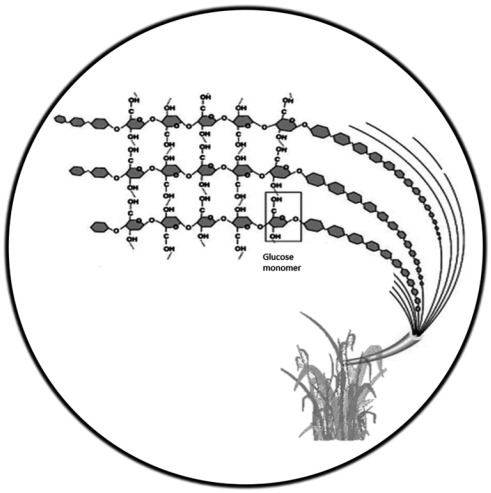
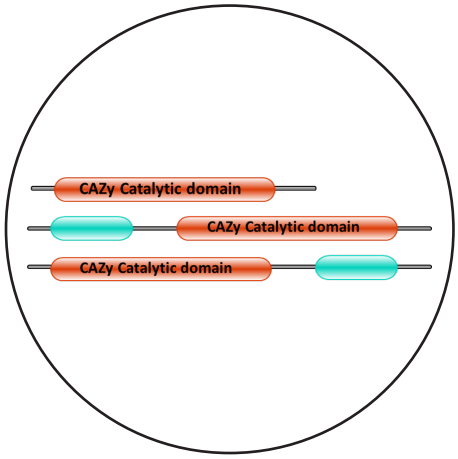
Unraveling modular architecture and domain engineering of carbohydrate-active enzymes: key insights for sustainable bio-based processes

Sidar, A.



Chapter 1

General Introduction



1.1 Introduction

In recent years, there has been a growing global focus on the substitution of fossil-based chemicals with biomass-derived alternatives. This shift is motivated by both the adverse environmental effects of non-degradable chemicals and the dwindling availability of fossil fuels resources such as oil, coal, and gas (Antar et al., 2021; Vogt and Weckhuysen, 2024). These fossil fuels are formed over millions of years through the decomposition of organic materials such as ancient plants and animals. However, the rate of consumption far exceeds the rate of natural replenishment, leading to concerns about depletion (Antar et al., 2021; Balat and Ayar, 2005). Fossil fuels are finite resources, meaning they cannot be replenished within the timespan of human civilization. Once depleted, they are irrevocably lost. This reality underscores the importance to transition toward alternatives for sustainable energy sources. Moreover, the extraction and combustion of fossil fuels release large quantities of greenhouse gases, primarily carbon dioxide (CO2) into the atmosphere (Vogt and Weckhuysen, 2024). This contributes significantly to global warming and climate change, as CO2 is a major driver of the greenhouse gas effect.

Carbohydrate biopolymers derived from biomass offer a promising alternative to replace fossil resources, presenting a renewable source and environmentally friendly option for the production of bioenergy as well as various biochemicals. By degrading polysaccharides within these biopolymers, valuable compounds such as oligosaccharides and monosaccharides are produced. These mono- or oligomeric sugars serve as building blocks for producing biofuel, bioplastics, many other biochemicals essential to pharmaceutical and food industries (Baranwal et al., 2022; Mujtaba et al., 2023; Reshmy et al., 2021), as illustrated in Fig. 1. In that regard, utilizing carbohydrate biopolymers contributes an effort to combat climate change and is paving the way towards a more sustainable future (Wang et al., 2021). Carbohydrate biopolymers encompass a diverse array of biomass materials such as lignocellulose, starch, pectin, chitin, and glycogen. This thesis specifically focuses on enzymatic degradation of carbohydrate biopolymers derived from starch and lignocellulose biomass, which are among the most abundant biopolymers in nature (Dahmen et al., 2019; Yu et al., 2021).

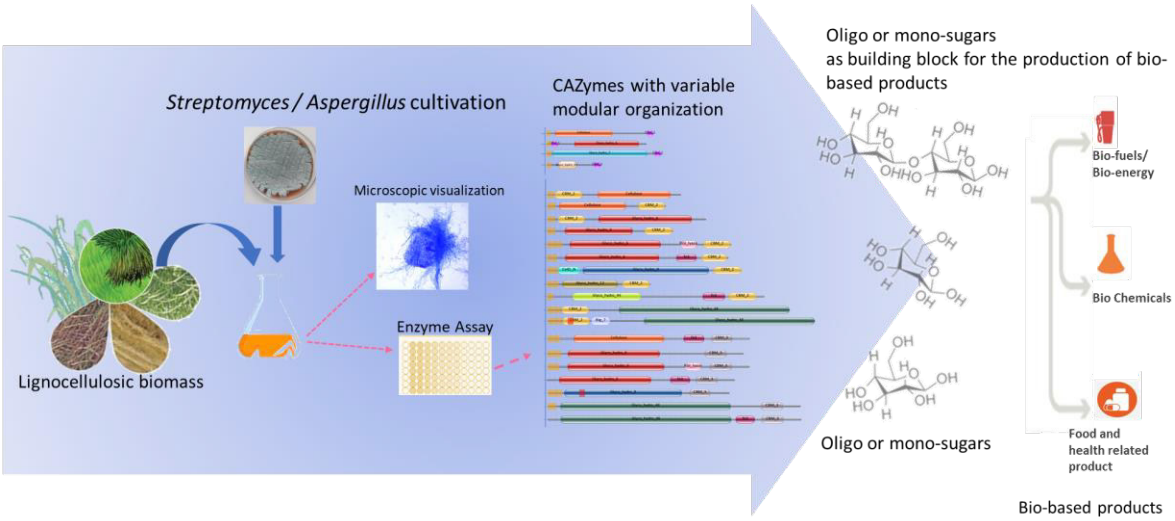


Fig. 1. Schematic process of utilizing lignocellulosic biomass through enzymatic degradation by enzymes from microorganisms, leading to the sustainable production of various value-added products.

1.2 Carbohydrate biopolymers from lignocellulose and starch

Lignocellulose biomass is a complex carbohydrate biopolymer primarily comprised of cellulose, hemicellulose, and lignin. Within lignocellulose structure, cellulose is the major polysaccharides constituent, comprising glucose units linked together via β-1,4-glycosidic bonds, and thus serves as the structural backbone of lignocellulose (Zoghlami and Paës, 2019). Furthermore, hemicellulose is a polysaccharide existing as a cross-linking matrix surrounding the cellulose structure, while lignin is an aromatic compound encapsulating cellulose to provide rigidity and making the plant cell wall waterproof. Lignin is also considered as a barrier hindering the enzymatic degradation of cellulose (Isikgor and Becer, 2015).

Agricultural by-products, such as rice straw, wheat straw, bagasse, and corn stover yields substantial volumes of lignocellulosic biomass. Amongst others, rice straw stands out as the most abundant agricultural waste globally due to the widespread cultivation of rice across the globe, presenting an ideal feedstock for bioethanol production (Akshaya et al., 2023; Saini et al., 2015). Rice is the staple food for nearly half the world's population, and thus, rice straw production is also increased tremendously over the years as a result of growing more rice for the increasing global population (Logeswaran et al., 2020; Singh et al., 2021). Rice straw biomass is rich in lignocellulose content with cellulose and

hemicellulose being the first and second major polysaccharide component, respectively (Sarkar et al., 2012; Sharma et al., 2023). Breaking down the (hemi)cellulose within rice straw to its constituent mono- or oligo-sugars presents a promising avenue for bioethanol production, underscoring its potential as a valuable resource. Research related to the use of rice straw as lignocellulosic substrate in enzymatic release of these sugars is addressed in **Chapter 3** and **Chapter 5**.

Starch is another complex carbohydrate polymer found in nature. In general, starch exists in semi-crystalline granules consisting of two types of glucose polymers, amylose and amylopectin (Apriyanto et al., 2022; Waterschoot et al., 2015). Amylose served as the backbone of starch, comprising of linear glucose units linked together by α-1,4 glycosidic bonds (Cornejo-Ramírez et al., 2018). Amylose forms a relatively straight chain structure, which contributes to the semi-crystalline nature of starch. Conversely, amylopectin is a branched polymer of glucose units linked by α-1,4 glycosidic bonds with occasional α-1,6 glycosidic bonds at branch points (Cornejo-Ramírez et al., 2018). This branched structure imparts greater solubility and more amorphous properties compared to amylose (Cornejo-Ramírez et al., 2018; Waterschoot et al., 2015). Research related to the use of various starches as substrate for enzymatic release on mono-sugars is presented in **Chapter 4** of this thesis.

1.3 Carbohydrate-Active Enzymes (CAZymes)

To harness the biomass potential for producing biofuels, biochemicals, and many other biomaterials, the carbohydrate polymers within lignocellulose and starch must first undergo enzymatic degradation into mono- or oligosaccharide sugars to be subsequently used for the production of valuable bio-based products (Zhu et al., 2016). This degradation typically requires the coordinated actions of multiple specialized carbohydrate degrading enzymes, known as Carbohydrate-Active Enzymes (CAZymes) which act together as a cocktail with synergistic activities (Bredon et al., 2018; Guo et al., 2023). Based on the CAZy classification system (CAZy database: www.cazy.org), CAZymes involved in the degradation of complex carbohydrate biopolymer are classified into several classes, including glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and auxiliary activities (AAs). The CAZy database presents a consistent nomenclature for CAZymes and is continually growing

and updated with new discoveries of sequences, protein structures and biochemical characterizations (Andlar et al., 2018; Drula et al., 2022). According to the amino acid sequence similarity and protein structure, each CAZyme class is subdivided into a number of families. Amongst CAZymes classes, GHs consist of the largest number of families. Currently, as per May 2024, GHs comprised of 189 families, while PLs, CEs, and AAs consist of 43, 20, and 17 families, respectively (CAZy database: www.cazy.org) (Drula et al., 2022).

The GHs enzymes are responsible for the hydrolysis of glycosidic bonds within the carbohydrate molecules. GH activity on polysaccharides can be categorized as endo- or exo-, referring to their ability to cleave polysaccharides, either within the chain or from the end of the chain, respectively (Andlar et al., 2018). Additionally, CEs play a crucial role in assisting GHs function by removing methyl, acetyl and phenolic esters of hemicelluloses, thus increasing the GHs accessibility to the polysaccharide chains (Duan et al., 2021; van den Brink and de Vries, 2011). Moreover, polysaccharides can also be depolymerized by the action of PLs on glycosidic linkage through β-elimination mechanisms (Lombard et al., 2010). Furthermore, AAs represent the most recently defined CAZymes class introduced into the CAZy database (Levasseur et al., 2013). AAs are characterized by their ability to mediate redox reactions with polysaccharide or non-polysaccharide substrates like lignin. Some well-known example of AAs enzymes include laccase, Lytic polysaccharide monooxygenases (LPMO) and copper radical oxidases (Bissaro et al., 2018; Chirania et al., 2022). Overall, these CAZymes across classes and families work together synergistically, each playing a specific role in cleaving complex carbohydrate biopolymers.

1.4 Modular architecture of CAZymes

The architecture of CAZymes typically exhibits a modular organization, comprising a catalytic domain responsible for substrate cleavage, accompanied by one or multiple a non-catalytic so-called accessory domains that generally contribute to substrate binding as well as overall structural stability of the enzyme (Berlemont et al., 2022). The accessory domain may be present at the C and/or N-terminus of the catalytic module, occurring as a single copy or in multiple copies of the same module or in a combination of different modules. These accessory domains are connected to the catalytic domains of CAZymes via peptide linkers (Armenta et al., 2017). Different organisms exhibit a variety of

CAZyme families and modular domain configurations, with or without accessory domains for the CAZyme families described in this thesis (Table 1, **Chapter 2,5,6**). Various accessory domains have been identified, including carbohydrate binding modules (CBMs), fibronectin type three domain (Fn3), and several types of Immunoglobulin-like domains (Han et al., 2013; Lv et al., 2023).

Table 1. Distribution of modular domain organization of main CAZymes family involved in lignocellulose and starch degradation identified in *Aspergillus* (Asp) and *Streptomyces* (Strep). The CAZymes are limited to those families present in both *Aspergillus* and *Streptomyces* genera. Numbers in brackets indicate the total number of different protein sequences from the corresponding GH family.

CAZyme Domain Organizations	β-Glucosidase		Cellulases						α-Amylase		Laccase	
	GH3		GH5		GH6		GH12		GH13		AA1 (laccase)	
	Asp (987)	Strep (1658)	Asp (241)	Strep (923)	Asp (107)	Strep (2303)	Asp (51)	Strep (91)	Asp (292)	Strep (1261)	Asp (708)	Strep (1619)
CAZy Catalytic domain	6 %	7 %	70 %	33 %	59 %	83 %	100 %	49 %	84 %	54 %	100 %	100 %
Single/multiple N-terminal accessory domain + CAZy Catalytic domain	0 %	19 %	0 %	20 %	41 %	12 %	0 %	0 %	0 %	9 %	0 %	0 %
CAZy Catalytic domain + Single/multiple C-terminal accessory domain	94 %	75 %	30 %	47 %	0 %	5 %	0 %	51 %	16 %	37 %	0 %	0 %

1.5 Carbohydrate Binding Modules (CBM)

Amongst accessory domains, CBM domains are the most abundant modules recognized for their ability to target and bind polysaccharides (Gilbert et al., 2013). The interaction between the binding sites of the CBMs and their substrate result in bringing the catalytic domain of CAZymes in a closer proximity toward its substrate. Hence, CBMs are commonly suggested to have a role in enhancing the enzymatic activities by targeting and proximity effects (Bernardes et al., 2019a; Chalak et al., 2019; Reyes-Ortiz et al., 2013). Furthermore, several studies have reported that certain CBMs can also interfere with cellulose crystallinity and thus increase enzymatic degradability (Bernardes et al., 2019a; Horn et al., 2012).

The CBM modules were originally defined as cellulose-binding domains (CBDs) due to their affinity for binding to crystalline cellulose (Van Tilbeurgh et al., 1986). As further related modules were identified with a variety of amino acid sequence and affinity to not only cellulose substrate but to various carbohydrate substrates, the nomenclature

carbohydrate binding module (CBM) was introduced to accommodate this broad range of domains. Family 1 CBM (CBM1) was used for the first discovered fungal CBDs and subsequently also applied to modules with similar amino acid sequences (Tomme et al., 1988). Eventually, CBMs from various origins have been categorized into families with a variety of substrate specificities as well as protein structural features based on their protein sequences (Drula et al., 2022; Shi et al., 2023). As per July 2024, 103 families of CBM from a wide range of organisms have been recorded in CAZy database (http://www.cazy.org/, accessed in July 2024). Based on the CBM domain characteristics, enzyme engineering strategies that provide CAZymes with new or additional CBM moieties have recently received significant research interest (Forsberg and Courtade, 2023; You et al., 2024). These strategies are particularly aimed at generating CAZymes with the enhanced potential to degrade polysaccharides (Y. Li et al., 2022a; Shi et al., 2023).

1.6 Enzyme domain engineering

Although many natural carbohydrate degrading enzymes have been discovered, obtaining enzymes with excellent improved properties for degrading complex biomass substrate remains challenging. To address this, numerous genetic and protein engineering techniques have been developed to discover improved enzymes being more efficient in degrading complex substrate, particularly recalcitrant polysaccharides (Y. Li et al., 2022a; You et al., 2024). One effective approach involves modification of modular CAZymes by fusing accessory domain such as CBMs to the N- or C-terminus of catalytic domain to create new modular CAZymes with enhanced catalytic properties (Hu et al., 2021; Pan et al., 2016).

1.7 Filamentous microorganisms in enzyme production

Enzyme-producing microorganisms are extensively employed in various industries for the production of heterologous engineered enzymes. To meet the demands of industrial-scale production, microorganism strains with high protein production capabilities are often used for the expression of the genes encoding these engineered CAZymes. Among

of them, filamentous microorganisms have been widely favored as hosts (Liu et al., 2023; Ntana et al., 2020; Sharma et al., 2021).

Filamentous microorganisms are characterized by their growth in the form of filaments known as hyphae. These hyphae often exhibit extensive branching, creating a complex network called mycelium (Barka et al., 2015; Riquelme et al., 2018). This extensive mycelium network provides a large surface area for interacting with substrates in their environment, including complex carbohydrates, allowing them to efficiently breakdown complex carbohydrates to be used as a nutrient source for growth and survival through the different enzymes they secrete (Jo et al., 2023).

In terms of enzyme production, filamentous fungi are significant reservoirs of carbohydrate-active enzymes (CAZymes) (Gruben et al., 2017; Gudynaite-Savitch and White, 2016). They also offer an interesting potential for enzymatic hydrolysis and are excellent cell factories for industrial-scale protein production (Fischer and Glass, 2019; Peter J. Punt et al., 2002; Wang et al., 2020). Several species of filamentous fungi such as *Aspergillus niger* and *Trichoderma reesei* are generally regarded as safe (GRAS) and exhibit superior protein secretory capability. Hence, many commercially available enzymes including those degrading starch and (hemi-)cellulose substrates are being produced by these organisms (Arnau et al., 2019; Wei et al., 2021). Over the past few decades, research has focused on improving *Aspergillus* species as cell factories for recombinant proteins production based on their inherent high secretion capacity (Meyer et al., 2011; Ntana et al., 2020). Also Actinomycete bacteria, such as *Streptomyces* genera are renowned for their ability to produces a vast array of metabolites and enzymes including CAZymes (Alam et al., 2022; Kumar et al., 2020; Sharma et al., 2021). Various hydrolytic enzymes such as extracellular amylase have been identified from *Streptomyces* (Kashiwagi et al., 2014; Lakshmi et al., 2020). Amylase are being widely used in the textile, food, pharmaceutical, and brewing industries (Farooq et al., 2021; Kandra, 2003). Moreover, cellulase and xylanase from Actinomycetes are well-known for converting lignocellulosic plant biomass into simple sugars (Talamantes et al., 2016; Ventorino et al., 2016). Notably, several *Streptomyces* produce laccases, a key enzyme in lignin biodegradation (Janusz et al., 2017). Lignin is a complex and rigid polymer in lignocellulose, making lignocellulose recalcitrant to enzymatic hydrolysis (Kirui et al., 2022).

1.8 Outline of the research presented in this thesis

Efficient depolymerization of complex carbohydrate biopolymers into mono- or oligosaccharides via enzymatic degradation is crucial step for producing various bio-based products, such as biofuel and many other biomaterials. However, the complex structure of carbohydrate polymers presents challenges for enzymatic degradation, leading to reduced efficiency in the complete degradation of polysaccharides. Enzyme engineering strategies that provide CAZymes with CBMs derived from filamentous microorganisms have recently received significant research interest due to their effect on enzymatic hydrolysis performance. One of the approaches towards novel and improved enzyme engineering may rely on the striking variety of domain architectures of these CAZymes in filamentous microorganisms. Following the introductory chapter **(Chapter 1),** the subsequent chapters will describe the exploration and engineering strategies related to the modular domain architecture of CAZymes from filamentous microorganisms involved in complex carbohydrate degradation, specifically lignocellulose and starch. The goal of this research is to improve the degradation of recalcitrant polysaccharides, a fundamental step in generating various valuable bio-based products through bioprocessing.

The variation of modular CAZymes especially those containing CBM domains affect enzymatic hydrolysis performance. Therefore, exploring the modular architecture of enzymes can offer insight for designing novel modular CAZymes with enhanced properties, enabling improved degradation of carbohydrate biomass resources. As a basis of this exploration, a comprehensive overview exploring the modular architectures of CAZymes with a focus on the variation and distribution of CBM domains associated with GH families involved in cellulose and starch degradation was made **(Chapter 2)**. This exploration specifically focused on the diversity of modular enzymes sourced from filamentous microorganism, represented by *Aspergillus* and *Streptomyces* genera which are renowned for their excellent capacity to secrete a diverse array of CAZymes.

In general, bacteria exhibit a large variation in CAZymes associated with accessory domains (Berlemont, 2017; Talamantes et al., 2016). In **Chapter 3**, the genome mining of actinomycetes represented by several *Streptomyces* strains was performed to analyze the variation of CAZymes along with their domain architecture. In addition, these *Streptomyces* strains were screened for their capacity to degrade lignocellulosic rice straw. A more detailed enzymatic and proteomic analysis were conducted on the best-

performing strains to identify specific sets of CAZymes potentially involved in the degradation of rice straw. Rice straw is mainly composed of cellulose and is available worldwide as a waste product of rice harvesting process. Unfortunately, burning rice straw waste still remains a common practice, resulting in harmful CO2 emissions (Singh et al., 2021). Therefore, utilizing rice straw biomass for bioethanol production not only contribute to fossil fuel replacement but also reduces waste as well as CO2 emissions from burning.

Furthermore, examples of enzyme domain engineering are explored by expressing newly designed enzymes in *Aspergillus niger*, where the attention is focused on designing modular enzymes with novel domain organization. These new modular enzymes were generated to improve degradation of various starches (**Chapter 4**) and lignocellulosic rice straw biomass (**Chapter 5**).

In **Chapter 4**, a unique domain architecture of an *A. niger* GH13 α-amylase carrying an N-terminal CBM20 was engineered. The substrate binding properties and hydrolytic efficiency of amylase with this new domain architecture toward starches substrate were investigated. The purpose of this research is to improve the degradation of the complex carbohydrate polymer within starch granules. Starch granules not only carry carbohydrate polymers but also small amounts of non-carbohydrate components such as lipids, proteins, and phosphate, which can hinder enzymatic hydrolysis of starch (Dhital et al., 2017). The efficient degradation of starch is essential for various industrial applications, including foods, pharmaceuticals, textiles, papers, and biofuel production (Adewale et al., 2022; Amaraweera et al., 2021).

Moreover, the domain engineering of a modular laccase and its impact on lignocellulosic rice straw degradation was explored in **Chapter 5**. In CAZy database, Laccase is categorized as Auxiliary Activity family 1 (AA1). These enzymes play a role in catalyzing the oxidation of lignin, a complex and rigid polymer present in lignocellulose . Typically, laccase consisted of three copper-oxidase (Cu-oxidase) domains (Hakulinen and Rouvinen, 2015; Mehra et al., 2018). However, in Actinomycetes such as *Streptomyces*, a unique domain arrangement of laccase comprising two Cu-oxidase domains is present, called small laccase (SLAC) (Kaur et al., 2022; Machczynski et al., 2004). This type of two-domain laccase has not been identified in *A. niger* or any other fungal species (The UniProt Consortium, 2023). Considering the unique domain arrangement and characteristics of the two-domain laccase SLAC from *Streptomyces* and its absence in *A.*

niger, research was focused on examining the expression of engineered SLAC in *A. niger* to enhance fungal enzyme cocktails for lignocellulose degradation. In this research, SLAC with and without addition of a CBM1 domain was investigated.

As a third example of CAZymes engineering, in **Chapter 6,** a screening system based on the *A. niger* β-glucosidase gene (*bglA*) was developed to study the modular domain architecture within the GH3 β-glucosidase family from filamentous microorganisms. β-glucosidase plays a vital role in breaking down cellobiose and cello-oligosaccharides into glucose, the end product in complete cellulose degradation (Srivastava et al., 2019).

Overall, the research addressed in this thesis offers the prospect to uncover new insights into protein domain organization and enzyme functionality improvement for a more efficient degradation of complex carbohydrate substrate. The discussion on the findings across all the work, alongside their implications for future prospectives is presented in **Chapter 7**.