TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

ENGINEERING LIGNOCELLULOLYTIC ENZYMES FOR INDUSTRIAL APPLICATIONS

Engineering of laccase and endoglucanase

Veera Nikoskelainen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. F OSA – TOM. 52 | TECHNICA – INFORMATICA | TURKU 2025

University of Turku

Faculty of Technology
Department of Life Technologies
Biochemistry
The Doctoral Programme in Technology (DPT)

Supervised by

Docent, Jarmo Niemi Department of Life Technologies University of Turku Turku, Finland	Klara Birikh, Ph.D. R&D Director MetGen Geleen, Netherlands

Reviewed by

Docent, Maarit Lahtinen Department of Food and Nutrition University of Helsinki Helsinki, Finland	Docent, Max Lönnfors Åbo Akademi University Turku, Finland

Opponent

Docent Nina Hakulinen
Department of Chemistry and
Sustainable Technology
University of Eastern Finland
Joensuu, Finland

ISBN 978-952-02-0074-9 (PRINT)
ISBN 978-952-02-0075-6 (PDF)
ISSN 2736-9390 (Painettu/Print)
ISSN 2736-9684 (Sähköinen/Online)
Painosalama, Turku, Finland 2025

UNIVERSITY OF TURKU
Faculty of Technology
Department of Life Technologies
Biochemistry
VEERA NIKOSKELAINEN: Engineering lignocellulolytic enzymes for industrial applications: Engineering of laccase and endoglucanase
Doctoral Dissertation, 194 pp.
Doctoral Programme in Technology (DPT)
February 2025

ABSTRACT

Energy consumption, especially electricity demand, is continually rising, while non-renewable fossil energy sources such as petroleum, natural gas, and coal are depleting. Fossil energy production also causes greenhouse gas emissions, which accelerate climate change. Efforts are being made to reduce reliance on these energy sources by increasing wind, water, and geothermal energy production. However, additional measures are still needed to fill the resulting energy gap. This shortfall can be addressed by biofuels, which can be produced from biomass. The world's most common biomass is lignocellulose, which consists of wood, forest residues, peat, straw, and agricultural biomass.

Lignocellulose has a complex molecular structure, making it difficult to break down. After pretreatment, it is broken into its main components—cellulose, hemicellulose, and lignin—which are further broken down into molecular fractions or fully soluble molecules. These can then be processed to produce bio-based chemicals and products. This process is called lignocellulose valorization, where the biomass is catalytically "upgraded" into new raw materials or products. Waste streams from existing industrial processes using lignocellulose are also being repurposed more efficiently. For example, lignin waste from paper mills, previously burned for energy, is now used as raw material for new products.

Enzymes are proteins that act as biological catalysts in chemical reactions. Using enzymes in industry can reduce energy and chemical consumption, making processes more sustainable, economically profitable, and environmentally friendly. Industrial processes are large-scale, so enzyme production must be as efficient as possible to keep production costs and industrial use affordable. Additionally, industrial processes require remarkably stable and resilient enzymes, especially under extreme heat and pH conditions. Enzyme engineering can improve production yield and enzyme properties to meet industrial needs.

This work focused on developing two bacterial enzymes - laccase (from *Bacillus wacoensis*) and endoglucanase (from *Spirochaeta thermophila*) - and their further utilisation in industrial processes. Public databases were searched for gene sequences of enzymes that were expected to have the desired catalytic activity. These sequences were synthesised, cloned, and produced using the standard industrial microbe *Escherichia coli* for the first time. The enzymes were characterised to determine their specific activities and thermal and pH stability under different conditions. The laccase enzyme was stable in alkaline, high-pH conditions and retained its activity, but its soluble yield in *E. coli* was very low. The goal of enzyme engineering was to increase the solubility of laccase and further improve its activity. As a result of the engineering, both the yield and activity were significantly improved, allowing the enzyme to be utilised in lignin depolymerisation at high pH and to separate small molecular lignin fractions of different sizes. Thanks to the successful development process, the enzyme's industrial use became feasible, and it remains unique for its laccase properties in lignin depolymerisation.

The second enzyme studied was endoglucanase, which had a good production yield in *E. coli* but was not stable enough at high temperatures for use in many industrial processes. Its engineering aimed to improve thermal stability and further increase activity. As a result of the engineering, its specific activity was significantly increased, and its substrate specificity was altered. However, no improvement in thermal stability was achieved. Endoglucanase is currently used as part of an enzyme cocktail for cellulosic fibre modification processes, but its engineered properties still need to be improved to replace existing alternative enzymes sustainably.

KEYWORDS: Lignocellulosic biomass, enzyme engineering, laccase, endoglucanase, biomass valorization, lignin depolymerisation, thermostability, MetZyme® LIGNO™, MetZyme® BRILA™, industrial biocatalysts, renewable energy, METNIN technology.

TURUN YLIOPISTO
Teknillinen tiedekunta
Bioteknologian laitos
Biokemia
VEERA NIKOSKELAINEN: Engineering lignocellulolytic enzymes for industrial applications: Engineering of laccase and endoglucanase
Väitöskirja, 194 s.
Teknologian tohtoriohjelma (DPT)
Helmikuu 2025

TIIVISTELMÄ

Energian kulutus, etenkin sähkön kulutus, kasvaa jatkuvasti ja samanaikaisesti uusiutumattomat fossiiliset energialähteet, kuten maaöljy, maakaasu ja hiili, ovat ehtymässä. Fossiiliset energian tuotantomuodot aiheuttavat lisäksi kasvihuonepäästöjä, jotka osaltaan kiihdyttävät ilmastonmuutosta. Näiden käyttöä pyritään vähentämään lisäämällä tuuli-, vesi- ja maalämmön tuotantoa mutta samanaikaisesti tarvitaan vielä lisätoimia paikkaamaan syntyvää energiavajetta. Tämä vaje voidaan korvata biopolttoaineilla, joita voidaan valmistaa biomassasta. Maailman yleisin biomassa on lignoselluloosa, joka koostuu mm. puustosta, metsätähteestä, turpeesta, oljesta ja viljelybiomassasta.

Lignoselluloosa on molekyylirakenteeltaan monimutkainen ja vaikeasti hajotettava. Esikäsittelyn jälkeen se hajotetaan pääkomponenteikseen selluloosaksi, hemiselluloosaksi ja ligniiniksi, jotka edelleen hajotetaan osin molekyylifraktioiksi tai täydellisesti liukeneviksi molekyyleiksi, jotka pystytään jatkojalostamaan biopohjaisten kemikaalien ja tuotteiden valmistukseen. Tätä kutsutaan lignoselluloosan valorisoinniksi eli biomassa katalyyttisesti ”rikastetaan” uusiksi raaka-aineiksi tai tuotteiksi. Myös olemassa olevien lignoselluloosaa hyödyntävien teollisuusprosessien jätevirtoja pyritään tehokkaammin hyödyntämään, kuten paperitehtaiden jäljelle jäävä ligniini, joka energiaksi polttamisen sijaan pyritään hyödyntämään tuotteiden raaka-aineina.

Entsyymit ovat proteiineja, jotka toimivat biologisina katalyytteinä kemiallisissa reaktioissa. Hyödyntämällä entsyymejä teollisuudessa, voidaan vähentää energian ja kemikaalien tarvetta tehden prosesseista kestävämpiä, taloudellisesti kannattavampia ja ympäristöystävällisempiä. Teollisuuden prosessit ovat mittakaavaltaan suuria, jonka takia entsyymien tuoton pitää olla mahdollisimman tehokasta, jotta tuotantokustannukset ja siitä seuraava kustannus teollisuuskäytössä on kannattavaa. Lisäksi teollisuuden prosessit tapahtuvat olosuhteissa, jotka vaativat entsyymeiltä erityistä kestävyyttä sekä lämmön- että pH- ääriolosuhteiden takia. Entsyymikehityksen avulla pyritään parantamaan sekä entsyyminen tuotannon saantoon sekä entsyymin ominaisuuksiin vastaamaan teollisuuden tarpeita.

Tässä työssä keskityttiin kahden bakteeriperäisten entsyymin, lakkaasin (*Bacillus wacoensis*) ja endoglukanaasin (*Spirochaeta thermophila*), kehitysprosessiin ja niiden edelleen hyödyntämiseen teollisuudessa käytettyihin prosesseihin. Julkisista tietokannoista etsittiin oletettavasti halutun katalyyttisen aktiivisuuden omaavien entsyymien geenisekvenssit, jotka syntetisoitiin, kloonattiin ja tuotettiin ensimmäistä kertaa yleisesti käytössä olevalla teollisuusmikrobin *Escherichia coli*-bakteerilla. Tuotetut entsyymit karakterisoitin määrittämällä niiden spesifiset aktiivisuudet ja olosuhdekestävyydet. Lakkaasi-entsyymi oli hyvin kestävä alkalisissa, korkean pH:n, olosuhteissa ja säilytti hyvin aktiivisuutensa mutta sen liukoinen tuotto *E.coli*- bakteerissa jäi hyvin alhaiseksi. Entsyymikehityksen tavoitteena oli lisätä lakkaasin liukoisuutta sekä edelleen parantaa sen aktiivisuutta. Kehitystyön ansioista lakkaasin tuottoa *E.coli*- bakteerilla pystyttiin kasvattamaan ja samanaikasesti sen aktiivisuutta saatiin parannettua huomattavasti. Paranneltua lakkaasi-entsyymiä hyödynnettiin korkeassa pH:ssa liuotetun ligniinin depolymerointiin ja eri kokoisia pienimolekyylisiä ligniinifraktiota pystyttiin tämän johdosta erottelemaan toisistaan. Kehitysprosessin ansioista entsyymin hyödyntäminen teollisissa prosesseissa tuli kannattavaksi ja se on edelleen ainutlaatuinen lakkaasi ominaisuuksiltaan ligniinin depolymeroinnissa.

Toinen tutkittava entsyymi oli endoglukanaasi jonka saanto *E.coli*- bakteerissa oli hyvä mutta se ei ollut lämmönkestävyydeltään riittävä moniin teollisiin prosesseihin. Sen kehitystavoitteena oli parantaa lämmönkestävyyttä sekä lisätä edelleen myös sen aktiivisuutta. Kehitystyön ansiosta endoglukanaasin ominaisuuksista spesifinen aktiivisuus nousi sekä sen substraattispesifisyys muuttui. Lämmönkestävyyttä ei pystytty näillä toimilla parantamaan. Endoglukanaasia hyödynnetään tällä hetkellä osana entsyymikoktailia, jota käytetään selluloosakuidun muokkausprosesseissa. Työssä kehitetty endoglukaasi ei ole ominaisuuksiltaan kuitenkaan vielä sitä tasoa, että se kannattavasti korvaisi jo olemassa olevia vaihtoehtoisia entsyymejä.

AVAINSANAT: Lignoselluloosabiomassa, entsyymin kehitys, lakkaasi, endoglukanaasi, biomassan valorisointti, ligniinin depolymerointi, lämmönkestävyys, MetZyme® LIGNO™, MetZyme® BRILA™, METNIN teknologia, teolllinen biokatalyysi, uusiutuva energia.

Table of Contents

Abbreviations

AA	Amino acid
ABTS	2,2′-Azino-bis(3-ethylbenzthiazoline-6-sulfonic acid)
BLAST	Basic Local Alignment Search Tool
CSF	Canadian Standard Freeness
CMC	carboxymethyl cellulose
DMP	2,6-dimethoxyphenol
DNA	deoxyribonucleic acid
DNS	3,5-dinitrosalicylic acid
dNTP	nucleoside triphosphate
IPTG	isopropyl β-D-1-thiogalactopyranoside
kDa	kilodalton
LB	Luria-Bertani broth
LMS	laccase-mediator system
NT	nucleotide
OD	optical density
ORF	open reading frame
PCR	polymerase chain reaction
PDB	Protein Data Bank
PSI-BLAST	Position-Specific Iterated BLAST
RNA	ribonucleic acid
RT	room temperature
SB	super broth
SEC	specific energy consumption
SDM	site-directed mutagenesis
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SGZ	syringaldazine
TB	terrific broth
tRNA	transfer RNA
WT	wild type

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Birikh K.B.& Hämäläinen* V. ALKALINE LACCASE VARIANTS WITH IMPROVED PROPERTIES. WO20193452A1

II Hämäläinen V.*, Grönroos T., Suonpää A., Heikkilä M.W., Romein B., Ihalainen P., Malandra S., Birikh K.B. Enzymatic Processes to Unlock the Lignin Value. Frontiers in Bioengineering and Biotechnology., 2018 Mar 22;6:20.

III Birikh K.B. & Hämäläinen* V. ENDOGLUCANASE VARIANTS WITH IMPROVED PROPERTIES. WO2019145288.

IV Hämäläinen* V., Dios Barajas-López J., Berlina Y., Álvarez-Rafael R.& Birikh K.B. New thermostable endoglucanase from Spirochaeta thermophila and its mutants with altered substrate preferences. Applied Microbiology and Biotechnology, 2021 Feb;105(3):1133-1145

V Birikh K.B., Heikkila M.W., Michine A., Mialon A., Grönroos T., Ihalainen P., Varho A., Hämäläinen* V., Suonpää A. and Rantanen S-P. From the book Biocatalysis: An Industrial Perspective. CHAPTER 11, MetGen: Value from Wood – Enzymatic Solutions. The Royal Society of Chemistry 2018: Catalysis Series No. 29

*Nikoskelainen

The original publications have been reproduced with the permission of the copyright holders.

1 Introduction

Energy consumption and energy demand are constantly increasing. Non-renewable energy resources, especially fossil fuels, are running out, raising serious concerns about climate change and the accumulation of waste. Renewable energy sources can be an alternative to overcoming the energy shortage by using environmentally friendly and sustainable means. While wind, water, solar, and geothermal can supplement the energy industry, biomass is needed to replace fuel production and the chemical industry (Raj *et al.*, 2022).

Lignocellulosic biomass is the most abundant and renewable biomass on earth. It is a promising and economical alternative to producing biofuels, biochemicals, and bioproducts that can reduce CO_2 emissions and air pollution. (Isikgor and Becer, 2015a). Lignocellulosic biomass consists of forest and agricultural biomass and food crops produced by photosynthesis from readily available atmospheric CO_2, water, and sunlight. (Zhou *et al.*, 2011).

First-generation biofuels are based on the fermentation of sugar cane and corn starch. They have been shown to reduce greenhouse gas emissions. Still, as they require large land areas for cultivation and compete with food crops, they are neither economically nor environmentally sustainable (Hill *et al.*, 2006). Second-generation biofuels are based on non-food feedstocks such as lignocellulosic biomass and industrial by-products and, therefore, do not compete with the food supply (Balat, Balat and Öz, 2008; Thompson and Meyer, 2013). The abundance of available lignocellulosic biomass makes it a cost-effective resource for producing energy, chemicals, and bio-based polymers using novel biotechnological approaches. However, due to the recalcitrant nature of lignocellulose and high processing costs, it is necessary to develop better lignocellulolytic enzymes to make the hydrolysis of lignocellulosic biomass economical and sustainable (Chukwuma *et al.*, 2020).

The use of enzymes in industrial applications makes processes more sustainable and environmentally friendly, as fewer chemicals and less energy are required. The improved methods and applications also lead to innovations and products. However, for this to be economically feasible, the process benefits need to compensate for the cost of the enzyme and modifications of the process. In industrial-scale processes, a large volume of enzymes is required. The enzyme's production must be optimized to minimize costs, ensuring maximum yield. With enzyme engineering, one can increase productivity, enhance enzyme activity, and improve stability under extreme temperature and pH conditions (Contreras et al., 2020).

This work involves applying lignocellulolytic enzymes and engineering two enzymes, laccase and endoglucanase. The novel enzymes were cloned and produced for the first time, and their specific activities and stabilities were characterized. Laccase had excellent stability in the alkaline pH range but very low soluble expression. The goal of engineering was to increase the enzyme's solubility and activity. After successful improvement, this laccase was used to depolymerize lignin at very alkaline pH, and low molecular weight lignin fractions of different sizes could be separated. Another recently discovered enzyme investigated in this work was endoglucanase. After characterization, the goal for engineering was both to improve activity and to increase thermal stability. Both rational and semi-rational engineering approaches were used, the enzyme activity was improved, and further engineering was used to change the substrate specificity of the enzyme.

2 Review of the Literature

2.1 Lignocellulosic biomass

Lignocellulosic biomass consists of forest residues, hardwood and softwood, agricultural biomass, crops, and perennial grass. The structural composition of lignocellulosic biomass varies according to plant components (even for the same species), biomass type, nutrient status, soil fertility, harvesting method, and geographical region. There are mainly three types of polymers in lignocellulosic biomass: Cellulose (35-50%), hemicellulose (25-30%) and lignin (25-30%) along with small amounts of other components such as acetyl groups, minerals and phenolic substituents, as shown in Figure 1 (Wongwilaiwalin et al., 2010; Zhou et al., 2011). Cellulose and hemicellulose are different sugars, while lignin is an aromatic polymer. These polymers are linked together to form a complex, non-uniform three-dimensional structure whose composition varies depending on the type of lignocellulosic biomass (Isikgor and Becer, 2015a). These structural characteristics and complexity make lignocellulosic biomass recalcitrant to degradation and depolymerization, making the development of processes to convert it into fuels and value-added chemicals a challenge (Galbe and Zacchi, 2007).

Cellulose is the major component of lignocellulosic biomass and is responsible for its mechanical strength (Anwar, Gulfraz, and Irshad, 2014). It is estimated that nearly half of the organic carbon in the biosphere is present in the form of cellulose (Zhou *et al.*, 2011). Cellulose is a linear polymer composed of glucose linked by β-1,4-linkages. The properties of cellulose depend on its degree of polymerization (DP), which is defined as the number of monomeric glucose units in a polymer molecule. Its structure consists of extensive intramolecular and intermolecular hydrogen bonding networks that firmly link the glucose units together. The glucose monomers are rotated 180 degrees along the polymer, meaning the repeating unit is the disaccharide cellobiose. Cellulose occurs both in crystalline and amorphous form in lignocellulosic biomass. It is completely insoluble in water and very difficult to hydrolyze under natural conditions. (Malherbe and Cloete, 2002; Nanda, A. Kozinski and K. Dalai, 2015)

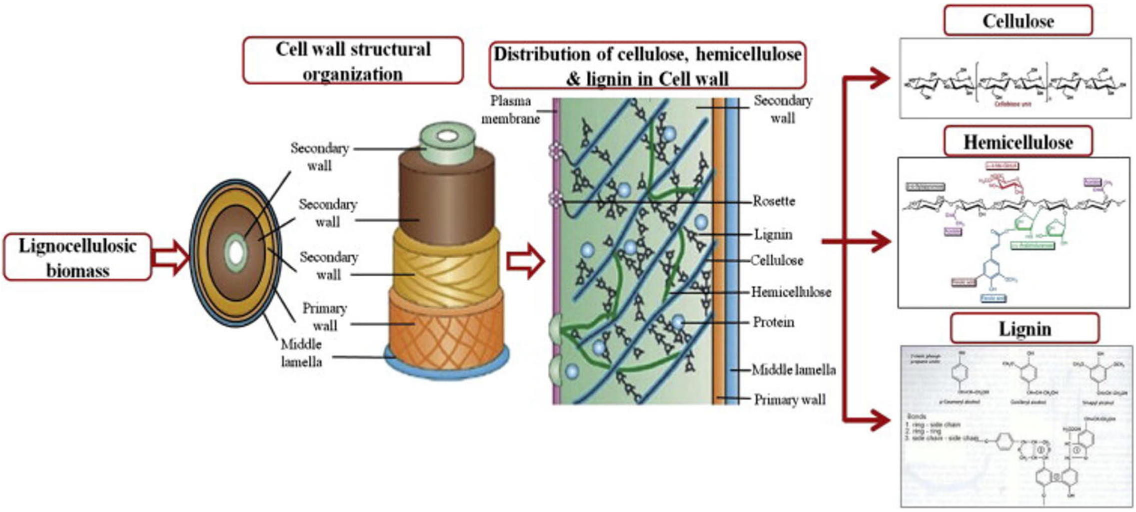


Figure1. Diagrammatic illustration of the framework of lignocellulose. Cellulose, hemicellulose, and lignin polymers are linked together, creating a complex structure that makes lignocellulose challenging to break down for bioconversion processes.(Menon and Rao, 2012) .

Hemicellulose is the second most common polymer with a random and amorphous structure. It is a branched heteropolymer consisting of pentose (D-xylose and D-arabinose) and hexose sugars (D-mannose, D-glucose, and D-galactose). Hemicelluloses differ in their composition depending on their origin; hardwood hemicelluloses mainly contain xylans, while softwood hemicelluloses mainly contain glucomannans (Isikgor and Becer, 2015b). Hemicelluloses are embedded in plant cell walls and form a complex network of bonds that provide structural strength by linking cellulose fibers in microfibrils and cross-linking with lignin. (Malherbe and Cloete, 2002; Somerville et al., 2010; Fatma et al., 2018).

Lignin is a class of complex, cross-linked, three-dimensional biopolymers with aromatic properties. It is a heterogeneous and hydrophobic biopolymer composed of phenylpropanoid subunits, namely p-hydroxyphenyl (H-type), guaiacyl (G-type), and syringyl (S-type) units which are linked together via various C–C or C–O bonds (Zhou *et al.*, 2023). Lignin is unusual as a biopolymer due to its heterogeneity and lack of a defined primary structure. It provides strength to plant tissue and fibers and stiffness to the cell wall, functioning as the cellular glue. (Malherbe and Cloete, 2002; Menon and Rao, 2012)

2.2 Lignocellulolytic enzymes and their applications

2.2.1 Lignocellulolytic enzymes

Proteins are biological macromolecules that are essential parts of the structure and function of all organisms. They are encoded by genes that are translated into amino acid polypeptide chains that fold into a specific three-dimensional structure. Enzymes are one type of protein essential for all life forms, as their catalytic power enables life processes and is an essential component of every living organism. They break down and build molecules such as nutrients, accelerating all vital biological processes. Enzymes have numerous applications in the food, medical, chemical, and pharmaceutical industries. More than 5000 enzymes are known, and about 200 enzymes of microbial origin are used commercially (Yoo *et al.*, 2017).

Enzymes are highly selective biological catalysts that accelerate chemical reactions by lowering the activation energy required to convert substrate molecules into various product molecules. Enzymes can accelerate conversion rates by a factor of 10^5-10^8- fold compared to spontaneous chemical reactions. One of the most important properties of enzymes is their activity, which determines how fast the reaction occurs, and it is influenced by their structure and their environment. The International Union of Biochemistry (IUB) categorizes enzymes into seven different classes based on the enzyme's mechanism of action. The seven enzyme classes are ligases, isomerases, oxidoreductases, lyases, transferases, translocases, and hydrolases. (Rigoldi *et al.*, 2018; Wilson, David B.; Sahm, Hermann; Stahmann, Klaus-Peter; Koffas, 2020).

Lignocellulolytic enzymes are a group of biocatalysts that are able to break down lignocellulosic biomass into its various monomer units, making them accessible for bioconversion into various products and have applications in a wide range of industries. Lignocellulolytic enzymes include ligninolytic (peroxidases and oxidases) and hydrolytic enzymes (cellulases, hemicellulases, pectinases, chitinases, amylases, proteases, esterases and mannanases) (Chukwuma et al., 2020). A wide variety of enzymes are required to break down all the components of lignocellulose, as it is very recalcitrant due to the complex chemical composition of lignin, hemicellulose, and acetyl groups, as well as the high crystallinity and degree of polymerization of cellulose (Contreras *et al.*, 2020).

2.2.1.1 Cellulases

Cellulases are among the most important industrial enzymes known to date. They belong to the glycoside hydrolase (GH) family, which can hydrolyze the β-1,4-Dglucan bonds in cellulose to yield glucose, cellobiose and oligosaccharides. The complete hydrolysis of cellulose requires the interaction of three types of cellulases: Endoglucanase (EG), which randomly cleaves the internal β-1,4-glycoside bonds; cellobiohydrolase (CBH), which acts processively on the chain ends to release cellobiose; and β-glucosidase (BGL), which hydrolyzes cellobiose to glucose. The synergistic action of these three cellulase components is essential for opening the reducing and non-reducing ends (Zheng et al., 2018). Some cellulases are modularly organized and consist of a catalytic domain (CD) linked to a carbohydrate-binding domain (CBD), which supports enzymatic action on solid substrates and influences substrate specificity (Contreras et al., 2020).

Endoglucanases are key elements in various industrial applications. In cellulose hydrolysis, endoglucanases provide entry points for the most abundant cellulolytic enzymes - exo-glucanases. It has been shown that endoglucanase activity can be a limiting factor in the hydrolysis of cellulose by natural fungal cellulase cocktails (Gao et al., 2017). Endoglucanases are also used in the pulp and paper industry, as they can improve the properties of recycled pulp and biological deinking processes. By modifying the cellulose fibers with endoglucanase, the microfibrils are realigned and the carbohydrate chains are better aligned, increasing the paper's strength (Chiriac et al., 2010; Gao et al., 2017). In the production of nanocellulose, endoglucanases break down or weaken the fibers in the amorphous areas, thus enabling energy savings in the refining process (Hu *et al.*, 2018; Ribeiro *et al.*, 2019).

2.2.1.2 Hemicellulases

Hemicellulases comprise a diverse group of enzymes that catalyze the hydrolysis of hemicelluloses to monomeric sugars, mainly xylose, arabinose, mannose, and galactose. These enzymes target the glycosidic bonds present in the hemicellulose polymers and cleave them into soluble oligosaccharides and, finally, into monosaccharides. Hemicellulases often work synergistically with other enzymes, such as cellulases and ligninases, to efficiently degrade lignocellulosic biomass. Synergistic interactions between xylanases, arabinofuranosidases, mannases, and galactanases facilitate the complete hydrolysis of hemicellulose polymers (Benatti and Polizeli, 2023). Several factors influence the efficiency of hemicellulose hydrolysis, including enzyme concentration, pH, temperature, substrate composition, and pretreatment methods. Optimization of these parameters is essential to maximize enzyme activity and substrate conversion rates. Protein engineering and directed evolution techniques are currently being used to improve the catalytic efficiency, substrate specificity, and stability of hemicellulases   (Juturu and Wu, 2012).

Hemicellulases play a crucial role in the enzymatic hydrolysis of lignocellulose, as the degradation of hemicellulose improves the overall yield of biomass conversion. Simultaneous depolymerization of cellulose and hemicellulose is crucial to achieve complete saccharification of lignocellulosic biomass. The selective hydrolysis of hemicelluloses yields platform sugars that can be converted into a wide range of bioproducts. In a lignocellulosic biorefinery, hemicellulose polymers can be molded into membranes, coating films, and hydrogels as well as hemicellulose sugars that can serve as platform chemicals (Østby and Várnai, 2023).

2.2.1.3 Lignin-modifying enzymes

Lignolytic enzymes are specialized enzymes that degrade lignin, a complex and recalcitrant polymer in lignocellulosic biomass. These enzymes are primarily produced by fungi (white-rot and brown-rot fungi), bacteria, and some actinomycetes (Pollegioni, Tonin and Rosini, 2015). They play a crucial role in breaking down lignin into smaller phenolic compounds, enabling its utilization in industrial and environmental applications. Lignin has a high molecular weight and is insoluble in water, which makes it structurally complex and contributes to it being difficult to break down. In nature, lignin is degraded by oxidative enzymes such as ligninolytic peroxidases and laccases, which are secreted by microbes. Peroxidases rely on hydrogen peroxide (H_2O_2) as a co-substrate, which, on an industrial scale, can lead to increased costs and safety concerns. For this reason, laccase is preferred as it requires dioxygen (O_2) as a co-factor, which is cheap and ubiquitous, and the chemical environment is milder (Chukwuma *et al.*, 2020).

Laccases are multicopper oxidases that catalyze the oxidation of various phenolic substrates using molecular oxygen as an electron acceptor. A small-molecule mediator [e.g., 1-hydroxybenzotriazole (HBT), 4-hydroxybenzoic acid, 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid), ABTS], can be paired with a laccase to develop a laccase mediator system (LMS) that enables the oxidation of non-phenolic components (Zhou *et al.*, 2023). This expands the laccase's substrate range and promotes lignin degradation. After laccase oxidation by electron abstraction, the small-molecule mediator diffuses out of the laccase's catalytic pocket and oxidizes the lignin's non-phenolic subunit. (Cañas and Camarero, 2010). Laccase typically targets the ether bonds between lignin units, particularly the β-O-4 linkages, which are the most abundant and flexible. The enzyme’s action leads to the depolymerization of lignin, breaking it into smaller phenolic compounds or oligomers that are more suitable for valorization. Depolymerized lignin fragments can be further processed into value-added products such as aromatic chemicals, biofuels, and biopolymers, contributing to sustainable biorefinery platform development (Chukwuma et al., 2020). Laccases can also be used in a variety of different applications, e.g. as medical diagnostic tools and biosensors, as cleaning agents in water treatment systems, and as catalysts in the production of drugs (Rigoldi *et al.*, 2018).

2.2.2 Application in the pulp and paper industry

The pulp and paper industry processes huge quantities of lignocellulosic biomass every year. The use of enzymes in the pulp and paper industry has great potential for improving process efficiency and reducing environmental impact. The pulp and paper industry uses harsh chemicals and bleaching agents and, at many stages, produces harmful by-products that have a significant environmental impact. Using enzymes in pulp production reduces energy consumption, and fewer chemicals can be used. Common enzymes used in bio-pulping include cellulases, laccases, and xylanases (Singh *et al.*, 2016). Enzymes are used in several applications like bio bleaching, fiber modification, wastewater treatments, deinking and removal of pitch, shives, and slime and help to reduce energy consumption in some large-scale applications (Bajpai, 1999).

The pre-treatment of wood chips with enzymes before refining opens the cell wall and thus separates the fibers. After enzymatic treatment, less chemicals and energy are required for refining. As the wood chips are treated at high temperatures and alkaline pH values, the enzymes used must be active and stable under very harsh conditions. Lignin residues are responsible for the color of the pulp, and removing lignin from pulps is called bleaching. Various enzymes are used in the bleaching process. Xylanases are widely used as they attack the hemicellulose and facilitate the release of lignin from the cellulose; laccases oxidize and depolymerize lignin, and cellulase treatment improves the dewatering of the pulp, all resulting in improved gloss and brightness of the paper (Benatti and Polizeli, 2023). Xylanases and endoglucanases have also been utilized to enhance pulp fibrillation, improve water retention, and reduce the defibration time of virgin pulp, which is made directly from freshly harvested trees or plants. As the supply of virgin pulp for paper production has decreased significantly, the paper industry uses more and more recycled paper and paper waste. With recycled fibers, enzymes like cellulases and hemicellulases help break down the surface layers of fibers where ink particles are trapped. Endoglucanases and other cellulases also modify fiber surfaces by creating new fibrils, increasing the surface area available for hydrogen bonding, resulting in enhanced pulp quality with better water retention and improved paper sheet formation. (Singh *et al.*, 2016).

2.2.3 Valorization of lignocellulose

As the components of lignocellulose are linked together in a complex way, biocatalytic degradation requires pretreatment to break the interaction between lignin and polysaccharides and allow solubilization of cellulose, hemicellulose, and lignin, thus improving the digestibility of the lignocellulosic material. There are physical, chemical, and physicochemical pretreatment strategies, all of which are very energy and cost-intensive. The overall compatibility of raw materials, enzymes, and organisms to be used should be considered when choosing the pre-treatment strategy (Galbe and Zacchi, 2007; Menon and Rao, 2012).

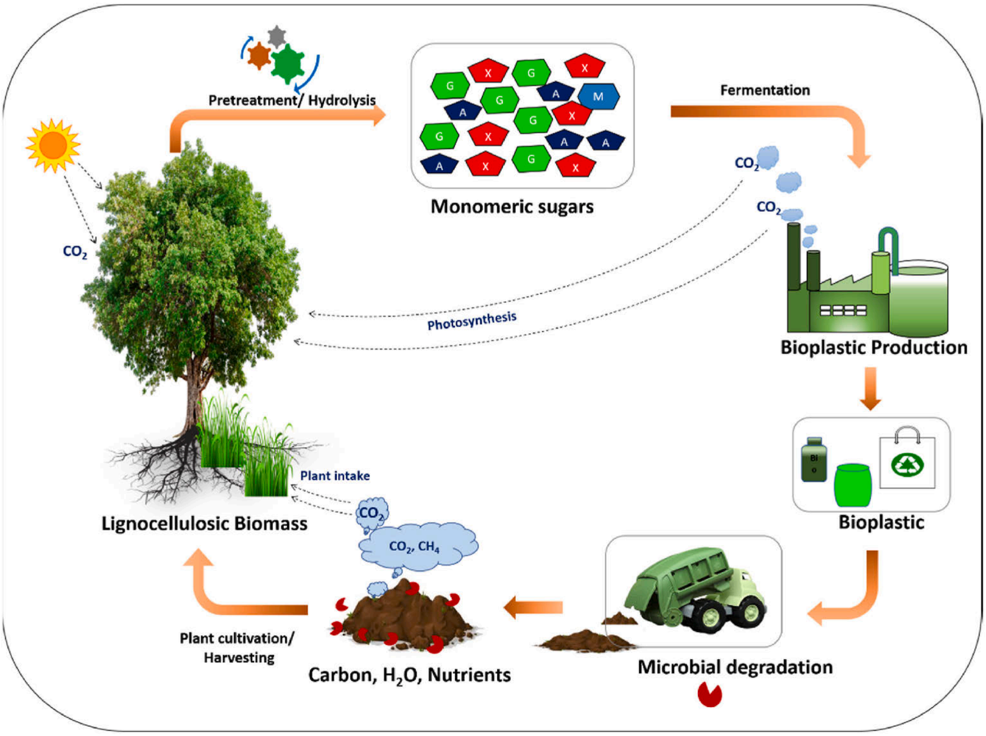


Figure 2. Closed loop biorefinery approach for bioplastic production. Sugars from the enzymatic hydrolysis of lignocellulosic biomass can be fermented or biocatalytically converted into renewable platform chemicals, which can be processed into bioplastics. Bioplastics are biodegradable, eco-friendly, and align with green chemistry principles, enhancing environmental sustainability. (Raj *et al.*, 2022)

The production of biofuels is one of the important alternatives to fossil fuels since they are renewable and produce less CO_2, thus slowing down climate change. Second-generation biofuels are made from inedible lignocellulosic biomass, including agricultural and food processing residues. After pretreatment, the efficient degradation of lignocellulosic biomass requires the synergistic action of several lignocellulolytic enzymes (Contreras et al., 2020). Since the structural composition of lignocellulosic biomass varies depending on plant components, biomass type, nutrient conditions, soil fertility, harvesting method, and geographical region, complete hydrolysis of the material requires thorough investigation and optimization to find the most suitable enzyme cocktail for hydrolysis. Sugar is further fermented to produce biofuels such as biogas, bioethanol, or biohydrogen (Benatti and Polizeli, 2023). The sugars obtained after the enzymatic hydrolysis of lignocellulosic biomass can also be converted through fermentation and biocatalytic processes into promising renewable platform chemical molecules such as lactic acid, succinic acid, ethanol, formic acid, and 5-(hydroxymethyl)-furfural, that can be further processed into bioplastics (Chukwuma et al., 2020, Wongwilaiwalin *et al.*, 2010). Bioplastics are biodegradable, environmentally friendly, and naturally comply with green chemistry principles while improving environmental benefits. Figure 2 shows a green approach in the context of a circular biorefinery.

Depolymerization is considered one of the critical challenges in lignin valorization (Raj *et al.*, 2022). Native lignin is referred to as the original lignin structure in lignocellulose, but it does not exist as such as it is in nature (Chio, Sain, and Qin, 2019; Iram, Cekmecelioglu and Demirci, 2021). When lignin is extracted from biomass, the structure is always altered and differs from the original form. It is considered a by-product of the Pulp and paper industry, and most of it is still burned for energy. The challenge in using lignin as a raw material for low molecular weight chemicals is its heterogeneity since the proportions of the constitutive building blocks and the cross-linking patterns vary based on the type of biomass and extraction methods. Lignin depolymerization involves breaking various types of bonds, a process heavily influenced by the chosen method (Abdelaziz et al., 2016a). Successful depolymerization of lignin leads to a series of substituted phenols that can be further converted into various high-value products, including aromatic chemicals and other marketable products such as adhesives, resins, bioplastics or carbon fibers (Iram, Cekmecelioglu and Demirci, 2021). The lignin valorization process is visualized in Figure 3.

2.3 Developing industrial enzymes

Industrial enzymes play a role in many areas, including the pharmaceutical industry, the chemical industry, the food industry, and other processes based on biochemical reactions. The use of enzymes as biocatalysts in industry has both environmental and production cost benefits. Enzymes can replace various harmful chemicals and can be used as detoxifying agents. Production can occur under milder conditions, saving energy and chemicals (Benatti and Polizeli, 2023). Since the applications are on a

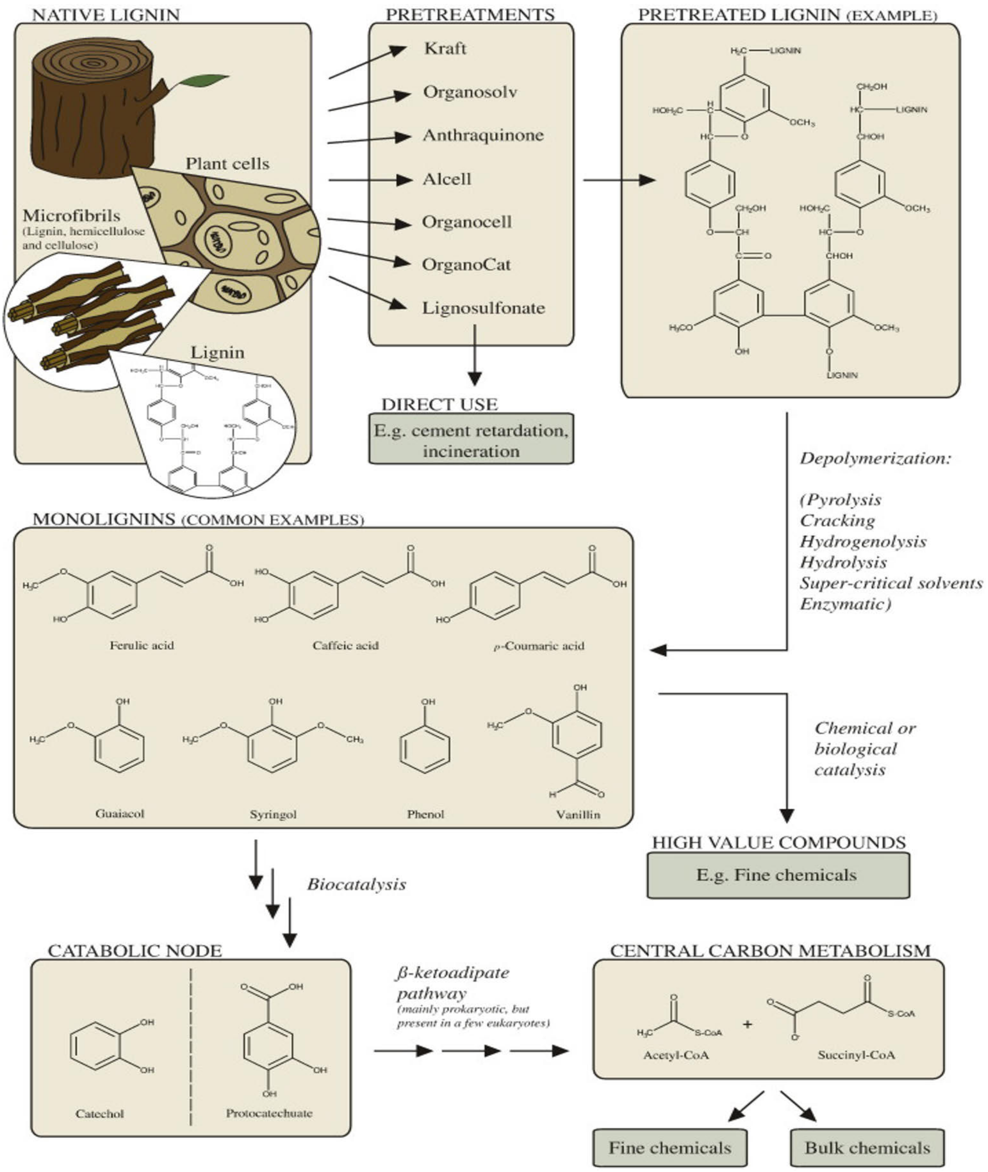


Figure 3. Lignin valorization to chemicals. A central process in the proposed strategy for lignin valorization is the breakdown of the lignin feedstock into compounds that can be further bioconverted by microorganisms. The depolymerization will need to yield low molecular weight lignins (mono- and oligomers), as only compounds of this size can serve as substrates for further cellular assimilation (Abdelaziz *et al.*, 2016)

very large scale, the volume of enzymes needed is also large. Large-scale operations may require hundreds of kilograms to several tons of enzyme product per month, depending on production volume and process needs. To make processes economically feasible, the production of enzymes needs to be cost-effective. This is accomplished by optimizing high production yield and simultaneously developing the enzymes to be as stable and active as possible. The enzymes produced by microbial strains are economical due to high production levels, ease of growth, inexpensive culture media, and short fermentation cycles (Rigoldi *et al.*, 2018).

The use of enzymes for industrial applications often requires harsh conditions. Usually, industrial enzymes that are stable in such environments originate from extremophiles whose habitats are characterized by high temperature, pressure, pH, or salt concentration, such as volcanic vents, hot springs, or the deep sea (Wilson, David B.; Sahm, Hermann; Stahmann, Klaus-Peter; Koffas, 2020). To improve the efficiency, thermostability, enzyme-substrate specificity, or activity of enzymes, genetic engineering technology has been applied to change the amino acid sequence of enzymes (Yoo *et al.*, 2017). Strategies like directed evolution, rational or semi-rational design, and metabolic engineering are used to refine the efficiency and stability of natural enzymes (Fasim, More and More, 2021).

2.3.1 Protein Production Optimization

There are several ways to improve protein yield, such as selecting the production strain type, optimizing the medium and nutrients needed during the process, temperature and oxygenation, pH, induction system and time, and genetic modifications that affect the actual expression and protein folding. (Wilson et al., 2020). The goal is to get as much biomass as possible from the raw materials and, at the same time, get individual cells to produce as much recombinant protein as possible. Parameters like optimal pH, temperature, agitation speed, and incubation time need to be systematically optimized to maximize the metabolite yield (Singh *et al.*, 2017). With medium optimization, the appropriate medium components like energy, carbon, nitrogen, and minerals can be defined (Singh *et al.*, 2017; Fasim, More and More, 2021). On an industrial scale, lower cost chemically undefined complex growth medium is usually used (Wilson, David B.; Sahm, Hermann; Stahmann, Klaus-Peter; Koffas, 2020). In aerobic fermentation processes, the oxygen rate needs to be sufficient to satisfy the growth needs of the organism. Aeration is one of the most critical factors for the efficiency of the fermentation process and is achieved by supplying the bacterial culture with adequate efflux of oxygen, resulting in enhanced bacterial growth and production of recombinant protein (Wilson, David B.; Sahm, Hermann; Stahmann, Klaus-Peter; Koffas, 2020).

The role of the environment in recombinant protein production is just one aspect of the optimization, as various genetic elements also contribute to the process. These include codon optimization, induction system selection, cell line modifications, targeting signals, and possible chaperone molecules to improve folding. All cell types naturally tend to use a certain type of codon. Codons are nucleotide triplets that code for certain amino acids. There are 64 different codons, including a few codons for initiation and termination of translation, and the rest coding for 20 different amino acids. Even though the genetic language is universal, the codon usage and the ratio of matching tRNAs may still vary a lot between cell types, and this can affect the recombinant gene expression. If there is a high number of so-called rare codons, the translation can slow down dramatically (Angov, 2011). Codon optimization is often used when the recombinant gene produced is from another organism. This means that the sequence is formatted in such a way that it is codon usage is optimal for the production strain. Sometimes rare codons might be beneficial to slow down the translation rate so that the protein has time to fold correctly (Chaney et al., 2017)

Recombinant protein production can be driven by constitutive promoter or an inducible promoter. A promoter is an element responsible for the initiation and regulation of gene transcription and can directly control the level of recombinant gene expression in a host cell. There are three different types of promoters commonly used. Constitutive promoters remain active under in vivo conditions, inducible promoters require external stimuli to initiate gene expression, and stationary-phase promoters become active during the stationary phase while showing little to no activity during the growth phase.(Firat Duzenli and Okay, 2020).

With inducible systems, the timing is critical for optimal expression, and with the inducing molecule or with thermoinducible systems, the temperature change must be done in the exponential growth phase with predefined density (e.g. OD 600 nm) (Larentis *et al.*, 2014). If the induction is done too early, the cells focus more on protein production than growth, and the overall biomass yield remains low, causing low protein yields. Then again, if the induction is done too late, the cells have already used most of the nutrients for growth and cannot produce that much protein anymore. With optimal timing, cells produce protein, but some growth remains. (Choi, Keum and Lee, 2006; Rosano and Ceccarelli, 2014; Wilson, David B.; Sahm, Hermann; Stahmann, Klaus-Peter; Koffas, 2020). Although inducible promoters show effective overexpression of the gene of interest, they are not widely accepted in industrial protein production because of the cost of the inducer and associated host-toxic properties (Firat Duzenli and Okay, 2020). Several different induction systems are used in procaryotic and eucaryotic cells. Constitutive promoters are also used in yeast and mammalian cells so that protein production is initiated right after transfection, and no additional change in culture is needed. The most commonly used inducers for bacterial cells are IPTG or lactose, which activates the lac promoter, and arabinose, which induces the pBAD promoter.

Autoinduction is an option to be used to induce *lac* operon with lactose instead of IPTG (isopropyl-b-D-thiogalactopyranoside). In the autoinduction method, the presence of glucose inhibits transcription of the recombinant genes from lactose-inducible promoters while it also serves as the carbon and energy source for the culture medium (Studier, 2005; Ding *et al.*, 2017). Depletion of glucose causes induction of *lac* operon and the consecutive utilization of lactose. Glycerol is added to the sugar mixture to be utilized in parallel with lactose as a supporting energy and carbon source. The oxygenation state of the culture affects the preference order for carbon sources and should be considered when optimizing production using autoinduction. Oxygen-limited conditions contribute to earlier onset of lactose consumption and protein expression (Ukkonen *et al.*, 2013). The benefit of the autoinduction system is that there is no need to follow the cell cultures to determine the correct induction window in recombinant protein production, which is essential when screening mutant enzyme libraries during the engineering process.

Production is also affected by temperature. Usually, the temperature is decreased after the induction which is a way of controlling the growth and production rate. Recombinant proteins have complex structures, and it takes some time to achieve the correct fold. The protein folding is further improved by lowering the temperature so that the folding peptide chain has time to find the proper structure. But then again, if the temperature is too low, the process will be slowed down, increasing the cost. There might also be "catabolite repression" that may prevent the induction, for example, the high concentrations of glucose, and at least in a process where glucose is continuously fed, the concentration must be kept low enough not to cause repression. Dissolved oxygen also affects the sugar metabolism of *Escherichia coli* and must be considered in process optimization. (Choi, Keum and Lee, 2006; Wilson et al., 2020)

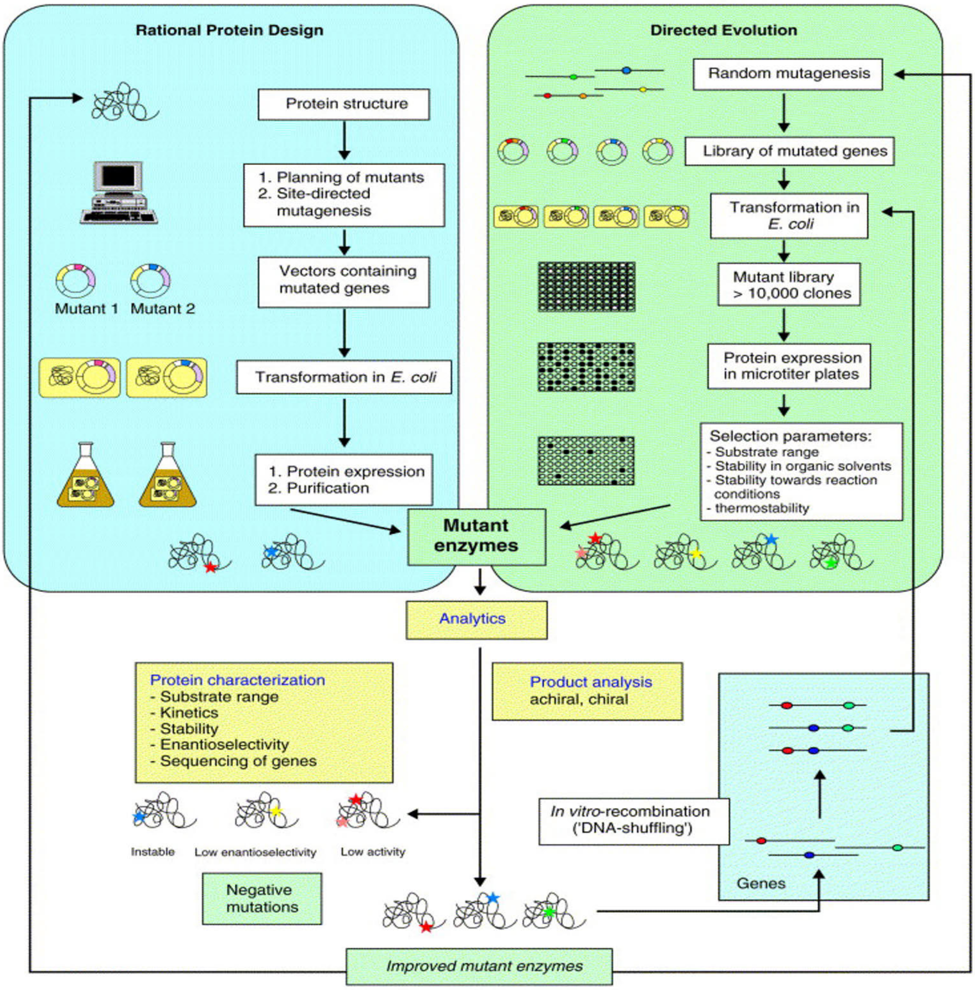
The localization of the produced protein inside the cells may also impact productivity. With a targeting genetic element called signal peptide, the recombinant protein production can be addressed to the selected environment. Some proteins are more likely to be better produced and folded inside the cytosolic space of the bacterial cells, while some prefer periplasmic space. This cannot be predicted, and it needs to be tested for each recombinant protein. Chaperones are proteins or parts of proteins that are used to improve recombinant protein folding during expression. A chaperone is usually cloned in front of a recombinant protein and expressed as a fusion protein (Wilson et al. 2020). If the recombinant protein structure is known to have an internal cysteine bridge, internal bonds between two cysteine amino acids, this needs to be considered. The cytoplasm of wild-type *E. coli* is not permissive to forming disulfide-bonded proteins due to numerous reductases and reducing agents such as glutathione. A protein that requires disulfide bonds for proper folding must be secreted to the periplasm, where oxidoreductases facilitate the pairing of cysteines, or it can be produced in genetically modified strains that co-express oxidoreductases in the cytosol. Disulfide bonds play structural roles that are important for the folding and stability of a protein, and lack of proper fold may lead to lower yield and aggregation of the protein. This method offers the advantage of simplified purification, as only the outer membrane needs to be lysed. Osmotic shock involves exposing bacterial cells to a rapid change in osmotic pressure, typically by transferring them from a high-salt or isotonic solution to a low-salt or distilled water solution, which causes lysis and the release of periplasmic contents. Several strains like Gram-positive bacterial as well as yeast and mammalian cells can be used to produce recombinant protein secreted in the culture medium (Wilson et al., 2020)

Production yields can also be improved with cell line engineering. This means that genetic alterations are made in the production strain. Modifications can include either protease knockouts or beneficial genes being knocked in. Proteases can cause recombinant protein proteolysis before the product is collected, and many of the production strains have been modified by knocking out the protease genes. (Wilson, David B.; Sahm, Hermann; Stahmann, Klaus-Peter; Koffas, 2020)

2.3.2 Enzyme engineering

Enzyme engineering means modifying the protein's genetic code to improve either the productivity or making desired modifications to the enzyme’s functionality. Alterations can be minor, such as only changing single nucleotides to optimize codons based on the relativity of production strain codon usage to improve protein folding and productivity. Single amino acid changes are achieved by changing whole codon triplets to code for desired amino acids, and in the case of mutation libraries, several single or multi-amino acid changes are introduced into the selected residues simultaneously. Also, more significant structural changes can be achieved by making sequence deletions, insertions, or replacements. Protein engineering is widely used for creating enzymes with outstanding performance like improved activity, selectivity, or stability, which are essential for industrial applications. Protein engineering approaches are directed evolution, semi-rational design, rational re-design, and even novel, totally synthetic proteins designed with computer software. (Bansal and Kundu, 2022; Xu *et al.*, 2023). Figure 4 presents the difference and the workflow of rational design and directed evolution in protein engineering.

To increase the recombinant protein's expression yield, altering the production strain and optimizing the gene sequence can be assessed. If the expression levels are good enough but the product is insoluble, meaning it aggregates in inclusion bodies to the cytosol, improved solubility can be achieved with engineering. Rational and computational approaches or semi-rational designs have become more favorable for engineering enzyme-substrate specificity because of the constant accumulation of enzyme structural data and knowledge of enzyme catalysis. This approach provides comprehensible information to understand the structure-function relationships involved in enzyme specificity (Yoo *et al.*, 2017).



Current Opinion in Chemical Biology

Figure 4. Comparison of directed evolution and rational design when using *E. coli* as a host strain. In rational design, mutants are planned based on protein structure, prepared via site-directed mutagenesis, expressed in a host (e.g., *E. coli*), purified, and analyzed for desired traits. Directed evolution involves creating mutant libraries through random mutagenesis, expressing them in a host, and screening for desired properties. DNA shuffling or in vitro recombination can refine results. Both approaches can be iterated or combined to develop biocatalysts with optimized properties. (Bornscheuer and Pohl, 2001)

2.3.2.1 Rational and semi-rational approaches

Rational design methods are computational bioinformatics based methods on knowledge, utilizing large sequence datasets, protein homology modeling, and existing crystal structures. Approaches use data on protein sequence, structure, and/or function, along with computational predictive algorithms, to identify potential target sites for protein engineering. Rational design relies on identifying critical amino acid residues that govern protein function. Usually, the basis of the design is initiated by comparing the target sequence to a pool of homologous sequences. The modifications may be site-directed mutagenesis with single amino acid changes or small libraries when focusing on specific amino acid positions. The approach is usually much faster than random mutagenesis or massive library screenings. (Wilson, 2015; Korendovych, 2018; Xu *et al.*, 2023). A semi-rational approach that combines the rational and random approach can generate synergistic effects by taking advantage of the rational and random approach by reducing the number of mutations and screening time because the libraries generated are more focused on selected positions of the sequence (Yoo et al., 2017).

In multiple sequence alignment (MSA) the conservation of amino acid residues can be assessed through aligning sequences of homologous proteins. This is used to evaluate the possible so-called “hot spots”, meaning the sequence positions where the sequence of the target enzyme is different in the conserved areas of the sequence set. In this approach, it is believed that a residue appearing more frequently at a given position is more likely to contribute to the folding and stability of the enzymes. Rational mutations are then introduced to enhance desired properties or create novel functions. If all other protein sequences in the set have the same amino acid variation as the target enzyme, a point mutation is applied, and if there are several different amino acids in that position among the alignment, a library of various amino acids (semi-rational approach) can be considered. (T. Wang et al., 2020; Bansal and Kundu, 2022)

Protein solubility is also affected by protein surface charges and hydrophilicity of the residues’ side chains. Usually, hydrophobic areas are well-packed in the protein core, and more hydrophilic amino acids are located on the protein surface. Thus, several hydrophobic residues close to each other, i.e., hydrophobic batches, on the protein surface can cause aggregation (Ventura, 2005). MSA can be utilized for both improving solubility through enhanced folding and stability of the enzymes. Increasing rigidity is usually aimed at introducing disulfide bridges, hydrogen bonding, or salt bridges between the residues for stabilizing protein structure. For example, in mobile loops, intramolecular bonding reduces the freeness of mobility (Rigoldi *et al.*, 2018). Selection of data sets used in characterization is usually selected based on the desired characteristics (Katoh and Standley, 2013). For example, if we want to improve the stability in alkaline conditions, we are more likely to succeed if we compare our protein to a group of alkaline stable proteins and try to figure out which of the areas in sequence are common in stable proteins and differing from the protein to be engineered (T. Wang et al., 2020). Zheng et al., 2018 used MSA when engineering cellulase GtCel5 from *Gloeophyllum trabeum* CBS 900.73. They were able to identify a key residue with MSA at position 233 that was found to have effects on both thermal adaptation and catalytic efficiency of GtCel5 (Zheng *et al.*, 2018). Alignment-based design was also used in a study where SOD (superoxide dismutase) from *Thermus thermphilus* HB27 was engineered to alter binding specificity (Wang *et al.*, 2009). Based on MSA, the substituting mutations D167E and I166L-D167E were designed, which changed the binding specificity of SOD from Mn to Fe ion.

Very often the engineering goal is to improve the enzyme activity so that less enzyme is needed to catalyze the modifications of the substrate. In this case, it is good to study all already known modifications of similar kinds of enzymes. Sometimes, beneficial modifications can be detected and applied to the other enzymes in the same protein family, especially if they are very homologous. Computer modelling-based methods are used when engineering enzyme activity or substrate specificity. Computational methods, such as molecular docking, homology modelling, and protein-protein interaction predictions, are used to guide the rational design of protein variants with improved performance. There are already >200 000 experimentally determined and >1 000 000 computationally defined three-dimensional (3D) structures in The Protein Data Bank (PDB), which is the single global structural data archive of biological macromolecules (Burley *et al.*, 2019). Very often, with novel proteins, the structure of the protein is not available, and engineering starts by doing a homology model of the protein structure. Several software for homology modelling is available, like Chimera Modeller (Pettersen *et al.*, 2004), Swiss-model server (Waterhouse et al., 2018), and AlphaFold (Jumper *et al.*, 2021).

To define the substrate binding site, the active site of the modelled enzyme, the structure can be superimposed with a known homologous enzyme that has a crystal structure of enzyme bound with the substrate. This can be done with one or several structures, and with good software, the relevant residues for binding are detectable. With these comparisons, the individual point mutations, libraries, or bigger alterations in the sequence can be designed to modify either enzyme activity or even substrate specificity.

Rational design can also be based on changes in the protein domains. Protein domains are the basic units of proteins that can fold, function, and evolve independently (Wang et al., 2021).Many enzymes have several domains, which may have specific functions, such as binding and catalytic domains. Domains can sometimes be successfully inserted, deleted, or changed to another domain from the same protein family. Chiriac et al., 2010 designed and characterized a series of truncated derivatives of endoglucanase Cel9B from *P. barcinonensis,* which have two different cellulose-binding domains (CBD). They did variants were either one of the CBDs or both were deleted. The deletion of other CBD had a major impact on cellulose binding in both amorphous and crystalline substrates, while the other only had an impact on the latter. The stabilities of all the truncated forms were reduced (Chiriac *et al.*, 2010)

2.3.2.2 Directed evolution

Directed evolution is a method used in protein engineering that mimics the natural evolution process and does not demand in-depth knowledge of the structural information (Tracewell, Cara, Arnold, 2009). In this method, the parental gene is manipulated by random mutations via error-prone mutagenesis, DNA shuffling, ultraviolet radiation, or by using chemicals. Clones having novel or improved features are selected, and the cycle of mutation screening is repeated until a particular activity or another feature is reached. The campaign begins with creating a diverse pool of genetic variants to introduce changes in the target protein sequence. The aim is to explore a wide range of mutations to identify those that confer improved functionality. The size and quality of the library are crucial for its success. Since the resulting library size in a directed evolution campaign can be very large, it is essential to have automated high-throughput screening methods, such as plate-based assays or fluorescence-based detection, to identify and isolate those mutants with the desired properties. The directed evolution campaign typically involves multiple library generation, screening, and characterization rounds. Beneficial mutations identified in early cycles serve as templates for further rounds of mutagenesis or recombination. After several rounds of directed evolution, the best-performing variant is further optimized and validated for the intended application. This may involve scaling up expression, optimizing production conditions, or conducting application-specific tests. (Farinas, Bulter and Arnold, 2001; Waldo, 2003; Wang, Si and Zhao, 2012) Figure 5 shows the workflow of the directed evolution.

In an industrial context, enzyme thermoresistance is crucial for sustaining high-performance levels under harsh processing conditions. Stephens et al. (2007) applied directed evolution to enhance the thermal stability and activity of the endo-β-1,4-xylanase (XynA) from *Thermomyces lanuginosus.* In the first round of mutagenesis, beneficial mutants were identified, exhibiting enhanced enzymatic properties compared to the wild-type. These mutants were subjected to further mutagenesis to identify second-generation variants. While most exhibited the expected compromise between stability and activity, one second-generation mutant displayed simultaneous improvements in both metrics. The results underscored the potential of directed evolution to generate industrially relevant enzymes capable of maintaining efficiency at elevated temperatures. (Stephens *et al.*, 2007)

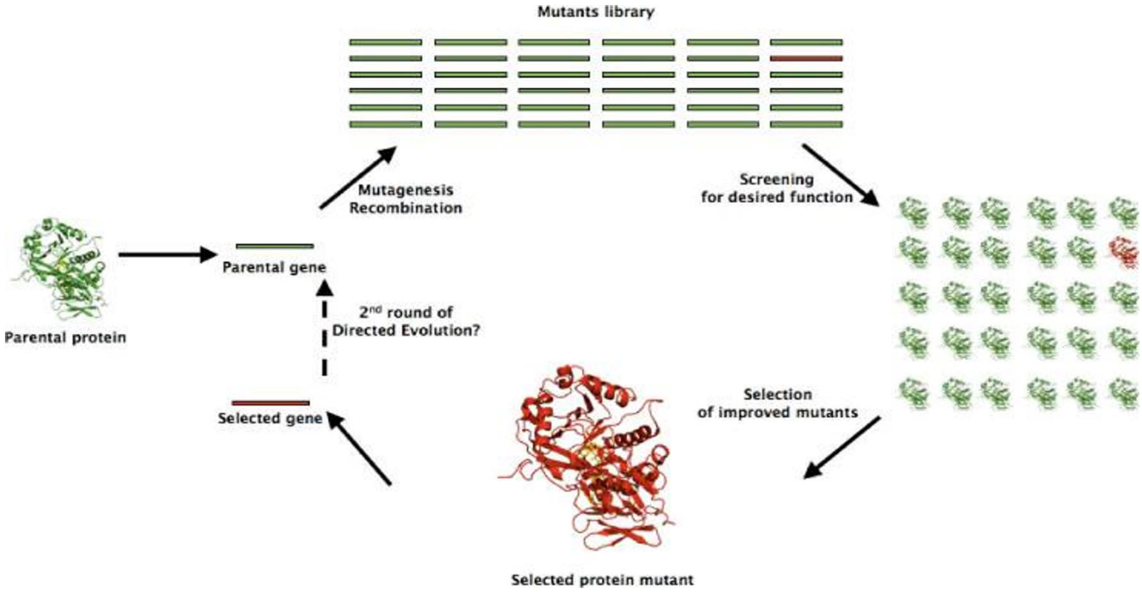


Figure 5. Directed evolution involves an iterative cycle of mutagenesis and screening. From the generated mutant library, mutants are screened, clones having novel or improved features are selected, and the cycle of mutation screening is repeated until a certain activity or another feature has been improved enough (Rigoldi *et al.*, 2018)

In another study, Lin et al. (2009) used error-prone polymerase chain reaction (PCR) and DNA shuffling, two common techniques of directed evolution, to improve the catalytic efficiency of the Bacillus subtilis endo-β-1,4-glucanase (Cel5A). Several variants exhibited increased activity toward sodium carboxymethyl cellulose (CMC), a common model substrate for endoglucanase testing. One of the variants exhibited both enhanced activity and also improved pH tolerance and thermostability. These improvements broadened the enzyme's applicability in various industrial settings, including biofuel production and textile processing, where adaptability to diverse pH environments and temperatures is vital. (Lin *et al.*, 2009)

These studies illustrate the potential of directed evolution in enzyme engineering for industrial applications. Both cases highlight the balance between stability and activity in enzyme optimization. Directed evolution, by mimicking natural selection, remains a powerful strategy for overcoming these trade-offs, enabling the development of tailored biocatalysts for diverse industrial applications.

3 Aims of the study

The aim of this work was to develop two industrial enzymes. The first is a thermostable laccase from *Bacillus wacoensis*, engineered for activity in highly alkaline conditions for lignin processing. The resulting product from this laccase is MetZyme® LIGNO™, which is a core component of the METNIN technology for lignin valorization. The second enzyme engineered and described in this thesis is a thermostable endoglucanase from *Spirochaeta thermophila*, used as part of enzyme cocktails called MetZyme® BRILA for cellulosic fiber modification.

This work was performed at an industrial biotech company, METGEN, and was part of the implementation of METGEN's mission to enable biobased industries via efficient and robust biocatalysts - enzymes. In order to make the enzymes instrumental in biomass valorization, they need to meet industry requirements - efficient and cost-effective production and the ability to withstand harsh industrial conditions. Thermostability is critical for process and storage stability and cost-effective production, as it facilitates downstream processing.

Naturally thermostable enzymes found in extremophile microorganisms are often unable to fold correctly in standard industrial microorganisms, *Escherichia coli*, and end up in inactive insoluble aggregates. Alternatively, - the native enzyme is folded correctly and expressed in *Escherichia coli* but lacks sufficient thermostability. Both challenges can be approached from the genetic engineering point of view using rational and semi-rational molecular design. These approaches were used in this work. Methods utilized included multiple sequence alignment-based site-directed mutations and libraries, homology model-based comparisons of structures: sequence conservations analysis, mutations and deletion design, and structural changes. Novel prototypes were sourced from public protein sequence databases. Respective recombinant wild-type proteins were produced in *E.coli*, characterized, and further improved to reach the desired properties. This work resulted in the development of enzymes that were further produced on an industrial scale and commercialized.

4 Materials and Methods

4.1 Enzyme engineering

4.1.1 Basic local alignment search tool (BLAST) and Sequence alignment

Protein sequences were used as queries in BLAST ®- server while searching through databases for similar proteins (Camacho *et al.*, 2009). Collections of sequences were selected based on set threshold parameters. MAFFT Multiple alignment program for amino acid sequences server was used to align protein sequences based on their homologous structures and key residues (Katoh and Standley, 2013). MAFFT alignment was further refined with ClustalIX2- program focusing on conserved domain area differences between sequences (Larkin *et al.*, 2007).

The focus in the multiple sequence alignment was to find the sequence positions ("hot spots") where the residues of the target enzyme differ among the conserved areas of the sequence set. It was assumed that the protein features, like folding, functionality, and stability, might be determined by those conserved areas. The set of sequences was chosen based on the focus feature. For example, if thermostability needs to be improved, the selection of sequences is based on proteins of thermophilic origin.

4.1.2 Homologous modelling and structural analysis

Since no crystal structures were available for the laccase and endoglucanase used in this study, 3D structures were predicted using the Chimera software interface Modeller. Theoretical models were generated using a sequence alignment, and at least one known related structure is used as a template. For the active site detection and substrate binding site detection, crystal structures of close homologous proteins were used, and structures were matched using Chimera Match Maker. The software also enables the importation of multiple sequence alignments to visualize the conserved and semi-conserved residues in the structure with colors. This may improve the prediction of which residues are important to the enzyme properties focused on development. Some modelling and structure alignments were also done using the Swiss-model server (Waterhouse *et al.*, 2018). When the experimental work was performed, AlphaFold (Jumper *et al.*, 2021) did not yet exist. Currently, an AlphaFold-predicted model of *Bacillus wacoensis* JCM 9140 Spore coat protein A, AF-W4PYT0-F1, is also available. This model was further utilized to visualize the mutated residues in this thesis.

4.2 Molecular Biology Techniques

4.2.1 Gene synthesis, PCR, and cloning

A synthetic gene fragment of Spirochaeta thermophila endoglucanase was purchased from Twist Biosynthesis, and the Bacillus wacoensis laccase used in these studies was purchased from BaseClear. They were cloned in a modified pET28- plasmid with either periplasmic or cytosolic signals. This was done with restriction digest and ligation between NcoI and BamHI- sites in plasmid and synthesized gene.

All the modifications in this study were implemented with full plasmid PCR. Primers were designed so that the reverse primer was homologous to the target gene and contained 20 nucleotide complementary sequences to the forward primer. Forward primer had the first 20 nucleotides homologous to the template and complementary to the reverse primer. Sequence modifications (mutations, library, replacement, deletion) were included in the forward primer after the reverse complement part and before the last N-terminal 15 nucleotides homologous to the template DNA responsible for primer annealing to the template and initiating the PCR reaction.

The PCR reactions were done using Phusion™ High- Fidelity DNA Polymerase (ThermoFisher Scientific, Waltham, Massachusetts, USA). The PCR reaction mixture consisted of 0.25 µl template DNA, 5 µl 10xHF buffer, 2 µl 0.5 nM dNTP mix, 1 µl 10 µM each primer, 0.25 µl Phusion DNA polymerase and 15,5 µl sterilized RO water. The standard conditions of PCR were: a pre-PCR denaturing step of 10 sec at 98°C, 30 cycles of denaturation for 30-sec at 94°C, annealing for 30-sec at 60 °C, and extension for 5 min at 72°C, and then a final post-PCR extension for 5 min at 72°C. Template DNA was digested by adding 0,5 µl DpnI- restriction enzyme (Thermo Scientific™) into the PCR reaction mixture and incubating for ~5 hours at 37 °C. DpnI specifically cleaves at its recognition site only when the site is methylated, as in the original template DNA extracted from BL21 cells, leaving the PCR product intact. 1 µl of the PCR reaction mixture was used to electroporate electrocompetent E. coli BL21(DE3) cells (Merck, Darmstadt, Germany). Due to the 20 nt homologous sites in linear PCR product, BL21(DE3) bacterial cells are able to recombine it into the circular plasmid DNA with the use of endogenous RecA enzyme. Afterward, all the beneficial mutations were verified by sequencing.

4.2.2 SDS-PAGE

Soluble and insoluble fractions of proteins were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in a 12% (w/v) separating gel with a 5% (w/v) stacking gel using an electrophoresis chamber (Bio-Rad Mini-PROTEAN Tetra Vertical Electrophoresis Cell with PowerPac basic). For insoluble protein samples, 30 µl of well-mixed total cell lysates were centrifuged at 4000 rpm for 5 minutes. The supernatant was discarded, and the pellet was washed with 500 µl of milli-Q water. The sample was further centrifuged at 4000 rpm for 5 minutes; the supernatant was discarded, and the pellet was mixed in 30 µl of water. For soluble samples, 200 µl of cell lysate was centrifuged for 4000 pm 5 minutes, and 30 µl of supernatant was transferred into a new Eppendorf tube. 10 µl of 4X Sample buffer [250 mM Tris, 4 % (w/v) SDS, 40 % (v/v) glycerol, few grains of bromophenol blue, 20 % (v/v) β-Mercaptoethanol] were added to each sample and mixed with pipet. Samples were heated for 5 minutes at 100 ºC before loading into gels. Thermo Scientific™ PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa, was used as a molecular weight ladder. 10 µl of samples were loaded to each well of the gel and run in 1 x running buffer 100V for 60 minutes. Gels were stained with Coomassie blue superstain (10 % acetic acid, 50 % methanol, and 0,1 % Coomassie stain R250) by incubating at room temperatures, slowly shaking overnight, and washed twice with Coomassie destain (60 % milli-Q water, 10 % acetic acid, 3 % methanol) heating for 10 seconds in microwave oven followed by incubation for 10 minutes at RT.

4.3 Protein production

In this thesis, most primary screenings are done by comparing non-purified soluble recombinant proteins in the total cell lysates. To compare them, the cells must be cultured in conditions where both cell growth and recombinant protein production are optimal. Thorough studies were done before this thesis, testing different inducers, sugar concentrations, volumes, temperatures, shaking speeds, and incubation times. Conditions for recombinant protein production were selected based on repeated constant levels of maximum enzyme activity yield among several parallel wild-type enzyme productions. Cultures were inoculated from single colonies from LB-agar plates, seeded in 3 ml of SB medium (3.0% Tryptone, 2.0% Yeast extract, 1.0% MOPS free acid) supplemented with autoinduction sugar cocktail (final concentration glycerol 0,25 % (v/w), glucose 0,025 %(v/w), lactose 0,1 % (v/w)) and with possible cofactors (like copper in the case of laccase) and were cultivated in 24 well plates 250 rpm 24 hours. Six separate single colonies were generally seeded for protein production in single amino acid mutation screenings. When library screenings were done, at least six times the total variant number of colonies were picked and seeded. During all the protein production, three parallel single colony wild-type productions were done and used as a comparison in activity tests.

Some of the recombinant proteins were leaked or liberated to the culture medium due to cell death, and that is why the entire suspension, including both the medium and cells, was analyzed further. The cells were lysed in the culture medium, adding 1:10 v/v of 10x Lysis buffer (200 mM Citric acid, 10 % Triton, pH 7.5 with or w/o 5 mM $CuCl_2$, pH adjusted with 5 M NaOH) into the suspension after protein production phase. Lysis was continued by incubating 250 rpm o/n at RT. Whole suspension lysis was used in the previous studies compared to more traditional, only cell-lysed samples and the product yields were clearly better when the concentrated lysis buffer was added to the medium. The supernatants from lysed cell solutions were used for enzyme-specific activity and stability tests. To obtain the supernatants, protein production plates or tubes were centrifuged at 4000 rpm for 10 minutes. The supernatants were either used directly or transferred to new plates or tubes for further analysis. For the final product activity measurements, the recombinant proteins were purified from the supernatant according to the manufacturer's protocol of HisPur Ni-NTA Resin (Thermo Scientific, Waltham, USA), described in publication II.

4.4 Activity assays

4.4.1 Laccase-specific activity assays

4.4.1.1 Standard substrates, pH optimum, and thermal stability

Relative laccase activity assays using standards substrates, 2,2′-Azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) (ABTS), Syringaldazine (SGZ), and 2,6-Dimethoxyphenol (DMP) at various pH, are described in publication I. Laccase stability at different pHs' was determined by pre-incubation of the enzyme at a specific temperature and pH, followed by measurement of residual activity with ABTS. pHis optimum for each mentioned standard substrate was defined by the reactions using Britton & Robinson buffers (pH 2–11) (Otto, Schlosser, and Reisser, 2010). Thermal stability studies were done by incubating the enzyme first in the desired temperature ranging from 50 - 80 °C. Immediately after incubation preparation, the reaction mixture was diluted to achieve the final enzyme concentration, ensuring an immediate shift to the substrate-specific reaction temperature.

Specific enzyme activity was determined using ABTS as a substrate and measuring reaction absorbance at 405 nm every 2 minutes. One unit of laccase activity was defined as the enzyme amount oxidizing 1 µmol of ABTS per minute under these conditions (whereas the extinction coefficient of ABTS at 405 nm is 36,800 M−1 cm−1). One microkatal is the amount of enzyme oxidizing one µmol of substrate per second; hence, one µkat equals 60 U.

4.4.1.2 Lignin depolymerization and fractionation

Lignin studies were made with hardwood lignin extracted by an extrusion-based process obtained from an industrial source (http://sweetwater.us/process/). Lignin was first solubilized in 0.25 M NaOH, and supernatants were dried due to convenience in storage and obtaining reproducible results and then depolymerized using a laccase-mediator system with syringaldehyde as the mediator. Actual laccase depolymerizations were done in bioreactors, and syringaldehyde was used as a reaction mediator. The reaction was carried out for 21 hours at 50°C with constant aeration and pH control set to 10.5. A control reaction was conducted without the enzyme. Details of reactions and lignin depolymerization conditions can be found in publication II.

4.4.2 Endoglucanase-specific activity assays

Endoglucanase activity was determined using several methods with high- and low-molecular-weight substrates. Carboxymethyl cellulose (CMC, ThermoFisher Scientific, Waltham, USA) is a standard high molecular weight substrate for endoglucanase activity measurement. The second high molecular weight substrate was azurine-crosslinked barley β-glucan (AZCL-β-glucan, Megazyme, Wicklow, Ireland). Low molecular weight substrates used were CellG5 (Megazyme, Wicklow, Ireland) and p-nitrophenyl- β-D-cellobioside (pNPC, Sigma-Aldrich, St. Louis, Missouri, USA). A more detailed description of activity assays is found in publication IV.

Primary library screenings with more extensive libraries were done using the colony picker QPix2™ (Molecular Devices, San Jose, USA) and seeding colonies on CMC-containing LB agar petri plates. After incubation, plates were dyed with Congo Red to determine active colonies defined by a clear halo around the colony. Primary screening of the libraries involved picking 3000 clones with a Qpix colony picker and seeding them onto CMC-containing LB agar plates. The most active clones, characterized by the most significant halos, were further analyzed. These clones were subsequently seeded from replica plates generated during colony picking, and recombinant soluble proteins were produced as described earlier. Secondary screening was done mostly using barley β-glucan, and the best-performing clones were sequenced.

For the thermostability test, enzyme solutions were heated to 50 - 80 °C, and samples were taken at various time points, ranging from 0 to 60 minutes. The activity of the heat-treated enzymes was determined using Beta-glucazyme tablets according to the Megazyme assay protocol.

Some stability measurements were also done in reaction buffers (Publication IV). To analyze the pH optimum and profile of endoglucanase, a CMC-DNS assay was used using Britton Robinson buffer pH 3-10. No activity for crystalline cellulose was detected, which was expected since endoglucanase cannot hydrolyze crystalline cellulose.

5 Results and discussion

5.1 Laccase

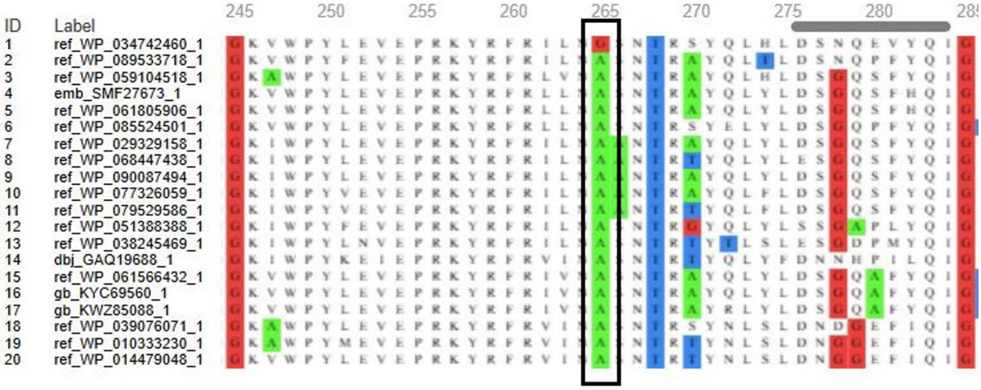
Two different laccases have been used in the publications related to this thesis, both part of MetZyme® LIGNO™- products. The best-studied bacterial laccase is a component of the outer coat layers of bacillus endospore laccase from *Bacillus subtilis* (CotA). Several components of the *B. subtilis* spore coat are enzymes, which possibly play a role in the final resistance properties of the spore structure. CotA is a versatile enzyme that belongs to the multicopper oxidase family. It catalyzes the oxidation of phenolic and non-phenolic lignin subunits and a variety of organic and inorganic substrates, using molecular oxygen as an electron acceptor. This property makes *Bacillus subtilis* laccase particularly valuable for eco-friendly applications in the pulp and paper industry. It is also highly thermostable, making it suitable for industrial processes under thermal stress. (Martins *et al.*, 2002)

Engineered CotA has been used in different applications in acidic conditions. However, several different laccases were shown to oxidize phenolic compounds in alkaline conditions. It was found that laccase from *Bacillus wacoensis* was remarkably stable at high pH. This laccase is notable for its stability and activity in highly alkaline conditions, making it particularly useful in various industrial processes, including the pulp and paper industry. The laccase used in these studies had already been modified and the enzyme referred to as Lac1 in this study is *Bacillus wacoensis* laccase, which contains two mutations, W109S and T260A, that improved the ABTS-specific activity of the wild-type laccase by approximately twofold. The expression level of the variant was high (based on SDS-PAGE), but most of the protein remained in insoluble fractions as aggregates. This double mutant variant W109S&T260A was selected as the starting point for further mutations. It was used as a control in most of this thesis's activity and characterization studies. A more detailed correlation of the mutations with the wild-type laccase is described in publication I (patent WO20193452A1).

A rational approach was used to engineer the laccase, selecting positions for site-directed mutagenesis. The selection of these sites is described in detail in the next chapter. Multiple sequence alignment was used to detect and select suitable mutation sites.

With the alkaliphilic laccase, the main goal was to improve protein solubility in recombinant production. This was initiated by searching for similar sequences using BLAST protein search, with the laccase sequence as the query, via PSI-BLAST (Position-Specific Iterated BLAST) for primary sequence set selection. A total of 100 sequences were selected for further refinement. All sequences were downloaded, and the MAFFT online server was used to align them. Primary multiple sequence analysis was performed using ClustalX2.1. The alignment was refined by deleting non-fitting or mismatching sequences and removing gaps until the conserved parts of the sequences were clearly aligned. After all outliers were removed, 64 sequences remained in the alignment, and the focus shifted to identifying mismatched amino acid positions in the target sequence compared to most or all the others. Even a single amino acid change can significantly affect protein folding due to side chain charges and chemical bonding. The impact of such a mutation is often impossible to predict in advance.

Figure 6 shows part of the multiple sequence alignment and some positions selected for site-directed mutagenesis. Based on this, 25-point mutations were initially chosen from conserved positions for site-directed mutagenesis, with another 25 selected later.



ID	Label	
1	ref_WP_034742460_1	G K V W P Y L E V E P R K Y R F R I L N G S N T R S Y Q L H L D S N Q E V Y Q I G
2	ref_WP_089533718_1	G K V W P Y F E V E P R K Y R F R I L N A S N T R A Y Q L T L D S N Q P F Y Q I G
3	ref_WP_059104518_1	G K A W P Y L E V E P R K Y R F R L V N A S N T R A Y Q L H L D S G Q S F Y Q I G
4	emb_SMF27673_1	G K V W P Y L E V E P R K Y R F R L L N A S N T R A Y Q L Y L D S G Q S F H Q I G
5	ref_WP_061805906_1	G K V W P Y L E V E P R K Y R F R L L N A S N T R A Y Q L Y L D S G Q S F H Q I G
6	ref_WP_085524501_1	G K V W P Y L E V E P R K Y R F R L L N A S N T R S Y E L Y L D S G Q P F Y Q I G
7	ref_WP_029329158_1	G K I W P Y L E V E P R K Y R F R I L N A A N T R A Y Q L Y L D S G Q S F Y Q I G
8	ref_WP_068447438_1	G K I W P Y L E V E P R K Y R F R I L N A A N T R T Y Q L Y L E S G Q S F Y Q I G
9	ref_WP_090087494_1	G K I W P Y L E V E P R K Y R F R I L N A A N T R A Y Q L Y L D S G Q S F Y Q I G
10	ref_WP_077326059_1	G K I W P Y V E V E P R K Y R F R I L N A A N T R A Y Q L F L D S G Q S F Y Q I G
11	ref_WP_079529586_1	G K I W P Y V E V E P R K Y R F R I L N A A N T R T Y Q L F L D S G Q S F Y Q I G
12	ref_WP_051388388_1	G K I W P Y F E V E P R K Y R F R I L N A S N T R G Y Q L Y L S S G A P L Y Q I G
13	ref_WP_038245469_1	G K I W P Y L N V E P R K Y R F R I L N A S N T R T Y T L S L E S G D P M Y Q I G
14	dbj_GAQ19688_1	G K I W P Y K E I E P R K Y R F R V I N A S N T R T Y Q L Y F D N N H P I L Q I G
15	ref_WP_061566432_1	G K V W P Y L E V E P R K Y R F R I V N A S N T R A Y Q L Y L D S G Q A F Y Q I G
16	gb_KYC69560_1	G K V W P Y L E V E P R K Y R F R I V N A S N T R A Y Q L Y L D S G Q A F Y Q I G
17	gb_KWZ85088_1	G K V W P Y L E V E P R K Y R F R I V N A S N T R A Y R L Y L D S G Q A F Y Q I G
18	ref_WP_039076071_1	G K A W P Y L E V E P R K Y R F R V I N A S N T R S Y N L S L D N D G E F I Q I G
19	ref_WP_010333230_1	G K A W P Y M E V E P R K Y R F R V I N A S N T R T Y N L S L D N G G E F I Q I G
20	ref_WP_014479048_1	G K V W P Y L E V E P R K Y R F R V I N A S N T R T Y N L S L D N G G E F I Q I G

Figure 6. Part of multiple sequence alignment showing conserved regions of aligned laccases (*Bacillus wacoensis* laccase (WP_034742460.1). Differing amino acids in the protein sequence were altered to those conserved in the other sequences. These so-called “hot spots” were detected through the whole gene sequence, but due to several sequences in alignment and long sequence, only part of it is shown as an example of the mutation site detection. Mutation G253A is aligned to other sequences framed (numbering in alignment does not match *Bacillus wacoensis* laccase).

5.1.1 Primary screening of individual mutations

All 50 primer pairs for mutations defined by MSA were designed and cloned. Proteins from six clones of each mutation and three Lac1 controls were produced in 24-well plates. After protein production, cells were lysed to release proteins into the suspension, and ABTS activity analysis was performed using the supernatants of the lysed samples. Since many mutations needed to be tested, the primary screening was conducted in several batches of mutants, each with its control samples.

The background value from blank samples was first subtracted from the raw data. Inactive clones, likely caused by improper recombination of the linear PCR product into a circular plasmid, were clearly identified, and their values were discarded from the primary screening. Each screening was compared to the controls of that batch's wild-type control production. Some of the primary screenings were repeated if there were no colonies or if most colony activities were at the same level as the control laccase. Figure 7 shows the primary screening of the combined average relative activity compared to control Lac1. The average activity of six clones from each mutation was compared to the average activity of three control Lac1 samples.

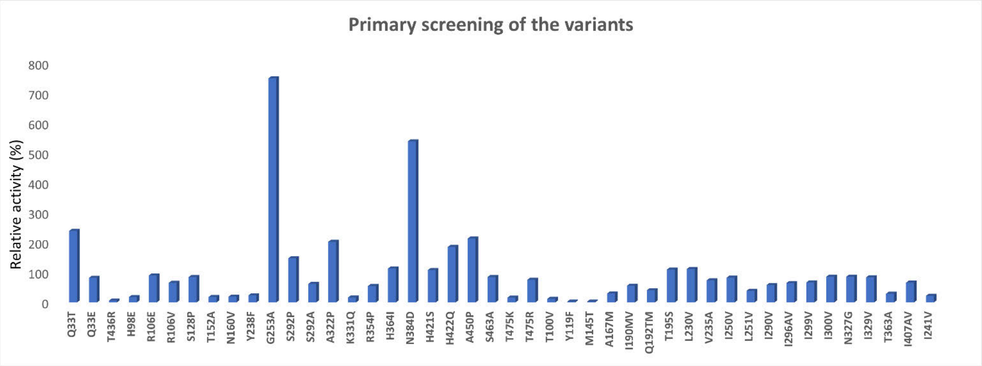


Figure 7. The combined results of primary screening where the relative activity of variants was compared to Lac1- enzyme, with its activity set at 100%. In total 50 different mutations were successfully implemented and analyzed. The primary screening was conducted once, using three parallel clones for each mutation, and the average activities were used for the comparison. Most of them had no effect or reduced the activity, but 9 of the mutations caused increased activity compared to control Lac1. The most improved activity was detected with mutations G253A and N364D.

The effects of each mutation were quite different. Most of them did not seem to have a very significant impact on protein based on the enzyme activity. However, several had reduced activity. This might be caused by either the misfolding of the protein or the loss of critical substrate-binding interactions. Non-beneficial mutations were not further studied, so the impact of the mutation is impossible to predict. Nine mutations were beneficial, and the correct implementation of these mutations was further confirmed by sequencing the three individual mutant clones. The best mutants verified were then re-produced and compared. Figure 8 shows the relative activities of these mutants compared to control Lac1. The selected mutants all had beneficial effects on activity in the second comparison. Most of them at least doubled the relative activity, and mutation G253A had a 650 % increase in activity. Based on the modelled structure, mutations are located all around the enzyme. At this point, no further speculations were made since the goal was to increase solubility and activity as much as possible, combining mutations and only afterward analyzing which can be combined, gaining further benefit.

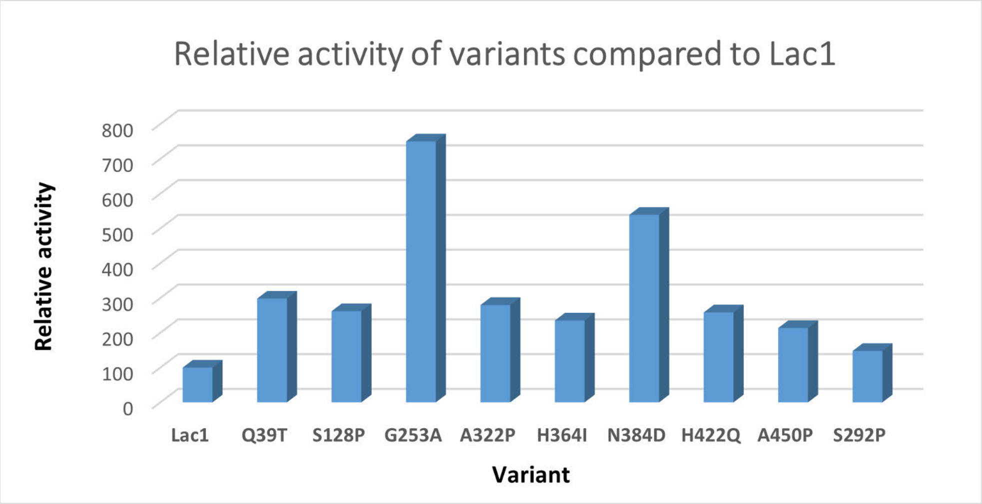


Figure 8. The best mutants, based on laccase-specific activity using ABTS as the substrate, after primary screening and re-production of enzymes. A clear improvement in activity was observed in all selected mutations, with increases ranging from 50% to 650%.

5.1.2 Combining the beneficial mutations

After several beneficial mutations were identified, the next step was combining them. The best mutant in each round was selected as the template for further mutations. If an additional mutation caused a decrease in activity, it was not included in the next round. All mutations were implemented separately on the mutant template (for the first combining round, it was G253A). The workflow remained unchanged: PCR, DpnI treatment to digest the template plasmid, transformation, selection of six colonies from each group, along with at least three wild-type control colonies, and inoculation into separate wells. Protein production was carried out under the same culture conditions, and after production and lysis, ABTS activity analysis was used for screening. All the combined double-mutant activities are listed in Publication I, Table 2, and all combined activities are shown in Table 3. These activities were measured from 1-liter fermentation-scale productions. In this thesis, all data were obtained from small-scale productions, typically in 24-well plate productions, where the final bacterial count (OD 600 nm) is approximately ten times lower than in fermentor-scale productions, and the downstream enzyme production process differs significantly. As a result, enzyme concentrations in the final products vary considerably, making direct comparisons impossible. Tables 1 and 2 present the naming scheme of the variants used in the subsequent studies.

Table 1. Mutation IDs are used to name the variants.

MUTATION ID	MUTATION
A	G253A
B	S128P
C	N384D
D	H364I
E	S292P
F	A450P

Table 2. The naming scheme of the variants is further used in characterization studies.

VARIANT	STUDY ID
Wild type B. wacoensis laccase T260A + W109S	Lac1
G253A	Lac1_A
G253A + S128P	Lac1_AB
G253A + N384D	Lac1_AC
G253A + S128P + N384D	Lac1_ABC
G253A + S128P + N384D + H364I	Lac1_ABCD
G253A + S128P + N384D + H364I + S292P	Lac1_ABCDE

5.1.3 Variants’ characterizations

Meanwhile, as beneficial mutations were combined, the characteristics and effects of each mutation and initial combinations were studied. For the variants that showed improved activity in the primary screening, further studies on solubility, activity, and stability were conducted. Due to the significant improvements in solubility and activity with the Lac1_ABC variant, the following upstream production was carried out. This upscaled enzyme production was studied in Publication II and used in lignin treatment applications. Final activity studies and comparisons in this thesis were performed with the variant Lac1ABCDE.

The beneficial mutations' insoluble and soluble samples were run on an SDS-PAGE gel. The results for the first set of advantageous mutations are shown in Figure 9. It is evident that the G253A mutation significantly affects solubility when comparing the soluble samples of this variant to Lac1. Visually, other variants (Q33T, S292A, and N384D) do not appear to impact solubility. Combined mutation samples are shown in Figure 10, which were also run on SDS-PAGE.

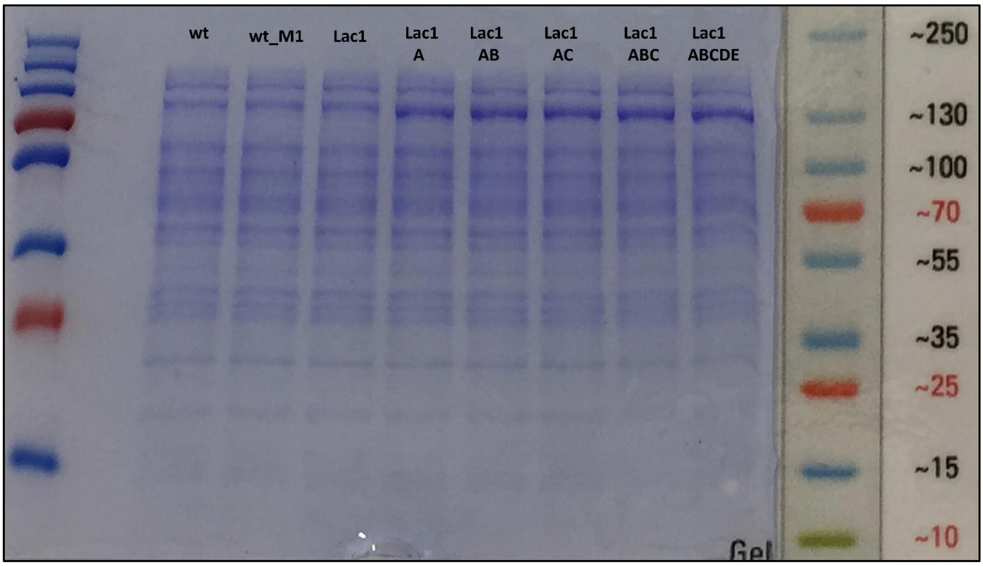


Figure 9. Effect of first mutations on solubility of the protein. The G253A mutation positively impacts solubility, as observed in the soluble protein samples. This effect is particularly evident when comparing Lac1 and Lac1A, where the protein band becomes significantly more intense in Lac1A. None of the other tested mutations had that much effect solely on the activity and no visible impact on solubility when analysed with SDS-PAGE.

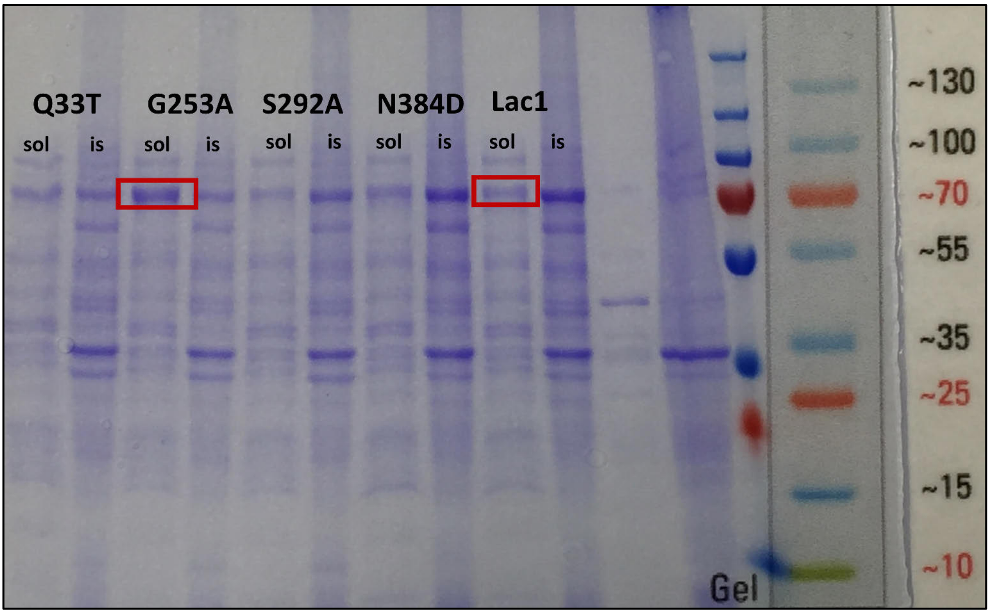


Figure 10. Solubility improvement of the enzyme with different variants. Both insoluble (IS) and soluble (SOL) samples from each mutation were analyzed using SDS-PAGE. The most significant impact was observed with the first mutation, G253A, compared to Lac1, where the soluble band was much more intense, and the insoluble band was less visible.

The specific activities of the variants were tested at different pH levels using DMP and syringaldazine. The enzyme was already known to be very alkaliphilic, and the substrate ABTS can only be used below neutral pH. The G253A mutation had the most significant impact on ABTS activity, while with DMP and syringaldazine, the effect was lower. This is likely because the wild-type enzyme has the highest specific activity with ABTS. Mutations greatly impact the enzyme’s solubility, and increased enzyme amount causes the biggest activity increase with ABTS. However, combining the first three mutations - G253A, S128P, and N384D - increases the enzyme's specific activities with all the substrates used (see Figure 11).

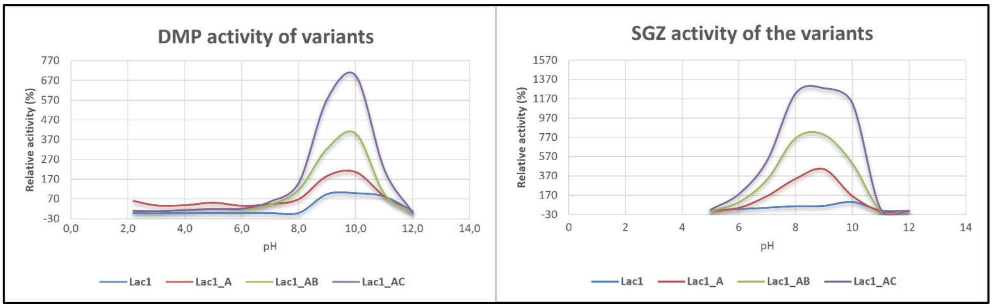


Figure 11. Specific DMP and syringaldazine activity of variants in different pHs. The effect of mutations in the specific activity of Lac1 is also clearly seen with other substrates.

For the final activity comparison in this thesis, the selected variants were produced in duplicate (two parallel productions of each in a 24-well plate) and compared to the wild type and the starting variant laccase. Final activity comparisons are shown in Figure 12. Table 3 displays the specific ABTS activity increases of the variants. This comparison was conducted using 1-liter-scale fermented recombinant protein productions. At this point, Lac1_ABC was already produced on a large scale and used in real substrate applications, and it would have needed thorough studies to ensure additional mutations didn’t alter the stability of the enzyme; they would have reduced stability effect if they were not used in further studies.

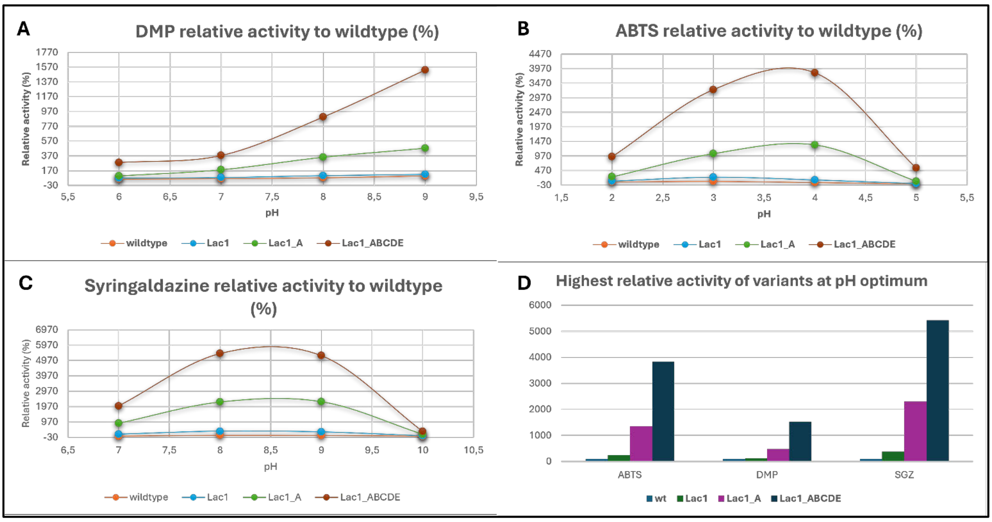


Figure 12. Specific activities of the engineered laccase variants at different pH for 2,6-dimethylphenol (**A**), 2,2′-Azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) (**B**), and syringaldazine (**C**), and relative highest activity on pH optimum (**D**).

Table 3. The final activities µkat/l are defined by specific activity method.

VARIANT	ACTIVITY (µKAT/L)
WILDTYPE	6,1
WT T260A	6,9
LAC1	35,8
LAC1_A	210
LAC1_AB	363,2
LAC1_AC	274,4
LAC1_ABC	221,6
LAC1_ABCDE	329,2

5.1.3.1 Mutations effect on enzyme stability

Stability at different pH levels and temperatures was studied using single mutation variants and some combinations. Initial stability measurements were conducted at 60°C with the best single mutant (G253A) and double mutant (G253A + S128P). Enzyme samples were taken at 0, 20, 40, and 60-minute time points. Figure 13 shows that the G253A mutation affects pH stability. The addition of the S128P mutation appears to compensate for this reduction in stability, resulting in a noticeable improvement in pH stability at pH 11. There is no clear explanation for this. All amino acids, glycine, alanine, proline, and serine, are neutral in charge, and adding a small side chain to glycine should not have much impact on the pH stability of the enzyme. Residue G253 is located in the core of the enzyme, while S128P is on the surface; the compensation in stability is difficult to explain at this point without further studies.

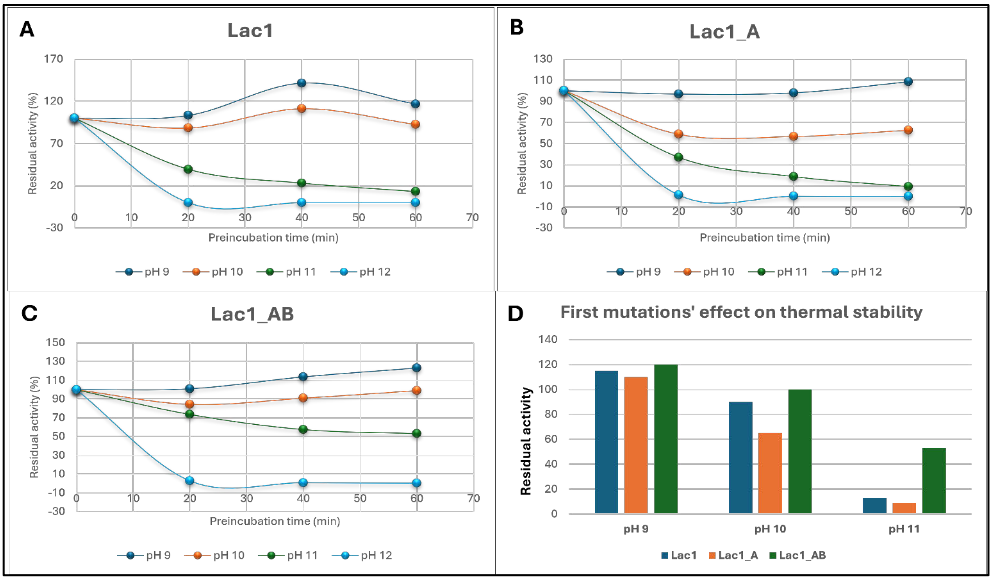


Figure 13. Stability of the variants in alkaline pH and 60°C. Even single mutant G253A (Lac1_A) greatly improves the solubility of the laccase it also seems to have some effect on reducing the pH stability of the enzyme incubated at 60°C. When mutation S128P is added (Lac1_AB), the reduction in stability becomes compensated and the pH stability is clearly improved.

The thermal stability of laccase at alkaline pH was further investigated. In the next stability test, the pH stability of single, double, and triple mutants was assessed with different incubation times at 70°C, the standard sterilization temperature for products used in lignocellulose valorization processes. As shown in Figure 14A, the single G253A mutation reduces pH stability at 70°C. This destabilization may result from structural changes affecting the enzyme's ability to withstand alkaline conditions. The double mutant G253A + N384D (Figure 14B) exhibits a similar reduction in pH stability, likely due to destabilizing effects introduced by the N384D mutation, such as disruptions to hydrogen bonding or electrostatic interactions essential for maintaining structural integrity at high pH.

Interestingly, the S128P mutation appears to compensate for the destabilizing effect of G253A (Figures 14C, D, E, and F), likely by enhancing structural stability under alkaline conditions. The rigid cyclic structure of proline can restrict backbone flexibility, providing conformational stability that helps maintain the enzyme's tertiary structure and prevents denaturation. The combination of S128P and G253A in various variants (Figures 14C, D, E, and F) restored and slightly improved pH stability. This suggests that S128P plays a compensatory role, possibly by stabilizing critical regions of the enzyme structure. Triple mutants demonstrated similar pH stability to the double mutant, indicating that the additional mutations (N384D, P450D, S292A, and H364I) did not further compromise stability.

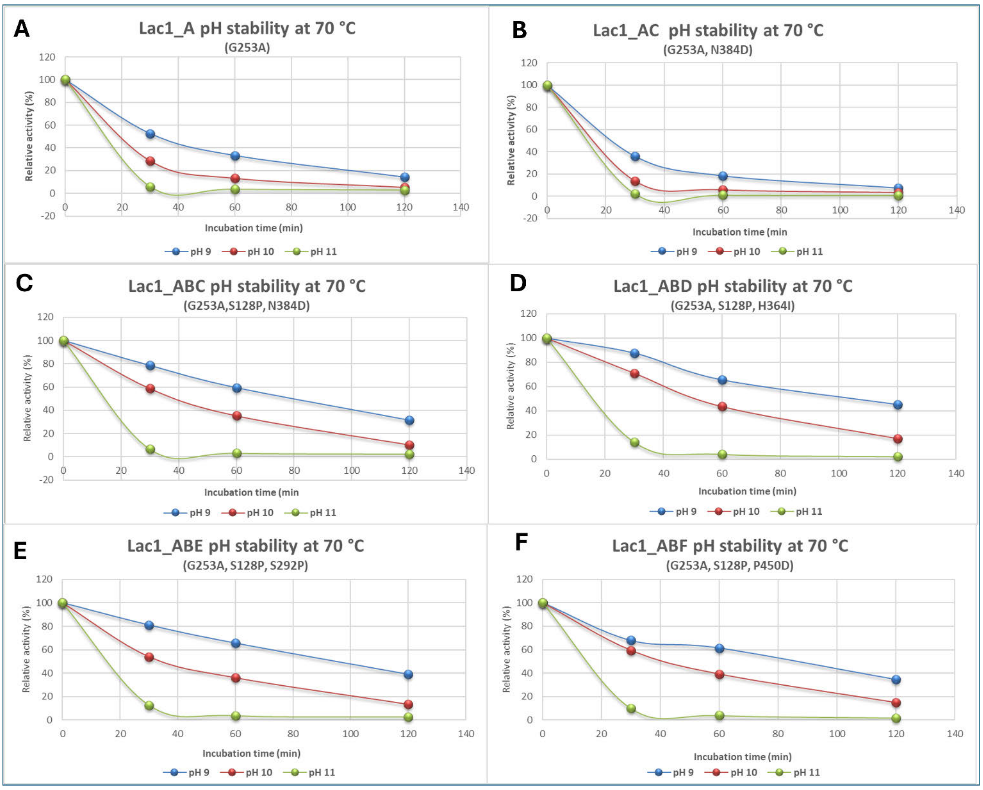


Figure 14. Stability of the variants in alkaline pH and 70°C. Only mutation G253A affects the enzyme's pH stability at 70°C (A), and it is still reduced when combined with mutation N384D (B). When combined with S128P, the stability is restored, and the additional mutations do not affect the pH stability (C-F).

5.1.3.2 Combined mutations - Lac1ABCDE

While combining the mutations, the improved variants with less mutations were already taken forward for upstreamed production and application testing. The mutation combination Lac1ABCDE was selected as the final variant for characterization studies in this thesis. The enzyme yield and specific activities increased significantly, with the highest increase in specific activity being over 60-fold with ABTS at pH 4. In contrast, the enzyme retained both thermal and high pH stability. The beneficial mutations included in Lac1ABCDE are illustrated on the homology model of *Bacillus wacoensis* laccase, as predicted by AlphaFold (model AF-W4PYT0-F1), shown in Figure 15.

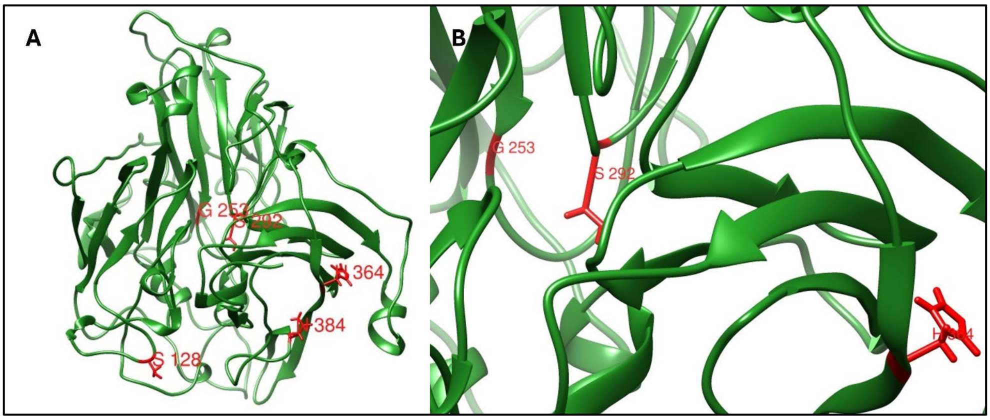


Figure 15. Mutations in Lac1ABCDE are shown in the homology model of *Bacillus wacoensis* laccase (15A). Only residue G253 is located buried inside the enzyme (15B). Residue S292 is located close to the substrate binding site but seems not to be in straight binding contact with the substrate (15B). The rest of the beneficial mutations are located on the protein surface.

The exact reason for the improvement is not entirely clear. The first mutation with the most significant impact on solubility, G253A, is buried inside the enzyme. Figure 16 shows the *Bacillus wacoensis* laccase model overlaid with the endospore coat laccase CotA from *Bacillus subtilis* (crystal structure 4YVU) to visualize the location of G253 near the trinuclear copper center (TNC). The mutation improved the activity but also had the biggest impact on solubility which could explain the improvement in activity. One possible explanation is that the high expression levels led to protein aggregation, forming inclusion bodies. The substitution of GGT (Gly) with GCT (Ala) may have reduced the codon usage ratio, thereby slowing the translation rate, and allowing the enzyme more time to fold properly, which in turn increased its solubility.

Interestingly, none of the beneficial mutations are directly in contact with the substrate. The S292P mutation is close to the substrate binding site, but no clear explanation for the improved activity can be concluded. This mutation may affect the folding of the active site in a way that influences substrate entry. This hypothesis could be further investigated through molecular docking simulations and studies, which would provide deeper insights into the structural and functional changes resulting from these mutations. Such simulations could help elucidate how these mutations impact enzyme-substrate interactions and stability at the molecular level.

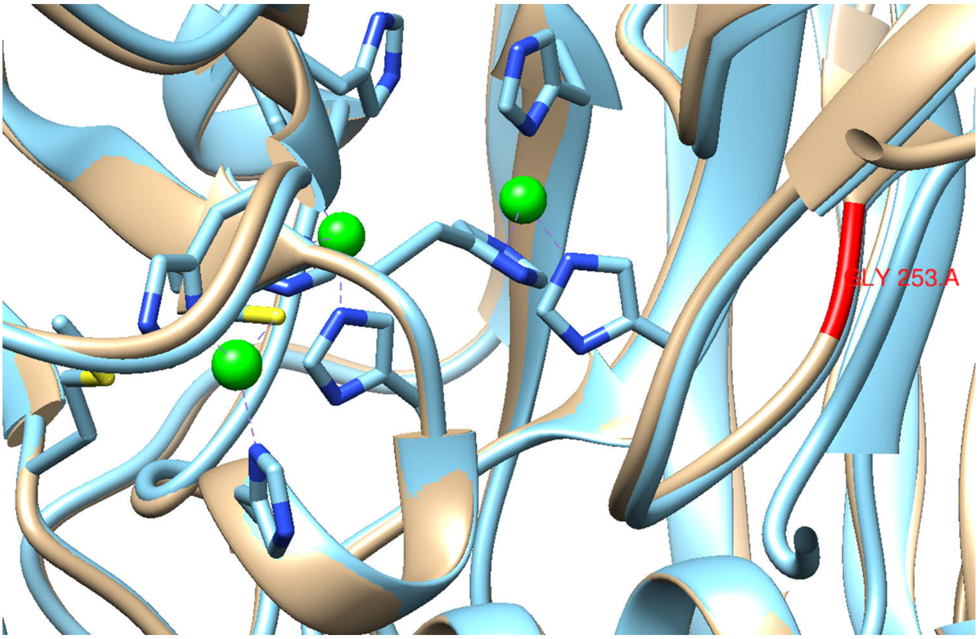


Figure 16. *Bacillus wacoensis* laccase model is overlayed with endospore coat laccase CotA from Bacillus subtilis crystal structure 4YVU to visualize the location of G253 close to the trinuclear copper center (TNC).

The beneficial mutations S128P, N384D, and H364I, located on the surface of Lac1ABCDE, likely have a positive impact on protein stability by influencing surface charge and protein-protein interactions. While these mutations do not visibly affect solubility, they may reduce aggregation by modifying how the protein interacts with its environment and other proteins. Aggregation often occurs due to hydrophobic patches or strong electrostatic interactions between proteins. By altering the surface properties of the protein—such as introducing a rigid proline (S128P), a negatively charged aspartate (N384D), and a hydrophobic isoleucine (H364I)—these mutations can help reduce unwanted protein-protein interactions that promote aggregation. For instance, the proline substitution could make the protein less prone to structural misfolding, while the aspartate and isoleucine changes could influence electrostatic interactions, reducing the likelihood of aggregation. Consequently, these mutations optimize protein surface characteristics, enhancing its stability and minimizing aggregation without directly affecting solubility.

5.1.4 Laccase applications

Metgen's lignin-oxidizing enzyme product is called MetZyme® LIGNO™. Two different laccases from this product group were used in the application studies included in this thesis. The previously developed laccase was used in mechanical pulping applications (Publication V). In contrast, the alkaline laccase—whose development process is described in detail in this thesis—was used in lignin fractionation studies (Publication II). The selected applications for the enzymes differed due to their distinct characteristics.

5.1.4.1 Lignin fractionation

Lignin solubility is influenced by the pH, with alkaline conditions playing a significant role in modifying its structure and increasing its solubility. This is why the lignin depolymerizations were also performed at high alkaline pH, and the alkaline laccase developed in this thesis was selected for the application due to its stability and activity at high pH. For lignin valorization applications, it is the main goal is to be able to collect unique fractions with specific molecular weight distributions. This is crucial for utilizing these fractions with specific molecular weights in various applications, replacing oil-based products, and developing new fine chemicals. Since lignin depolymerization is reversible, treatment conditions such as aeration and pH must be optimized. To prevent re-polymerization, it is necessary to couple the membrane separation system with the depolymerization system to enable fraction separation during the reaction.

Lignin was first solubilized in NaOH and then depolymerized using a laccase-mediator system with syringaldehyde as the mediator. After depolymerization, enzyme-treated lignin and control samples were titrated with hydrochloric acid. It was observed that the enzyme-treated lignin exhibited more extensive buffering at high pH and began to precipitate at a much more acidic pH compared to the control. The enzyme-treated lignin remained fully soluble even at pH 5.0, whereas only 35% of the lignin in the control sample remained in solution at this pH. Treated samples were also ultracentrifuged to separate fractions of different molecular weights. Figure 4 of Publication II shows a clear difference in particle size distributions, indicating increased solubility and better dispersibility in the enzymatically treated lignin. This suggests extensive lignin depolymerization with the laccase-treated samples. The study also demonstrated that unique fractions with specific molecular weight distributions could be collected. This was also because this laccase could be used at such a high pH.

The same laccase development described in this thesis was also applied in another study where a solvent fractionation method was first used to obtain lignin fractions with reduced structural heterogeneity and more defined characteristics. These fractions were then used in enzymatic oxidation and polymerization experiments in the lignin-soluble state (pH 10), aided by treatment with the alkaliphilic laccase. The study demonstrated that laccase significantly catalyzed oxidation and polymerization, as evidenced by increased molar mass. (Wang et al., 2021)

5.1.4.2 Energy saving

Laccase can be applied in the mechanical pulping process to enhance pulp quality and decrease energy consumption. Mechanical pulping is a process in the paper industry where wood fibers are physically separated from lignin and other cell wall components, often using mechanical forces like grinding or refining. Chip treatments were conducted under neutral or acidic conditions.

The first engineered laccase in the MetZyme® LIGNO™ products was derived from *Bacillus subtilis* CotA laccase and primarily used for energy savings in refining. The engineering process of this laccase is not described in this thesis. This enzyme could withstand temperatures over 80 °C and was operating at pH between 4 and 8., making it suitable for mechanical pulping, where the application pH is slightly acidic. Publication V describes two different treatment applications for mechanical pulping. In both applications, chips were first treated in an impregnation step, where water and extractives were pressed out of the chips before they were transferred to an impregnation tank containing the impregnation fluid. The sudden expansion of the chips in the impregnator allows them to absorb the impregnation fluid quickly. When the enzyme is dissolved in the fluid, it also penetrates the chips, leading to lignin oxidation during the incubation in the impregnator. After treatment, the chips are normally refined into pulp, and the specific energy consumption is compared to that of a non-enzyme-treated pulping process.

CotA laccase was incorporated into the mechanical pulping process in the first example described in Publication V. The enzyme treatment was conducted during impregnation, with the chips being treated for 60 minutes at 65°C. After treatment, the chips were refined into mechanical pulp as usual. The specific energy consumption (SEC) decreased by 300-500 kWh/t compared to a reference, where only impregnation liquid was used, depending on the pulp freeness level, measured by Canadian Standard Freeness (CSF).

In the second example, enzyme treatment was incorporated into the thermomechanical pulping process. To maximize its impact, the enzyme was added to the impregnation fluid and a small molecular weight mediator (not revealed due to trade secrecy). The impregnated chips were incubated for 1 hour at pH 6 and 70°C before primary refining. Continuous addition of the impregnation solution maintained a constant liquid level in the impregnation tower. For a target CSF of 180 ml, the energy consumption was reduced from 2.02 to 1.87 MWh/t with enzyme-treated samples compared to the reference (non-enzyme-treated) samples.

Energy savings in both applications are evident, with energy consumption reduced by 10-20%, which is clearly lowering the operational expenses. In both examples, adding the enzyme to the process required no additional modifications, making it easy to incorporate into any mechanical pulping system. Similar energy savings with reduced steam requirements have been reported in other case studies

(Bajpai P.K., 2011). The volume of enzyme required is also substantial because industrial processes are on a large scale, 10000-100000L. Industrially, enzyme dosages typically range between 1–10 kg per ton of pulp for lignin modification or energy-saving treatments, depending on the lignin content, pulp type, and desired outcome (e.g., brightness or energy reduction).

5.2 Endoglucanase

5.2.1 3D structure modelling and structural comparison

A GH9 family endoglucanase from *Spirochaeta thermophila* (egSth) had been previously synthesized and characterized, demonstrating excellent expression levels. Consequently, the engineering focus shifted towards enhancing its activity. The enzyme was modelled using the crystal structure of *Alicyclobacillus acidocaldarius* endoglucanase Cel9A (AaCel9A) in complex with cello-oligosaccharides (PDB Structure ID 3H3K) as a template. The modelling process was conducted through the Modeller interface in Chimera software. The modelled egSth structure (in grey), overlaid with the AaCel9A structure (in cyan), is shown in Figure 17. The active site of egSth exhibits the typical groove configuration of GH9 family enzymes, with the substrate (depicted in yellow) fitting into this groove.

The comparison of the modelled egSth structure with AaCel9A revealed the presence of two metal binding sites: one for a calcium ion (shown in orange) and another for a zinc ion (shown in magenta). The zinc-binding site is conserved in some, but not all, members of the GH9 family, while the calcium-binding site is found in all GH9 crystal structures, suggesting that it is conserved across the entire family. In the case of AaCel9A, both the zinc and calcium binding sites play a crucial role in maintaining the enzyme’s structural integrity (Eckert et al., 2009). The study by Younesi et al. (2016) also demonstrated that calcium played a significant role in enhancing both the catalytic efficiency and the thermal stability of AaCel9A. This suggests that the presence of calcium ions can have a dual function in maintaining structural integrity and boosting enzymatic performance, a property likely shared across similar GH9 family enzymes.

In our previous studies, the effects of zinc and calcium on the thermal stability of egSth were explored by supplementing the growth medium with varying concentrations of these metal ions. The results showed that adding 10 mM calcium had a clear positive impact on enzyme stability, while 0.5 nM zinc produced only a minor improvement. Interestingly, when both calcium and zinc were added simultaneously, the combined effect on stability was lower than that of calcium alone. This suggests that calcium is more critical in stabilizing egSth, with zinc offering only a slight contribution.

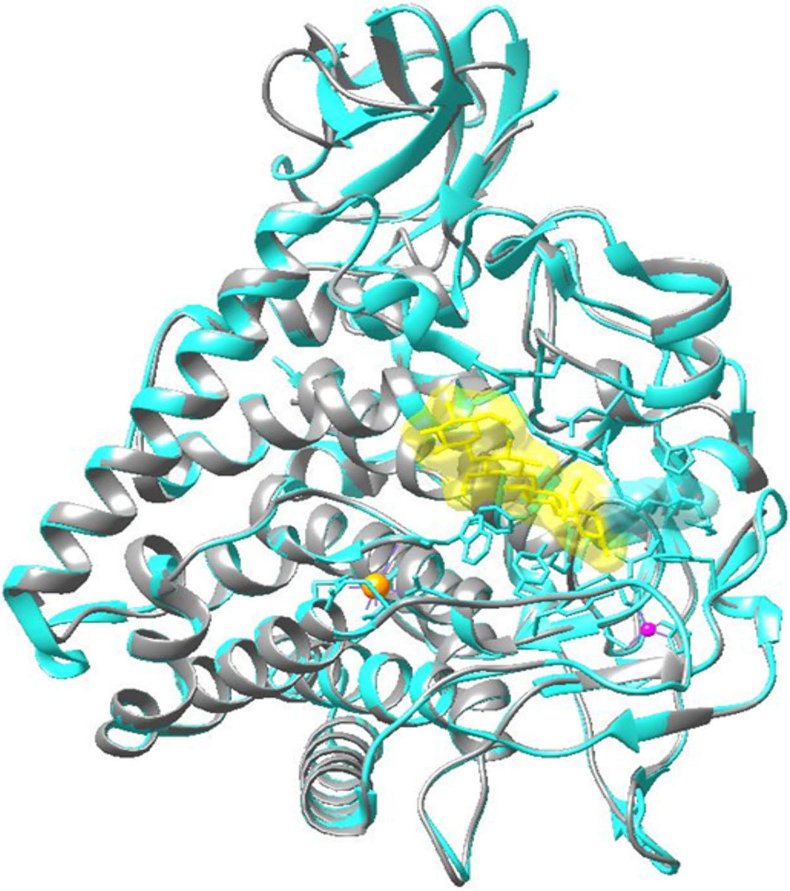


Figure 17. Homology model of endoglucanase from *Spirochaeta thermophila* egSth structure (grey) overlayed with *Alicyclobacillus acidocaldarius* endoglucanase AaCel9A (cyan, PDB Structure ID 3H3K), one for a calcium ion (shown in orange) and another for a zinc ion (shown in magenta).

5.2.2 Semi-rational engineering approach

Before the studies described in this thesis, some initial engineering efforts had already been carried out using both rational (site-directed mutagenesis) and semi-rational (site-directed libraries) approaches. One such effort involved testing mutations on the calcium-binding residues to stabilize the loop via hydrogen bonds, potentially bypassing the need for calcium ions to maintain stability. Despite these modifications, when wildtype and mutant variants were produced and tested for thermal stability, no significant differences were observed between the variants and the wildtype enzyme. This indicated that altering the calcium-binding residues did not substantially impact the enzyme's thermal stability.

In addition to the previous efforts, some libraries were designed and tested based on structure and sequence alignments. Figure 18A illustrates the modeled egSth structure with conservation coloring derived from multiple sequence alignments: highly conserved residues are depicted in red, semi-conserved residues in white, and non-conserved residues in blue. The substrate binding site is highly conserved, while some residues involved in substrate binding display semi-conservation. These semi-conserved residues were chosen as initial hot spots for site-directed mutagenesis and library construction, aiming to enhance activity without altering the most conserved residues, which were preserved to maintain the enzyme's core functions. This approach focused on improving activity while minimizing the risk of compromising the enzyme’s structural integrity.

During the PCR implementation of the libraries, unintended mutations occurred, notably at residue G656S, which surprisingly resulted in approximately a 100% increase in activity. This outcome was unexpected, given that alignment data indicated G656 was highly conserved among the closest homologous genes. Figure 19 presents the partial alignment, highlighting the conservation of this residue. Despite its high conservation, the G656S mutation significantly boosted enzyme performance, demonstrating how even slight changes in conserved regions can lead to substantial functional improvements. This serendipitous discovery opened new avenues for further exploration of activity-enhancing mutations in the enzyme.

Figure 18B shows the modelled enzyme *Spirochaeta thermophila* endoglucanase (grey) aligned with the structure of *Alicyclobacillus acidocaldarius* cellulase CelA in complex with cellobiose (light grey, PDB ID: 3H2W). The substrate (cellobiose)

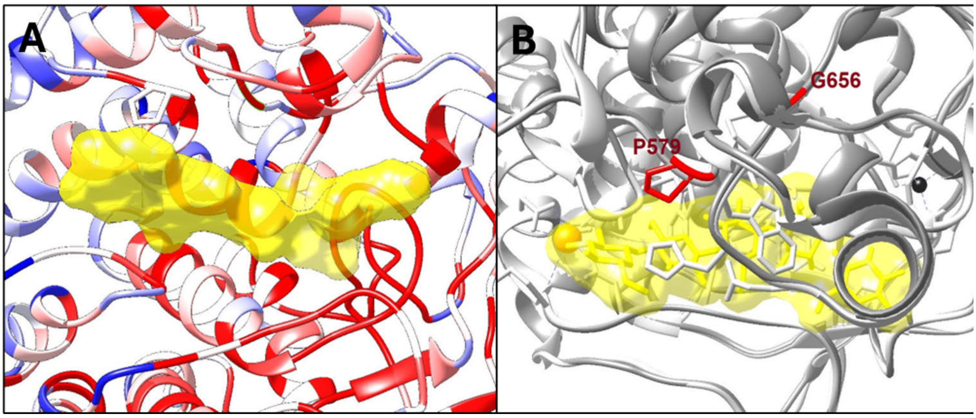


Figure 18. A) Part of the modelled structure of Spirochaeta thermophila (egSth) with multiple sequence alignment-based conservation coloring showing most conserved residues red, semi-conserved white and non-conserved blue. **B**) modelled structure of Spirochaeta thermophila (egSth) superimposed with Structure of A. acidocaldarius cellulase CelA in complex with cellobiose (PDB ID:3H2W) to visualize the part of active site and residues involved with substrate binding.

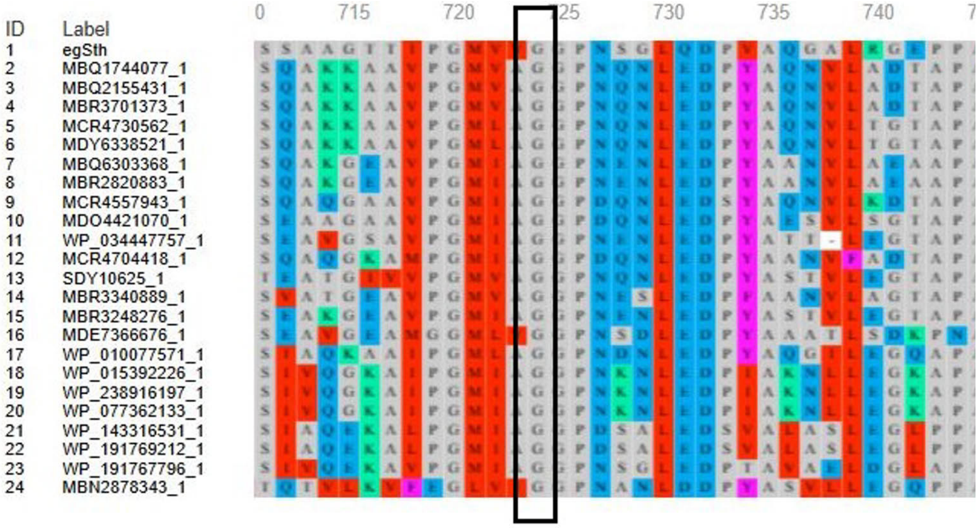


Figure 19. Partial alignment of egSth sequence with closest homologs. Random mutation appeared in residue that is highly conserved.

is depicted in yellow, and key residues (P579 and G656) involved in substrate binding, as described in the patent WO2020/193452 A1, are highlighted in red. As indicated in the patent, these residues have undergone beneficial mutations that enhance enzyme activity. The calcium ion (orange) and zinc ion (black), essential for structural integrity and stability, are also displayed. These mutations improve activity while preserving the enzyme's stability, showing potential for industrial applications.

Due to the promising findings with the G656 mutation, a library was designed to test all 19 other natural amino acids at this residue. This approach revealed similar improvements in activity with mutations such as G656V and G656A (as shown in Table 2 of patent IV). A library targeting residue P579 also demonstrated significant benefits, with many replaced amino acid mutants increasing specific activity by 50-100% (also detailed in Table 2 of patent IV). Combining these beneficial mutations led to a notable overall improvement, with one of the best combinations achieving approximately a 170% increase in activity.

However, in previous studies, it was observed that while some mutations in the binding area (including unintended insertions from PCR) improved activity, they also dramatically altered the enzyme's thermostability. This instability was hypothesized to be due to an open active site cleft, which, while facilitating substrate binding and release, resulted in misfolding of the covering loop and a consequent loss of enzyme stability. This raises the question of whether rationally designed loop deletions could enhance activity while maintaining stability. Further investigation into targeted loop modifications could potentially offer a solution to balance activity improvements with the preservation of thermal stability.

5.2.3 Loop deletions

The structure of egSth was compared to a range of thermostable endoglucanases, revealing that the loop region between Asp 657 and Cys 677 in egSth is highly variable in length and structure. Many thermostable endoglucanases have a significantly shorter loop. In the crystal structure of one of the most potent bacterial endoglucanases, CelA from *Caldicellulosiruptor bescii* (PDB ID: 4DOE), this loop consists of only five amino acids (Cela *et al.*, 2013). This is clearly illustrated in Figure 20, where the modelled structure of egSth (in grey) is aligned with the crystal structure of the thermostable enzyme *Caldicellulosiruptor bescii* CelA (in light grey, PDB ID: 4DOE). The figure shows the substrate binding pocket and highlights the loop (in green) that is significantly shorter in the thermostable CelA.

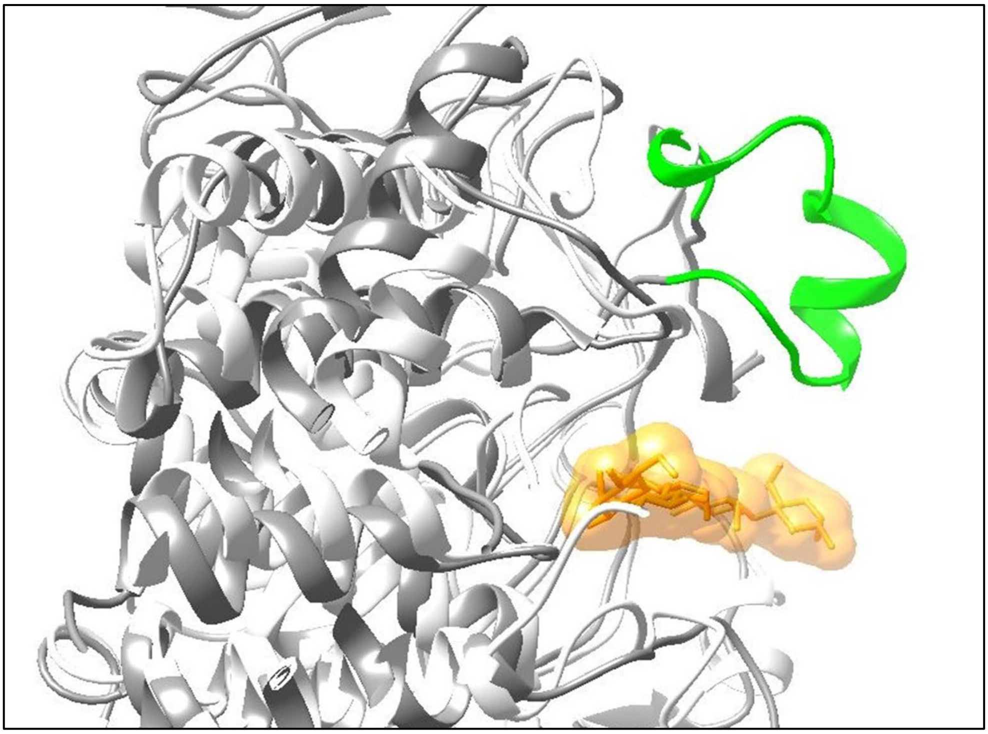


Figure 20. The modelled structure of egSth (dark grey) matched with *Caldicellulosiruptor bescii* CelA (PDB code 4DOE) structure (light grey). Substrate beta-cellotriose and alpha-cellotriose are shown in orange. The mutagenesis region, extended loop area Asp 657-Cys 677, around the active site of egSth are shown in light green, and the loop area that was used in replacement is shown in pink (Gly 427-Ser 431).

Based on the observed structural differences, a deletion was designed to remove the extended loop region in *Spirochaeta thermophila* egSth, compared to *Caldicellulosiruptor bescii* CelA. The deletion involved replacing the extended loop region (highlighted in green in Figure 20) with glycine residues (GG). This modification was implemented using full plasmid PCR and recombination techniques. After confirming the correct deletion in the gene, recombinant proteins were produced, including both the deletion variant and the wild-type enzyme, with three parallel samples prepared under standard conditions. The activity and thermostability of the proteins were tested with a 20-minute incubation at 70 °C. The deletion variant exhibited a 50% decrease in activity compared to the wild-type enzyme, although its stability was retained (Figure 21).

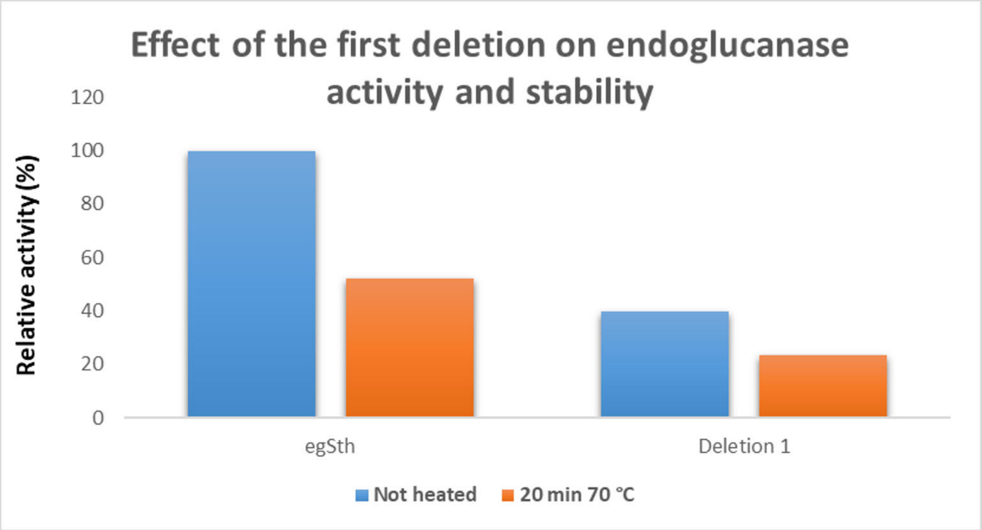


Figure 21. The relative activity and stability of the first deletion variant. Even if the activity is clearly reduced, the stability of the enzyme is not altered.

Even though the activity was decreased, this deletion demonstrated that removing certain parts of the loop is feasible without altering the enzyme’s stability. Additionally, beneficial modifications related to activity while maintaining stability had already been achieved (as described in the patent). Therefore, we adopted a new rational approach by replacing the egSth loop (green in Figure 20, residues Asp 657-Cys 677) with the loop sequence from *C. bescii* CelA. The sequence to be replaced, NSGLQDPVAQGALRGEPPAKC, was substituted with GSDDS, as shown in the alignment in Figure 22.

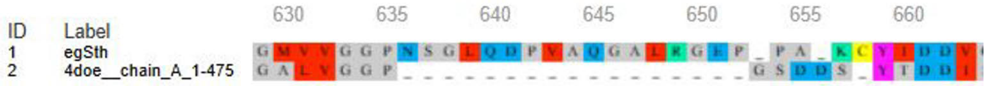


Figure 22. Loop sequence alignment of *Spirochaeta thermophila* (egSth) and C.bescii CelA.

The enzyme variant with the loop replaced by the C. bescii loop was named egSthCA. After implementing the loop replacement using full plasmid PCR and transformation, three clones were selected for miniprep DNA preparation. Following verification of the correct clones with the intended modification, proteins were produced in a 24-well plate, with three parallel samples for the wildtype and each variant (egSthCA, egSth_656S, and egSth_656S&P579V). After lysing the cells and collecting the soluble protein supernatants, the activities were analyzed using barley β-glucan tablets (Megazyme) with Britton-Robinson buffer (pH 3-8) to determine the enzymes' relative activities and pH optima. Thermal stabilities were assessed by incubating enzyme solutions at 65°C for 0, 10, 20, 30, 40, 50, and 60 minutes. Activity analyses were performed for the post-heated samples as described previously.

Figure 23 presents the results of relative activities, pH optima, and thermal stability. Replacing the loop with the *C. bescii* loop led to a 3.5-fold increase in enzyme activity, likely due to creating a more open substrate-binding cavity. Previous variants were also included in the analysis. The single mutation G656S improved relative activity by 2-fold, while the double mutant G656S&P579V increased activity by nearly 300%. Residue P579 is involved in substrate binding so that the mutation might impact substrate interaction. However, the G656V mutation is more likely to influence loop folding, resulting in a more open substrate binding cavity and enhancing activity. The thermal stability of the variants shows only a minor decrease compared to the wild type, primarily observed at shorter incubation times.

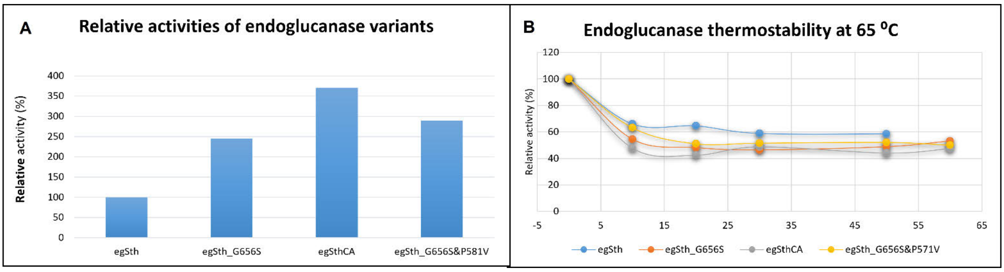


Figure 23. Wildtype egSth compared to variants egSth_G656S, egSthCA and egSth_G656S&P581V. The activity of the variants is improved, and thermal stability is retained.

5.2.4 Loop library

Based on structure alignment of the most similar endoglucanases to egSth with known 3D structures using SwissModel, it was observed that a loop overhanging the active site is highly variable in length and structure within this group (Fig. 4b, Publication IV). The loop corresponding to amino acids 657–677 in length was replaced with a library of short (5 amino acids) loops at the respective positions of CelA homologs. Since Aspartic acid at position 4 of the loop was conserved among the homologs, this residue was preserved in the library. For the other residues, only partially randomized codons were used to encode desired sets of amino acids.

These libraries were synthesized as oligonucleotides and introduced into the target gene by full plasmid PCR. After primary screening with the CMC-Congo Method, the most active clones, characterized by the most significant halos, were further analyzed. These clones were subsequently seeded from replica plates generated during colony picking, and recombinant soluble proteins were produced as described earlier. Secondary screening was performed using long substrates, and the best-performing clones were sequenced. Fourteen unique clones were identified.

These clones were further characterized and compared, with detailed results presented in Publication IV. The clone with the highest volumetric activity towards long substrates was selected for further purification and characterization. It was designated as egSthL6 (with the loop sequence NSADG, compared to the wildtype GSDDS). This clone demonstrated a four-fold increase in activity for long substrates compared to the parent short loop variant, egSthCA. However, both variants exhibited reduced thermal stability compared to the wild-type enzyme egSth. Despite the design being based on a thermostable enzyme structure, further engineering around the loop area may be required to improve proper folding and intermolecular interactions to enhance structural stability. Nonetheless, these modifications alter substrate specificity and reaction rates.

5.2.5 Endoglucanase application example

Endoglucanases can be utilized in a variety of applications. If the goal is to maximize hydrolysis, several different enzymes must be included in the enzyme mixture. Individual endoglucanases can be used when the objective is to modify cellulose fibers by introducing specific breaks in a cellulose strand. Enzyme cocktails must be tailored to suit the process conditions and the substrate. MetZyme® BRILA™ is one of the Metgen-developed products, including cellulases like endoglucanase.

In the pulp and paper industry, recycled paper is an essential raw material. Lower-grade fibers can be used instead of more expensive virgin fibers, but there are several challenges, such as the drainability of recycled pulp and reduced strength. One major issue with recycled fiber is its high relative surface area of fines, leading to lower dewatering rates than virgin fiber. Enzyme treatment can help address this problem by reducing the need for extensive refining to achieve the desired strength. Additionally, enzyme treatment can improve properties such as the paper's tensile index and overall strength. (Verma, Bhardwaj and Singh, 2015; Deng *et al.*, 2016)

Reduced refining results in fewer fines, which improves pulp dewatering during papermaking. This, in turn, allows for faster drying of the paper and enables increased paper machine speeds, thereby boosting production rates. Publication V presents several examples of improved dewatering by treating wastepaper and recycled fiber with MetZyme® BRILA™ products. The benefits of using enzymes in tissue paper machines were evident, with increased strength and tensile index and significant energy savings in refining (35-60%). Additionally, enzyme treatment improved machine runnability by reducing break times.

In the study by Verma et al. (2015), the use of endoglucanase significantly improved pulp drainage (14.5–22.8%). However, a mixture with lower endoglucanase and higher cellobiohydrolase activity did not show the same improvement. Additionally, the study observed increases in both tensile index and smoothness. Another study demonstrated that endoglucanase-mediated modification of kraft pulp enhanced the tensile index by reorienting microfibrils and improving the alignment of carbohydrate chains. (Deng et al., 2016).

6 Summary/Conclusions

Enzymatic valorization represents a versatile and sustainable approach to harnessing the potential of lignocellulosic biomass for producing biofuels, biochemicals, and bioproducts. By leveraging the catalytic power of enzymes, this process enables efficient biomass conversion while minimizing environmental impact. As research advances and technological innovations continue to emerge, enzymatic valorization holds the key to unlocking lignocellulosic biomass's full potential as a bio-based economy's cornerstone.

In this thesis, two widely used enzymes for lignocellulose modification have been improved through engineering. A novel alkaline-stable laccase was engineered to enhance its soluble expression and increase its activity, resulting in an overall ~100-fold increase in specific activity. These improvements are highly valuable and make the enzyme suitable for industrial-scale applications. This laccase can be utilized in lignin valorization, improving the depolymerization and fractionation of lignin into smaller molecular weight fractions for various applications. The laccase developed and described in this thesis is unique in the market, as no other fungal laccases have demonstrated such thermostability and activity under alkaline conditions. It is a key component of METNIN technology (https://www.metgen.com/lignin-products/), which is applied in producing lignin fractions for paperboard, polyols for insulation panels, and lignin-based resins and adhesives. The enzyme is recycled within the METNIN process, making its stability a critical factor.

The other enzyme engineered in this thesis was endoglucanase. The engineering improved the enzyme's activity three-fold. Although the goal was also to enhance thermostability, this was not achieved despite radical modifications to the active site, including replacing the existing loop with a much shorter loop from a thermostable endoglucanase. By implementing a library based on the conserved residues of thermostable enzymes, we could modify the enzyme's substrate specificity. This enzyme is now part of some enzyme cocktails called MetZyme® BRILA (https://www.metgen.com/fibre-solutions/), used for fiber modification to improve strength. Although competing fungal alternatives are less thermostable, their production is more cost-effective, and due to their high activity, the retention time –

the duration an enzyme remains in contact with the substrate - allows for their use. To make this enzyme a more advantageous option, further optimization is needed to increase activity and production yield. Further studies could focus on improving thermostability through additional modifications by directed evolution for heat-tolerant variants. Investigating the enzyme's kinetics across different temperatures and pH levels, as well as exploring co-factors or activators to enhance activity, could optimize performance.

Acknowledgements

The experimental work for this study was conducted between 2017 and 2019 at Metgen Oy, an industrial biotech company in Kaarina, Finland. Although I began writing this thesis during that time, progress was delayed as I started working at Orion in 2019, and I faced several personal challenges that year and afterward. As a result, the thesis was put on hold for several years. In 2024, I finally reached a point where I could take a 5.5-month study leave from work, which allowed me to complete the writing of this thesis.

I would like to express my sincere gratitude to my supervisors, Adjunct Professor Jarmo Niemi and Dr. Klara Birikh, for their invaluable guidance and support throughout these years. While working at Metgen, Klara consistently offered insightful advice, mentorship, and practical solutions to the challenges I faced along the way. Jarmo served as my connection to the University of Turku, always providing support and guidance. I am also truly grateful to Alex Michine for giving me the opportunity to grow professionally and pursue my studies alongside my work.

I would like to extend my heartfelt thanks to all my co-authors—Klara Birikh, Toni Grönroos, Anu Suonpää, Matti Heikkilä, Bas Romein, Petri Ihalainen, Sarah Malandra, Juan De Dios Barajas Lopez, Yana Berlina, Rafael Álvarez, Alex Michine, Antoine Mialon, Antti Varho, and Sami-Pekka Rantanen. Many of them were not only invaluable collaborators but also wonderful colleagues during my years at Metgen.

My deepest gratitude goes to my family. I am incredibly proud of my three wonderful children, Emilia, Juuso, and Julia. They are the heart of my life, bringing me immense happiness and stability, which made it possible for me to complete this thesis. Even without realizing it, they provided the kind of support I needed the most.

I also want to thank my mum for always being there for me. She has been a true role model, showing how to balance a career and motherhood with wisdom, strength, and determination. I am equally grateful to my three big brothers, who have made me mentally stronger and taught me not to take life too seriously.

Special thanks to Janne for all the love, joy, and stability you have brought into my life and for your constant support.

Lieto, February 2025

Veera Nikoskelainen

References

Abdelaziz, O.Y. *et al.* (2016) ‘Biological valorization of low molecular weight lignin’, *Biotechnology Advances*, 34(8), pp. 1318–1346. Available at: https://doi.org/10.1016/j.biotechadv.2016.10.001.

Bajpai, P. (1999) ‘Application of enzymes in the pulp and paper industry’, *Biotechnology*, (15), pp. 147–157concomitant.

Balat, M., Balat, H. and Öz, C. (2008) ‘Progress in bioethanol processing’, *Progress in Energy and Combustion Science*, pp. 551–573. Available at: https://doi.org/10.1016/j.pecs.2007.11.001.

Bansal, S. and Kundu, B. (2022) *Protein engineering: Methods and applications*, *Advances in Protein Molecular and Structural Biology Methods*. Elsevier Inc. Available at: https://doi.org/10.1016/B978-0-323-90264-9.00039-8.

Benatti, A.L.T. and Polizeli, M. de L.T. de M. (2023) ‘Lignocellulolytic Biocatalysts: The Main Players Involved in Multiple Biotechnological Processes for Biomass Valorization’, *Microorganisms*, 11(1). Available at: https://doi.org/10.3390/microorganisms11010162.

Bornscheuer, U.T. and Pohl, M. (2001) ‘Improved biocatalysts by directed evolution and rational protein design’, pp. 137–143.

Burley, S.K. *et al.* (2019) ‘Protein Data Bank: The single global archive for 3D macromolecular structure data’, *Nucleic Acids Research*, 47(D1), pp. D520–D528. Available at: https://doi.org/10.1093/nar/gky949.

Camacho, C. *et al.* (2009) ‘BLAST+: Architecture and applications’, *BMC Bioinformatics*, 10, pp. 1–9. Available at: https://doi.org/10.1186/1471-2105-10-421.

Cañas, A.I. and Camarero, S. (2010) ‘Laccases and their natural mediators: Biotechnological tools for sustainable eco-friendly processes’, *Biotechnology Advances*, 28(6), pp. 694–705. Available at: https://doi.org/10.1016/j.biotechadv.2010.05.002.

Cela, C. *et al.* (2013) ‘Revealing Nature ’ s Cellulase Diversity :’, *Science*, 342(December), pp. 1513–1516.

Chaney, J.L. *et al.* (2017) ‘Widespread position-specific conservation of synonymous rare codons within coding sequences’, *PLoS Computational Biology*, 13(5), pp. 1–19. Available at: https://doi.org/10.1371/journal.pcbi.1005531.

Chiriac, A.I. *et al.* (2010) ‘Engineering a family 9 processive endoglucanase from Paenibacillus barcinonensis displaying a novel architecture’, *Applied Microbiology and Biotechnology*, 86(4), pp. 1125–1134. Available at: https://doi.org/10.1007/s00253-009-2350-8.

Choi, J.H., Keum, K.C. and Lee, S.Y. (2006) ‘Production of recombinant proteins by high cell density culture of Escherichia coli’, *Chemical Engineering Science*, 61(3), pp. 876–885. Available at: https://doi.org/10.1016/j.ces.2005.03.031.

Chukwuma, O.B. *et al.* (2020) ‘Lignocellulolytic enzymes in biotechnological and industrial processes: A review’, *Sustainability (Switzerland)*, 12(18), pp. 1–31. Available at: https://doi.org/10.3390/su12187282.

Contreras, F. *et al.* (2020) ‘Engineering robust cellulases for tailored lignocellulosic degradation cocktails’, *International Journal of Molecular Sciences*, 21(5). Available at: https://doi.org/10.3390/ijms21051589.

Deng, T. *et al.* (2016) 'Endoglucanase enzymatic modification of kraft pulp during recycling', *Biotechnology Letters*, 38(7), pp. 1139–1145. Available at: https://doi.org/10.1007/s10529-016-2088-1.

Ding, N. *et al.* (2017) 'Increased glycosylation efficiency of recombinant proteins in Escherichia coli by auto-induction', *Biochemical and Biophysical Research Communications*, 485(1), pp. 138–143. Available at: https://doi.org/10.1016/j.bbrc.2017.02.037.

Farinas, E.T., Bulter, T. and Arnold, F.H. (2001) 'Directed enzyme evolution', *Current Opinion in Biotechnology*, 12(6), pp. 545–551. Available at: https://doi.org/10.1016/S0958-1669(01)00261-0.

Fasim, A., More, V.S. and More, S.S. (2021) 'Large-scale production of enzymes for biotechnology uses', *Current Opinion in Biotechnology*, 69, pp. 68–76. Available at: https://doi.org/10.1016/j.copbio.2020.12.002.

Fatma, S. *et al.* (2018) 'Lignocellulosic Biomass: A Sustainable Bioenergy Source for Future', *Protein & Peptide Letters*, 25. Available at: https://doi.org/10.2174/0929866525666180122144504.

Firat Duzenli, O. and Okay, S. (2020) 'Promoter engineering for the recombinant protein production in prokaryotic systems', *AIMS Bioengineering*, 7(2), pp. 62–81. Available at: https://doi.org/10.3934/bioeng.2020007.

Galbe, M. and Zacchi, G. (2007) 'Pretreatment of lignocellulosic materials for efficient bioethanol production.', *Advances in biochemical engineering/biotechnology*, 108, pp. 41–65. Available at: https://doi.org/10.1007/10_2007_070.

Hill, J. *et al.* (2006) 'Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels', *Proceedings of the National Academy of Sciences of the United States of America*, 103(30), pp. 11206–11210. Available at: https://doi.org/10.1073/pnas.0604600103.

Hu, J. *et al.* (2018) 'Enzyme mediated nanofibrillation of cellulose by the synergistic actions of an endoglucanase, lytic polysaccharide monooxygenase (LPMO) and xylanase', *Scientific Reports*, 8(1), pp. 4–11. Available at: https://doi.org/10.1038/s41598-018-21016-6.

Iram, A., Cekmecelioglu, D. and Demirci, A. (2021) 'Ideal feedstock and fermentation process improvements for the production of lignocellulolytic enzymes', *Processes*, 9(1), pp. 1–26. Available at: https://doi.org/10.3390/pr9010038.

Isikgor, F.H. and Becer, C.R. (2015a) 'Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers', *Polymer Chemistry*, 6(25), pp. 4497–4559. Available at: https://doi.org/10.1039/c5py00263j.

Isikgor, F.H. and Becer, C.R. (2015b) 'Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers', *Polymer Chemistry*, 6(25), pp. 4497–4559. Available at: https://doi.org/10.1039/c5py00263j.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. Available at: https://doi.org/10.1038/s41586-021-03819-2.

Juturu, V. and Wu, J.C. (2012) 'Microbial xylanases: Engineering, production and industrial applications', *Biotechnology Advances*, 30(6), pp. 1219–1227. Available at: https://doi.org/10.1016/j.biotechadv.2011.11.006.

K, B.P. (2011) '2011_Issue_1_IPPTA_Articel_16', 23(1), pp. 181–186.

Katoh, K. and Standley, D.M. (2013) 'MAFFT multiple sequence alignment software version 7: Improvements in performance and usability', *Molecular Biology and Evolution*, 30(4), pp. 772–780. Available at: https://doi.org/10.1093/molbev/mst010.

Korendovych, I. V. (2018) 'Rational and semirational protein design', *Methods in Molecular Biology*, 1685, pp. 15–23. Available at: https://doi.org/10.1007/978-1-4939-7366-8_2.

Larentis, A.L. *et al.* (2014) 'Evaluation of pre-induction temperature, cell growth at induction and IPTG concentration on the expression of a leptospiral protein in E. coli using shaking flasks and microbioreactor', *BMC Research Notes*, 7(1), pp. 1–13. Available at: https://doi.org/10.1186/1756-0500-7-671.

Larkin, M.A. *et al.* (2007) 'Clustal W and Clustal X version 2.0', *Bioinformatics*, 23(21), pp. 2947–2948. Available at: https://doi.org/10.1093/bioinformatics/btm404.

Lin, L. *et al.* (2009) ‘Improved catalytic efficiency of Endo-β-1,4-glucanase from Bacillus subtilis BME-15 by directed evolution’, *Applied Microbiology and Biotechnology*, 82(4), pp. 671–679. Available at: https://doi.org/10.1007/s00253-008-1789-3.

Malherbe, S. and Cloete, T.E. (2002) ‘Lignocellulose biodegradation: Fundamentals and applications’, *Reviews in Environmental Science and Biotechnology*, 1(2), pp. 105–114. Available at: https://doi.org/10.1023/A:1020858910646.

Martins, L.O. *et al.* (2002) ‘Molecular and biochemical characterization of a highly stable bacterial laccase that occurs as a structural component of the Bacillus subtilis endospore coat’, *Journal of Biological Chemistry*, 277(21), pp. 18849–18859. Available at: https://doi.org/10.1074/jbc.M200827200.

Menon, V. and Rao, M. (2012) ‘Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept’, *Progress in Energy and Combustion Science*, 38(4), pp. 522–550. Available at: https://doi.org/10.1016/j.pecs.2012.02.002.

Nanda, S., A. Kozinski, J. and K. Dalai, A. (2015) ‘Lignocellulosic Biomass: A Review of Conversion Technologies and Fuel Products’, *Current Biochemical Engineering*, 3(1), pp. 24–36. Available at: https://doi.org/10.2174/2213385203666150219232000.

Østby, H. and Várnai, A. (2023) ‘Hemicellulolytic enzymes in lignocellulose processing’, *Essays in Biochemistry*, 67(3), pp. 533–550. Available at: https://doi.org/10.1042/EBC20220154.

Otto, B., Schlosser, D. and Reisser, W. (2010) ‘First description of a laccase-like enzyme in soil algae’, *Archives of Microbiology*, 192(9), pp. 759–768. Available at: https://doi.org/10.1007/s00203-010-0603-7.

Pettersen, E.F. *et al.* (2004) ‘UCSF Chimera?A visualization system for exploratory research and analysis’, *Journal of Computational Chemistry*, 25(13), pp. 1605–1612. Available at: https://doi.org/10.1002/jcc.20084.

Pollegioni, L., Tonin, F. and Rosini, E. (2015) ‘Lignin-degrading enzymes’, *FEBS Journal*, pp. 1190–1213. Available at: https://doi.org/10.1111/febs.13224.

Raj, T. *et al.* (2022) ‘Lignocellulosic biomass as renewable feedstock for biodegradable and recyclable plastics production: A sustainable approach’, *Renewable and Sustainable Energy Reviews*, 158(April 2021), p. 112130. Available at: https://doi.org/10.1016/j.rser.2022.112130.

Ribeiro, R.S.A. *et al.* (2019) ‘Production of nanocellulose by enzymatic hydrolysis: Trends and challenges’, *Engineering in Life Sciences*, 19(4), pp. 279–291. Available at: https://doi.org/10.1002/elsc.201800158.

Rigoldi, F. *et al.* (2018) ‘Review: Engineering of thermostable enzymes for industrial applications’, *Citation: APL Bioengineering*, 2, p. 11501. Available at: https://doi.org/10.1063/1.4997367.

Rosano, G.L. and Ceccarelli, E.A. (2014) ‘Recombinant protein expression in Escherichia coli: Advances and challenges’, *Frontiers in Microbiology*, 5(APR), pp. 1–17. Available at: https://doi.org/10.3389/fmicb.2014.00172.

Singh, G. *et al.* (2016) ‘Enzymes: Applications in Pulp and Paper Industry’, in *Agro-Industrial Wastes as Feedstock for Enzyme Production: Apply and Exploit the Emerging and Valuable Use Options of Waste Biomass*. Elsevier Inc., pp. 157–172. Available at: https://doi.org/10.1016/B978-0-12-802392-1.00007-1.

Singh, V. *et al.* (2017) ‘Strategies for fermentation medium optimization: An in-depth review’, *Frontiers in Microbiology*, 7(JAN). Available at: https://doi.org/10.3389/fmicb.2016.02087.

Somerville, C. *et al.* (2010) ‘Feedstocks for Lignocellulosic Biofuels’, *Science*, 329(5993), pp. 790–792. Available at: https://doi.org/10.1126/science.1189268.

Stephens, D.E. *et al.* (2007) ‘Directed evolution of the thermostable xylanase from Thermomyces lanuginosus’, *Journal of Biotechnology*, 127(3), pp. 348–354. Available at: https://doi.org/10.1016/j.jbiotec.2006.06.015.

Studier, F.W. (2005) ‘Protein production by auto-induction in high density shaking cultures.’, *Protein expression and purification*, 41(1), pp. 207–234. Available at: https://doi.org/10.1016/j.pep.2005.01.016.

Thompson, W. and Meyer, S. (2013) ‘Second generation biofuels and food crops: Co-products or competitors?’, *Global Food Security*, pp. 89–96. Available at: https://doi.org/10.1016/j.gfs.2013.03.001.

Tracewell, Cara, Arnold, F. (2009) ‘Directed enzyme evolution: climbing fitness peaks one amino acid at a time’, *Current Opinion in Chemical Biology*, 13(1), pp. 427–427. Available at: https://doi.org/10.1016/j.cbpa.2009.01.017.

Ukkonen, K. *et al.* (2013) ‘Use of slow glucose feeding as supporting carbon source in lactose autoinduction medium improves the robustness of protein expression at different aeration conditions’, *Protein Expression and Purification*, 91(2), pp. 147–154. Available at: https://doi.org/10.1016/j.pep.2013.07.016.

Ventura, S. (2005) ‘Sequence determinants of protein aggregation: Tools to increase protein solubility’, *Microbial Cell Factories*, 4(February). Available at: https://doi.org/10.1186/1475-2859-4-11.

Verma, P.K., Bhardwaj, N.K. and Singh, S.P. (2015) ‘Selective hydrolysis of amorphous cellulosic fines for improvement in drainage of recycled pulp based on ratios of cellulase components’, *Journal of Industrial and Engineering Chemistry*, 22, pp. 229–239. Available at: https://doi.org/10.1016/j.jiec.2014.07.015.

Waldo, G.S. (2003) ‘Genetic screens and directed evolution for protein solubility’, *Current Opinion in Chemical Biology*, 7(1), pp. 33–38. Available at: https://doi.org/10.1016/S1367-5931(02)00017-0.

Wang, M., Si, T. and Zhao, H. (2012) ‘Biocatalyst development by directed evolution’, *Bioresource Technology*, 115, pp. 117–125. Available at: https://doi.org/10.1016/j.biortech.2012.01.054.

Wang, T. *et al.* (2009) ‘Changing the metal binding specificity of superoxide dismutase from thermus thermophilus HB-27 by a single mutation’, *Molecular Biotechnology*, 42(2), pp. 146–153. Available at: https://doi.org/10.1007/s12033-009-9149-9.

Wang, T. *et al.* (2020) ‘Small design from big alignment: engineering proteins with multiple sequence alignment as the starting point’, *Biotechnology Letters*, 42(8), pp. 1305–1315. Available at: https://doi.org/10.1007/s10529-020-02914-0.

Wang, Y. *et al.* (2021) ‘Protein domain identification methods and online resources’, *Computational and Structural Biotechnology Journal*, 19, pp. 1145–1153. Available at: https://doi.org/10.1016/j.csbj.2021.01.041.

Waterhouse, A. *et al.* (2018) ‘SWISS-MODEL: Homology modelling of protein structures and complexes’, *Nucleic Acids Research*, 46(W1), pp. W296–W303. Available at: https://doi.org/10.1093/nar/gky427.

Wilson, C.J. (2015) ‘Rational protein design: Developing next-generation biological therapeutics and nanobiotechnological tools’, *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, 7(3), pp. 330–341. Available at: https://doi.org/10.1002/wnan.1310.

Wilson, David B.; Sahm, Hermann; Stahmann, Klaus-Peter; Koffas, M. (2020) *Industrial Microbiology*.

Xu, S.Y. *et al.* (2023) ‘Recent advances in structure-based enzyme engineering for functional reconstruction’, *Biotechnology and Bioengineering*, 120(12), pp. 3427–3445. Available at: https://doi.org/10.1002/bit.28540.

Yoo, Y.J. *et al.* (2017) *Fundamentals of enzyme engineering*, *Fundamentals of Enzyme Engineering*. Available at: https://doi.org/10.1007/978-94-024-1026-6.

Zheng, F. *et al.* (2018) ‘Enhancing the catalytic activity of a novel GH5 cellulase GtCel5 from Gloeophyllum trabeum CBS 900.73 by site-directed mutagenesis on loop 6’, *Biotechnology for Biofuels*, 11(1), pp. 1–13. Available at: https://doi.org/10.1186/s13068-018-1080-5.

Zhou, C.H. *et al.* (2011) ‘Catalytic conversion of lignocellulosic biomass to fine chemicals and fuels’, *Chemical Society Reviews*, 40(11), pp. 5588–5617. Available at: https://doi.org/10.1039/c1cs15124j.

Zhou, M. *et al.* (2023) ‘Laccase-catalyzed lignin depolymerization in deep eutectic solvents: challenges and prospects’, *Bioresources and Bioprocessing*, 10(1). Available at: https://doi.org/10.1186/s40643-023-00640-9.