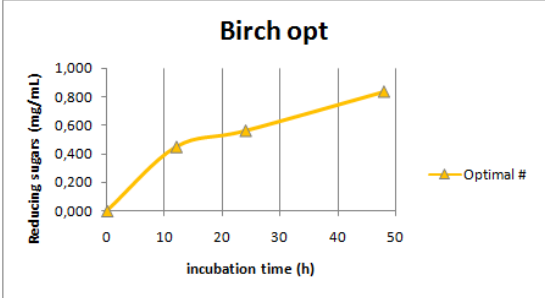
TRS opt: **EG5** (0.200) **EG7** (0.196) **CBH6** (0.356) **CBH7** (0.248), Yield: **1.00049 mg/mL**

Figure 8.5. Time-course hydrolysis using the enzyme combination that was predicted to lead to the highest Total Reducing Sugars production. Experimental yield appear to be lower than the theoretically predicted one, but close to the combination #20, that that produced the highest amount of sugars. The rate of hydrolysis is the highest one during the first 12h of incubation, but later decreases, due to the recalcitrance of the forest material.

Core enzyme mixtures were tested for their hydrolysis performance on hydrothermally pretreated with sulphuric acid spruce. All enzyme loadings were based on equivalent bovine serum albumin BCA based measurement. *Mt*GH61 and the xylanase cocktail were added as accessory enzymes, as described above. Xylanases' concentration was 2% and was added in combination with 3% *Mt*Man26a, as spruce belongs to softwoods and its main hemicellulose is glucomannan. *Mt*BGL3 was loaded in excess (10% of the total cellulose loading), in order to prevent cellobiose inhibition. The trends for both total reducing sugars (TRS) and glucose (Glc) yields among different enzyme mixtures were dependent on the unique enzyme combinations. % rate of hydrolysis was calculated from the following equation, assuming that the glucose content of wheat straw is 40.7, according to the literature data (Hakkila, 1989; Korhonen, 1997; Sjöström, 1993):

$$\% \text{ hydrolysis rate} = \frac{\text{TRS (mg / mL)}}{\text{DM (mg/mL)} * 0.395 * 1.111} * 100$$

	Enzyme proportions						
# React	EG5	EG7	CBH6	CBH7	RS (mg/mL)	% Hydrolysis	Glucose (mg/mL)
1	0,06	0,07	0,65	0,21	1,664	15,165	1,729
2	0,20	0,05	0,53	0,22	1,678	15,297	1,621
3	0,01	0,40	0,20	0,39	1,867	17,017	1,865
4	0,12	0,33	0,24	0,31	2,201	20,061	2,203
5	0,20	0,40	0,20	0,20	2,070	18,870	1,913
6	0,01	0,05	0,47	0,47	1,715	15,628	1,685
7	0,01	0,08	0,21	0,70	1,685	15,363	1,633
8	0,09	0,19	0,52	0,20	1,874	17,083	1,766
9	0,03	0,19	0,35	0,44	2,027	18,473	1,876
10	0,01	0,40	0,20	0,39	1,831	16,686	1,821
11	0,01	0,24	0,21	0,54	1,925	17,546	1,780
12	0,01	0,05	0,47	0,47	1,751	15,958	1,715
13	0,01	0,33	0,46	0,20	2,063	18,803	2,070
14	0,20	0,40	0,20	0,20	2,056	18,737	1,933
15	0,20	0,21	0,20	0,38	2,230	20,325	2,097
16	0,13	0,05	0,20	0,62	2,077	18,936	1,787
17	0,01	0,33	0,46	0,20	2,048	18,671	2,060
18	0,06	0,07	0,65	0,21	1,685	15,363	1,736
19	0,11	0,24	0,37	0,28	2,302	20,987	2,217
20	0,20	0,05	0,37	0,38	2,070	18,870	1,855
BLANK					0,176	1,601	0,174

Table 8.7. TRS and Glc yields for each enzyme combination used for the hydrolysis of pretreated spruce. The marked mixture indicates the reaction that produced the highest amount of sugars.
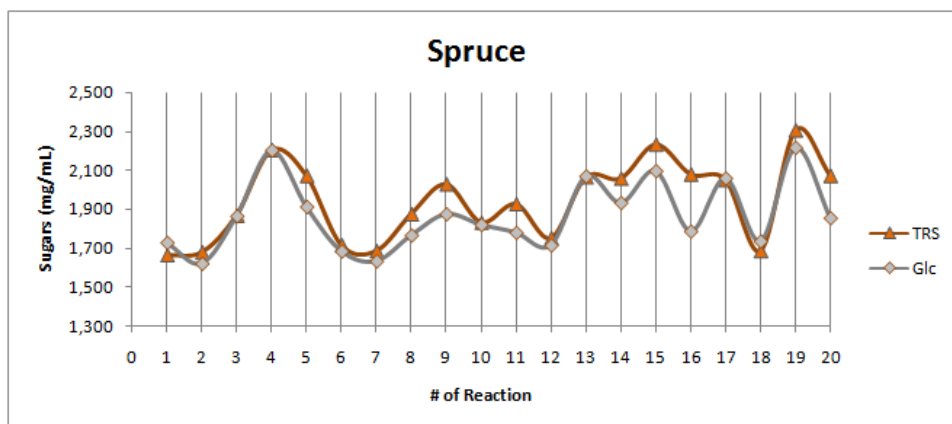
Spruce

Spruce. Total Reducing Sugars (TRS): quadratic, p< 0.0001, R^2= 0.9463

TRS =   - 5.97838 * EG5

       - 0.90743 * EG7

       + 0.25214 * CBH6

       + 0.70942 * CBH7

       + 13.95667 * EG5 * EG7

       + 8.95833 * EG5 * CBH6

       + 15.45688 * EG5 * CBH7

       + 7.90828 * EG7 * CBH6

       + 4.67673 * EG7 * CBH7

       + 3.67017 * CBH6 * CBH7

TRS opt: **EG5** (0.178) **EG7** (0.200) **CBH6** (0.200) **CBH7** (0.422), Yield: **2.10181 mg/mL**

*Spruce, glucose (Glc):* quadratic, p=0.0060, R^2= 0.8364

Glc =   - 5.80615 * EG5

       - 0.42048 * EG7

       + 0.73850 * CBH6

       + 0.59326 * CBH7

       + 12.76669 * EG5 * EG7

       + 7.65752 * EG5 * CBH6

       + 13.82491 * EG5 * CBH7

       + 6.18348 * EG7 * CBH6

       + 4.43330 * EG7 * CBH7

       + 2.83733 * CBH6 * CBH7

Glc opt: **EG5** (0.124) **EG7** (0.275) **CBH6** (0.276) **CBH7** (0.325), Yield: **1.93566 mg/mL**
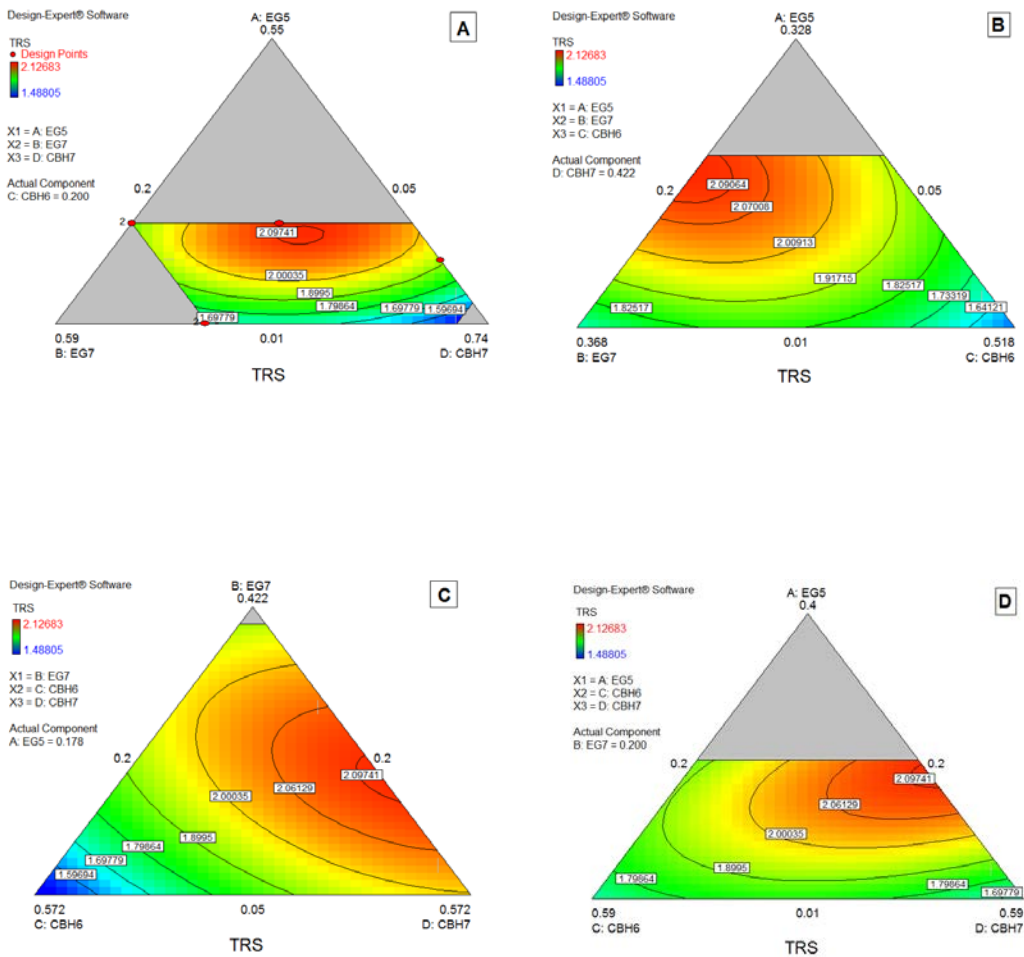
Figure 8.9. Ternary plots showing predicted final Total Reducing Sugars (TRS) yields from spruce hydrolysis, as a function of three out of four "core" enzymes (X1, X2, X3) content. For each plot, the forth enzyme ("actual component") has been fixed to the proportion of the point resulting in the optimal TRS yield, as predicted by the model. *Mt*EG5 and *Mt*CBH7 are the key enzymes for maintaining the highest TRS yield, as highlighted by the optimal domains.
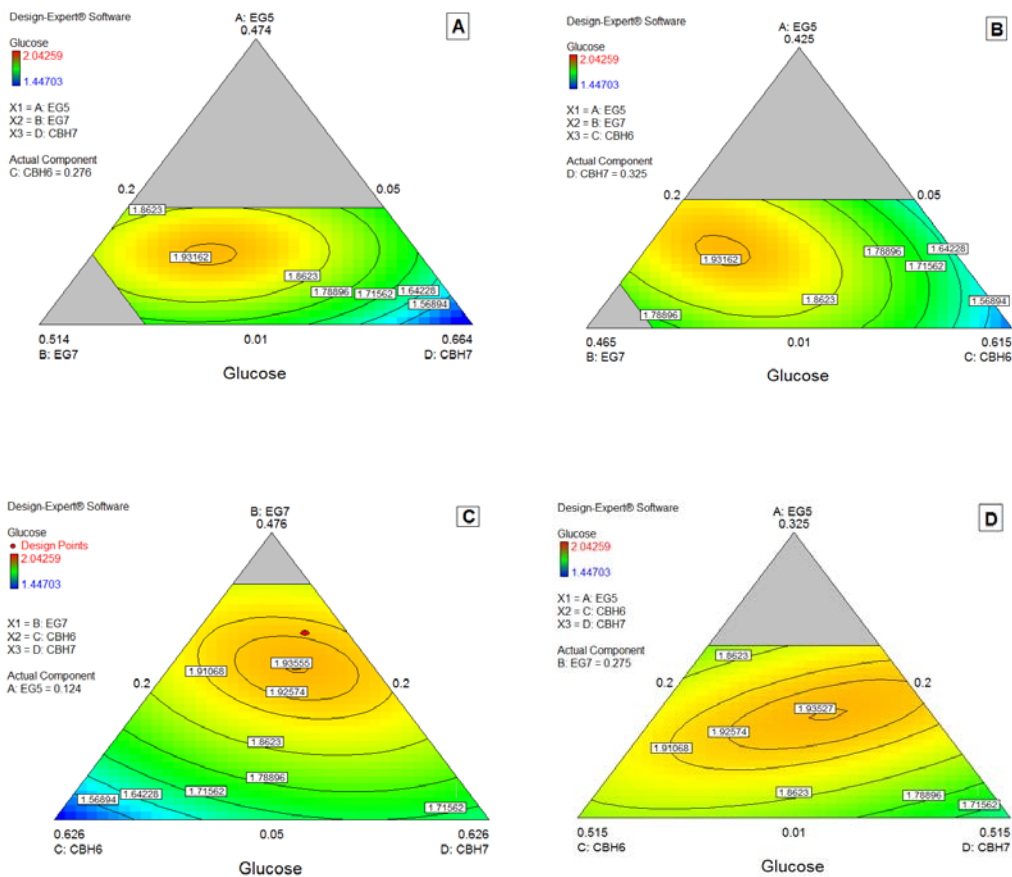
Figure 8.10. Ternary plots showing predicted final glucose (Glc) yields from spruce hydrolysis, as a function of three out of four "core" enzymes content (X1, X2, X3). For each plot, the forth enzyme ("actual component") has been fixed to the proportion of the point resulting in the optimal glucose yield, as predicted by the model. *Mt*EG7 is an enzyme with crucial role for the optimal glucose yield, as the optimal domain in ternary plots A-C is located where higher proportions of this enzyme are used in the enzyme mixture.

time course reaction (TRS optimal)		
incubation time	RS (mg/mL)	% Hydrolysis
0	0,000	0,000
12	1,090	9,938
24	1,519	13,841
48	2,426	22,112

Spruce opt

Reducing sugars (mg/mL) vs incubation time (h)

— Optimal #
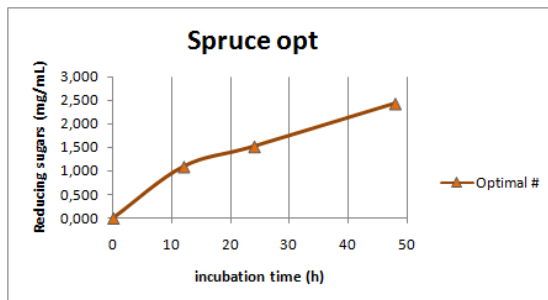
TRS opt: **EG5** (0.178) **EG7** (0.200) **CBH6** (0.200) **CBH7** (0.422), Yield: **2.10181 mg/mL**

Figure 8.5. Time-course hydrolysis using the enzyme combination that was predicted to lead to the highest Total Reducing Sugars production. Experimental yield appear to be higher than the theoretically predicted one, but close to the combination #20, that that produced the highest amount of sugars. The rate of hydrolysis is the highest one during the first 12h of incubation; later it is maintained with a slight decrease, due to the recalcitrance of the forest material.

Core enzyme mixtures were tested for their hydrolysis performance on hydrothermally pretreated with sulphuric acid pine. All enzyme loadings were based on equivalent bovine serum albumin BCA based measurement. *Mt*GH61 and the xylanase cocktail were added as accessory enzymes, as described above. Xylanases' concentration was 2% and was added in combination with 3% *Mt*Man26a, as spruce belongs to softwoods and its main hemicellulose is glucomannan. *Mt*BGL3 was loaded in excess (10% of the total cellulose loading), in order to prevent cellobiose inhibition. The trends for both total reducing sugars (TRS) and glucose (Glc) yields among different enzyme mixtures were dependent on the unique enzyme combinations. % rate of hydrolysis was calculated from the following equation, assuming that the glucose content of wheat straw is 40.7%, according to the literature data (Kilpeläinen *et al.*, 2003.; Korhonen, 1997; Viikki, 1995; Saarela *et al.*, 2005):

$$\% \text{ hydrolysis rate} = \frac{\text{TRS (mg / mL)}}{\text{DM (mg/mL)} * 0.407 * 1.111} * 100$$

	Enzyme proportions						
# React	**EG5**	**EG7**	**CBH6**	**CBH7**	**RS (mg/mL)**	**% Hydrolysis**	**Glucose (mg/mL)**
1	0,06	0,07	0,65	0,21	0,544	4,808	0,467
2	0,20	0,05	0,53	0,22	0,578	5,116	0,492
3	0,01	0,40	0,20	0,39	0,671	5,938	0,608
4	0,12	0,33	0,24	0,31	0,674	5,964	0,596
5	0,20	0,40	0,20	0,20	0,555	4,911	0,458
6	0,01	0,05	0,47	0,47	0,555	4,911	0,532
7	0,01	0,08	0,21	0,70	0,445	3,935	0,442
8	0,09	0,19	0,52	0,20	0,564	4,988	0,502
9	0,03	0,19	0,35	0,44	0,561	4,962	0,496
10	0,01	0,40	0,20	0,39	0,674	5,964	0,610
11	0,01	0,24	0,21	0,54	0,549	4,860	0,454
12	0,01	0,05	0,47	0,47	0,558	4,937	0,525
13	0,01	0,33	0,46	0,20	0,555	4,911	0,471
14	0,20	0,40	0,20	0,20	0,544	4,808	0,457
15	0,20	0,21	0,20	0,38	0,605	5,348	0,486
16	0,13	0,05	0,20	0,62	0,587	5,193	0,499
17	0,01	0,33	0,46	0,20	0,549	4,860	0,469
18	0,06	0,07	0,65	0,21	0,549	4,860	0,471
19	0,11	0,24	0,37	0,28	0,642	5,681	0,520
20	0,20	0,05	0,37	0,38	0,573	5,065	0,469
BLANK					0,059	0,519	0,055

Table 8.8. TRS and Glc yields for each enzyme combination used for the hydrolysis of pretreated pine. The marked mixture indicates the reaction that produced the highest amount of sugars.
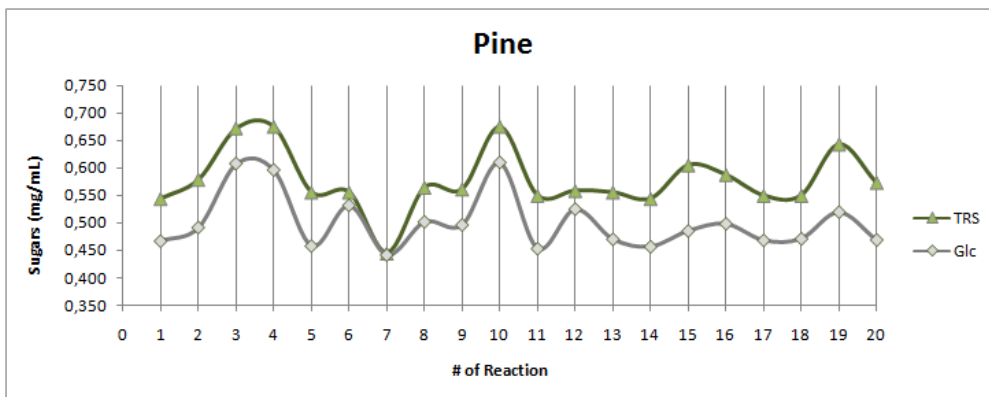
Pine

*Pine, Total Reducing Sugars (TRS):* quandratic, p= 0.0002, R^2=0.9213

TRS =   - 5.65174 * EG5

       + 1.12947 * EG7

       + 0.16687 * CBH6

       - 0.026177 * CBH7

       + 5.41898 * EG5 * EG7

       + 8.20012 * EG5 * CBH6

       + 9.16821 * EG5 * CBH7

       - 0.98642 * EG7 * CBH6

       + 0.47835 * EG7 * CBH7

       + 1.57099 * CBH6 * CBH7

TRS opt: **EG5** (0.073) **EG7** (0.400) **CBH6** (0.200) **CBH7** (0.327), Yield: **0.647236 mg/mL**

*Pine, Glucose (Glc):* special cubic p=0.0002, R^2= 0.9868

Glc =   - 5.34359 * EG5

       + 2.40264 * EG7

       - 8.44275 * CBH6

       + 0.12096 * CBH7

       - 3.37416 * EG5 * EG7

       + 10.06396 * EG5 * CBH6

       + 8.08666 * EG5 * CBH7

       - 2.63309 * EG7 * CBH6

       - 2.36182 * EG7 * CBH7

       + 1.88699 * CBH6 * CBH7

       + 9.58033 * EG5 * EG7 * CBH6

       + 18.79645 * EG5 * EG7 * CBH7

       - 9.96154 * EG5 * CBH6 * CBH7
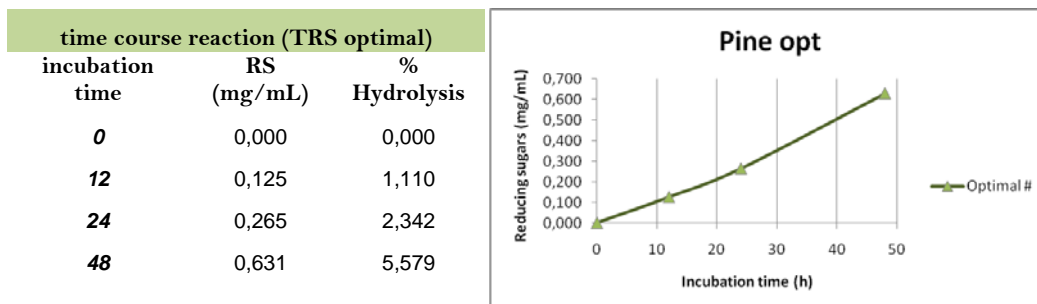
       - 1.17278 * EG7 * CBH6 * CBH7

Glc opt: **EG5** (0.076) **EG7** (0.400) **CBH6** (0.200) **CBH7** (0.324), Yield: **0.611436 mg/mL**
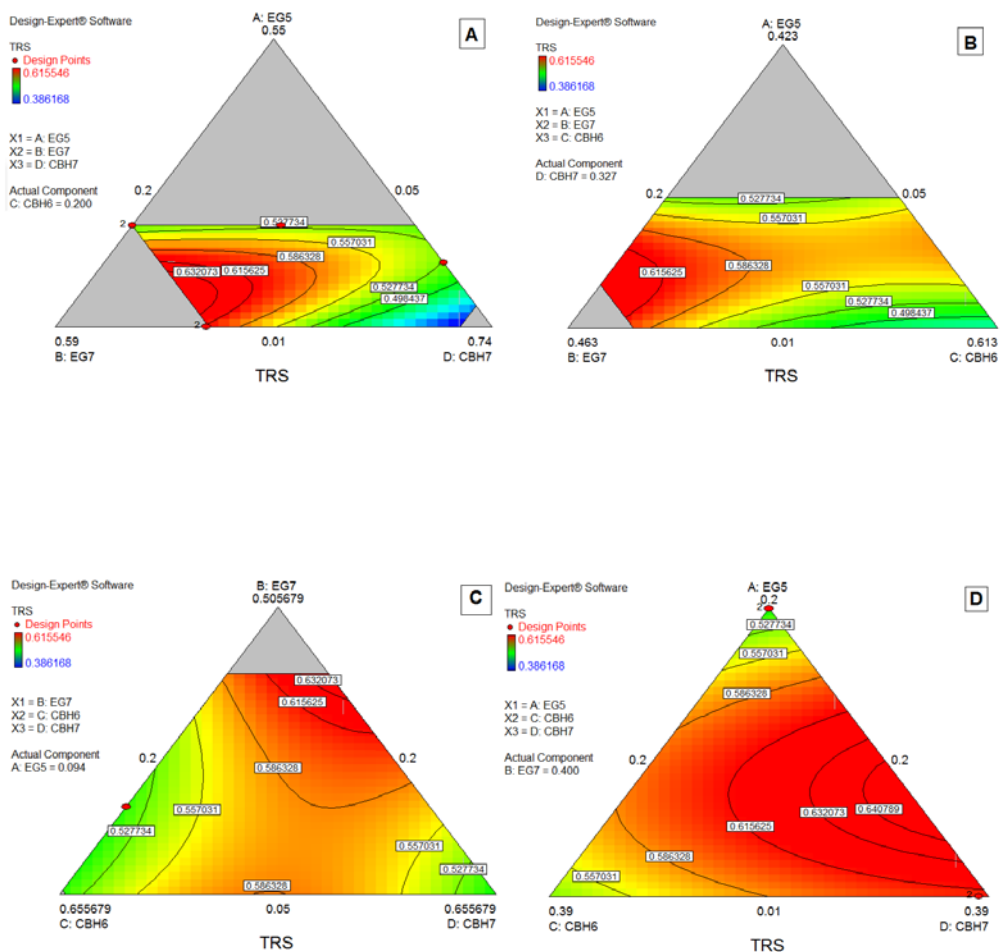
Figure 8.11. Ternary plot showing predicted final Total Reducing Sugars (TRS) yields from pine hydrolysis, as a function of three out of four "core" enzymes (X1, X2, X3) content. For each plot, the forth enzyme ("actual component") has been fixed to the proportion of the point resulting in the optimal TRS yield, as predicted by the model. As illustrated at (A), when *Mt*EG7 decreases, there is a clear reduction of the TRS yield and this reduction is more profound when there is combined decrease of *Mt*EG7 and *Mt*CBH7 enzymes. Within some limits, when moving from points with lower *Mt*CBH7 and higher *Mt*EG7 proportion, hydrolysis yields remain stable; this indicates that the one enzyme can partially compensate for the other one.
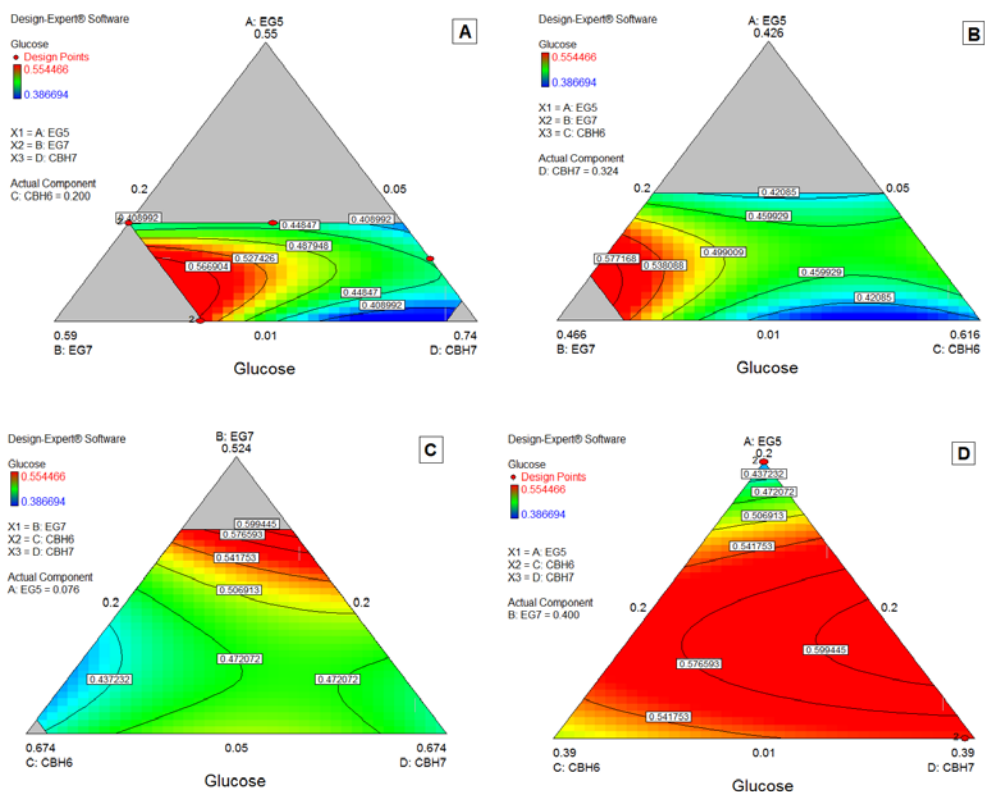
Figure 8.12. Ternary plot showing predicted final glucose (Glc) yields from pine hydrolysis, as a function of three out of four "core" enzymes content (X1, X2, X3). For each plot, the forth enzyme ("actual component") has been fixed to the proportion of the point resulting in the optimal glucose yield, as predicted by the model. In ternary plot (D), there is a clear optimal domain for moderate *Mt*EG5 proportion. When moving upwards along the AD axis within the optimal domain, when CBH6 remains stable, *Mt*EG5 compensates very well for decreasing *Mt*CBH7 ratios. This can be explained at some point in the partial overlapping activities of the two enzymes, as *Mt*EG5 is a processive enzyme, ability that can also attributed to cellobiohydrolases. The same tension can be found at the plot for glucose yields, with a more stretched optimal domain.

time course reaction (TRS optimal)		
incubation time	RS (mg/mL)	% Hydrolysis
0	0,000	0,000
12	0,125	1,110
24	0,265	2,342
48	0,631	5,579

TRS opt: **EG5** (0.073) **EG7** (0.400) **CBH6** (0.200) **CBH7** (0.327), Yield: **0.647236 mg/mL**

Figure 8.13. Time-course hydrolysis using the enzyme combination that was predicted to lead to the highest Total Reducing Sugars production. Experimental yield appear to be similar to the theoretically predicted one. The rate of hydrolysis remains almost stable during the first 48h of hydrolysis.

*8.6. The non-ionic surfactant effect in enzymatic hydrolysis*

Several factors prevent an effective utilization of lignocellulose raw materials in a bioethanol process. The hydrolysis rate rapidly decreases during the time course of hydrolysis which leads to decreased yields and long process times. High enzyme concentrations are needed to reach high cellulose conversion, and enzyme recycling is difficult due to adsorption of enzymes to residual lignocellulose. The addition of surfactants increases yield and the rate of enzymatic hydrolysis, leading to reduced cellulase dosage for the hydrolysis of lignocellulosic biomasses (Borjesson *et al.*, 2007; Kumar and Wyman, 2009; Yang *et al.*, 2011). Different explanations to the surfactant effect on cellulose hydrolysis have been proposed until now, including the effect they could have on enzyme–substrate interactions leading to more effective conversion of cellulose, as well as the ability of surfactants to increase enzyme stability and prevent denaturation of enzymes during hydrolysis.

Lignocellulose conversion to sugar monomers on a commercial scale is hampered by the inhibitory effect of lignin (Nakagame *et al.*, 2010; Lee *et al.*, 2009).

Lignin provides a physical barrier limiting the accessibility of cellulolytic enzymes to the substrate, and the residual lignin could block the removal of the cellulase from the cellulose chain (Alvira *et al.*, 2010). In addition, the non-productive adsorption of lignin on cellulolytic enzymes reduces the productive hydrolysis of the substrate (Kumar *et al.*, 2009). Lignin may also directly inhibit the activities of cellulolytic enzymes (Dyk and Pletschke, 2012). Therefore, studies are focusing on additives that improve the conversion of lignocellulosic feedstock. A large number of reports have stated that surfactants, especially non-ionic surfactants, were the most suitable additives for improving the saccharification of lignocellulose and the recovery of cellulolytic enzymes (Zhou *et al.*, 2013; Cao and Aita, 2013; Ekcard *et al.*, 2013). The central role of surfactants concerns the enzyme-substrate interactions, has no effect on the catalytic mechanism of cellulolytic enzymes, but *affects the adsorption of enzymes on the substrate.* Yang *et al.* reported that non ionic Tween 80 may decrease the non-productive adsorption of cellulase to cellulose (Yang *et al.*, 2011). Another study reports that the increase in conversion when surfactant was added coincided, in all hydrolysis experiments, with a decrease in cellobiohydrolase Cel7A adsorption to the substrate (Eriksson *et al.*, 2002). A likely explanation to these results is that the hydrophobic part of the surfactant binds through hydrophobic interactions to lignin on the lignocellulose fibers and the hydrophilic head group of the surfactant prevents unproductive binding of cellulases to lignin. The hydrophilic parts of the non-ionic surfactants are composed of short ethylene oxide chains. Thus, the adsorption of surfactant on lignin surfaces will prevent unproductive binding of cellulases on lignin. The carbohydrate binding domains (or modules) of most of fungal cellulases have hydrophobic amino acids exposed on the surface, e.g. tyrosines (Kovacs *et al.*, 2009); the presence of these residues on enzyme surfaces is likely to lead to unspecific adsorption to lignin surfaces. Unspecific adsorption of enzymes on lignin could have a stronger role with pretreated substrate compared to native wood samples due to increased exposure of lignin surfaces at the pretreatment process.

One of the typical characteristics of a surfactant is that it can *stabilize the surface tension* in a solution, so it can decrease the surface tension in supernatants, which reduces the energy consumption of hydrolysis and protects the cellulolytic enzymes from deactivation on the surface of the liquid phase, acting as an accelerant for lignocelluloses hydrolysis in a solid-liquid phase system (Feng *et al.*, 2013). As shown

in **Figure 8.4**, the hydrophilic group can combine the cellulolytic enzymes in the liquid supernatant, while the hydrophobic group can bond the solid substrates. Enzymes and substrates might be attracted together by the surfactant, which enhances the adsorption of enzymes and the accessibility of substrates. The surfactant also appears to promote the release of enzymes binding on the substrate by its hydrophobic interaction with lignin, thus enhancing the adsorption and desorption between the enzymes and the substrate, which is one reason for the improved hydrolysis of lignocellulose.
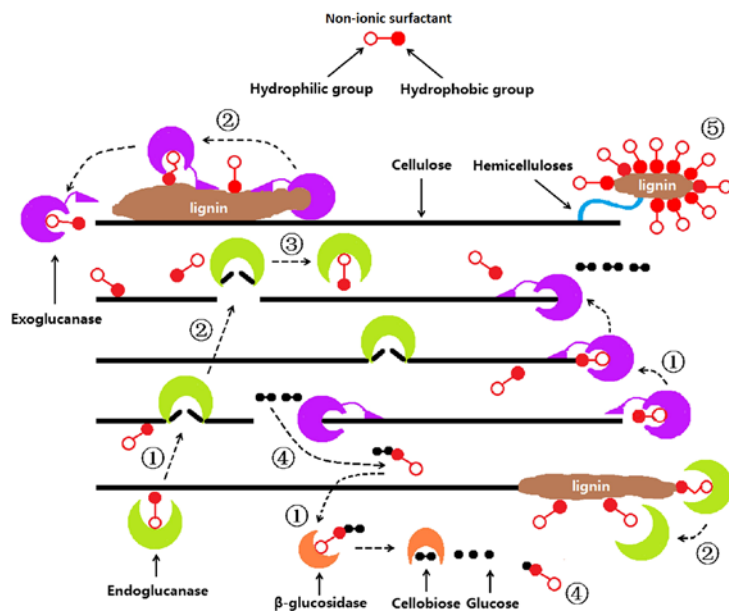
Figure 8.14. Schematic diagram representing the process of lignocellulose hydrolysis with the participation of non-ionic surfactant. Enzymes and substrates are attracted together by the surfactant (1), which promotes the release of enzymes non-productively binding on the substrate (2) and leads to a higher recovery of enzymes at the end of hydrolysis (3). The surfactant homogenizes organic matter in solution with its hydrophilic and hydrophobic groups (4). The hydrophobic interaction between the surfactant and lignin can also enhance the hemicellulose hydrolysis (5) (Feng *et al.*, 2013).

Apart from the effect on enzyme-substrate interactions, surfactants have been reported to improve the stability of the enzymes during hydrolysis. The activities of

many enzymes implicated in the degradation of lignocelluloses (Reese and Mandels, 1980; Gunjikar *et al.*, 2001; Ye *et al.*, 2012) are known to be decreased by shear stress from agitation, either by rotary shaker or impeller, during hydrolysis reaction. The increase in the hydrolysis rate by addition of non-ionic Tween 80 was observed under the agitated condition only, suggesting that the non-ionic surfactant may prevent the decrease of protein concentration brought about by agitation stress (Okino *et al.*, 2013). The most profound effect was observed regarding the cellobiohydrolase CBH2. The decrease in the protein concentration was likely due to the precipitation observed during the experiment, indicating that protein aggregated and precipitated by agitation, resulting in the decrease of protein concentration in the solution. Since aggregation of proteins correlates to their hydrophobicity, charge, and secondary structure propensity (Chiti *et al.*, 2003; Zbilut *et al.*, 2003), the addition of surfactant stabilized the enzyme by reducing the hydrophobicity of enzyme surface in hydrolysis yield under agitated conditions. The addition of non-ionic surfactants in the medium of cellulase production fermentation to stabilize instable components during cellulase production has also been reported, underlining the crucial role of surfactants during cellulase production fermentation as well as in hydrolysis (Kruszewska *et al.*, 1990; Reese and Maguire, 1969).
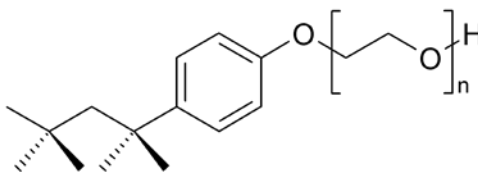
Triton X-100

Triton X-100 ($C_{14}H_{22}O(C_2H_4O)_n$) is a non-ionic surfactant that was used as an additive to the hydrolusis experiments described in this Chapter. It has a hydrophilic polyethylene oxide chain (on average it has 9.5 ethylene oxide units) and an aromatic hydrocarbon lipophilic or hydrophobic group. The hydrocarbon group is a 4-(1,1,3,3-tetramethylbutyl)-phenyl group. Although, together with Tween, Triton surfactants have showed the best improvements of lignocellulose conversion, they are not suitable for large-scale use because of the environmental effects due to the presence of the aromatic ring in the surfactant. Biosurfactants, surface-active substances synthesised

by living cells, are becoming more and more popular for their high efficiency and avirulence (Feng *et al.*, 2013).

8.8. Conclusions

Hydrolysis of the pretreated lignocellulose to simple sugars typically uses a complex of enzymes containing high levels of cellulases, together with lower amounts of enzymes that attack non-cellulosic polysaccharides such as hemicellulose and pectin. Attempts to improve the hydrolytic efficiency of such enzyme complexes have traditionally focused on their component cellulases because cellulose is the most abundant polysaccharide component in lignocellulose. However, it is now recognized that the hydrolytic efficiency of fungal cellulose complexes determined using a model cellulosic substrate (e.g. Avicel, PASC) does not provide a reliable indication of its performance on pretreated lignocellulose (Berlin *et* al., 2007, Kabel *et al.*, 2005). Evidently, other components in pretreated biomass, particularly hemicellulose and lignin, exert significant restraints on cellulose hydrolysis. For example, one mechanism whereby lignin seems to reduce hydrolytic performance is by binding enzyme components non-productively. Consequently enzyme mixtures with similar cellulase activity may show differences in performance on lignocellulose if they differ in affinity for lignin (Berlin *et al.*, 2007). Similarly, it is probable that hemicelluloses restrict the access of cellulolytic enzymes by coating cellulose fibers. In some lignocelluloses, pectin could exert a similar effect. Consequently, enzyme mixtures with similar cellulose activity may show differences in performance on lignocellulose if they differ in hemicellulase composition (Berlin et al., 2005, 2006). However, it should be possible to compensate for deficiencies in these so-called "accessory" enzymes by supplementation of cellulase mixtures with appropriate activities.

As first step towards evaluating this approach for enzyme improvement, we sought to increase the hydrolytic activity of biomass derived from agricultural residues such as *wheat straw*, and forest materials from boreal forests, such as softwoods (*spruce* and *pine)* and hardwoods (*birch*) show significant quantitative and qualitative differences in their non-cellulosic polysaccharide components. A statistical model was set up to search for optimized enzymatic mixtures containing four core enzymes, in the

presence of other four "accessory" enzymes, all encoded by *M. thermophila*'s genes. For industrial applications, more enzymes and higher substrate content than those applied here are needed. It was therefore of interest to test an optimized cellulose mixture under such conditions. The present results suggest that *Mt*CBH7 and *Mt*EG7 are enzymes of major importance for optimized final TRS and Glc yields during the hydrolysis of pretreated *wheat straw* and *pine*, while *Mt*CBH7 plays a crucial role in case of *spruce*. Cellobiohydrolases *Mt*CBH6 and *Mt*CBH7 act in combination and are key enzymes for the hydrolysis of the hardwood (birch); the synergism of these enzymes with xylanases have been reported in the literature (Alvira *et al*., 2011; Qing and Wyman, 2011). For the hydrolysis of the pure substrate (PASC), high proportions of *Mt*EG7 are needed for efficient yields.

Future studies must show if these findings are also true for conditions of high dry matter content and addition of accessory enzymes, conditions which are relevant for enzymatic hydrolysis in industrial applications. It is possible that hydrolysis ratios of the optimized mixtures can be improved when other enzyme components are added. It was shown for instance that xylanases from different families (10 and 11) act synergistically and that their simultaneous presence leads to improvement of glucose yields (Banerjee *et al*., 2010a; Banerjee *et al*., 2010b; Gao *et al*., 2011). In our experiments we used a cocktail of partially purified xylanases from *M. thermophila* culture broth, at very low concentration (2-3%) of the total enzyme loading. Though the addition of enzymes with xylanolytic activity would lead only to minor improvements on steam-exploded wheat straw, as this substrate contains only very little xylan , it can be hypothesized that an additional xylanase (as well as β-xylosidase) would raise the release of reducing sugars from birch (xylan is the dominant hemicelluloses component) and other forest materials. Another candidate for further hydrolysis improvement is the use of feruloyl esterase, as a synergistic effect between cellulases, FAEs and xylanases for the hydrolysis of wheat straw have been proven (Tabka *et al*., 2006 Selig *et al*., 2008). Ferulic acid is the most abundant hydroxy cinnamic acid in the cell wall (Mueller-Harvey and Hartley, 1986) and is covalently cross-linked to arabinoxylans by ester bonds and to components of lignin mainly by ether bonds (Akin *et al*., 1996). Accessory enzymes such as feruloyl esterases should also act in synergy with xylanases by cleaving diferulic bridges between xylan chains, opening the

structures and releasing lignin (Yu *et al.*, 2002; Faulds and Williamson, 1995). The synergistic action of xylanase with cellulases has already been demonstrated in earlier studies using corn stover (Alvira *et al.*, 2011; Qing and Wyman, 2011). Much of the synergism between cellulases and xylanolytic enzymes is believed to expose the cellulose microfibril core, by either removing the hemicellulose or the hemicellulosic side chains (Yu *et al.*, 2003) or reduce the inhibitory effect of xylan and xylo-oligomers on the activity cellulolytic enzymes (Qing and Wyman, 2011).

References

Akin, D.E., Morrison, W.H., Rigsby, L.L, Gamble, G.R., Sethuraman, A., Eriksson, K-EL. 1996. Biological delignification pf plant components by the white rot fungi *Ceriporiopsis subvermispora* and *Cyathus stercoreus. Anim Feed Sci Technol* 63:305–21.

Alén, R, 2000. Structure and chemical composition of wood. In: Stenius P (ed.) Forest Products Chemistry. Jyväskylä, Finland: Gummerus Printing; 2000. p. 11-57.

Alvira P, Tomás-Pejó E, Ballesteros M, Negro MJ. 2010 Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Biores Technol* 101:4851–4861.

Alvira P, Tomas-Pejo E, Negro MJ, Ballesteros M. 2011. Strategies of xylanase supplementation for an efficient saccharification and cofermentation process from pretreated wheat straw. *Biotechnol Progr* 27:944-950.

Banerjee, G., Car, S., Scott-Craig, J.S., Borrusch, M.S., Bongers, M., Walton, J.D. 2010a. Synthetic multi-component enzyme mixtures for deconstruction of lignocellulosic biomass. *Bioresour. Technol.* 101:9097-9105.

Banerjee, G., Car, S., Scott-Craig, J.S., Borrusch, M.S., Walton, J.D. 2010b. Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. *Biotechnol Biofuels* 3:22.

Berlin, A., Maximenko, V., Gilkes, N., Saddler, J. 2007. Optimization of enzyme mixtures for lignocellulose hydrolysis. Biotechnol. Bioeng. 97, 287–296.

Borjesson, J., Engqvist, M., Sipos, B., Tjerneld, F., 2007. Effect of poly(ethylene glycol) on enzymatic hydrolysis and adsorption of cellulase enzymes to pretreated lignocellulose. *Enzyme Microb. Technol.* 41, 186–195.

Cao, S., Aita, G.M. 2013. Enzymatic hydrolysis and ethanol yields of combined surfactant and dilute ammonia treated sugarcane bagasse. *Biores Technol* 131:357–364.

Chiti, F., Stefani, M., Taddei, N., Ramponi, G., Dobson, C.M., 2003. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424, 805–808.

Dyk, JSV, Pletschke BI. 2012. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes factors affecting enzymes, conversion and synergy. *Biotechnol Adv* 30:1458–1480.

Eckard, A.D., Muthukumarappan, K., Gibbons, W. 2013. Enzyme recycling in a simultaneous and separate saccharification and fermentation of corn stover: A comparison between the effect of polymeric micelles of surfactants and polypeptides. *Biores Technol*, 132:202–209.

Eriksson, T., Borjesson, J., Tjerneld, F., 2002. Mechanism of surfactant effect in enzymatic hydrolysis of lignocellulose. *Enzyme Microb. Technol.* 31, 353–364.

Faulds, C.B. and Williamson, G.1995. Release of ferulic acid from wheat bran by a ferulic acid esterase (FAE-III) from *Aspergillus niger. Appl Microbiol Biotechnol* 43:1082–7.

Feng, Y., Jiang, J., Zhu, L., Yue, L., Zhang, J., Han, S. 2013. Effects of tea saponin on glucan conversion and bonding behaviour of cellulolytic enzymes during enzymatic hydrolysis of corncob residue with high lignin content. *Biotechnol Biofuels.* 6:161.

Gao, D.H., Uppugundla, N., Chundawat, S.P.S., Yu, X.R., Hermanson, S., Gowda, K., Brumm, P., Mead, D., Balan, V., Dale, B.E. 2011. Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. *Biotechnol Biofuels,* 4:5.

Gunjikar, T.P., Sawant, S.B., Joshi, J.B., 2001. Shear deactivation of cellulase, exoglucanase, endoglucanase, and beta-glucosidase in a mechanically agitated reactor. *Biotechnol. Prog.* 17, 1166–1168.

Hakkila, P, 1989. Utilization of residual forest biomass. Springer, Germany.

Kilpeläinen, A, Peltola, H, Ryyppö, A, Sauvala, K, Laitinen K, Kellomäki S, 2003. Wood properties of scots pines (*Pinus sylvestris*) grown at elevated temperature and carbon dioxide concentration. Tree Physiology 23(13):889-97.

Korhonen, P, 1997. Kemiallisten pääkomponenttien jakautuminen puuaineksessa ja puukuidun soluseinässä [Master's thesis]. Department of Chemistry, University of Jyväskylä, Finland.

Kovacs, K., Szakacs, G., Zacchi, G. 2009. Comparative enzymatic hydrolysis of pretreated spruce by supernatants, whole fermentation broths and washed mycelia of *Trichoderma reesei* and *Trichoderma atroviride. Biores Technol* 100:1350–1357.

Kruszewska, J., Palamarczyk, G., Kubicek, C.P., 1990. Stimulation of exoprotein secretion by choline and Tween 80 in Trichoderma reesei QM 9414 correlates with increased activities of dolichol phosphate mannose synthase. *Microbiology* 137, 1293–1298.

Kumar, P, Barrett DM, Delwiche MJ, Stroeve P. 2009 Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production. *Ind Eng Chem Res*, 48:3713–3729.

Kumar, R., Wyman, C.E., 2009. Effect of additives on the digestibility of corn stover solids following pretreatment by leading technologies. *Biotechnol. Bioeng.* 102, 1544–1557.

Lee, SH, Doherty TV, Linhardt RJ, Dordick JS. 2009. Ionic liquid-mediated selective extraction of lignin from wood leading to enhanced enzymatic cellulose hydrolysis. *Biotechnol Bioeng* 102:1368–1376.

Mueller-Harvey, I., Hartley, R.D. 1986. Linkage of p-coumaroyl and feruloyl groups to cell-wall polysaccharides of barley straw. *Carbohyd Res* 148:71–85.

Nakagame, S., Chandra RP, Kadla JF. 2010. The characterization and possible role of lignin from steam and organosolv pretreated substrates on enzymatic hydrolysis. In 32nd Symposium on Biotechnology for Fuels and Chemicals 8–20.

Nurmi J 1997. Heating values of mature trees. *Acta Forestalia Fennica* 256:28p.

Nurmi, J. 1993. Heating values of the above ground biomass of small-sized trees. *Acta Forestalia Fennica* 236:30p.

Okino, S., Ikeo, M., Ueno, Y., Taneda, D. 2013. Effects of Tween 80 on cellulase stability under agitated conditions. *Bioresour Technol.* 142:535-9.

Qing, Q., Wyman, C.E. 2011. Supplementation with xylanase and beta-xylosidase to reduce xylo-oligomer and xylan inhibition of enzymatic hydrolysis of cellulose and pretreated corn stover. *Biotechnol Biofuels* 4:18.

Reese, E.T., Maguire, A., 1969. Surfactants as stimulants of enzyme production by microorganisms. *Appl. Microbiol.* 17, 242–245.

Reese, E.T., Mandels, M., 1980. Stability of the cellulase of Trichoderma reesei under use conditions. *Biotechnol. Bioeng.* 22, 323–335.

Saarela, K.E., Harju, L., Rajander, J., Lill, J.O., Heselius, S.J., Lindroos, A., Mattson, K., 2005. Elemental analyses of pine bark and wood in an environmental study. *Science of the Total Environment* 343(1–3):231-4.

Schulein, M.1997. Enzymatic properties of cellulases from *Humicola insolens. J Biotechnol* 57: 71-81.

Selig, M.J., Knoshaug, E.P., Adney, W.S., Himmel, M.E., Decker, S.R. 2008. Synergistic enhancement of cellobiohydrolase performance on pretreated corn stover by addition of xylanase and esterase activities. *Bioresour Technol*, 99:4997-5005.

Sjöström, E., 1993. Wood chemistry: Fundamentals and applications. Academic Press, USA.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., Klenk, D.C. 1985. Measurement of protein using bicinchoninic acid. *Anal Biochem.* 150, 76-85.

Szijarto, N., Siika-aho, M., Sontag-Strohm, T., Viikari, L. 2011. Liquefaction of hydrothermally pretreated wheat straw at high-solids content by purified Trichoderma enzymes. *Bioresour Technol* 102:1968-1974.

Tabka, M.G., Herpoel-Gimbert I., Monod, F., Asther, M., Sigoillot, J.C. 2006. Enzymatic saccharification of wheat straw for bioethanol production by a combined cellulase xylanase and feruloyl esterase treatment. *Enz Microb Technol* 39, 897–902.

Viikki, A., 1995. Puun polysakkaridien käyttäytyminen sulfaattikeitossa [Master's thesis]. Department of Chemistry, University of Jyväskylä, Finland.

Willför, S., Sundberg, A., Hemming, J., Holmbom, B., 2005. Polysaccharides in some industrially important softwood species. *Wood Science and Technology* 39(4):245-57.

Xiros, C, Topakas, E., Christakopoulos, P. 2013. Hydrolysis and fermentation for cellulosic ethanol production. *WIREs Energy Environ* 2, 633-654.

Yang, M., Zhang, A., Liu, B., Li, W., Xing, J., 2011. Improvement of cellulose conversion caused by the protection of Tween-80 on the adsorbed cellulase. *Biochem. Eng. J.* 56, 125–129.

Ye, Z., Hatfield, K.M., Eric Berson, R., 2012. Deactivation of individual cellulose components. *Bioresour. Technol.* 106, 133–137.

Yu P, Maenz DD, McKinnon JJ, Racz VJ, Christensen DA. 2002. Release of ferulic acid from oat hulls by *Aspergillus* ferulic acid esterase and *Trichoderma* xylanase. *J Agric Food Chem* 50:1625–30.

Zbilut, J.P., Colosimo, A., Conti, F., Colafranceschi, M., Manetti, C., Valerio, M., Webber Jr., C.L., Giuliani, A., 2003. Protein aggregation/folding: the role of deterministic singularities of sequence hydrophobicity as determined by nonlinear signal analysis of acylphosphatase and Abeta(1–40). *Biophys. J.* 85, 3544–3557.

Zhang, Y-HP, Lynd, L.R. 2005. Determination of the number-average degree of polymerization of cellodextrins and cellulose with application to enzymatic hydrolysis. *Biomacromolecules.* 6: 1510-1515.

Zhou H, Lou H, Yang D, Zhu JY, Qiu X. 2013. Lignosulfonate to enhance enzymatic saccharification of lignocelluloses: role of molecular weight and substrate lignin. *Ind Eng Chem Res*, 52:8464–8470.