Characterization and engineering of carbohydrate-active enzymes for biotechnological applications

Noor Hassan
Akademisk avhandling som med tillstånd av KTH i Stockholm framlägges till offentlig granskning för avläggande av teknisk doktorsexamen tisdagen den 26 maj 2015 kl. 13:00 i sal FB55, Albanova, Roslagstullsbacken 21, Stockholm.
ABSTRACT

Extremozymes are enzymes produced by microorganisms that live in extreme habitats. Due to their higher stability, extremozymes is attracting interest as biocatalysts in various industrial processes. In this context, carbohydrate-active extremozymes can be used in various processes relevant to the paper, food and feed industry.

In this thesis, the crystal structure, biochemical characterization and the capacity to synthesize prebiotic galacto-oligosaccharides (GOS) were investigated for a β-glucosidase (HoBGLA) from the halothermophilic bacterium Halothermothrix orenii. The wild-type enzyme displays favorable characteristics for lactose hydrolysis and produces a range of prebiotic GOS, of which β-D-Galp-(1→6)-D-Lac and β-D-Galp-(1→3)-D-Lac are the major products (Paper I).

To further improve GOS synthesis by HoBGLA, rational enzyme engineering was performed (Paper II). Six enzyme variants were generated by replacing strategically positioned active-site residues. Two HoBGLA variants were identified as potentially interesting, F417S and F417Y. The former appears to synthesize one particular GOS product in higher yield, whereas the latter produces a higher yield of total GOS.

In Paper III, the high-resolution crystal structure and biochemical characterization of a hemicellulase (HoAraf43) from H. orenii is presented. HoAraf43 folds as a five-bladed β-propeller and displays α-L-arabinofuranosidase activity. The melting temperature of HoAraf43 increases significantly in the presence of high salt and divalent cations, which is consistent with H. orenii being a halophile.

Furthermore, the crystal structures of a thermostable tetrameric pyranose 2-oxidase from Phanerochaete chrysosporium (PcP2O) were determined to investigate the structural determinants of thermostability (Paper IV). PcP2O has an increased number of salt links between subunits, which may provide a mechanism for increased stability. The structures also imply that the N-terminal region acts as an intramolecular chaperone during homotetramer assembly.

Keywords: β-glucosidase, pyranose oxidase, carbohydrates, lactose conversion, galacto-oligosaccharides, thermostability, propeptide
SAMMANFATTNING


I denna avhandling presenteras kristallstrukturbestämning, biokemisk karakterisering och förmågan att syntetisera prebiotiska galaktoligosackarider (GOS) hos ett β-glukosidas (HoBGLA) från den halotermofila bakterien Halothermothrix orenii. Vildtypsenzymet uppvisar gynnsamma egenskaper med avseende på hydrolys av laktos och produktion av prebiotiska GOS, där de främsta produktarna utgörs av β-D-Galp-(1→6)-D-Lac och β-D-Galp-(1→3)-D-Lac (Paper I).

I syfte att ytterligare förbättra den GOS-producerande förmågan hos HoBGLA tillämpades rationell enzymdesign (Paper II). Sex enzymvarianter genererades genom utbyte av strategiskt placerade aminosyror i det aktiva sätet. Två enzymvarianter framstod som potentiellt intressanta, F417S och F417Y. Den förstnämnda verkar producera en särskild GOS-produkt med högre utbyte, medan den sistnämnda producerar ett högre utbyte av total mängd GOS.


Vidare bestämdes kristallstrukturer för ett termostabilt tetramert pyranos 2-oxidas från Phanerochaete chrysosporium (PcP2O) i syfte att undersöka de strukturella faktorer som bidrar till termostabiliteten (Paper IV). PcP2O uppvisar ett ökat antal saltbindningar mellan subenheter vilket kan utgöra en mekanism för att öka stabiliteten. Strukturerarna antyder även att den N-terminala regionen underlättar hopsättningen av homotetrameren.

Nyckelord: β-glukosidas, pyranosoxid, kolhydrater, laktosomvandling, galakto-oligosackarider, termostabilitet
List of publications

This thesis is based on the following publications, which are referred to by their roman numerals:

**Paper I**
Noor Hassan*, Thu-Ha Nguyen, Montira Intanon, Lokesh D. Kori, Bharat K. C. Patel, Dietmar Haltrich, Christina Divne, Tien Chye Tan

**Paper II**
Noor Hassan*, Barbara Geiger*, Rosaria Gandini, Bharat K. C. Patel, Roman Kittl, Dietmar Haltrich, Thu-Ha Nguyen, Christina Divne, Tien Chye Tan

**Paper III**
Noor Hassan*, Lokesh D. Kori, Rosaria Gandini, Bharat K. C. Patel, Christina Divne, Tien Chye Tan

**Paper IV**
Noor Hassan*, Tien-Chye Tan, Oliver Spadiut, Ines Pisanelli, Laura Fusco, Dietmar Haltrich, Clemens K. Peterbauer, Christina Divne
Crystal structures of *Phanerochaete chrysosporium* pyranose 2-oxidase suggest that the N-terminus acts as a propeptide that assists in homotetrameric assembly. FEBS Open Bio, 2013; 3:496–504.

Paper I, II and IV are published with Open Access. Copyright © The Authors.
Contents

1 Introduction 1
  1.1 Plant biomass 1
    1.1.1 Cellulose 2
    1.1.2 Hemicellulose 3
      1.1.2.1 Galactooligosaccharides 3
  1.2 Enzymes in lignocellulose degradation 5
    1.2.1 Glycoside hydrolases 5
      1.2.1.1 β-glucosidases (BGLs) 8
      1.2.1.2 β-galactosidases (BGALs) 10
      1.2.1.3 α-L-arabinofuranosidases (ARAFs) 11
    1.2.2 Oxidative enzymes 12
      1.2.2.1 Polysaccharide monooxygenases 12
      1.2.2.2 GMC oxidoreductases 12
  1.3 Thermostability and halophilicity of proteins 16

2 Present investigation 19
  2.1 Objectives 19
  2.2 Methodology overview 21
    2.2.1 Ligation-independent cloning 21
    2.2.2 Site-directed mutagenesis 22
    2.2.3 Protein expression and purification 23
    2.2.4 β-D-glucosidase/galactosidase assay 24
    2.2.5 p-nitrophenyl sugar assay 25
    2.2.6 Analysis of GOS synthesis and products 25
    2.2.7 Protein crystallization 26
    2.2.8 Structure determination and refinement 27
  2.3 Summary of Paper I 29
  2.4 Summary of Paper II 31
  2.5 Summary of Paper III 34
  2.6 Summary of Paper IV 38

3 Concluding remarks 42

Abbreviations 43
Acknowledgements 44
References 45
1 INTRODUCTION

Lignocellulose is a major renewable natural biomass resource that consists of sugar-based polymers, i.e., cellulose and hemicellulose, embedded in a complex network of aromatic alcohols (monolignols). Extensive efforts are made to develop strategies for efficient biomass utilization to harness energy through the conversion of biomass to fermentable sugars by environmentally sustainable enzymatic processes and the production of valuable products such as biofuel (Himmel et al., 2007). The major structural element in plant cell walls, lignocellulose, constitutes about 70% of the plant biomass (Pauly et al., 2008). Carbohydrates carry out a wide variety of biological functions including adhesion, motility, immune response, and pathogen interaction (Ohtsubo & Marth, 2006). Monosaccharides are the simplest sugar building blocks. Disaccharides are formed when two monosaccharides are joined together through a dehydration reaction resulting in the formation of a glycosidic bond, while oligo- and polysaccharides result in molecules structurally and functionally far more diverse than peptides and nucleic acids.

1.1 Plant biomass

Plant polysaccharides are found in a variety of sources such as agricultural residues, woods, and different types of waste products. The major component of this lignocellulosic biomass is cellulose (Table 1).

<table>
<thead>
<tr>
<th></th>
<th>Cellulose</th>
<th>Hemicellulose</th>
<th>Lignin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn fiber</td>
<td>15</td>
<td>35</td>
<td>8</td>
</tr>
<tr>
<td>Corn stover</td>
<td>40</td>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td>Rice straw</td>
<td>35</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>Wheat straw</td>
<td>30</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>Sugarcane bagasse</td>
<td>40</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>Switch grass</td>
<td>25</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>Coastal Bermuda grass</td>
<td>25</td>
<td>35</td>
<td>6</td>
</tr>
</tbody>
</table>

Table adapted from Saha, 2003.
1.1.1 Cellulose

Cellulose is the most important polysaccharide in plants and an inexhaustible polymeric raw material offering an environment friendly and biodegradable natural resource (Klemm et al., 2002; Klemm, 2005; Kaplan 1998). Cellulose constitutes nearly one half of both soft wood and hard wood (Fengel & Wegener, 1984). It is a homopolymer composed of repeating D-glucose units (Fig. 1) covalently linked via $\beta$-1,4-glycosidic linkages. In plants, cellulose is synthesized by cellulose synthases followed by assembly of cellulose chains into microfibrils (Somerville et al., 2006, review). In the plant cell wall, parallel cellulose chains interact through extensive inter- and intra-molecular hydrogen bonds as well as van der Waals forces such that hydrogen-bond ordered (crystalline) regions are interspersed by less ordered (amorphous) regions along the microfibrils (Nieduszynski & Preston, 1970; Gardner & Blackwell, 1974). Cellulose microfibrils are typically about 3-nm thick, comprising 36 crystalline, parallel $\beta$-1,4-glucan chains, each probably including 8,000-15,000 glucose molecules. Thus, the cellulose microfibril is one of the largest biomolecules known in nature. Microfibrils associate into fibrils, and ultimately into cellulose fibers (Somerville et al., 2006, review).

![Figure 1. Schematic drawing of the molecular architecture of cellulose fibers formed by assembled microfibrils. Microfibrils are formed by 36 parallel cellulose chains.](image-url)
1.1.2 Hemicellulose

Hemicellulose is the second most common polysaccharide group, typically accounting for 15–35% of the plant biomass. In hardwood, about 20 to 30% of the biomass is hemicellulose, and in plants of herbaceous origin up to 50% (Ebringerova et al., 2005). Hemicellulose is a heterogeneous polymer of xylose, arabinose, mannose, glucose, and galactose and other sugar-based compounds such as glucuronic acids (McMillan et al., 1993). The major hemicellulose in soft wood are glucomannan and galactoglucomannan, which are composed of linear chains of β(1→4)-linked D-mannopyranose and D-glucopyranose units decorated with α(1→6)-linked D-galactopyranose residues and xylan with backbone of β(1→4)-linked xylopyranose units with via α(1→2)-linked 4-methylated D-glucuronic acid, and α(1→3)-linked L-arabinose units. Other hemicelluloses in soft wood include arabinogalactan, xyloglucan, and pectin components including galactans, arabinans and rhamnogalacturonan. The main hemicellulose in hardwood is xylan (Fengel & Wegener, 1989) with a backbone of β(1→4)-linked xylopyranose units with inserted acetyl groups at regular intervals (every 10th residue).

1.1.2.1 Galactooligosaccharides

The term galacto-oligosaccharide (GOS) is used for any β-linked oligosaccharide composed of galactose units with glucose or galactose at the non-reducing end, and with a degree of polymerization (DP) between 2–10 units per molecule. GOSs are generated from incomplete hydrolysis of galactose polymers (galactans) present in hemicelluloses and occur naturally in honey and various fruits such as banana, as well as vegetables including onions, garlic, soybeans and chicory (Angus et al., 2005). GOSs are also found in mammalian milk, but only in trace amounts in bovine and human milk where it is exclusively composed of β-linked galactose and glucose (Kunz et al., 2000; Tao et al., 2008). These GOSs are non-digestible functional food additives with beneficial physiological effects. GOSs are prebiotic compounds, and have been termed the "Bifidus growth factor" (Bielecka et al., 2002), which selectively stimulate the growth of lactobacilli and bifidobacteria living commensally in the human gut. The bifidobacteria are well known for their beneficial effects on the production of B vitamins, including B1, B2, B6, and B12, along with biotin, nicotinic acid and folic acid (Tomomatsu, 1994; Hoover, 1993).
However, GOSs not only promote health, but also improve the quality of food by modifying taste and physicochemical properties (Sako et al., 1999). Additionally, GOSs are sweet in taste with a relative sweetness of 0.3-0.6 compared with sucrose, and can be used as a replacement for sucrose to avoid dental caries and biofilm formation by *Streptococcus mutans* (Bowen & Lawrence, 2005; Cvitkovitch et al., 2003; Moynihan, 2005). Another potential advantage of GOSs relates to infectious disease. Pathogenic bacteria and pathogenic toxins recognize specific cell-surface oligosaccharides to establish infection (Kulkarni et al., 2010), and it has been shown that GOSs are able to neutralize these pathogens by blocking their adhesion to cell-surface oligosaccharides (Shoaf et al., 2006; Searle et al., 2010).

About 150-200 million tons of lactose is generated annually as a waste product from liquid whey (Smithers, 2008). Milk is the only significant natural source of lactose, in which lactose is the dominant carbohydrate. The estimated annual global market consumption of GOSs is 50,000-80,000 tons (Crittenden & Playne, 1996), of which about 40% is synthesized enzymatically from lactose through transgalactosylation by β-galactosidases. A number of scientific reports are available that describe GOS production via trans-galactosylation using β-galactosidase (for examples see: Playne et al., 1996; Crittenden et al., 1996; Sako et al., 1999; Ladero et al., 2001; 2002; Jurado et al., 2002; Sener et al., 2006; Haider & Husain, 2008). The products from lactose hydrolysis (glucose and galactose) are converted into various types of GOSs by enzymatic transglycosylation reactions where lactose is used as an acceptor, and the lactose hydrolysis products (glucose and galactose) can serve as donors. More than 30 different GOSs of various lengths are typically generated through enzymatic transgalactosylation.

The galactose units may be linked through β(1→3), β(1→4), or β(1→6) linkages, with the β(1→4) linkage being the predominant bond, while β(1→4) linkages are mainly present between galactose and glucose units (Oku, 1996; Tzortzis & Vulevic, 2009; Macfarlane et al., 2008). When lactose is used for GOS production, glucose is present at the reducing end of the synthesized oligosaccharide. During enzymatic synthesis of GOS the precise saccharide composition, regiochemistry and degree of polymerization depends on the enzyme used (Gosling et al. 2010).

Besides the value of GOS production, a large number of people in the
world are intolerant to lactose, which creates a high demand for lactose-free milk and dairy products. Enzymatic lactose hydrolysis by β-galactosidases is the principal means by which to produce lactose-free dairy products.

1.2 Enzymes in biomass degradation

1.2.1 Glycoside hydrolases
Glycoside hydrolases (GHs; EC 3.2.1.x) comprise a diverse group of enzymes that hydrolyze glycosidic bonds between sugar units, or between a sugar (glycone) and a non-sugar moiety (aglycone). Many GHs display more than one substrate specificity. To achieve efficient decomposition of lignocellulose, cellulolytic microorganisms generate a variety of GHs, which are mainly cellulases and hemicellulases. This is reflected by the large number of existing protein sequences for cellulose and hemicellulose-active enzymes in the CAZy database, where the glycoside hydrolases have been classified into 133 different GH families, corresponding to 14 unique clans (http://www.cazy.org; Lombard et al., 2014). Each GH family contains enzymes with similar tertiary structures and mechanisms, but typically different substrate specificities.

Cellulose is hydrolyzed by the synergistic action of different cellulases comprising three principal ways to approach the complex substrate (Payne et al., 2015, review): (i) endo-1,4-β-glucanases (EC 3.2.1.4), which depolymerize cellulose by random hydrolysis of internal β-(1,4)-glycosidic bonds to produce shorter chains and cellobioheteromers; (ii) exo-1,4-β-glucanases (CBH, cellobiohydrolase; EC 3.2.1.91) that hydrolyze cellulose polymers from either of the free cellulose chain ends (reducing or non-reducing end) to generate cellobiose (i.e., the β-1,4-linked disaccharide of glucose); and (iii) β-glucosidases (BGL; EC 3.2.1.21) that hydrolyze cellobiose to glucose. The β-glucosidases also play an important role to relieve the effect of product inhibition by cellobiose on cellulases (Enari et al., 1987; Yazaki et al., 1997).

Hemicellulose is composed of different types of carbohydrates including β-glucan, xylan, xyloglucan, arabinoxylan, mannann, galactomannan, arabinan, galactan, and polygalacturonan; and therefore
complete degradation requires a complex battery of synergistic enzyme activities, \textit{e.g.}, \( \beta \)-glucanase, xylanase, xyloglucanase, mannanase, arabinase, galactanase, polygalacturonase, glucuronidase, acetyl xylan esterase, as well as other enzymes (Sjöström, 1993). Hemicellulases not only help to expose cellulose chains to make them more accessible for cellulases, but also convert hemicellulose into valuable saccharides.

Glycosidic bond cleavage by GHs leads to either of two possible stereochemical outcomes of the hydrolytic reaction, \textit{i.e.}, inversion or retention of configuration at the anomic carbon atom, giving the \( \alpha \)- or \( \beta \)-anomer, respectively (Koshland, 1953). The \textit{inverting mechanism} (Fig. 2a) is a bimolecular reaction that involves a single displacement at the asymmetric carbon and the direct displacement of the leaving group by a nucleophilic water molecule. Here, the nucleophile and leaving group are positioned on opposite sides of the bond to be cleaved such that the attack proceeds from the “back”. The catalytic amino acids are both carboxylic acids, which are strategically positioned approximately 10 Å apart. The nucleophilic water molecule is activated by the carboxylic acid functioning as a base, supported by the other carboxylic acid acting as a general acid that protonates the glycosidic oxygen.

In the \textit{retaining mechanism} (Fig. 2b), which is a double-displacement reaction (Koshland, 1953; Sinnott, 1991), the two carboxylic-acid residues are closer, typically only 5.5 Å apart and involves two displacements at the asymmetric carbon. In the first step, the nucleophile attacks the anomic carbon assisted by the other carboxylate residue that acts as general acid to protonate the glycosidic bond oxygen. The displacement results in inversion of configuration and the formation of a covalent glycosyl-enzyme intermediate. In the second step a water molecule attacks at the anomic carbon from the opposite side, leading to another inversion at the asymmetric center and net retention of configuration. In this step, the carboxylic acid previously acting as an acid now works as a general base abstracting a proton from the water.
Carbohydrate-binding modules (CBMs) are non-catalytic sugar-binding protein domains associated with carbohydrate-active enzymes. The principal role of CBMs is to bind to various polysaccharide structures in order to increase the catalytic efficiency of the enzyme on soluble and insoluble substrates by keeping the catalytic module in close proximity of the substrate (Tomme et al., 1995). Based on their amino-acid sequence similarities (Fujimto, 2013, review), the CBMs are currently classified into 71 CBM families in the CAZy database. The CBMs are further
grouped into seven fold families based on their three-dimensional (3D) structure, and into three CBM types based on their sugar recognition modes (Boraston et al., 2004).

1.2.1.1  β-glucosidases (BGLs)

The β-glucosidases (BGLs; EC 3.2.1.21) are found in most living organisms and comprise a diverse group of GHs that hydrolyzes glycosidic linkages between sugar units, or between a sugar (glycone) and a non-sugar moiety (aglycone). BGLs have been classified by the use of various criteria, as there seems to be no single, well-defined strategy for classification. The two classification strategies found in the literature are based on substrate specificity and sequence similarity. Based on substrate specificity, BGLs are classified as (i) aryl β-glucosidases (enzymes that act on aryl glycosides), (ii) true cellobiases which act on cellobiose, and (iii) broad-substrate specificity β-glucosidases (acting on broad spectrum of substrates). Most BGLs belong to the third category but fall into different GH families in the CAZy database.

BGLs are currently placed in CAZy families GH1 and GH3 (Lombard et al., 2014). Family GH1 comprise enzymes with varying substrate specificities, and from all domains of life (Bhatia et al., 2002, review). The typical GH1 enzyme displays broad substrate specificity, and hydrolyzes the β-glycosidic linkages between glucose and another moiety (glucose or other) via a double-displacement mechanism (Withers et al., 1986; Kempton & Withers, 1992; Withers, 2001) using two glutamic acid residues where one functions as a catalytic acid/base and the other as a nucleophile. An exception is myrosinase, which lacks the acid/base residue and instead uses ascorbate as a co-factor that replaces the catalytic base (Burmeister et al., 2000).

GH1 BGLs belong to clan GH-A with an (β/α)8-barrel fold (i.e., TIM barrel fold). The active site is a 15 to 20 Å wide cleft where the two glutamate residues are placed on opposite sides of the glycosidic bond to be cleaved. The acid/base catalyst and nucleophile are positioned at the ends of β-strands 4 and 7, respectively, and the enzymes are therefore sometimes referred to as clan 4/7 enzymes (White et al., 1994; McCarter et al., 1994). The substrate-binding cleft can be divided into subsites that bind individual sugar units. The non-reducing end sugar unit binds in
subsite –1, also referred to as the glycone subsite, and the additional sugar units in the substrate occupy subsites +1, +2, +3 etc. The glycosidic bond to be cleaved is positioned between subsites –1 and +1 (Davies et al., 1997), where the +1 subsite is referred to as the aglycone subsite.

As for GH1 BGLs, the BGLs belonging to family GH3 are retaining enzymes, but use an aspartate residue as nucleophile instead of a glutamate (Dan et al., 2000). Besides BGLs, the GH3 family includes a range of other enzyme activities, for instance, α-L-arabinofuranosidases, β-D-xylopyranosidases and N-acetyl-β-D-glucosaminidases (Harvey et al., 2000). These enzymes are produced by a variety of organisms including bacteria, fungi and plants where they are involved in diverse biological processes such as cellulose degradation, plant and bacterial cell-wall remodeling and metabolism and pathogen defense mechanisms (Lee et al., 2002) where they hydrolyze the terminal glycosyl residue form the non-reducing end from a broad range of glucooligosaccharides. Crystal structures are available for a number of GH3 enzymes, and beyond the catalytic core domain, they typically contain additional domains of diverse architecture (Varghese et al., 1999; Balcewich et al., 2009; Pozzo et al., 2010; Yoshida et al., 2010; Suzuki et al., 2013).

Various industrial applications exist for BGLs; for instance, BGL has proved beneficial as feed supplement for single-stomached animals such as pigs and chickens (Leclerc et al., 1987; Zhang et al., 1996) to promote cellulose degradation and enhanced nutrient utilization (Coenen et al., 1995). Flavonoids and isoflavonoid glucosides occur naturally in tea, red wine, soya beans, fruit and vegetables and when hydrolyzed by BGLs, the released aglycone compounds were shown to have potent therapeutic and health-promoting effects making them useful as added-value compounds (Matsuda et al., 1994). In addition, the bitterness in citrus-based fruit juices can be removed by BGL-catalyzed hydrolysis of naringin to prunin (Roitner et al., 1984). Other applications of BGLs include cellobiose hydrolysis during cellulose degradation by relieving product inhibition (i.e., cellobiose) of endo- and exoglucanases (Mandels, 1982). BGLs are also used for the synthesis of surfactants such as alkyl glycosides (Turner et al., 2007; Gräber et al., 2010; Lundemo et al., 2013; Ojha et al., 2013; Lundemo et al., 2014). Alkyl glycosides have various attractive properties such as good biodegradability, antimicrobial activity and low toxicity, and can be used in cosmetics, food, and pharmaceuticals (Matsumura et al., 1990; von Rybinski & Hill, 1998; Le Maire et al., 2000).
1.2.1.2 β-galactosidases (BGALs)

β-galactosidases (BGALs) are produced by a wide range of organisms including, animal, plants, bacteria and fungi (Rubio-Texeira et al., 2006). Based on the functional similarities, BGALs are classified into four clan-A GH families (CAZy database; Lombard et al., 2014). Except for the BGALs from K. lactis and K. marxianus, which belong to GH2 together with prokaryotic BGALs (Juers et al., 2000), the BGALs from eukaryotic organisms belong to family GH35. The BGALs typically fold into five individual domains where one domain is responsible for the catalytic function.

The GH2 enzymes are widespread and include members with varying specificity, including β-galactosidases, β-glucuronidases, β-mannosidases, exo-β-glucosaminidases, endo-β-mannosidases (Fowler et al., 1978). GH2 BGALs catalyze the substrate hydrolysis via a retaining reaction following Koshland’s double-displacement mechanism (Koshland, 1953). As for most other retaining GHs, two catalytic glutamate residues provide the function as nucleophile and acid/base catalyst (Gebler et al., 1992). The first crystal structure of a GH2 BGAL to be determined was the β-galactosidase LacZ from E. coli (Jacobson et al., 1994), which unlike the GH1 enzymes forms a tetramer with each monomer consisting of multiple domains. The third domain is the catalytic domain that folds as a classical clan GH-A TIM barrel. Similar to GH1 enzymes, the acid/base catalyst and nucleophile are located at the ends of β-strands 4 and 7, respectively. The enzyme requires Mg$^{2+}$ and Na$^+$ as cofactors for optimum activity (Wallenfels et al., 1972). The enzyme from Lactobacillus delbrueckii is known to be dimeric (Adams et al., 1994).

A majority of the world’s population is intolerant to lactose owing to low levels of BGAL in the intestine (Harju et al., 2012), and as discussed above (section 1.1.2.1), BGALs are used for industrial hydrolysis of milk lactose to make lactose-free dairy products (Oliveira et al., 2011), but also for GOS production, which exploits the inherent transglycosylation activity of the enzymes (Gänzle et al., 2008; Gänzle, 2012). However, the usefulness of BGALs is compromised by product inhibition by glucose and/or galactose, which leads to incomplete hydrolysis (Hatzinikolaou et al., 2005). Several factors need to be considered during lactose conversion such as pH, temperature, time and the cost of enzyme.
production. The BGAL from *K. lactis* is a major commercial enzyme source for hydrolysis of lactose in milk, but suffers from poor stability at higher temperatures (Ganeva *et al*., 2001), which means that lactose conversion has to be performed at 35-45°C, which unfortunately is a temperature range where microbial contamination and growth is a problem. Thus, thermostable BGALs are desirable to allow high yields at higher temperatures where the risk of microbial contamination is reduced (Zeikus *et al*., 1998).

1.2.1.3 **α-L-arabinofuranosidases (ARAFs)**

Arabinoxylan is an important plant hemicellulose composed of a linear backbone of β-1,4-linked D-xylopyranosyl units containing O-2 and/or O-3 substituted L-arabinofuranosyl units (Lagaert *et al*., 2014, review). The polymer is degraded enzymatically by endoxylanases, β-xylosidases and arabinofuranosidases. Arabinofuranosidases are found in five GH families, i.e., GH3, GH43, GH51, GH54 and GH62. The GH3, GH51 and GH54 enzymes hydrolyze the glycosidic bond with retention of anomeric configuration, while the GH43 enzymes proceed by inversion of configuration. The reaction mechanism for GH62 has not been elucidated. Some GH43 glycosidases use a single substrate (typically arabinofuranosides or xylohexosides), whereas other show bifunctional activity, for instance, combined arabinofuranosidase/xylosidase activity. Besides α-L-arabinofuranosidases and β-xylosidases, family GH43 also contains arabinanases, galactan 1,3-β-galactosidases, xylanases, α-1,2-L-arabinofuranosidases, exo-α-1,5-L-arabinofuranosidases, exo-α-1,5-L-arabinanase, and β-1,3-xylosidases.

The fold signature of GH43 enzymes is a catalytic domain with a five-bladed β-propeller fold where each of the blades contains four β-strands (Nurizzo *et al*., 2002). The catalytic amino acids involved include a general base (aspartate), a general acid (glutamate) and a third carboxylic acid that serves to modulate the *pKₐ* of the general acid and helps to orient the substrate (Nurizzo *et al*., 2002). Despite different activity and substrate specificity, the three catalytic residues are strictly conserved within the family (Nurizzo *et al*., 2002; Brüx *et al*., 2006).

α-L-arabinofuranosidases, or ARAFs (EC 3.2.1.55) hydrolyze the non-reducing terminal unit of α-L-arabinofuranoside linkages in arabino-
furanose-containing polysaccharides, and have potential applications in agroindustrial processes, such as the processing of fruits, vegetables, and cereals, as well as the conversion of hemicellulose to fuels and chemicals (Van Laere et al., 2000; Aristidou & Penttilä, 2000).

1.2.2 Oxidative enzymes

The microorganisms capable of solubilizing lignin are mainly white- and brown-rot filamentous fungi that besides hydrolytic enzymes also produce an elaborate machinery of extracellular oxidative enzymes (Payne et al., 2015, review). The more well-characterized oxidative enzymes include lignin peroxidase (EC 1.11.1.14), laccase (EC 1.10.3.2, and manganese peroxidase (EC 1.11.1.13) that are mainly involved in the depolymerization of lignin (Martinez et al., 2005; review Martinez et al., 2009; Martinez et al., 2011). The importance of oxidative enzymes for the degradation of the cellulose and hemicellulose components in lignocellulose has received considerable attention lately, most importantly a fungal enzyme system involving small lytic polysaccharide monooxygenases (LPMOs) and cellobiose dehydrogenases (CDHs), but also various metabolic pyranose-oxidizing enzymes such as pyranose 2-oxidases (P2Os).

1.2.2.1 Polysaccharide monooxygenases

Historically, the hydrolytic enzymes were given all credit for the efficient degradation of crystalline cellulose and hemicellulose. This changed dramatically by the recent discovery and characterization of fungal copper-dependent LPMOs that catalyze redox-mediated cleavage of glycosidic bonds in crystalline polysaccharides such as cellulose and chitin (Vaaje-Kolstad et al., 2010) and the coupling of LPMOs to CDHs (Horn et al., 2012; Payne et al., 2015, reviews).

1.2.2.2 GMC oxidoreductases

The oxidoreductases glucose dehydrogenase from Drosophila melanogaster, choline dehydrogenase from Escherichia coli, glucose oxidase from Aspergillus niger, and methanol oxidase from yeast
Hansenula polymorpha were historically classified as related but diverse flavoproteins that shared certain sequence motifs and a Rossmann-like dinucleotide-binding $\beta\alpha\beta$ fold. Based on the identity of the original members, the family was named the Glucose-Methanol-Choline (GMC) family of oxidoreductases (Cavener et al., 1992). The GMC sugar oxidoreductases constitutes a subgroup of the GMC enzyme family that includes a large number of enzymes from various organism, but are particularly common in fungi (Van Hellemend et al., 2006; review Martinez et al., 2009a; Martinez et al., 2009b).

GMC non-sugar oxidoreductases include aryl-alcohol oxidases (EC 1.1.3.7), methanol oxidases, and glyoxal oxidases (EC 1.1.3.x); while GMC sugar oxidoreductases include glucose oxidase (EC 1.1.3.4), pyranose-2 oxidase (EC 1.1.3.10), and cellobiose dehydrogenase (CDH; EC 1.1.99.18). These GMC sugar oxidoreductases have been further classified as CAZy AA (auxiliary activities) enzymes, which currently include 10 families, AA1–AA10. The GMC sugar oxidoreductases are found in four subfamilies of family AA3, namely: AA3_1 (mainly cellobiose dehydrogenases); AA3_2 (glucose 1-oxidase and aryl alcohol oxidase); AA3_3 (alcohol oxidase); and AA3_4 (pyranose 2-oxidase).

CDHs are large flavocytochromes (Eriksson et al., 1974; Henriksson et al., 2000; Hallberg et al., 2000; Hallberg et al., 2002; Zamocky et al., 2006) that generate reducing equivalents by the flavin-dependent oxidation of cellobiose to provide electrons for the LPMO reaction to boost cellulose-degradation efficiency and rendering the substrate more accessible to hydrolytic enzymes (Harreither et al., 2010; Langston et al., 2011; Kittl et al., 2012). The flavin domain of CDH belongs to family AA3_1 and the cytochrome domain to AA8.

The subfamily AA3_4 (pyranose 2-oxidase; Janssen & Ruelius 1968) is a flavin adenine dinucleotide (FAD)-dependent GMC sugar oxidoreductase (Hallberg et al., 2004; Zamocky et al., 2004), and has been well characterized to date. The enzyme is co-expressed with a number of redox enzymes when the fungus is grown on lignin as carbon source (Daniel et al., 1994; De Koker et al., 2004). The enzyme catalyzes the oxidation of D-glucose, as well as most monosaccharides released during lignocellulose degradation, at the C2 position (Daniel et al., 1994; Freimund et al., 1998; Sucharitakul et al., 2010; Tan et al., 2011) to the corresponding ketoaldoses and $\text{H}_2\text{O}_2$. Occasionally, oxidation at C3 has
been observed (Freimund et al., 1998). The catalytic reaction of P2O can be subdivided into two half-reactions, a reductive (reaction 1) and an oxidative half-reaction (reaction 2).

\[
\begin{align*}
\text{FAD}_{\text{ox}} + \text{aldopyranose} & \rightarrow \text{FAD}_{\text{red}} + 2\text{-keto-aldopyranose} \quad \text{(reaction 1)} \\
\text{FAD}_{\text{red}} + \text{O}_2 & \rightarrow \text{FAD}_{\text{ox}} + \text{H}_2\text{O}_2 \quad \text{(reaction 2)}
\end{align*}
\]

The overall reaction is of ping-pong type (Prongjit et al., 2009), meaning that the first product (2-keto-sugar) is released prior to the reaction with the second substrate (O\_2). In the reductive half-reaction (reaction 1), two electrons are transferred as hydride from the C2 carbon of d-glucose to the oxidized flavin adenine dinucleotide (FAD\_ox) group thereby forming a 2-keto-sugar and FAD\_red. In the oxidative half-reaction (reaction 2), the flavin is re-oxidized by molecular oxygen via a covalent C4a-hydroperoxyflavin intermediate to eliminate hydrogen peroxide from the flavin intermediate (Sucharitakul et al., 2008). Whereas the C4a-hydroperoxyflavin intermediate is common in flavin monooxygenases, it is rare in flavoprotein oxidases. It is not known whether the intermediate is unique to P2O, or if it occurs also in other flavoprotein oxidases.

P2O is an intracellular homotetrameric enzyme (Fig. 3) that buries a large cavity at the center of the tetramer (Hallberg et al., 2004). Each monomer in the tetramer binds one FAD covalently via a histidine residue (Halada et al., 2003). The crystal structures of P2O from two sources have been determined, from Trametes multicolor (Hallberg et al., 2004) and Peniophora sp. (Bannwarth et al., 2004). Each monomer consists of a “body”, “arm”, and a “head” domain. The body contains the FAD-binding βαβ motif and a substrate-binding domain with a six-stranded β-sheet that forms the “bottom” of the active site. The two monomers of the two dimers (A/B and C/D) forming the tetramer are connected by “arms” that act as oligomerization loops. The function of the head domain is unknown, however, due to the extremely flat structure of the β-sheet it may serve as an interaction surface for a substrate surface, membrane, or another protein.

The active site in P2O is gated by a dynamic substrate-recognition loop that is highly sensitive to different substrates as well as mutations in
the active site (Kujawa et al., 2006; Spadiut et al., 2010; Tan et al., 2011; Tan et al., 2014). When sugar substrate is bound, the loop is either in a non-productive open conformation (Kujawa et al., 2006) or in a productive semi-open conformation relevant for oxidation at C2 (Tan et al., 2011). The loop shows pronounced conformational plasticity allowing the enzyme to accommodate not only various types of sugar substrates, but also different binding modes for the same sugar (Kujawa et al., 2006; Tan et al., 2011; Tan et al., 2014).

**Figure 3.** Ribbon representation of the homotetrameric structure of P2O from *Trametes multicolor* (Hallberg et al., 2004). With courtesy of Christina Divne.

Fungal P2Os have a number of applied uses. They are widely used for the oxidation of D-glucose to corresponding 2-keto-D-glucose, which is an important intermediate for the biosynthesis of the antibiotic cortalcerone (Baute et al., 1987; Giffhorn et al., 2000; Giffhorn, 2000), and for D-fructose production through the Cetus process (Neidleman et al., 1981). In the latter process, 2-keto-D-glucose is hydrogenated chemically to D-fructose. P2O has also been shown to function in biofuel cells when wired with osmium as redox mediator on graphite electrodes (Tasca et al., 2007; Nazaruk & Bilewicsz, 2007). Moreover, and relevant to this thesis, P2Os are potentially useful for the production of value-added compounds from lactose hydrolysis (Spadiut et al., 2010; Wongnate & Chaiyen, 2013; Tan et al., 2014).
1.3 Thermostability and halophilicity of proteins

A wealth of reports on crucial and essential determinants of thermostability can be found in the literature. However, in most cases these are “case reports” that draw conclusions from a very limited dataset of maybe only one or two proteins. It is common for these reports to pinpoint and highlight the contribution of only one type of interaction as a dominant determinant of thermostability. Clearly, there are a large number of structural determinants that together govern thermal stability.

A large number of seminal papers have been published by Brian Matthews and co-workers and others that describe systematic site-directed mutagenesis approaches to evaluate the importance of various determinants for protein stability, such as disulfide bonds (Wetzel et al., 1988; Matsumura et al., 1989a; Matsumura et al., 1989b; Matsumura et al., 1989c), hydrophobic core residues (Matsumura et al., 1988; Matsumura et al., 1989d), ionic links (Perutz et al., 1975; Anderson et al., 1990; Sun et al., 1991), helix-dipole stabilization (Nicholson et al., 1988; Bell et al., 1992; Hennig et al., 1995), and others. Thus, there appears to be no single, common determinant of thermostability (Matthews et al., 1974), however various analyses indicate that stability to thermal denaturation is mainly governed by a large number of smaller changes (Argos et al., 1979; Wallon et al., 1997).

Hydrophobicity is the major driving force for protein folding and stability. For soluble proteins, most hydrophobic residues are partitioned in the hydrophobic core, avoiding contact with polar solvents, and residues in the hydrophobic core are more conserved than residues related to other parts of the structure (Dill, 1990). Disulfide bridges are important determinants of thermostability, as evidenced by engineering of T4 lysozyme (Wetzel et al., 1988; Matsumura et al., 1989a; Matsumura et al., 1989b; Matsumura et al., 1989c) and subtilisin (Siezen & Leunissen 1997). Another example is provided by the serine protease from *Aquifex pyrophilus*, which contains eight cysteine residues (Choi et al., 1999). When treated with dithiothreitol, the reduction of disulfide bonds in the *A. pyrophilus* protease reduced the protein half-life of 90 hours at 85°C to less than two hours.

Hydrogen bonds are important stabilizing determinants for the folded state, contributing to the large enthalpic term. However, a large number of hydrogen bonds also exist in the unfolded state, and it is difficult to
assign particular importance to any individual hydrogen bond in the context of contribution to thermal stability. In the case of RNase T1, mutagenesis and unfolding experiments showed that 86 hydrogen bonds with an average length of 2.95 Å contributed by approximately a total of 110 kcal/mol (Shirley et al., 1992).

**Ionic interactions** are strongly implicated in protein stability at higher temperatures. The pioneering work of Perutz and Raid highlighted the importance of electrostatic interactions for increased thermal tolerance of proteins from thermophilic organisms (Perutz et al., 1975). Since then, structural comparison of hyper-thermostable proteins with their mesophilic counterparts have given considerable insight into ion-pair interactions and their contribution to thermostability, as for instance in the case of the icosahedral lumazine synthase capsids from the mesophile *Bacillus subtilis* and hyperthermophile *Aquifex aeolicus* (Zhang et al., 2001).

For the homologous protease pair *Bacillus amyloliquefaciens subtilisin* BPN9 and *Thermoactinomyces vulgaris* thermitase, the number of charged residues is the same, but the thermophilic thermitase contains eight additional ion pairs. Thus, the amino-acid distribution and context is probably more relevant to thermostability rather than composition, which indicates evolutionary relationship rather than an indication of thermostability (Teplyakov et al., 1990). In the case of T4 lysozyme, a single ion pair contributed 3 to 5 kcal/mol of stabilization (Anderson et al., 1990), and for barnase, a pair of salt bridges formed at the surface proved to be crucial for thermostability as evidenced by a set of site-directed mutagenesis of Asp8, Asp12 and Arg110 (Horovitz et al., 1990). The stabilizing contribution of the individual ion pairs was –1.25 kcal/mol for Asp12-Arg110 and –0.98 kcal/mol for the ion pair.

Ionic links that improve thermostability are often organized into networks on the protein surface or partially buried at domain or subunit interfaces near local symmetry axes of oligomeric assemblies. In the case of glutamate dehydrogenase from *Pyrococcus furiosus*, stability was attributed to an ion-pair network of 18 charged side chains displaying two-fold symmetrical pattern at a dimer interface (Yip et al., 1995). The disulfide oxidoreductase from *P. furiosus* is stabilized through a similar network of 12 ion-pairing side chains located at a dimer interface (Ren et al., 1998).
As a much-needed complement to the many studies on individual
proteins, a thorough structural genomics study on the hyperthermophilic
bacterium *Thermotoga maritima* has provided statistically significant
evaluation of the importance of various determinants of thermostability
(Robinson-Rechavi et al. 2006). The authors show that secondary
structure, hydrogen bonds and oligomeric state is of less importance for a
protein’s adaptation to function at high temperatures, whereas they found
a strong and significant correlation with the density of salt links and
compactness of the structure.

Halophilic are extremophilic microorganisms that are found in all
domains of life including archaea, bacteria and eukarya (Oren, 2006;
2013). They typically thrive in saline or hypersaline habitats where they
use specialized strategies to cope with osmotic stress. Halophiles can be
divided into slight halophiles that grow optimally at 3% (w/v) salt
concentration, moderate halophiles that grow at 3-15% (w/v) salt, and
extreme halophiles that grow at 15-30% (w/v) total salt (Ventosa et al.,
1998). Thus, the proteins produced by halophiles need to be able to
maintain their structural and functional integrity under high salt
conditions. A conclusive model that accurately explains the stability of
halophilic proteins is lacking, however adaptation of proteins to
halophilic conditions is characterized by an increase in the number of
acidic amino acids at the surface, smaller hydrophobic patches at the
surface, and the presence of salt bridges between acidic and strategically
placed basic residues (Lanyi, 1974; Eisenberg et al., 1992; Danson &
Hugh, 1997; Madern et al., 2000).

The rationale for the presence of excess number of acidic residues on
the protein surface is that a protective hydration shell is formed that
protects the protein from aggregation at high salt concentration (Elcock &
McGammon, 1998). Halophilic proteins also display a significant
reduction in the number of hydrophobic patches arising from reduction
in the number of reduced surface-exposed lysine residues (Oren, 2013). A
correlation has been established between the increase in number of
surface-exposed acidic amino acids and the solubility of halophilic
proteins (Tadeo et al., 2009), and the presence of salt has been shown to
be a prerequisite for the function of halophilic proteins (Mevarech et al.,
2000). Since salt reduces the water activity, halophilic enzymes are
suitable choices for reaction conditions carried out in non-aqueous media
(Sellek & Chaudhuri, 1999; Marhuenda-Egea & Bonete, 2002).
2 PRESENT INVESTIGATION

2.1 Objectives

The overall objective is to obtain increased knowledge of structure-function mechanisms of carbohydrate-active enzymes that can be used for biotechnological applications, including lactose conversion, biomass conversion for second-generation biofuels, and generation of added-value carbohydrate compounds from lignocellulosic waste. Such processes can help to support future environmental and societal sustainability. Knowledge acquired through structure-function studies provides not only general understanding of the natural mechanisms, but also paves the way for design of robust enzymes for existing or new applications.

Paper I

β-glucosidases are promiscuous enzymes that are useful for various applications. In this project, we wanted to characterize biochemically and structurally the β-glucosidase (HoBGLA) from a hyperthermophilic bacterium to evaluate the catalytic performance in applications such as lactose conversion and production of galacto-oligosaccharides.

Question at issue:

• Could the thermostable HoBGLA offer advantages over β-galactosidases currently used in lactose-conversion applications?

Paper II

Galacto-oligosaccharides are prebiotic compounds. In this project, we wanted to further optimize HoBGLA for improved production of GOS through rational design.

Question at issue:

• Is it possible to optimize, through rational design, HoBGLA for lactose conversion and GOS biosynthesis to serve as an alternative to β-galactosidases in industrial applications?
Paper III

Continuous efforts are being made to biochemically and structurally characterize extremozymes to harness their potential in biotechnological applications. In this project, we characterized biochemically and structurally a putative $\alpha$-L-arabinofuranosidase from *Halothermothrix orenii*.

Question at issue:

- A *Halothermothrix orenii* gene showed sequence similarity with GH43 enzymes and was annotated as an $\alpha$-L-arabinofuranosidase. Here, we wanted to characterize the gene product biochemically and structurally to investigate the function of the enzyme, and if possible, to expand the repertoire of extremozymes that can be used for biotechnological applications.

Paper IV

PcP2O is an enzyme of biotechnological interest. It has a fundamentally different primary structure than previously characterized P2Os, and is also more thermostable. By determining the structure of PcP2O we wanted to investigate how the differences in sequence translate into structural differences, and if possible, to pinpoint structural determinants responsible for the increased thermal stability. Binding of substrate by TmP2O has been extensively studied structurally. To evaluate if there are any differences in substrate binding, a crystal structure of PcP2O in complex with a slow substrate was determined.

Questions at issue:

- What are the structural manifestations of the different amino-acid sequence of PcP2O?
- Is it possible to rationalize the higher thermal stability of PcP2O at the structural level?
### 2.2 Methodology overview

#### 2.2.1 Ligation-independent cloning

Ligation-independent cloning (LIC) is a high-throughput cloning method for proteomics that makes use of 12 to 15 base-pair long single-stranded complementary overhangs on the vector and a PCR-generated insert. The primer for the insert is designed with a 12 base-pair long tail that is LIC compatible. The insert that is polymerized by PCR is treated with T4 DNA polymerase in the presence of only one dNTP in the reaction mix (dGTP).

![Vector map for pNIC28Bsa4](image)

Figure 4. Vector map for pNIC28Bsa4 (Savitsky et al., 2010).

T4 DNA polymerase removes nucleotides through its 3′→5′-exonuclease activity until it encounters dCTP in the complementary strand as it incorporates dGTP through its 5′→3′-polymerase activity. Similarly, a LIC-compatible plasmid, in this case pNIC28Bsa4 (Fig. 4; Savitsky et al., 2010), is linearized with T4 DNA polymerase in the presence of only dCTP in the reaction mixture, and modified as above using the exonuclease and polymerase activities, thus creating a short single-stranded tail on both insert and plasmid. Extra nucleotides are removed.
from both the PCR amplified insert and plasmid using a PCR cleanup kit. The T4 DNA polymerase treated plasmid (vector) and insert are annealed when incubated together, and transformed into competent host cells. The final phosphodiester bond between the short single-stranded tail of the insert and that of the plasmid is formed by the ligation machinery of the cell, thus eliminating the use of any independent ligation system (Aslanidis et al., 1990).

The amplified methylated DNA template is subjected to DpnI digestion and the product (recombinant plasmid) is subcloned by transformation into competent Mach1™ cells (Invitrogen). Transformed cells are then plated on Luria Bertani (LB) agar plates supplemented with the antibiotic kanamycin and 5% sucrose. Single colonies of Mach1 cells were inoculated into a small volume of LB culture (with kanamycin) and grown overnight at 37°C, followed by recombinant plasmid extraction using a QIAprep Spin Miniprep Kit (Qiagen). Following plasmid preparation, a glycerol stock of the recombinant plasmid was prepared by first transforming competent E. coli BL21(DE3) cells (expression strain) with recombinant plasmid, followed by spreading the cells on LB agar plates supplemented with kanamycin. A single colony of freshly transformed BL21(DE3) cells was used to inoculate a seed culture containing Terrific Broth medium supplemented with kanamycin. The culture was left to grow overnight after which aliquots were stored in 20% glycerol at –80°C.

### 2.2.2 Site-directed mutagenesis

Site-directed mutagenesis is a technique used to change the nucleotide sequence of a cloned DNA fragment by using primed amplification by PCR. The technique can be used to construct various types of changes in a gene such as single or multiple codon replacements, as well as insertions and deletions. The method involves the use of a nucleotide primer pair (reverse and forward) with internal mismatch codons corresponding to the desired mutation, complementary to template DNA, which is incubated with DNA polymerase (PfuUltra High-Fidelity DNA Polymerase with 3′→5′, 5′→3′) and dNTPs in the PCR machine. At the end of the first round of PCR, the newly synthesized DNA comprises one strand corresponding to the parental strand and the other strand carrying the desired mutation. Upon the completion of the second PCR run four
DNA pairs are produced corresponding to eight template strands for the third run, with two parental DNA strands and six strands with the desired mutation. Thus, each run doubles the number of mutant DNA strands for the subsequent PCR step. At the end of the specified number of PCR cycles (30 cycles), the DNA is subjected to DpnI digestion to degrade the parental DNA with methylated and hemimethylated base sequences, i.e. parental DNA, while the PCR-produced DNA remains intact due to absence of methylated bases (Zheng et al., 2004).

2.2.3 Protein expression and purification

To overexpress recombinant protein, a seed culture was grown by inoculating a single colony/glycerol stock of the transformed cells into TB medium supplemented with kanamycin. The seed culture was allowed to grow overnight at 37°C with constant shaking. The overnight seed culture (5--10 mL) was inoculated into 600 mL TB medium supplemented with kanamycin and the culture was allowed to grow at 37°C with constant shaking until reaching an optical density at 600 nm (OD$_{600}$) of 0.7. At this point, protein expression was induced by the addition of IPTG, and the cultivation was continued at 18°C for 16-18 hours. Cells were harvested by centrifugation and the bacterial cell pellet was re-suspended in lysis buffer (HEPES, pH 7.0). To break the cells, the sample was homogenized four times using a homogenizer and the resulting lysate collected in a beaker on ice, after which the lysate was centrifuged to remove cell debris.

As the first step of purification, Ni$^{2+}$-charged immobilized metal affinity chromatography (IMAC) was used to purify the recombinant protein. The recombinant protein was subjected to Tobacco Etch virus (TEV) protease to cleave off the hexa-histidine tag and buffer dialyzed overnight. Following TEV protease treatment, the protein was subjected to a second round of IMAC fractionation, and the flow through, now containing the non-tagged target protein, was collected. The protein sample was concentrated and further purified by size-exclusion chromatography (SEC). Purity of the fractions was assessed by SDS-PAGE and the suitable fractions were pooled and concentrated for subsequent crystallization and characterization.
2.2.4 β-D-glucosidase/galactosidase assay

The GOD/POD assay (Kunst et al., 1988) includes two coupled reactions. The first reaction involves the catalytic oxidation of β-D-glucose to D-glucono-δ-lactone and hydrogen peroxide by glucose oxidase, GOD (reaction 1).

\[
\text{GOD} \\
\beta-D\text{-glucose} + O_2 + H_2O \rightarrow D\text{-glucono-δ-lactone} + H_2O_2 \quad (1)
\]

In the second reaction, horseradish peroxidase (POD) is mixed with \( H_2O_2 \), \( p \)-hydroxybenzoic acid, and 4-aminoantipyrene (4-AAP), which results in the formation of a colored dye (Quinoeimine). The intensity of the Quinoeimine dye is measured spectrophotometrically at 546 nm. The amount of colored dye produced is stoichiometric with the amount of D-glucose present in the sample. The amount of glucose produced is estimated from a linear glucose standard curve.

\[
\text{POD} \\
2 \text{H}_2\text{O}_2 + p\text{-hydroxybenzoic acid} + 4\text{-AAP} \rightarrow \text{Quinoeimine dye} + \text{H}_2\text{O} \quad (2)
\]

The amount of released D-glucose from either cellobiose (β-glucosidase activity) or lactose (β-galactosidase activity) was determined using the GOD/POD assay outlined above. The GOD/POD assay solution (solution A) was prepared by adding GOD and POD to a final concentration of 2.41 U/mL and 1.45 U/mL, respectively to a 200-mL solution of 4 mM KH\(_2\)PO\(_4\), 6.4 mM 4-AAP, 11 mM phenol (pH 7.0). A glucose standard curve was produced by adding 60 \( \mu \)L D-glucose (in the range 0.28 to 3.9 mM) to 600 \( \mu \)L GOD/POD assay solution (solution A) and incubated at room temperature in the dark for 40 minutes. One unit of enzyme activity was defined as the amount of enzyme releasing 1 \( \mu \)mol of D-glucose per minute under the given conditions. The glucose standard curve was calculated using linear regression.

The cellobiose/lactose reaction (solution B) was prepared by taking adding 20 \( \mu \)L enzyme solution to 480 \( \mu \)L substrate solution (cellobiose or lactose in 20 mM Bis-Tris buffer pH 7.0). The reaction mixture was incubated at 50°C in a heat block for 5 minutes, and stopped by heating at
99°C for 5 minutes. The sample was centrifuged at 13,000 r.p.m. for one minute to pellet precipitated enzyme, and cooled to room temperature. To determine the amount of released D-glucose from cellobiose or lactose, 60 μL cellobiose/lactose reaction mixture (solution B) was added to 600 μL GOD/POD assay solution (mixture A). The A/B assay solution (660 μL) was incubated in the dark at room temperature for 40 minutes, and the absorbance measured at 546 nm. The amount of glucose produced could then be determined from the glucose standard curve. Since the amount of glucose formed in the cellobiose/lactose reaction (B) is determined only for 60 μL of the total volume of 500 μL, the glucose concentration needs to be multiplied by a dilution factor (50/6). The kinetic constants $V_{\text{max}}$, $K_m$ and $k_{\text{cat}}$ were obtained by non-linear regression using the software GraphPad Prism®.

### 2.2.5 $p$-nitrophenyl sugar assay

For assays performed with chromophoric substrates, *i.e.*, $o$- and $p$-nitrophenyl ($o$NP and $p$NP) sugars, a high-throughput screen was used were the enzyme reactions were performed in small volumes of 100 μL, including 90 μL buffer solution and 8 μL substrate. The buffer mixtures were incubated at a desired temperature for 30 min to equilibrate before adding 2 μL enzyme solution at an appropriate concentration. Samples were withdrawn at various time points and 100 μL sodium carbonate were added to stabilize the anionic 4-nitrophenolate form. The absorbance for 150 μL sample volumes was measured at 420 nm ($o$NP) or 405 nm ($p$NP) and concentrations derived using standard curves prepared for $o$NP and $p$NP. The $p$NP concentration range used was 0-80 mM. For $o$NP, the molar extinction coefficient of 4600 M$^{-1}$ cm$^{-1}$ was used to calculate the amount of generated $o$NP. Graphs were prepared using GraphPad Prism 6.0 for Windows.

### 2.2.6 Analysis of GOS synthesis and products

GOS product analysis was performed by thin-layer chromatography (TLC). For analysis directly on lysed cells, the crude cell extracts were incubated with 40% lactose (w/v) and 1 mM Mg$^{2+}$ at 70°C. Samples were then incubated at 95°C for 5 min and diluted 1:10 prior to loading on the TLC plates. The adsorbent was HPTLC Li Chrosper®Silica gel 60 F$_{254}$s
(Merck) and \((n\text{-butanol}-n\text{-propanol}-\text{ethanol}-\text{water} = 2:3:3:2)\) was used as eluent. Staining was performed by immersing the plate in a solution containing [0.5 g thymol, 95 mL 96% ethanol, 5 mL concentrated sulfuric acid]. Standards used for lactose conversion were a mixture of glucose, galactose and lactose (LGG); a purified GOS mixture with monosaccharides removed produced using \textit{Lactobacillus} spp. BGAL (Maischberger \textit{et al.} 2008); and Vivinal®GOS (Borculo Domo, NL). For cellobiose conversion, cellobiose and glucose were used as standards.

### 2.2.7 Protein crystallization

Growing protein crystals of suitable size is the first bottleneck in the crystal-structure determination process. Proteins remain in solution only up to a certain concentration. Once this limiting concentration is reached, the solution will no longer remain homogeneous and enters a new phase, which forms the basis of protein crystallization (McPherson, 1985). Crystallization is a phase transition phenomenon involving different steps: nucleation, growth and cessation of growth. Nucleation is the key to crystallization, and is a first-order phase transition where molecules go from a disordered to an ordered state through the formation of partially ordered intermediates (review Chayen, 2004; Chayen, 2007; Chayen & Saridakis, 2008).

The most widely used method for protein crystallization is the hanging or sitting drop vapor diffusion method (McPherson, 1982). Typically, equal volumes of a protein and precipitant solution (i.e., combinations of salt, buffer, organic polymer) are mixed and placed on a cover slip and turned upside down (hanging drop, Fig. 5a), or put in a depression in a crystallization plate (sitting drop). A volume of precipitant solution (reservoir) is placed separate from the drop, but within the same closed system to allow the difference in concentration between drop and reservoir to cause vapor to evaporate from the drop and thereby slowly increasing protein and precipitant concentration in the drop until the precipitant concentration in the drop reaches that of the reservoir. If successful, the movement in the solubility diagram will take place such that the drop system reaches the labile zone and protein crystal nuclei form spontaneously.
Crystals grow from a supersaturated non-equilibrium solution where the protein concentration exceeds its solubility limit. The supersaturation region can be divided into metastable, labile and precipitation zone (Fig. 5b). The labile zone is the region where the supersaturation is large enough to allow molecules to overcome a threshold energy barrier and start to form small stable aggregates called nuclei. Once nuclei are formed, the protein concentration decreases and the system enters the metastable zone where the nuclei can grow by providing surfaces for addition of new molecules. In the metastable zone, nucleation is unlikely to occur, but existing nuclei can continue to grow. If the supersaturation is too large disordered structures (aggregates or precipitate) may form.

**Figure 5.** Simplified solubility phase diagram for a protein solution.

### 2.2.8 Structure determination and refinement

Briefly, X-ray intensity data were collected on well-ordered protein crystals using synchrotron radiation at beamlines equipped optimally for this purpose. The resulting diffraction is recorded on photosensitive detector and the intensity and position of each “spot” determined by an integration procedure. These intensities represent amplitudes of
diffracted rays that do not contain wave phase information. The pattern of reflection makes up the Fourier transform of the image (i.e., the protein). Since there is no lens to recombine the scattered waves to reconstitute the three-dimensional image, the missing phase information has to be retrieved by other means.

If the 3D structure of a similar (homologous) protein exists, information derived from the atoms’ coordinate positions of the known structure can be used to approximate initial starting phases (calculated phases) for the unknown crystal structure. This method is called the molecular replacement method. In principle, the amplitudes for the known homologous 3D model and of the unknown crystal are used to calculate two vector maps, i.e., Patterson functions. These functions are compared, and if successful, a “hit overlap” is obtained, which provides information about the rotation and translation needed to place the known model in the unknown crystal. Calculated phases from the known model coordinates can then be used with the amplitudes for the unknown crystal, and Fourier synthesis performed. The Fourier synthesis is a summation of all structure factors for the crystal where each structure factor has a known amplitude and phase. The result is an electron density that represents a time-averaged image of the contents of the crystal’s unit cell.

\[
\rho_{(xyz)} = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F_{hkl} \cdot e^{-2\pi i(hx + ky + lz)}
\]

Electron density equation

where \( F_{hkl} \) is the structure factor for a reflection \( hkl \).

The electron density, which represents where and how the electron are distributed inside the crystal, is interpreted by the scientist and coordinates corresponding to the amino-acid sequence of the protein assigned to suitable positions inside the electron density. The result is the atomic model. This model contains errors and needs to be iteratively refined and adjusted before deposition with the Protein Data Bank. Both structure factors (raw data) and coordinates (derived model) are deposited.
2.3 Summary of Paper I

Biochemical and structural characterization of a thermostable β-glucosidase from *Halothermothrix orenii* for galacto-oligosaccharide synthesis

*Background and aim*

Galacto-oligosaccharides (GOSs) are β-linked non-digestible galactose oligomers found in human milk with a typical DP of 1-3 (Gosling *et al.*, 2010). Enzymatic synthesis of GOSs exploits the inherent ability of BGALs to carry out a transgalactosylation reaction, and is normally coupled to milk lactose conversion (Park *et al.*, 2010). GOSs are prebiotic compounds with health-promoting effects such as lowering of cholesterol levels and prevention of colon cancer (Shadid *et al.*, 2007). They also stimulate the growth of bifidobacteria in the human gut, which provides additional health benefits (Gibson *et al.*, 1995). GOS production through transgalactosylation is favored at higher substrate concentrations, and since lactose is more soluble at higher temperatures and the risk of microbial contamination decreases, thermostable enzymes are desirable (Gekas & Lopez-Leiva, 1985; Akiyama *et al.*, 2001). Besides BGALs, many BGLs also possess inherent transgalactosylation activity, and the use of BGLs rather than BGALs would be preferred considering that BGLs are smaller and more stable enzymes. An ideal GOS producer would be a thermostable BGL with little or no product inhibition that can operate at high substrate concentration and temperatures at 70–80°C.

In the first step of the galactosidase-catalyzed synthesis of GOS during milk lactose hydrolysis, the enzyme binds the galactosyl end of lactose. It then catalyzes the hydrolysis of the glycosidic bond between galactose and glucose to form an enzyme-galactosyl intermediate. In the second step, the galactosyl moiety is transferred either to water (hydrolysis) or to the hydroxyl of the free glucose or galactose of the lactose leading to formation of disaccharide β(1→6) allolactose or a trisaccharide consisting of two galactose unit and one glucose (Otieno, 2010; Hernández-Hernández *et al.*, 2011).

Here, a BGL from the halothermophilic bacterium *Halothermothrix orenii* was characterized biochemically and structurally, and evaluated for its potential usefulness for lactose conversion and GOS production.
Summary of results

The gene for wild-type HoBGLA was cloned and expressed in *E. coli*. In addition, three catalytically impaired active-site mutants were produced by site-directed mutagenesis. The activity of the HoBGLA wild type and variants on substrates relevant for both BGLs and BGALs were investigated. The temperature and pH profiles for wild-type HoBGLA were determined, as well as the ability of wild-type HoBGLA to convert lactose to GOS. To evaluate ligand binding and guide theoretical modeling of the GOSs produced by HoBGLA, the crystal structures of three ligand complexes were determined, *i.e.*, HoBGLA with bound thiocellobiose, 2-deoxy-2-fluoro-d-glucose, and glucose.

HoBGLA was found to have promising characteristics relating to GOS production compared with other GOS-producing BGLs reported to date. HoBGLA hydrolyzes both cellobiose and lactose, with β-glucosides being the preferred substrates. Although HoBGLA is quite non-specific with respect to its substrate preference, the enzyme displays significant galactosidase activity. HoBGLA shows a pH optimum of about 6 for lactose hydrolysis, and a broad optimal stability over the pH range of 4.5–7.5, with a temperature optimum for activity and stability in the range 65–70°C. These features argue in favor of HoBGLA as an attractive biocatalyst for lactose conversion. In transgalactosylation experiments, HoBGLA produced a range of GOSs with β-D-Galp-(1→6)-D-Lac (6GALA) and β-D-Galp-(1→3)-D-Lac (3GALA) being the two major products. The two major GOS products, 3GALA (Fig. 6a) and 6GALA (Fig. 6b), were modeled in the substrate-binding cleft of wild-type HoBGLA and shown to be favorably accommodated.

Figure 6. Theoretical models of HoBGLA in complex with 3GALA (a) and 6GALA (b).
2.4 Summary of Paper II

Engineering a polyextremophilic *Halothermothrix orenii* β-glucosidase for improved galacto-oligosaccharide synthesis

*Background and aim*

Non-digestible galacto-oligosaccharides (GOSs) are attracting increasing attention as health-promoting functional food additives and dietary supplements due to their ability to sustain and promote the growth of bifidobacteria in the intestinal tract. GOSs can be synthesized chemically, however, the chemical approach requires several reaction steps since selective protection of the sugar hydroxyl groups is needed. As discussed in Chapters 1.2 and 2.3 (Paper I), enzymatic synthesis of GOS is an attractive alternative to chemical synthesis where the inherent ability of BGALs and BGLs to perform transglycosylation reactions is used (a side reaction to the biologically relevant catalysis of glycosidic bond cleavage), which offers an environmentally friendly and economical approach (Gekas and Lopez-Leiva 1985). BGLs have a number of advantages over BGALs, i.e., BGLs are (i) smaller size in the range 50-60 kDa (BGALs are large, typically 120 kDa); (ii) easy to produce; (iii) typically do not require metal ions for activity; and (iv) not inhibited by the reaction product galactose.

Since the majority of enzymes used are from mesophilic organisms, their application range is restricted due to limited enzyme stability under industrial conditions. In this context, the use of enzymes produced by extremophiles, *extremozymes*, are attractive biocatalyst alternatives. Thermophilic and hyperthermophilic bacteria are among the most useful microbial producers of highly stable and robust enzymes for bioprocesses. For instance, *Halothermothrix orenii* is a heterotrophic, halophilic, thermophilic, obligate anaerobic bacterium (Cayol *et al*., 1994) that produces a thermostable β-glucosidase, *HoBGLA* (Mijts & Patel, 2001; Mavromatis *et al*., 2009; Kori *et al*., 2011; Paper I: Hassan *et al*., 2015). We have reported the biochemical and structural characterization of *HoBGLA* and shown that the enzyme has attractive properties relevant to GOS production (Paper I: Hassan *et al*., 2015), *e.g.*, the enzyme is stable at high temperatures (65–70°C) and within a broad pH range (4.5–7.5), and produces two major types of GOS (3GALA and 6GALA).
Efforts to further engineer mesophilic or extremophilic BGLAs and BGLs for improved transglycosylation have been reported in the literature (Hansson et al., 2001; Jørgensen et al., 2004). Typically, the approaches used to engineer an enzyme for improved performance under bioprocess conditions include rational structure-based design, directed evolution and semi-rational approach (Lutz, 2010; Bomarius et al., 2011; Bornscheuer et al., 2012; Steiner & Schwab, 2012).

In this work, the goal was to improve the transglycosylation properties of HoBGLA towards higher total GOS yield, but also higher yields of individual GOS products. HoBGLA was engineered by structure-based rational design guided by engineering strategies used for other BGLs to improve the ratio of hydrolysis-to-transglycosylation. The strategy mainly involves replacement of amino acids in the glycone and aglycone sugar-binding subsites.

Summary of results

In total, six HoBGLA variants were designed rationally by replacing side chains in the +1 and −1 subsites: N222F, N294T, Y296F, N406I, F417Y, and F417S. All variants except N406I were successfully expressed and purified. The kinetic parameters (V_{max}, K_m, k_{cat}/K_m) for HoBGLA wild type and variants were determined at 70°C using cellobiose and lactose as substrates. Compared with the wild type, all HoBGLA variants show significantly impaired catalytic activity with cellobiose and lactose as substrates. Particularly F417S is hydrolytically crippled with cellobiose as substrate with a 1000-fold decrease in apparent $k_{cat}$, but to a lesser extent affected when catalyzing hydrolysis of lactose (47-fold lower $k_{cat}$). This large selective effect on cellobiose hydrolysis is manifested as a change in substrate selectivity from cellobiose to lactose.

The variant Y296F shows 72% lactose conversion and produces a total GOS yield of 68% (12h) with 6GALA (133 g/L) as the predominant product, and the highest amounts of 3GALA (33 g/L) of the variants analyzed. F417S shows 79% lactose conversion (12h) and produces a total GOS yield of 71% of which 6GALA is obtained in higher amounts (136 g/L) than the wild-type and other variants. The variant F417Y shows good capacity to hydrolyze both cellobiose and lactose with the same relative substrate selectivity as the wild type, but with approximately 10-fold
lower turnover numbers. F417Y converts lactose efficiently (92% conversion, 2h) and produces a total GOS yield of 43% with mainly 6GALA as product, however at lower total yield (64 g/L after 2h) than obtained with F417S and Y296F.

For a comparison with commercial sources of enzymes used for lactose conversion and GOS synthesis, two products based on the *K. lactis* BGAL, *i.e.*, Maxilact® LGX 5000 (DSM) and Lactozym® 3000 L HP (Novozymes), produce GOS yields at 40°C of 41% (154 g/L after 15h) with 83% lactose conversion, and 42% (160 g/L after 10h) with 95% lactose conversion, respectively (Rodriguez-Colinas *et al.*, 2011). Another commercial source is available for *A. oryzae* BGAL, Enzeco® Fungal lactase (EDC), which shows 70% lactose conversion at 40°C with a total GOS yield of 27% (107 g/L, 7h) (Urrutia *et al.*, 2013). Thus, the preliminary results presented in this work show that the engineered *HoBGLA* variants perform very well with respect to lactose conversion and GOS production. A clear advantage with *HoBGLA* compared with the commercial sources is that *HoBGLA* operates at 70°C where the risk of bacterial contamination and growth is minimal.
2.5 Summary of Paper III

High-resolution crystal structure of a polyextreme GH43 glycosidase from *Halothermothrix orenii* with α-L-arabinofuranosidase activity

**Background and aim**

Hemicelluloses are heterogeneous matrix polymers built from various sugars such as xylose, glucose, galactose and arabinose. These heteropolymers contribute significantly to the insoluble nature of lignocellulose, which in turn complicates efficient biomass conversion. To improve the efficiency of biomass conversion, the use of naturally occurring biomass-converting enzymes has been investigated intensively with a main focus on glycoside hydrolases from mesophilic organisms. The total expenditure on enzymes, particularly industrial enzymes, increased to $3.3 billions by the end of 2011, and is expected to increase further to $4.4 billion by the end of 2015 (Delgado-Garcia *et al*., 2012). The demand for improved enzymes for biotechnological applications and the need to lower production costs to make enzyme-based applications economically viable have prompted major efforts in enzyme discovery and engineering.

In this context, the large number of sequenced microbial genomes has paved the way for the discovery of more robust enzymes produced by extremophilic organisms, so called *extremozymes*, which are of considerable interest for industrial settings that involve harsh conditions, *e.g.*, high temperature, high salt concentrations, high acidity, or other extreme reaction conditions (Gomes & Steiner, 2004; Gupta *et al*., 2014; Yin *et al*., 2014). Relevant to the applications discussed in this thesis is the fact that many extremophiles produce a large repertoire of genes encoding intracellular and extracellular GHs, some of which could prove useful in improved or new enzymatic bioprocesses. The anaerobic bacterium *Halothermothrix orenii* is a true halotermophile isolated from the salt lake Chott El Guettar in Tunisia (Cayol *et al*., 1994), and its genome sequence has revealed a number of GH genes encoding potentially useful enzymes. The GH1 *H. orenii* β-glucosidase HoBGLA has been discussed in Chapter 2.3 (Paper I) and Chapter 2.4 (Paper II) of this thesis.
Enzymes belonging to CAZy family GH43 are mainly involved in hydrolysis of various hemicelluloses, and microbes that degrade more complex plant biomass typically produce a large number of GH43 enzymes (DeBoy et al., 2008). Interestingly, the genome of H. orenii contains only a single GH43 gene. Based on sequence similarity to other GH43 members, the gene product was provisionally annotated as an α-L-arabinofuranosidase. The gene encoding this H. orenii GH43 glycosidase was previously cloned and the gene product expressed and purified as part of a larger effort to identify extremozymes from H. orenii (Kori, 2012).

The aim in this work was to determine the crystal structure and substrate specificity of the only GH43 glycosidase produced by H. orenii to obtain expanded knowledge about the hemicellulose-degrading enzyme system of this unusual marine microbe. We also wanted to further investigate structural determinants that provide adaptation to extreme habitats in general, and to evaluate the potential industrial usefulness of the H. orenii GH43 enzyme in particular.

Summary of results

The crystal structure of the H. orenii GH43 glycosidase was determined by molecular replacement and refined at 1.10 Å resolution. As expected from the primary sequence, the enzyme belongs to the CAZy family GH43 of inverting glycoside hydrolases with a characteristic five-bladed β-propeller fold (Fig. 7). As for related GH43 members, the conserved catalytic nucleophile is placed on blade 4 (Glu195), the catalytic base (Asp17) on blade 1, and the assisting residue Asp126 on blade 3 (Nurizzo et al., 2002). Three putative sugar binding sites were identified, subsites –1, +1 and +2, with a metal-binding site located on the central axis of the propeller near the –1 subsite.

The role of the metal-binding site was investigated by evaluating the effect of various cations on thermal stability to unfolding. It was found that the enzyme was stabilized in the presence of divalent cations, where Mn²⁺ at pH 7.0 was the most stabilizing condition corresponding to an increase in Tₘ value of 5.5°C (from 63°C to 68.5°C). The metal ions Ca²⁺ and Mg²⁺ resulted a Tₘ increase of 4.2°C and 2.0°C, respectively.
The influence of high salt concentration on thermal stability was also investigated. The *H. orenii* GH43 glycosidase showed a remarkable stabilization at 4 M NaCl corresponding to an increase in $T_m$ by 10°C (from 63°C to 73°C). Increased stability was closely correlated with increased salt concentration up to 4 M NaCl, whereas stability started to decrease again at 5 M NaCl. The combination of high salt and metal ion produced an even larger increase stabilization to thermal unfolding with 4 M NaCl in the presence of Mn$^{2+}$ at pH 6.0 being the most stabilizing condition with a $T_m$ value of 75-76°C compared with no salt or metal ion ($T_m$ 63°C).

Next, substrate screening was performed using ten different chromophoric aryl-glycoside substrates relevant to enzyme members of the GH43 family. Of these, only *p*-nitrophenyl-α-L-arabinofuranoside (*pNP*-αAraf) was hydrolyzed by the *H. orenii* GH43 glycosidase, with the highest activity found in the presence of 4 M NaCl and Mn$^{2+}$. The calculated specific activity was ~20 to 36 μmol min$^{-1}$ mg$^{-1}$ at 4 M NaCl and 2 mM Mn$^{2+}$ (pH 6.5). The specific activity is similar to that reported for a commercial source of α-L-arabinofuranosidase from Megazyme (*pNP*-αAraf; ~32 μmol min$^{-1}$ mg$^{-1}$). Thus, the assignment of the *H. orenii* enzyme as an α-L-arabinofuranosidase is justified, and the enzyme is
hereafter referred to as HoAraf43. TLC analysis (3h incubation in the presence of substrate) revealed no hydrolysis products from α-L-arabinofuranopentaose indicating that the natural polymeric substrate must be fundamentally different.

The amino-acid composition of thermophilic proteins has been studied quite thoroughly, but less has been done for halophilic proteins, and for proteins that are both thermophilic and halophilic, virtually no information is available. Therefore, we considered it of particular interest and importance to analyze the distribution of amino acids on the surface and interior of HoAraf43. The results show that HoAraf43 shares similarities to mesophiles, thermophiles and halophiles, but also displays unique features, such as more hydrophobic amino acids on the surface and fewer buried charged residues (Table 2). Since the statistics presented for halothermophilic proteins are based only on one protein, the values are not statistically significant, however, as more halothermophilic proteins become characterized, additional data will provide a more refined picture of the differences in amino-acid composition and distribution of proteins that function under these conditions.

Table 2. Comparison of amino-acid composition

<table>
<thead>
<tr>
<th>Class (%)</th>
<th>Mesophiles and non-halophiles</th>
<th>Thermophiles</th>
<th>Halophiles</th>
<th>Halo-thermophile (HoAraf43)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Int</td>
<td>Surf</td>
<td>Int</td>
<td>Surf</td>
</tr>
<tr>
<td>Apolar</td>
<td>45.6</td>
<td>18.1</td>
<td>46.9</td>
<td>20.2</td>
</tr>
<tr>
<td>Polar</td>
<td>15.3</td>
<td>21.0</td>
<td>13.4</td>
<td>15.6</td>
</tr>
<tr>
<td>Basic</td>
<td>6.0</td>
<td>16.7</td>
<td>7.1</td>
<td>20.1</td>
</tr>
<tr>
<td>Acidic</td>
<td>7.2</td>
<td>21.0</td>
<td>7.8</td>
<td>23.0</td>
</tr>
<tr>
<td>Other</td>
<td>25.9</td>
<td>3.1</td>
<td>24.8</td>
<td>21.1</td>
</tr>
</tbody>
</table>

1 Classification: apolar V, I, L, M, F, W, Y; polar N, Q, S, T; basic K, R; acidic D, E; other A, C, G, P, H; given as percentage in the interior (Int) and on the surface (Surf).
2 Values from Fukuchi et al., 2003.
3 Values that are different in the H. orenii GH43 enzyme with respect to other groups are highlighted in boldface.
2.6 Summary of Paper IV

Crystal structures of *Phanerochaete chrysosporium* pyranose 2-oxidase suggest that the N-terminus acts as a propeptide that assists in homotetramer assembly

Background and aim

Fungal P2Os are tetrameric enzymes that oxidize glucose and other pyranoses, including galactose. Galactose is present in the disaccharide lactose, *i.e.*, the major carbohydrate present in milk and dairy products. As discussed in Chapters 2.3 and 2.4, lactose is readily hydrolyzed to glucose and galactose by BGALs and BGLs. Whereas BGLs can convert the products from lactose hydrolysis to value-added GOS compounds, P2Os can be used to modify the same products by oxidation into valuable value-added sugar intermediates for further chemoenzymatic synthesis (Freimund *et al.*, 1998; Giffhorn *et al.*, 2000). For instance, oxidation of D-galactose at the C-2 position by P2O produces 2-keto-D-galactose which can be used as a platform intermediate for enzymatic production of D-tagatose, a non-carcinogenic low-calorie sweetener and prebiotic (Izumori & Tsuzaki 1988; Haltrich *et al.*, 1998; Manzoni *et al.*, 2001). P2O applications are however not limited to the synthesis of platform molecules, but can also be of potential use in biofuel cells (Spadiut *et al.*, 2009).

The white-rot fungus *Phanerochaete chrysosporium* is a well-studied fungal model for lignocellulose degradation (Martinez *et al.*, 2004). Like the previously well-characterized P2O from *Trametes multicoro*, P2O from *Phanerochaete chrysosporium* (*PcP2O*) is a flavin dependent homotetrameric protein with molecular mass of 270 kDa, expressed along with other redox-active enzymes when lignin is available as carbon source (De Koker *et al.*, 2004). Despite the functional similarity, the amino-acid sequences of *TmP2O* and *PcP2O* are only 40% identical. Especially the N-terminus is different, and it has been suggested that *PcP2O* may contain an N-terminal propeptide with unknown function (Nishimura *et al.*, 1996; Artolozaga *et al.*, 1997; Leitner *et al.*, 2001). *PcP2O* is also more stable than *TmP2O* at high temperature, as shown by a 17°C higher melting temperature, *T*<sub>m</sub> (Salaheddin *et al.*, 2010).
A large number of prokaryotic and eukaryotic proteins are synthesized as preproprotein precursors. The pre-region typically functions as signal peptide involved in transport of the protein, while the propeptide is responsible for correct folding or activation of the protein (Shinde & Inouye, 1993; 2000; Fu et al., 2000). Unlike a molecular chaperone, an intramolecular chaperone is covalently attached to the amino or carboxy terminal of the mature protein, and is referred to as a prosequence or propeptide that is not part of the functional protein (Inouye, 1991). The intramolecular chaperone is removed during protein folding by either autocatalytic cleavage as in the case of proteases, or by an external protease in the case of most other proteins (Eder & Fersht, 1995).

The role of an intramolecular chaperone is typically to assist protein folding and function. The first case of an intramolecular chaperone was reported for the extracellular protease subtilisin from Bacillus subtilis (Ikemura et al., 1987). The enzyme is secreted with a 77-residue long propeptide at the N-terminus preceded by a signal peptide of 29 residues (Zhu et al., 1989). Other examples include the P. chrysosporium lignin peroxidases LiP2 and LiP, which are secreted with a 28-residue long peptide sequence of which 7 amino acids comprise a putative propeptide (Schalch et al., 1989) similar in sequence to the propeptide of glucoamylase from Aspergillus awamori (Innis et al., 1985). Intramolecular chaperones can be classified according to their role in protein folding. Type-I intramolecular chaperone are mainly present as an N-terminal extension to assist in tertiary-structure formation, whereas Type-II intramolecular chaperone are typically located at the C-terminus to assist in assembly of the quaternary structure (Chen et al., 2008).

Here, the aim was to determine and analyze the crystal structure of PcP2O to shed light on structural features that differ between PcP2O and TmP2O, especially those that may account for increased thermostability, and to investigate the possible presence of a propeptide at the N-terminus.

Summary of results

The crystal structures of PcP2O expressed from three different gene constructs (Fig. 8) were determined: (i) wild-type PcP2O isolated from the fungal extract of P. chrysosporium strain K-3 (PcP2ONATWT); (ii)
PcP2O inserted into the expression vector pET21a(+) vector with an N-terminal T7-epitop tag and a non-cleavable C-terminal His\(_6\) tag, including the N-terminal region proposed to contain a propeptide function (PcP2O\(_{RECWT}\)); and (iii) the PcP2O variant H158A expressed as a construct with a cleavable N-terminal His\(_6\) tag (PcP2O\(_{RECH158A}\)). The His\(_6\) tag in PcP2O\(_{RECH158A}\) was removed by TEV protease leaving the amino acids \((-^{2}SM^{1}\) at the N-terminus from the TEV cleavage site (\(^{-23}\text{MHHHHHHS}\text{VDHTENLYFQSM}^{+1}\)). To reduce the number of additional residues at the N-terminal, the naturally occurring methionine in the PcP2O gene was replaced by the translation start-site methionine, thus carrying only one extra serine residue before the proposed propeptide sequence at the N-terminus (i.e., \((-^{1}\text{SMFLDTTPF}^{9})\) in PcP2O\(_{RECH158A}\).

![Diagram](image)

**Figure 8.** Expressed and crystallized PcP2O constructs.

Wild-type PcP2O from the natural source (PcP2O\(_{NATWT}\)) was crystallized in the absence of ligand, and the two recombinant PcP2O variants, PcP2O\(_{RECWT}\) and PcP2O\(_{RECH158A}\), were co-crystallized in the presence of a slow substrate, i.e., 3-deoxy-3-fluoro-β-D-glucose (3F\text{Glc}), that has previously been shown to bind to TmP2O in a productive binding mode for 2-oxidation. X-ray diffraction intensity data collection was performed at MAX-lab in Lund (Sweden) and the crystal structures of the three PcP2O forms determined and refined at a resolution of 1.80 Å, 1.80 Å, and 2.40 Å, respectively.
Despite the low amino-acid sequence identity between PcP2O and TmP2O, the overall structures are very similar, including the active site and the precise details of binding of the slow substrate 3FGlc. The most striking difference between the enzymes appears to be the thermal stability. PcP2O has a $T_m$ value of 75.4°C, whereas TmP2O has a $T_m$ value of 58.2°C. Based on the large body of evidence for the importance of ionic links as structural determinants for protein thermostability, the structures of PcP2O and TmP2O were analyzed using ESBRI (http://bioinformatica.isa.cnr.it/ESBRI/) to identify and compare inter- and intrasubunit salt bridges. The result was unexpected. Although TmP2O had a higher total number of salt links (247) compared with PcP2O (195), the number of links between subunits was only four in TmP2O compared with 20 in PcP2O. Based on these results, a likely hypothesis is that the higher thermostability of PcP2O is governed mainly by a larger number of ionic links between protein subunit of the homotetramer.

In the crystal structure of PcP2ONATWT, Pro13 was identified as the first amino acid from the N-terminus in all the four subunit; residues 1-12 may have been removed due to processing or simply flexible. In the crystal structure of PcP2ORECW1 the first visible residue was Ala10 in subunit B, but the first visible residue was Pro13 in the other subunits. Most likely, the N-terminal T7 tag forces the N-terminus into the surrounding solvent where it would become flexible and not visible in the electron density, or it may have been proteolytically removed. In the case of PcP2ORECH158A, subunit A shows well-defined density for the N-terminal sequence 2FLD4 that is part of the 7-residue long stretch assigned as a propeptide. For subunits B and E, the density is reasonable, but subunits C, D, F, G display poor density for the N-terminal sequence. The 2FLD4 is placed at the interface between A and B subunits where it appears to stabilize the 2-fold symmetrical A/B interface, forming a temporary hydrophobic core, which in the mature protein (as defined by PcP2ONATWT and PcP2ORECW1) is occupied by 66QFG68 of an oligomerization loop in the subunit B of PcP2ONATWT and PcP2ORECW1. This N-terminal propeptide segment in PcP2ORECH158A may serve the function of an intramolecular chaperone stabilizing the A/B interface prior to sequestering the 66QFG68 of the subunit B to assume its final position and conformation at A/B the interface of the mature enzyme.
3 CONCLUDING REMARKS

In this thesis, three enzymes with biotechnological potential have been studied biochemically and structurally. The 3D structures enable rational design to engineer the enzymes towards improved performance in relevant applications. The precise engineering approach will ultimately depend on the enzyme and the application in mind. In the case of HoBGLA, the inherent high thermostability and transgalactosylation activity (Paper I) offered an ideal starting point for further improvement towards improved lactose conversion and GOS production. This was further investigated in Paper II, and the results showed that there is considerable possibility to improve an already “good” enzyme. In Paper III, a novel hemicellulase was characterized and found to have α-L-arabinofuranosidase activity under the extreme conditions of 4M salt and elevated temperatures. In the case of PcP2O (Paper IV), the structure provided new insights into both thermostability and the possible role of an N-terminal intramolecular chaperone where a network of ionic links between subunits of the tetramer are likely determinants of thermostability for this enzyme.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>three-dimensional</td>
</tr>
<tr>
<td>3GALA</td>
<td>β-D-Galp-(1→6)-D-Lac</td>
</tr>
<tr>
<td>6GALA</td>
<td>β-D-Galp-(1→3)-D-Lac (3GALA)</td>
</tr>
<tr>
<td>4-AAP</td>
<td>4-aminoantipyrene</td>
</tr>
<tr>
<td>BGAL</td>
<td>β-galactosidase</td>
</tr>
<tr>
<td>BGL</td>
<td>β-glucosidases</td>
</tr>
<tr>
<td>CBM</td>
<td>carbohydrate-binding module</td>
</tr>
<tr>
<td>CDH</td>
<td>cellobiose dehydrogenase</td>
</tr>
<tr>
<td>GMC</td>
<td>glucose-methanol-choline</td>
</tr>
<tr>
<td>GOD</td>
<td>glucose oxidase</td>
</tr>
<tr>
<td>DP</td>
<td>degree of polymerization</td>
</tr>
<tr>
<td>FAD</td>
<td>flavin adenine dinucleotide</td>
</tr>
<tr>
<td>GH</td>
<td>glycoside hydrolase</td>
</tr>
<tr>
<td>GOS</td>
<td>galacto-oligosaccharide</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</td>
</tr>
<tr>
<td>IMAC</td>
<td>immobilized metal affinity chromatography</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
</tr>
<tr>
<td>LIC</td>
<td>ligation-independent cloning</td>
</tr>
<tr>
<td>LPMO</td>
<td>lytic polysaccharide monooxygenases</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>P2O</td>
<td>pyranose 2-oxidase</td>
</tr>
<tr>
<td>POD</td>
<td>horseradish peroxidase</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco Etch virus</td>
</tr>
</tbody>
</table>
Acknowledgements

I am extremely grateful to my supervisor Assoc. Prof. Christina Divne for accepting me as a PhD student in her group and providing guidance throughout the course of the study through continuous discussions, support and active supervision.

I would like to thank my co-supervisor Tien-Chye Tan (postdoc) for his active and continued supervision and for always being available for providing guidance in the lab.

I would also like to thank Rosaria Gandini (postdoc) in our group, who also always remained available for help in the lab whenever asked for.

Thanks to our collaborators Dietmar Haltrich, Oliver Spadiut, Ines Pisanelli, Laura Fusco, Clemens Peterbauer, Thu-Ha Nguyen, Montira Intanon, Barbara Geiger, Roman Kittl, Lokesh Kori, and Bharat Patel for their contribution to the research work.

Thanks to all the colleagues, friends and lab members at the Biophysics division at Karolinska Institute MBB floor 5, for their help and exchanging views and creating a healthy environment that proved very helpful in materializing this research work. Thanks also to colleagues at the Division of Industrial Biotechnology at the KTH School of Biotechnology.

Last but not the least I would like to thank my family, whose cooperation, love and affection always proved as a stimulant for the continuous and sustained efforts to achieve the desired results through out the course of this study.
References


Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-


