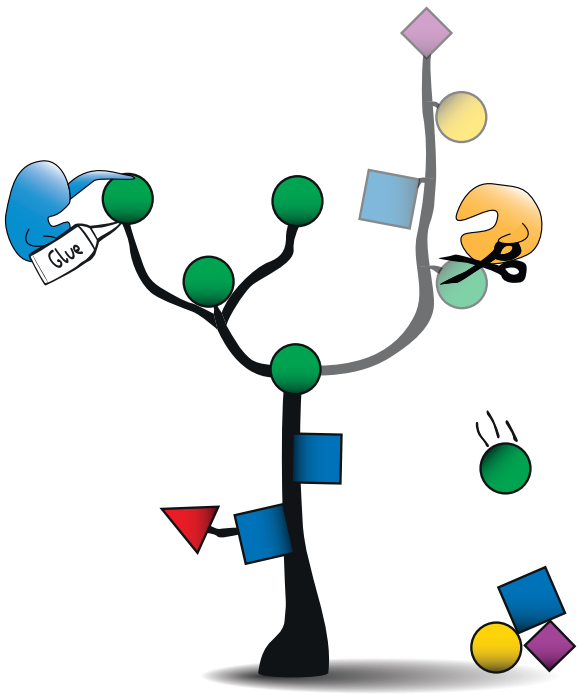
KTH ROYAL INSTITUTE
OF TECHNOLOGY

Doctoral Thesis in Biotechnology

Structural and biochemical insights into biosynthesis and degradation of *N*-glycans

TOM REICHENBACH



Stockholm, Sweden 2020

Structural and biochemical insights into biosynthesis and degradation of *N*-glycans

TOM REICHENBACH

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Doctor of Philosophy on Friday the 16th October 2020, at 10:00 a.m. in Kollegiesalen, KTH, Brinellvägen 8, Stockholm.

Doctoral Thesis in Biotechnology
KTH Royal Institute of Technology
Stockholm, Sweden 2020

ISBN 978-91-7873-660-7
TRITA-CBH-FOU-2020:41

Printed by: Universitetsservice US-AB, Sweden 2020

Abstract

Carbohydrates are a primary energy source for all living organisms, but importantly, they also participate in a number of life-sustaining biological processes, e.g. cell signaling and cell-wall synthesis.

The first part of the thesis examines glycosyltransferases that play a crucial role in the biosynthesis of *N*-glycans. Precursors to eukaryotic *N*-glycans are synthesized in the endoplasmic reticulum (ER) in the form of a lipid-bound oligosaccharide, which is then transferred to a nascent protein. The first seven sugar units are assembled on the cytoplasmic side of the ER, which is performed by glycosyltransferases that use nucleotide sugars as donors. The mannosyl transferase *Pc*ManGT is produced by the archaeon *Pyrobaculum calidifontis*, and the biochemical and structural results presented in the thesis suggest that the enzyme may be a counterpart to the glycosyltransferase Alg1 that participates in the biosynthesis of *N*-glycans in eukaryotes. Within the ER (in the lumen), activated dolichol-bound sugars are used as donor substrates instead of nucleotide sugars for glycosyltransferases that synthesize *N*-glycans. The glycosyltransferase dolichylphosphate mannose synthase (DPMS) catalyzes the formation of dolichylphosphate mannose, which is one of these dolichyl-bound sugars. The structure and function were studied for DPMS from *Pyrococcus furiosus* using protein X-ray crystallographic and biochemical methods and a new assay based on proteoliposomes was designed.

The second part of the thesis focuses on glycoside hydrolases from bacteria that break down oligo- and polysaccharides. In one of the studies, a bacterial glycoside hydrolase from the acne bacterium *Cutibacterium acnes* was characterized. The enzyme was shown to be able to break down the host's *N*-glycans, which can be used as nutrients or perhaps even evade detection of the immune system. This study also suggests a cytoplasmic biosynthetic pathway for the formation of *N*-glycans in the acne bacterium. In another study, a glycoside hydrolase from a bacterium living in the moose rumen was characterized. The enzyme was shown to be able to break down β-1,3-glucans, which is a property that can be used industrially for biomass treatment.

Sammanfattning

Kolhydrater utgör en primär energikälla för alla levande organismer, man deltar dessutom i mängd livsuppehållande biologiska processer, t.ex. cellsignalering och cellväggssyntes.

Den första delen av avhandlingen undersöker glykosyltransferaser som spelar en avgörande roll för biosyntes av N-glykaner. Förstadier till eukaryota *N*-glykaner syntetiseras i det endoplasmatiska nätverket (ER) i form av en lipidbunden oligosackarid som sedan överförs till ett nybildat protein. De första sju sockerenheterna sätts ihop på den cytoplasmatiska sidan av ER, vilket sker med hjälp av glykosyltransferaser som använder nukleotidsockerföreningar som kan donera sockerenheter. Mannosyltransferaset *Pc*ManGT produceras av arkebakterien *Pyrobaculum calidifontis* och de biokemiska och strukturella resultat som presenteras i avhandlingen tyder på att enzymet kan vara en motsvarighet till glykosyltransferaset Alg1 som deltar i biosyntes av *N*-glykaner i eukaryoter. Inuti ER (i lumen) används aktiverade dolikolbundna socker istället för nukleotidsockerföreningar som substrat för glykosyltransferaser som syntetiserar *N*-glykaner. Glykosyltransferaset dolikolfosfatmannossyntas (DPMS) katalyserar bildandet av dolikolfosfatmannos vilket är en av dessa dolikolbundna sockerföreningar. Struktur och funktion studerades för DPMS från *Pyrococcus furiosus* med hjälp av proteinröntgenkristallografi samt biokemiska metoder samt en ny analysmetid baserad på proteoliposomer.

Den andra delen av avhandlingen fokuserar på glykosidhydrolaser från bakterier som bryter ned oligo- och polysackarider. I en av studierna karakteriserades ett bakteriellt glykosidhydrolas från aknebakterien *Cutibacterium acnes*. Enzymet visade sig kunna bryta ned värdorganismens *N*-glykaner vilka kan användas som näring eller kanske även i syfte att undgå upptäckt av immunförsvaret. Studien föreslår även en cytoplasmatisk biosyntesväg för bildande av N-glykaner i aknebakterien. I en annan studie karakteriserades ett glykosidhydrolas från en bakterie som lever i älgvåmmen. Enzymet visade sig kunna bryta ned β-1,3-glukaner vilket är en egenskap som kan användas för i industriella tillämpningar för bearbetning av biomassa.

Zusammenfassung

Kohlenhydrate sind eine primäre Energiequelle für alle lebenden Organismen, aber vor allem sind sie an einer großen Anzahl von lebensnotwendigen biologischen Prozessen, z.B. Zellsignalwegen und Zellwandaufbau, beteiligt.

Der erste Teil dieser Doktorarbeit untersucht Glycosyltransferasen, welche entscheidend für die Synthese von *N*-Glycanen sind. Die Vorläufer der eukaryotischen *N*-Glycane werden im endoplasmatischen Retikulum (ER) als Lipid-gebundene Oligosaccharide synthetisiert, bevor sie auf das entstehende Protein transferiert werden. Die ersten sieben Zucker werden auf der zytoplasmatischen Seite der ER-Membran von Glycosyltransferasen, welche Nukleotidzucker als Donor nutzen, transferiert. Die Mannosyltransferase *Pc*ManGT wird von den Archaea *Pyrobaculum calidifontis* produziert, und die in dieser Doktorarbeit präsentierten biochemischen und strukturellen Resultate deuten darauf hin, dass das Enzym ein Pedant zu dem eukaryotischen Enzym Alg1 ist, welches Teil der eukaryotischen *N*-Glycan Biosynthese ist. Im ER (ER-lumen) werden aktivierte Dolichol-gebundene Zucker, statt Nukleotidzuckern als Donor, für *N*-Glycan synthetisierende Glycosyltransferasen bereitgestellt. Die Glycosyltransferase Dolichylphosphate Mannose Synthase (DPMS) katalysiert mit der Bildung von Dolichylphosphate Mannose, einen dieser Dolichol-gebundene Zucker. Die Struktur und Funktion der DPMS von *Pyrococcus furiosus* wurde mit Hilfe von sowohl X-ray Kristallographie als auch biochemischen Methoden untersucht und ein neuer Assay basierend auf Proteoliposomen designt.

Im zweiten Teil der Doktorarbeit fokussiert sich auf bakterielle Glycosidasen welche Oligo- und Polysaccharide abbauen. Charakterisiert wurde die bakterielle Glycosidase von dem Aknebakterium *Cutibacterium acnes*. Das Enzym kann Teile des Wirt *N*-Glycans abbauen, wodurch Nährstoffe bereitgestellt werden oder die Erkennung des Immunsystems vermieden wird. Zusätzlich wird auch ein potentieller *N*-Glycan Abbauweg des Aknebakterium vorgeschlagen. In einer weiteren Studie wurde eine Glucosidase von einem Elchmagenbakterium charakterisiert. Das Enzym war in der Lage β-1,3-Glucane abzubauen, welches Anwendung in der industriellen Biomassebehandlung finden könnte.

Popular Science Abstract

Carbohydrates occur naturally in an enormous variety and participate in a number of life-sustaining biological processes in all living organisms. For organisms they are important as major energy source, but they are also essential for cellular recognition processes, as structural component in cell-walls and as constituents of nucleotides.

This thesis focuses two types of carbohydrates and on enzymes that are involved in their synthesis and break-down. The first type of carbohydrates are sugar chains that are connected to proteins, also known as glycosylation. These glycosylations are crucial for the function and shape of proteins, and therefore a highly complex pathway is needed to ensure correct protein glycosylation. If this complex process is disturbed it can cause congenital disorder of glycosylation (CDG), which is a monogenic disease manifesting as a metabolic disorder. Children being diagnosed can suffer from severe consequences, like developmental delay, failure to thrive, stroke-like symptoms, seizures and cerebellar dysfunction. With that in mind it is important to understand the molecular basics for this disease in order to be able to find suitable treatments. This thesis provides insights for two enzymes that are regulating protein glycosylation and the results rationalize their importance in this process.

Immunglobulin G, as part of the immune system, is an example for a glycoprotein and its connected sugars are important for recognition of e.g. pathogens. However, commensal and pathogenic bacteria have evolved methods to avoid the recognition by the immune system. The thesis will give an example how the acne bacterium possesses a set of enzymes that can break-down these glycosylations and therefore stays *incognito* for the human immune system.

The second type of carbohydrates which will be part of this thesis, is used by brown algae to store energy. Bacteria and fungi have specialized in breaking-down those complex carbohydrates from brown algae in order to unlock simple sugars that are used for their nutrition.

In the industry those microorganisms and their enzymes are utilized for the treatment of forest biomass, to produce biofuel as a sustainable alternative to fossil fuels.

Public defence of dissertation

This thesis will be defended on October 16th 2020 at 10:00 a.m. in Kollegiesalen, Brinnelvägen 8, Stockholm, Sweden.

Respondant:

Tom Reichenbach

School of Engineering Science in Chemistry, Biotechnology and Health
Department of Industrial Biotechnology
KTH Royal Institute of Technology, Stockholm, Sweden

Faculty opponent:

Professor David Drew

Department of Biochemistry and Biophysics
Stockholm University, Stockholm, Sweden

Evaluation committee:

Professor Eva Nordberg Karlsson

Department of Chemistry
Lund University, Lund, Sweden

Associate Professor Uwe Sauer

Department of Chemistry
Umea University, Umea, Sweden

Associate Professor Gisela Brändén

Department of Chemistry & Molecular Biology
University of Gothenburg, Gothenburg, Sweden

Chairperson:

Professor Antonius Van Maris

School of Engineering Science in Chemistry, Biotechnology and Health
Department of Industrial Biotechnology
KTH Royal Institute of Technology, Stockholm, Sweden

Respondant's main supervisor:

Professor Christina Divne

School of Engineering Science in Chemistry, Biotechnology and Health
Department of Industrial Biotechnology
KTH Royal Institute of Technology, Stockholm, Sweden

List of publications and manuscripts

This Thesis is based on the following publications or manuscripts, which are referred to in the text by their roman numerals:

PAPER I

Structural basis for dolichylphosphate mannose biosynthesis.
Gandini R, **Reichenbach T**, Tan TC, Divne C
Nature Communications **2017**, *8*,120.

PAPER II

Is *Pyrococcus furiosus* dolichylphosphate mannose synthase moonlighting as a biogenic flippase for dolichylphosphate mannose?
Reichenbach T*, Gandini R*, Kalyani D, **Divne C**
Manuscript in preparation.

PAPER III

A transmembrane crenarchaeal mannosyltransferase is Involved in *N*-glycan biosynthesis and displays an unexpected minimal cellulose-synthase-like fold.
Gandini R, **Reichenbach T**, Spaduit O, Tan TC, Kalyani D, Divne C
J Mol Biol **2020**, *432* (16), 4658-4672

PAPER IV

Structural and biochemical characterization of the *Cutibacterium acnes* exo-β-1,4-mannosidase that targets the *N*-glycan core of host glycoproteins.
Reichenbach T*, Kalyani D*, Gandini R, Svatström O, Aspeborg H, Divne C
PLoS One **2018**, 13(9)

PAPER V

A homodimeric bacterial exo-β-1,3-glucanase derived from moose rumen 1microbiome shows a structural framework similar to yeast exo-β-1,3-glucanases.
Kalyani D*, **Reichenbach T***, Gandini R, Svatström O, Aspeborg H, Divne C
Submitted

Contributions to appended papers and manuscripts

PAPER I

Contribution to mutagenesis experiments, and preparation of manuscript.

PAPER II

Contribution to cloning experiments, gene expression, protein purification, enzyme assays, design and execution of experiments, analysis of structural and biochemical data and preparation of manuscript.

PAPER III

Contribution to gene expression, protein purification, enzyme assays, design and execution of experiments, and preparation of manuscript.

PAPER IV

Contribution to gene expression, protein purification and crystallization, heavy-atom screening and phasing, synchrotron data collection, structure determination, model refinement, design and execution of experiments, analysis of structural and biochemical data and preparation of manuscript.

PAPER V

Contribution to gene expression, protein purification and crystallization, heavy-atom screening and phasing, synchrotron data collection, structure determination, model refinement, analysis of structural and biochemical data and preparation of manuscript.

List of abbreviations

3D	three-dimensional
α	alpha
β	beta
CBMs	carbohydrate-binding modules
CDG	congenital disorder of glycosylation
CO_2	carbon dioxide
C-term	carboxy terminus
Dol-P	dolichol-phosphate
Dol-PP	dolichol-diphosphate
Dol-P-Man	dolichol-phosphate mannose
Dol-P-Glc	dolichol-phosphate mannose
DPMS	dolichylphosphate mannose synthase
EC	Enzyme Commission
ER	endoplasmic reticulum
ERAD	ER-associated degradation pathway
Fc	fragment crystallizable
GDP-Man	guanosine diphosphate *α*-D-mannose
GHs	glycoside hydrolases
Glc	glucose
GlcNAc	*N*-acetylglucosamine
GTs	glycosyltransferases
HMG	high-mannose *N*-glycan
IF	interface
ITC	isothermal titration calorimetry
LLO	lipid-linked oligosaccharide
Man	mannose
UDP-GlcNAc	uridine diphosphate *N*-acetylglucosamine
OST	oligosaccharyltransferase
PC	phosphatidylcholine
PDCs	protein detergent complexes
PE	phosphatidylethanolamine
PG	phosphatidylglycerol
PNG	protein *N*-glycosylation
*p*NP-sugars	4-nitrophenyl *β*-D-glycopyranosides
PRLs	proteoliposomes
PS	phosphatidylserine
TIM	triosephosphate isomerase
TLC	thin-layer chromatography
TM	transmembrane
TMH	TM helices
ΔT_m	melting temperature

Table of contents

1 Introduction

1.1 Carbohydrates

Carbohydrates are biomolecules mainly consisting of carbon, hydrogen and oxygen atoms with a general formula of $C_m(H_2O)_n$, but only the simplest sugars, monosaccharides, fulfill this formula. More complex carbohydrates will be formed by including other elements like nitrogen, phosphate or sulfur, or by linking monosaccharides to form oligosaccharides, i.e. contain three to ten monosaccharides, and polysaccharides, i.e. consist of more than ten monosaccharides, through glycosidic bonds. Those more complex carbohydrates will show slight variations from the general formula. Primarily, monosaccharides are the major energy resource in all organisms. Glucose (Glc) as the most abundant one can be derived during the process of photosynthesis in plants, by which carbon dioxide and water will be reduced to Glc under the use of sunlight [1].

Monosaccharides can be considered building blocks for all occurring carbohydrates. They contain at least three carbon atoms and will be named accordingly to the number of carbon atoms, e.g. triose, tetrose, etc., whereby most frequently pentoses and hexoses are found in nature. Structure wise most monosaccharides typically form a cyclic structure as five or six membered rings (furanose or pyranose) next to their acyclic appearance. Under physiological conditions hexoses appear primarily in their cyclic form either as alpha (α) or a beta (β) anomer (Figure 1 A). The cyclic form exists in two major conformations: chair and boat form, where the chair conformation is energetic favorable (Figure 1 B) [1-3].

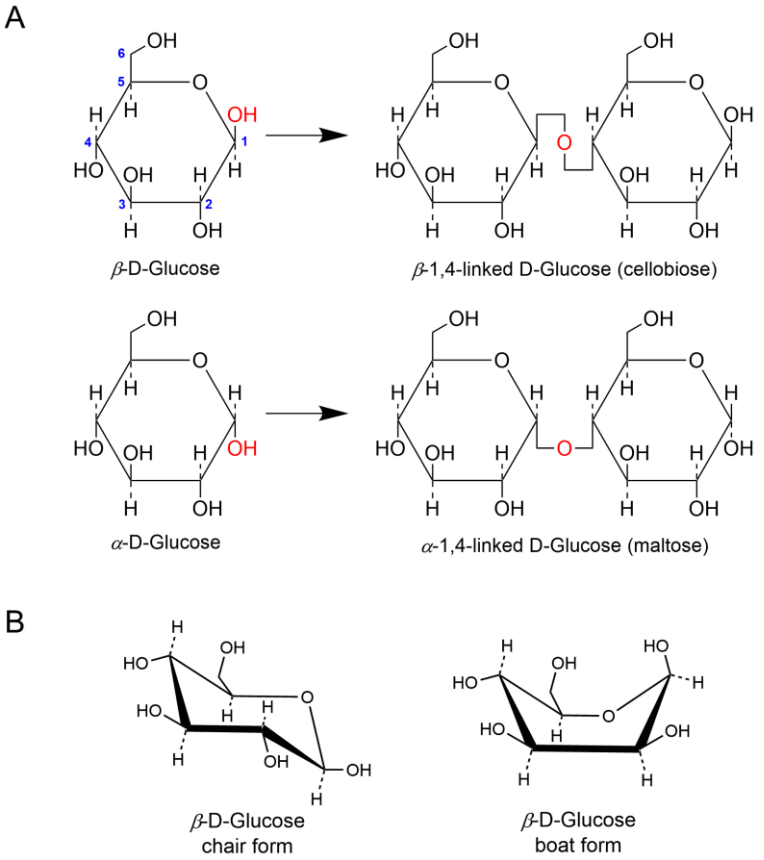


Figure 1: Selected structural properties of carbohydrates, A) Haworth projection of both Glc anomers (α and β) forming respective glycosidic bonds, B) pyranose conformation in chair and boat form of β-glucose [1, 4].

As mentioned above monosaccharides can be linked via glycosidic bonds to form oligo- and polysaccharides, and thereby links between different carbon atoms are formed, e.g. C1→C3, C1→C4, C1→C6. Based on the orientation of the hydroxyl group on the C1 atom it can either be an α or a β linkage (Figure 1 A). Consequently, a large variety of combination can be generated based on the choice of the monosaccharides and linkages, which will give rise to different structures and functions of the carbohydrate compounds. Additionally, the formation of different glycosidic bonds allows branching of the carbohydrates to increase diversity [4].

For example, oligosaccharides play a key role in cell-surface processes like cell-cell interactions and immune recognition. Antibodies, such as immunoglobulin G, are glycosylated proteins, i.e. a saccharide chain is linked to a protein, with highly branched oligosaccharides, and are essential for the adaptive immune response. They have a conserved glycosylation site in their fragment crystallizable (Fc) region, which defines the interaction with suitable Fc receptors [5].

Glycans, covalently bound oligosaccharides, are crucial features to increase the structural and functional diversity beyond the genetic V(D)J recombination. Those glycans represent two major types of protein glycosylation in eukaryotes. The glycosidic bonds are formed with either the side-chain nitrogen of an asparagine residue or the terminal oxygen of a serine or threonine residue to create *N*-glycans or respectively *O*-glycans [6, 7].

Longer carbohydrate chains (polysaccharides) are essential storage and structural components in all organisms. For example, the most abundant sugars found in nature are glycogen in animals and starch and cellulose in plants [8].

1.1.1 *N*-glycans

Nowadays it is known that more than 50% of all proteins are glycosylated. They can either be *N*-linked or *O*-linked, whereby *N*-linked glycans are attached to the side chain of an asparagine residue through an *N*-glycosidic bond to *N*-acetylglucosamine (GlcNAc) or *N*-acetylgalactosamine, while *O*-linked glycans feature sugars attached to the hydroxyl group of threonine or serine.

N-glycans show an enormous variety in oligosaccharide structure, but in eukaryotes, the core structure consists of five sugars defined by $(GlcNAc)_2$-Man(1→4)-Man(1→3)-Man(1→6). The eukaryotic *N*-glycans are classified as three types based on the antennae composition: *i*) high-mannose, less processed with only mannose (Man) units, *ii*) complex, processed *N*-glycan with two to four antennae, hybrid, which is partially processed show a mix of high-mannose and complex antenna (Figure 2) [9].

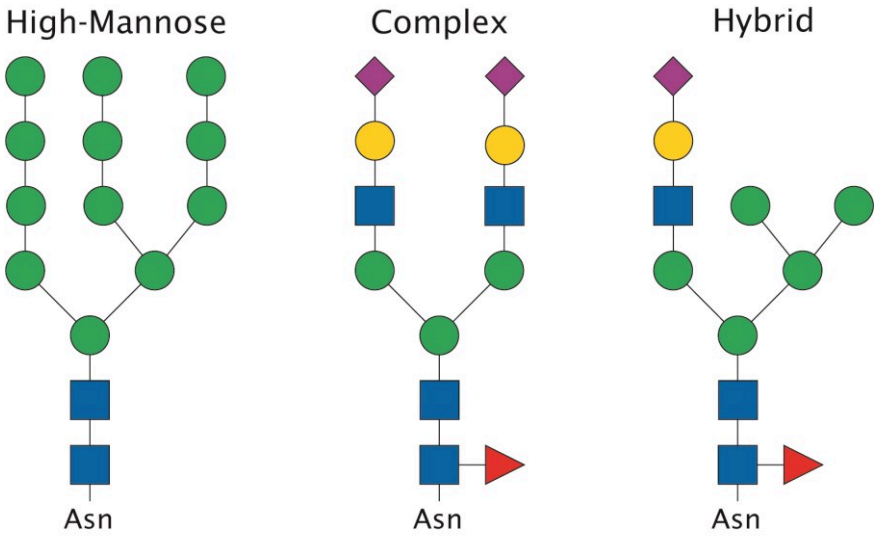


Figure 2: Eukaryotic *N*-glycan types in mature glycoproteins linked at Asn-X-Ser/Thr sequons: high-mannose, complex and hybrid *N*-glycan, illustrating the common core $(GlcNAc)_2(Man)_3$; green circle: mannose, yellow circle: galactose, blue square: *N*-acetylglucosamine, red triangle: fucose and purple diamond: sialic acid [9].

In other domains of life, prokaryotes and archaea, the glycans can differ in their structure. However, all domains have in common that the *N*-glycans are assembled on a lipid-linked intermediated, and not directly on the polypeptide chain. In eukaryotes and archaea a dolichylphosphate carrier is used as lipid carrier, while prokaryotes use undecaprenol phosphate as lipid carrier [10]. After maturation of the *N*-glycan, it will be transferred "en-bloc" to the polypeptide chain of the protein.

In eukaryotes, the synthesis of the lipid-linked intermediate will occur in the endoplasmic reticulum (ER). The first seven sugars will be transferred from a nucleoside diphosphate sugar to a dolichol-phosphate (Dol-P) facing the cytosolic site of the ER membrane. Transfer reactions will be catalyzed separately by specialized enzymes (glycosyltransferases). The first two *N*-acetylglucosamine will be donated from a uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) catalyzed by *N*-acetylglucosaminyl-transferase (Alg7 and Alg13/14). Subsequently, five mannose sugars are transferred from guanosine diphosphate α-D-mannose (GDP-Man) onto

the precursor by mannosyltransferases (Alg1, Alg2 and Alg11). This derived preliminary precursor, Dol-PP-$GlcNAc_2Man_5$, will be transported across the ER membrane from the cytosolic side to the luminal side, and further matured to a complete lipid-linked *N*-glycan precursor. Inside the ER lumen no nucleoside diphosphate sugars are available, since they cannot penetrate, and no suitable transporters are present in the ER membrane. Therefore, activated sugars are provided in form of dolichol phosphate mannose (Dol-P-Man) and dolichol phosphate glucose (Dol-P-Glc). Dol-P-Man is synthesized by the enzyme dolichyl phosphate mannose synthase (DPMS), which catalyzes the transfer of a mannose from GDP-Man to Dol-P, and similarly, dolichyl phosphate glucose synthase produces Dol-P-Glc from UDP-Glc and Dol-P. These dolichol-linked sugars are supposedly transported to the luminal side of the ER by yet unknown protein(s). Inside the ER lumen, the activated dolichyl sugars will act as sugar donors for mannosyltransferases (Alg3, Alg9, Alg12) and glucosyltransferases (Alg6, Alg8 and Alg10) which transfer the remaining mannose and glucose units to complete the lipid-linked precursor, Dol-PP-$(GlcNAc)_2$-Man_9Glc_3. The lipid-linked precursor will be transferred “en-bloc” onto the Asn-X-Ser/Thr sequons of nascent polypeptide chain via an oligosaccharyltransferase (OST) (Figure 3). [11, 12]

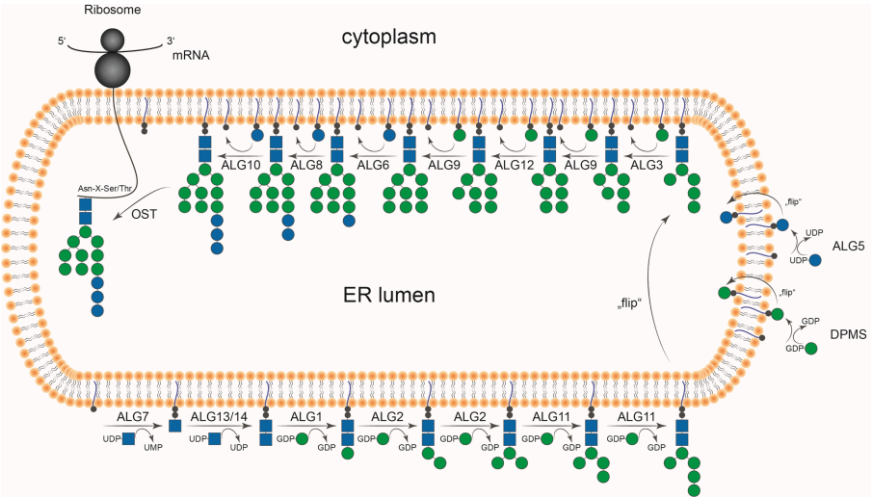


Figure 3: *N*-glycan biosynthesis in the eukaryotic ER. The picture was adapted from [12].

After the precursor is transferred to the polypeptide chain it will be further matured into either high-mannose, complex or a hybrid *N*-glycans (Figure 2). Therefore, the *N*-glycan precursor will be modified by different glycosidases and glycosyltransferases while the protein is transported via the secretory pathway. Firstly, the core will be trimmed by α-glucosidases I and II and an ER α-mannosidase to yield a high-mannose *N*-glycan structure. Trimming is crucial to access the folding status of the protein, since *N*-glycans can prevent aggregations and might interfere with protein folding. Proteins secreted from the ER will be subjected to ER quality control, where protein folding will be accessed by lectin chaperons, e.g. calnexin and calreticulin. If the protein is not in its native state it will undergo an additional folding cycle before being transported into the Golgi apparatus, or if it fails after multiple folding cycles, it will be degraded by the ER-associated degradation pathway (ERAD). These glycosylation reactions occurring in the ER are essential for proper protein folding, sorting, degradation and secretion [11, 13, 14]. The following maturation process in the Golgi apparatus will determine the destination and function of the glycoprotein. Thus, the *N*-glycans will be elongated and trimmed to fit the desired function of the protein [15].

Changes in the *N*-glycosylation machinery can cause severe problems for the organism. Since protein glycosylation is occurring co-translationally, it is most likely that multiple systems will be affected. For instance, congenital disorder of glycosylation (CDG) is a group of metabolic disorders caused by abnormal glycosylation of proteins and lipids. To date, more than 150 different CDG types have been, most of which originate from defects in *N*-glycosylation. Aberrant glycosylation patterns, that may affect the function of glycoproteins, can occur due to mutations in genes coding essential enzymes in the *N*-glycosylation pathway [16, 17]. Glycosylation patterns can also be used as a biomarkers for cancer diagnosis since the *N*-glycan maturation is altered in cancer patients, e.g. accumulation of high-mannose glycans, increased branching or increased core fucosylation [18].

1.1.2 Biomass

The mass of a living organisms includes sugars, fats and proteins and is referred to as biomass. Carbohydrates constitute a primary source when utilizing biomass in an industrial setting to yield valuable components, e.g. biofuel, biogas, food additives. Organisms capable of performing photosynthesis, e.g. plants, cyanobacteria or algae, generate sugars by reducing carbon dioxide (CO_2) and water with the use of sun light. This process fixes carbon in the form of carbohydrates, e.g. Glc. Monosaccharides like Glc are small and diffuse easily across cell membranes, hence storing them at high concentrations would increase the osmotic pressure on cells. To avoid this, monosaccharides are linked to polysaccharides. Depending on their function this can either be storage polysaccharides like starch in plants and glycogen in animal and microbial cells, or structural polysaccharides like cellulose in plants, glucan in fungi and laminarin in algae (Figure 4) [19, 20].

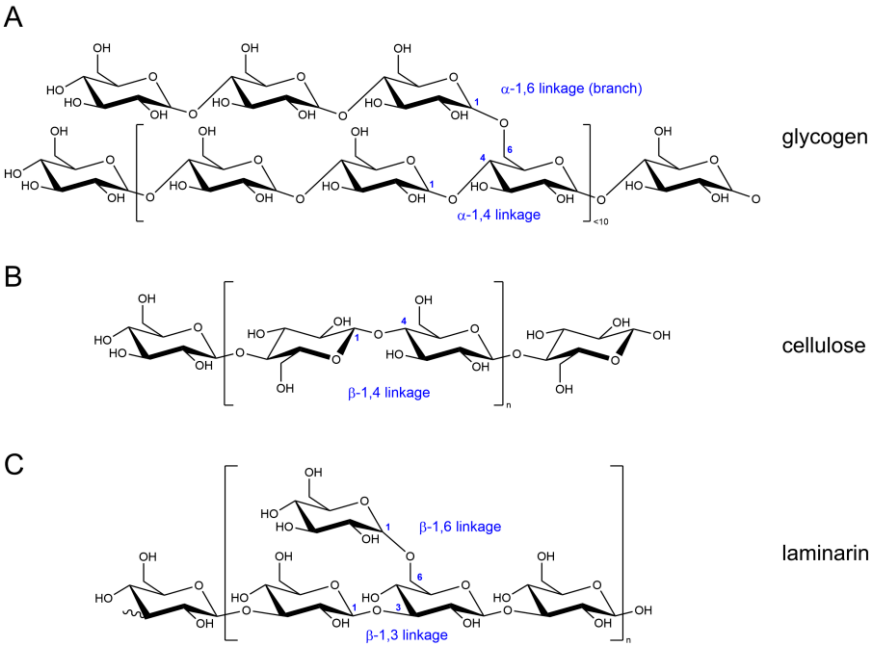


Figure 4: Storage polysaccharides represented by bacterial glycogen and structural polysaccharides by cellulose and laminarin [19-21].

Storage polysaccharides are used by the organisms as energy storage. To access them, they will be debranched and hydrolyzed into monosaccharides which will ultimately be oxidized into CO_2 and H_2O to provide energy. For example, bacterial glycogen is a largely branched water-soluble Glc homopolymer that accumulates in presence of carbon excess, where up to 12 Glc molecules will be combined to form an α-1,4-linked oligosaccharide (Figure 4A). This chain can adapt a helical structure and additionally connect with other chains to form α-1,6 linked branches [21].

Structural polysaccharides have the primary function to form the structural, supportive framework of cells. The lignocellulosic biomass of plants consists mainly of three components cellulose (40-60%), hemicellulose (20-40%) and lignin (10-25%) [22]. Cellulose is a *β*-1,4-linked glucose homopolymer consisting to up to several thousands of glucose units and tend to crystallize in the form of microfibrils (Figure 4B) [23]. Cellulose microfibrils are usually crosslinked by hemicelluloses, which are non-cellulosic polysaccharides. Those hemicelluloses can be grouped into classes based on the main sugar component of the backbone, e.g., xyloglucans, xylans, mannans, glucomannans and *β*-glucans. Hemicelluloses usually do not share any common structural features, and are usually difficult to extract without the use of chaotropic agents [19, 24]. An example of a hemicellulose that provides structural stability is laminarin, which is found as part of the cell wall in algae, especially brown algae. Laminarin is water-soluble *β*-1,3-1,6-glucan where the main chain consists of 1,3-*β*-linked glucose molecules and some *β*-1,6-linked branches (Figure 4C) [20].

1.2 Carbohydrate-active enzymes

An extensive set of glycosyltransferases is required to generate the remarkable diversity of carbohydrates, and an equally large number of enzymes (glycoside hydrolases) are required to degrade the carbohydrates. Carbohydrate-active enzymes have been classified in the CAZy database (www.cazy.org), and are therefore popularly referred to as CAZymes [25]. The CAZy database classifies enzymes based on their sequence, three-dimensional (3D) structure and function. First, the amino-acid sequences should have significant similarity to a biochemically characterized founding member of the family [26]. Second, the classification is modular, since enzymes are structures in domains which can either be catalytic or not. Therefore, a CAZyme can be assigned to different families if the constitutive modules belong to separate families. Third, the CAZy database only analyses protein sequence released by the GenBank to avoid misclassification [25]. Enzymes can be further classified based on different parameters that are of interest. For example, enzymes that catalyze a certain chemical reaction will be labeled with an Enzyme Commission (EC) number which is associated with a recommended name for the class of enzymes [27]. Enzymes use different mechanisms to catalyze chemical reactions, and in the case of glycosyltransferases (GTs) and glycoside hydrolases (GHs) either an inverting or a retaining mechanism is used depending on whether the product has the same (retained) or opposite (inverted) stereochemical configuration at the anomeric center [28]. Additionally, GHs can have different approaches to attack their substrates. Two principal modes are used, endo attack and exo attack: endo-acting GHs hydrolyze internal glycosidic bonds in polymeric carbohydrates, and exo-acting GHs cleave off monosaccharide units from either end of a carbohydrate chain. Another important feature used to classify CAZymes is the 3D structure [29, 30].

1.2.1 Glycosyltransferases

GTs can be defined as enzymes that catalyze the formation of glycosidic linkages by transferring a sugar from a donor substrate to an acceptor substrate. The donor substrate is an activated sugar with a phosphate as leaving group, most commonly a nucleotide sugar (e.g. GDP-Man), but it can also be a lipid phosphate (e.g. dolichylphosphate sugars), or unsubstituted phosphates. Nucleotide-dependent GTs are also known as Leloir enzymes[1]. The sugar can either be transferred to another sugar or to a lipid, protein, nucleic acid antibiotic or another small molecule [31, 32].

Considering the structural diversity of GTs, it is remarkable that all structures solved of Leloir enzymes belong to one of the two general folds GT-A and GT-B. The GT-A fold is reminiscent of two abutting Rossmann-like folds with an open-twisted β-sheet is surrounded by α-helices on both sides. Each Rossmann-like fold can be considered as one domain with a nucleotide-binding region and an acceptor-binding region. Typical for nucleotide-binding enzymes is the Asp-X-Asp (DXD) motif, where the sidechains of the two carboxylic acids coordinate a divalent cation and a ribose (Figure 5A). Similar to the GT-A fold, the GT-B fold also consists of two $\beta/\alpha/\beta$ Rossmann-like domains, where the two domains are loosely associated, and the active site is located at the cleft between both domains (Figure 5B). A third fold, GT-C, has also been described for large hydrophobic integral membrane proteins in the ER or plasma membrane using lipid phosphate-linked sugar donors (Figure 5C) [33].

1 In honor of Luis F. Leloir, who together with his team identified nucleotide sugars in 1948.

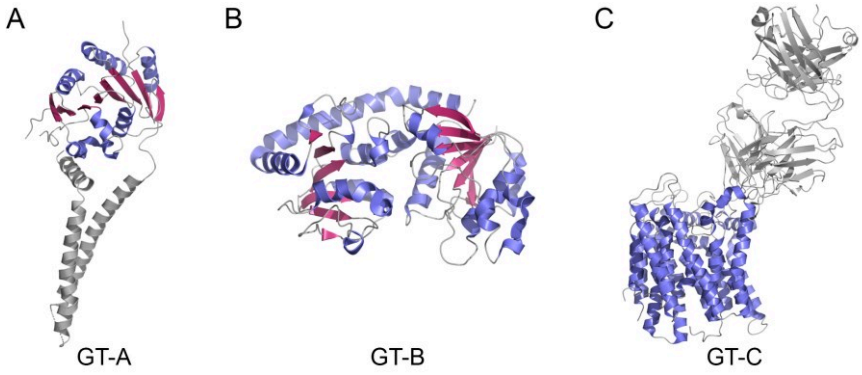


Figure 5: The three overall GT folds represented by A) GT-A, catalytic domain of GtrB from *Synechocystis* (PDB code 5EKP) [34], B) GT-B fold, by bacteriophage T4 β-glucosyltransferase (PDB code: 1JG7) [35], and C) TM domain ofALG6 from *Saccharomyces cerevisiae* (PDB code: 6SNH) [36]; relevant *β*-sheets and *α*-helices shown in red and blue respectively.

As mentioned above, GTs can be classified based on the stereochemical outcome of the reaction (retaining or inverting) [37], which is manifested in different anomeric configurations of the formed glycosidic bond with respect to the donor substrate. Taken together, GTs can be divided in four principal classes: inverting GT-A, retaining GT-A, inverting GT-B and retaining GT-B [31, 38, 39].

Inverting GTs catalyze a direct displacement *via* an S_N2-like reaction, i.e. a nucleophilic substitution-reaction mechanism involving two reacting species in the rate-limiting step. Here, the incoming acceptor will be deprotonated by an active-site residue, e.g. Asp or Glu, serving as a Brønsted base and forming an oxocarbenium-like ion transition state. This increases the nucleophility of the acceptor's hydroxyl group and facilitates the attack at the anomeric carbon of the sugar. The phosphate departs as leaving group supported either by a metal-ion (typically Mg^{2+} and Mn^{2+}) for GT-A or protonation from a Brønsted acid, any molecule that can donate a proton, for GT-B (Figure 6A) [40, 41].

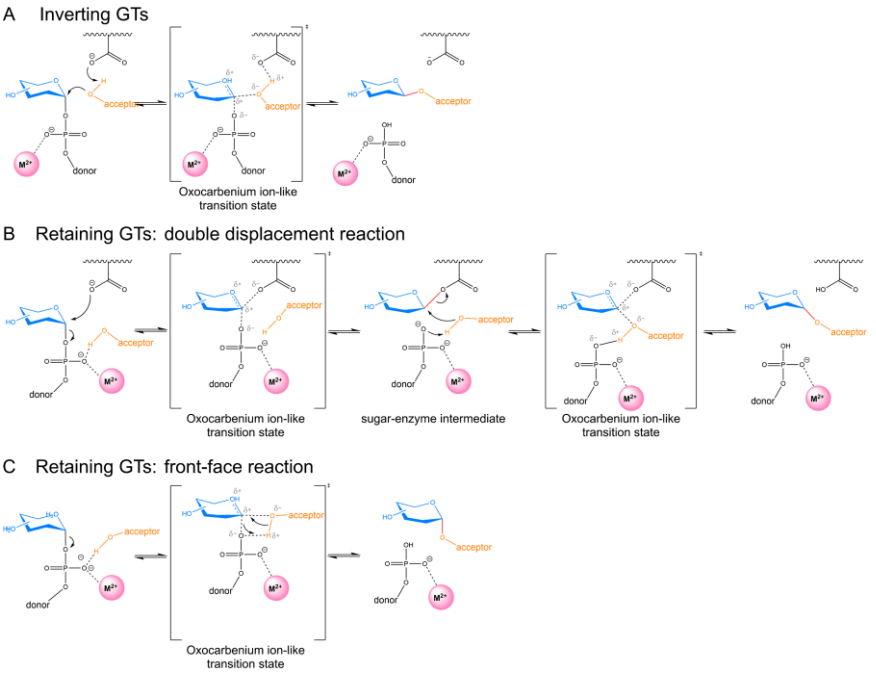


Figure 6: Reaction mechanism of GTs. A) Inverting GTs use a direct-displacement SN_2-like reaction mechanism, B) proposed double-displacement mechanism for retaining GTs involving a sugar-enzyme intermediate, and C) Alternative front-face mechanism for some retaining GTs [40-42].

Retaining GTs are considered to catalyze the reaction based on the classical Koshland double-displacement mechanism [43]. In the first displacement reaction, a carboxylic acid (Asp or Glu) performs a nucleophilic attack on the anomeric carbon of the donor sugar to generate a glycosyl-enzyme intermediate with an inverted configuration at the anomeric carbon. In the second step, the incoming acceptor is performing a nucleophilic attack at the anomeric carbon of the glycosyl-enzyme intermediate to form a new glycosidic bond with an overall retained configuration (Figure 6B). However, some GTs do not present a suitably positioned nucleophilic side chain in the active site, and a different mechanism, the front-face reaction, has been proposed for these GTs. Here, the nucleophilic attack of the acceptor at the anomeric carbon and the departure of leaving group take place on the same side. It is crucial for this mechanism that the leaving group acts as a catalytic base and

deprotonates the incoming nucleophile, i.e. a S_Ni-like mechanism (Figure 6C) [28, 42].

Although GTs do not show large structural and mechanistic diversity, they can generate a wide range of carbohydrates and glycoconjugate compounds. In recent times, there has been considerable interest in research on glycosyltransferases involved in protein glycosylation, as well as cellulase synthases [44].

1.2.1.1 Mannosyltransferases

The group of mannosyltransferases is involved in many different biological pathways that involve mannose. In this thesis, the focus lies on *N*-glycan biosynthesis. In eukaryotes, the mannosyl residues are added onto the lipid-linked oligosaccharide (LLO) intermediate by mannosyltransferases. These mannosyltransferases can be either active on the cytoplasmic side of the ER membrane (Alg1, Alg2 and Alg11) or inside the ER lumen (Alg3, Alg9 and Alg12). In the cytoplasm, the activated mannose donor is available as GDP-Man. Since there are no transporters for GDP-Man in the ER membrane, the activated mannose donor inside the ER lumen is provided as Dol-P-Man [45, 46].

1.2.1.2 Dolichylphosphate mannose synthase

In eukaryotes part of the glycosylation reactions take place on the luminal side of the ER, and as mentioned above, these reactions depend on activated sugar donors in the form of dolichyl carriers, i.e. Dol-P-Man and Dol-P-Glc. Dol-P-Man is synthesized by dolichylphosphate mannose synthase (DPMS, EC 2.4.1.83) that catalyzes the transfer of mannose from GDP-Man to Dol-P on the cytoplasmic face of the ER [47]. Following synthesis, Dol-P-Man is translocated to the luminal face where it can serve as mannosyl donor for non-Leloir GTs in the ensuing mannosylation reactions, including the biosynthesis of N-glycans, O-glycans [48], C-mannosylation [49] and glycosylphosphatidylinositol anchors [50-52].

DPMSs are transmembrane (TM) proteins with a catalytic extramembrane domain and a TM domain. These enzymes can be classified based on the complexity of the TM domain. Type-I is represented by DPMS from *Saccharomyces cerevisiae* [47], here the catalytic domain (DPM1) has one TM helix at the carboxy terminus (C-term). While type-II, like the human DPMS [52], is a complex of three separate gene products that forms a heterotrimer: DPM1, DPM2 and DPM3. DPM1 is the catalytic domain and does not contain a TM domain, whereas DPM2 and DPM3 are both small TM proteins of ~90 residues. While DPM1 retains catalytic activity in the absence of DPM2 and DPM3, activity is enhanced in their presence [52]. The membrane topology of type-III is similar to that of the fully assembled type-II complex, with the difference that the type-III DPMS is translated as a single polypeptide chain instead of three separate proteins. More details will be discussed later in this thesis.

1.2.2 Glycoside hydrolases

In nature, GHs catalyze hydrolysis of glycosidic bonds, although under specific conditions they can perform a synthetic reaction by reversing the hydrolysis reaction, i.e., transglycosylation. Some GHs display higher level of transglycosylation than hydrolysis activity, which can be explained by the active site being able to accommodate a non-water nucleophile which serves as an acceptor at the same time, e.g. trans sialidases [53].

GHs have been classified in the CAZy database based on the characteristics discussed above. Depending on the stereochemical outcome of the reaction GHs are catalyzing an inverting or retaining reaction mechanism. In both cases, GHs use a general acid/base and a nucleophile as catalytic residues [29]. Inverting GHs use a single displacement mechanism where the general base enables a water molecule to perform a nucleophilic attack at the anomeric carbon of the sugar and the final product is released (Figure 7A). The retaining mechanism involves instead two steps, glycosylation and deglycosylation. During the first step (glycosylation) the residue acting as acid protonates the glycosidic bond, which promotes departure of the leaving group, and simultaneously the nucleophilic residue attacks at the anomeric carbon of the sugar to form a sugar-enzyme intermediate. In some cases, the carbonyl group of an *N*-acetyl sugar at the site of bond cleavage (-1 subsite) can act as nucleophile instead of the active site residue, which is known as a substrate-assisted mechanism. In the following the second step (deglycosylation) a water molecule is deprotonated by the acid/base residue, which facilitates attack on the anomeric carbon and the final hydrolyzed product is released (Figure 7B) [54].

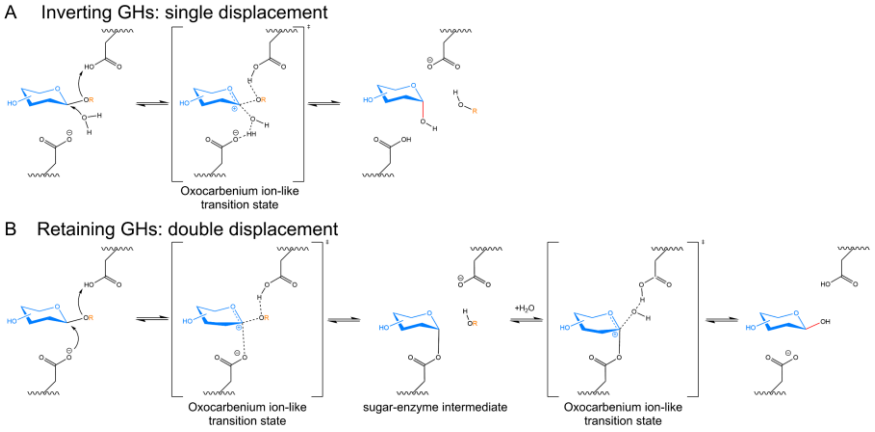


Figure 7: Reaction mechanisms for glycoside hydrolysases. A) inverting GHs via single-displacement mechanism and B) retaining GHs utilizing a double-displacement mechanism [54].

The mode of attack (exo or endo) determines the structural design of the active site. For true exo-acting enzymes the active site is shaped as a pocket, while a cleft is used by true endo-acting enzymes. GHs that act processively, i.e. does not release the substrate between cleavage events, often have a tunnel-shaped active site [30].

1.2.2.1 Mannan-degrading enzymes

In higher plants, the polymer of mannose, mannan, is an essential component of hemicellulose where it acts as a structural element, but also serves as a non-starch carbohydrate reservoir in the endosperm walls. Structurally, it consists of a β-1,4-linked backbone, occasionally with α-1,6-linked side branches. Mannans can be divided in four groups: linear mannan, glucomannan, galactomannan and galactoglucomannan. Enzymes that degrade mannose-containing polysaccharides and glycoconjugates are endo-1,4-β-mannosidases, also known as β-mannanases (EC 3.2.1.78) and β-mannosidases (EC 3.2.1.25) [55-57].

According to their sequence similarity, β-mannanases have been classified into CAZy families GH5, GH26 and GH113. They are endo-acting enzymes that catalyze the random hydrolysis of β-1,4-glycosidic bonds in mannans to yield primarily mannobiose and mannostriose as end products. Despite belonging to different families, they are all retaining enzymes that use a double-displacement mechanism. The

double-displacement mechanism allows some enzymes, especially fungal and archaeal β-mannanases in families GH5 and GH113, to generate new glycosides or oligosaccharide compounds through transglycosylation activity [58]. Considering the similarities in their tertiary structure, all β-mannanases have been classified into the superfamily referred to as clan A (GH-A). The 3D structure shows a catalytic domain with a $(\beta/\alpha)_8$ triosephosphate isomerase (TIM)-barrel fold. Additionally, some appear to be modular enzymes with accessory domains appended, e.g., one or a several carbohydrate-binding modules (CBMs), which facilitates binding to the substrate to significantly increase degradation efficiency [57, 59].

The β-mannosidases catalyze the hydrolysis of glycosidic bonds from the non-reducing end of mannan, i.e. in an exo-acting mode, to produce mannose as main end product. Based on sequence similarity, they belong to CAZy families GH1, GH2 and GH5, but exhibit the same characteristics regarding mechanism and structural classification as β-mannanases [55, 60].

1.2.2.2 Glucan-degrading enzymes

Some of the most widespread polysaccharides are glucans. They are glucose polymers used either as structural component in the cell wall or as energy reservoir. The structural type can be divided in three different types: α-glucan, β-glucan and mixed α,β-glucans. Regarding β-glucans, some of the most common representatives are β-1,4-glucans in higher plants (cellulose), β-1,3-glucans in fungi (pachyman) and β-1,3-glucans in algae (laminarin) [61]. Degradation of β-glucans involves mainly endo-β-glucanases and exo-β-glucosidases, and based on the linkage that they hydrolyze, they are classified differently.

For the degradation of cellulose, three principal types of enzymes are needed. Cellobiohydrolases catalyze the hydrolysis of cellulose in an exo-mode of action working from either the reducing (EC 3.2.1.176) or non-reducing end (EC 3.2.1.91) to release cellooligosaccharides. In contrast, endoglucanases (EC 3.2.1.4) hydrolyze internal glycosidic bonds in cellulose in a typically random and non-processive manner, although exceptions exist. β-glucosidases (EC 3.2.1.21) degrade cello-oligosaccharides from the non-reducing end to release glucose [62, 63].

For degrading β-1,3-glucans, organisms use endo-β-1,3-glucanases (EC 3.2.1.39) that catalyze internal cleavage in the β-linked polysaccharide chain to release new free chain ends, and β-1,3-glucosidases (EC 3.2.1.58) that release glucose from the non-reducing end of the β-glucan chain [63]. β-glucosidases are important for the metabolization of glycolipids and dietary glucosides, as well as their important role in signaling pathways. In plants, β-glucosidases are involved in a range of essential functions, for instance cell-wall catabolism, defense mechanisms and secondary metabolism. Microorganisms use β-glucosidases in more endogenous functions, such as production of glucose from plant biomass and cell-wall penetration for pathogenic or symbiotic relationships [64].

At the structural level, most β-glucosidases (GH1, GH5, GH30) belong to the GH-A clan with a catalytic domain folding as a $(\beta/\alpha)_8$ TIM-barrel. There are also members of the GH3 family, which have two domains that contribute to the active site, a TIM barrel and a $(\alpha/\beta)_6$-sandwich fold. Members of family GH8 adapt a different overall structure with a $(\alpha/\alpha)_6$ barrel fold [64].

1.3 Applications

After introducing the relevant enzymes and their preferred substrates and products it is worth considering their contribution to sustainable development by providing value for society as well as industrial and medical applied research. While the present thesis offers new knowledge of how the enzymes work, it also provides a basis for their possible use as biocatalysts in relevant applications.

The GTs of the *N*-glycan biosynthesis pathways can be recombinantly produced in prokaryotic (e.g. *Escherichia coli*) or eukaryotic (e.g. insect cells and yeast) expression systems. While the prokaryotic system is lacking some the protein-modification system, expression libraries for eukaryotic system are created to allow efficient production. The recombinantly produced enzymes are used to create *N*-glycan libraries, which later can be used to screening for glycan-binding proteins by high-throughput assays [65, 66]. It is equally important to gain understanding of the underlying principles that cause diseases, such as CDG. Current therapeutic strategies are focusing on treating the symptoms rather than the source. Enzyme therapy is one such approach, where the deficient enzyme is supplemented. However, obstacles include the blood-brain barrier breach and reaching the correct side of action. A more promising alternative could be gene therapy, which aims to restore the wild-type sequence of the mutated gene by using an Adeno-associated virus or CRISPR/Cas9 to replace nucleotides [17].

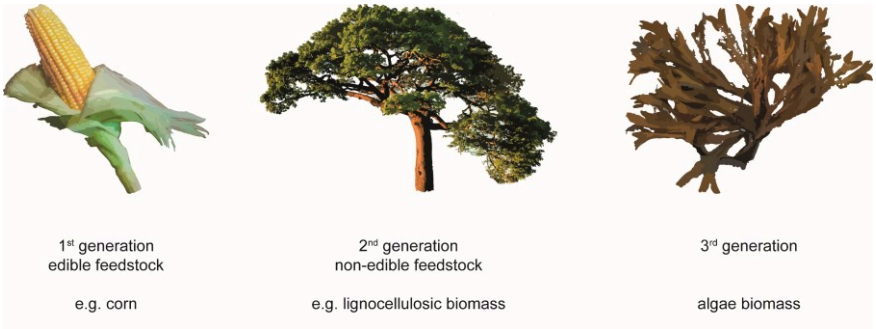


Figure 8: Overview of biofuel generations based on the utilized biomass as feedstock.

GHs are very popular in industrial biotechnological application where a major focus is on biofuel production. Several types of feedstock have been investigated for biofuel production (Figure 8). In the first-generation biofuel, edible feedstock like corn and sugarcane were used as starting material that was saccharificated and fermented to yield bioethanol. To overcome the competition between food and fuel, other resources were explored, which led to the second-generation biofuel where more recalcitrant lignocellulosic biomass like forest waste is utilized. However, this introduces new problems since lignocellulosic biomass requires extensive pretreatment. A promising alternative is the third-generation of biofuel, i.e., algae as renewable feedstock. Algal biomass could solve issues related to the first- and second-generation biofuels since it is not competing with other important, and limited, resources such as agricultural cropland needed for food or fresh water for drinking. Additionally, algae can utilize wastewater and CO_2 in their cultivation and no extensive pretreatment is needed [67-70].

2 Current studies

The aim of this thesis is providing new biochemical and structural data on enzymes that are active on carbohydrates. In the first part, glycosyltransferases that are involved in *N*-glycan biosynthesis will be discussed. Insights into their mechanisms and how they relate to diseases like CDG will be offered, as well as adding new knowledge to our current understanding of *N*-glycan biosynthesis. The second part of the thesis will present new glycoside hydrolases that are involved in the degradation of oligo- and polysaccharides to serve as nutrition for the microorganism, or even as a way to escape the immune system of the affected host. The use of the enzymes in possible biotechnological application will also be covered, primarily in the context of polysaccharide degradation (e.g. laminarin) for a sustainable production of biofuel from brown algae.

2.1 Biosynthesis and translocation of dolichylphosphate mannose (PAPER I & II)

In all domains of life, protein glycosylation is essential for protein folding, sorting, stability and function [71-73]. The biogenesis of nascent proteins and membranes occurs in biogenic membranes, i.e. the self-synthesizing ER membrane in eukaryotes and the cytoplasm cell membrane in archaea [74, 75]. This posttranslational modification requires activated mannose, but due to the lack of a suitable mannose transporter it cannot cross the ER membrane, and therefore mannose is provided in the form of the glycolipid Dol-P-Man [76]. At the inner face of the plasma membrane of archaea, or at the cytosolic face of the ER membrane in eukaryotes, the integral membrane protein dolichylphosphate mannose synthase (DPMS; EC 2.4.1.83) catalyzes the transfer of mannose from GDP-Man to Dol-P [47], after which the Dol-P-Man is translocated across the ER membrane to the luminal side where it becomes available as mannose donor [77].

Studies have shown that glycerophospholipids, like phosphatidylserine, are distributed asymmetrically between the outer and inner leaflets of the plasma membrane in eukaryotes. This asymmetry is mainly generated by enzyme-assisted translocation of lipids by flippases, floppases and scramblases [78]. Phospholipid flippases are ATP-dependent enzymes, P4-ATPases, that transport phospholipids from the exoplasmic to the cytoplasmic face of a membrane, while floppases facilitate the translocation in the opposite direction. Phospholipid scramblases translocate phospholipids bidirectional and are energy-independent and instead rely on the presence of Ca^{2+} ions [78, 79].

In PAPER I, the catalytic mechanism of a DPMS from the euryarchaeon *Pyrococcus furiosus* (*Pf*DPMS) was investigated by determining high-resolution crystal structures. Structures of *Pf*DPMS could be captured in complex with donor substrate and metal cofactor, and also with the reaction product Dol-P-Man. From these different complexes, distinct structural states could be identified rationalized several important structure-function relationships.

In PAPER II, the structure of *Pf*DPMS was crystallized from a reaction mixture incubated at low temperature (25°C), which allowed capture of *Pf*DPMS in complex with the reaction product (Dol55-*P*-Man) in a flipped orientation. The complex revealed a polar binding site inside the TM

domain, which suggested a possible mechanism for translocating Dol-P-Man across the archaeal plasma membrane.

2.1.1 Structural features of *Pf*DPMS

Before this study, two classes of DPMSs had been proposed based on the membrane topology (type I and II, see 1.2.1.2) [80]. However, since the structure of the TM domain of *Pf*DPMS was different, a third class was proposed (type III). The catalytic domain of *Pf*DPMS belongs to CAZy family GT2 including inverting GTs [25]. Like other GT2 enzymes, *Pf*DPMS has a catalytic domain that adopts a GT-A fold with a conserved DXD sequence motif [81]. The N-terminal catalytic domain is covalently bound to a C-terminal TM domain consisting of four TM helices (TMH), which are arranged as two TMH dimers (TMD1 and TMD2). TMD1 is roughly orthogonal to the membrane face and spans across the membrane. TMD2 is rotated by approximately 60° with respect to TMD1 making the TM domains resemble two "walking legs". The catalytic domain has two amphipathic interface (IF) helices (IFH1 and IFH2) that interact with the inner face of the cytoplasmic membrane.

The C-terminal ends of the parallel β-strands (β1-β4) form a pocket where the guanosine moiety of GDP interacts with the β-α loop regions. The extended canonical DXD motif, ^{89}DADLQH^{94}, coordinates the diphosphate group in GDP *via* the metal cation and forms a binding site in β4-α4 loop region. The active site is gated by a flexible loop (acceptor loop or A-loop), which is in a closed conformation when GDP or GDP-Man is bound in the active site. The A-loop and the residue Phe177 (whose conformation is secured by IFH2) are acting as a “front door” and “back door”, respectively, to control the donor's access to the active site (Figure 9).

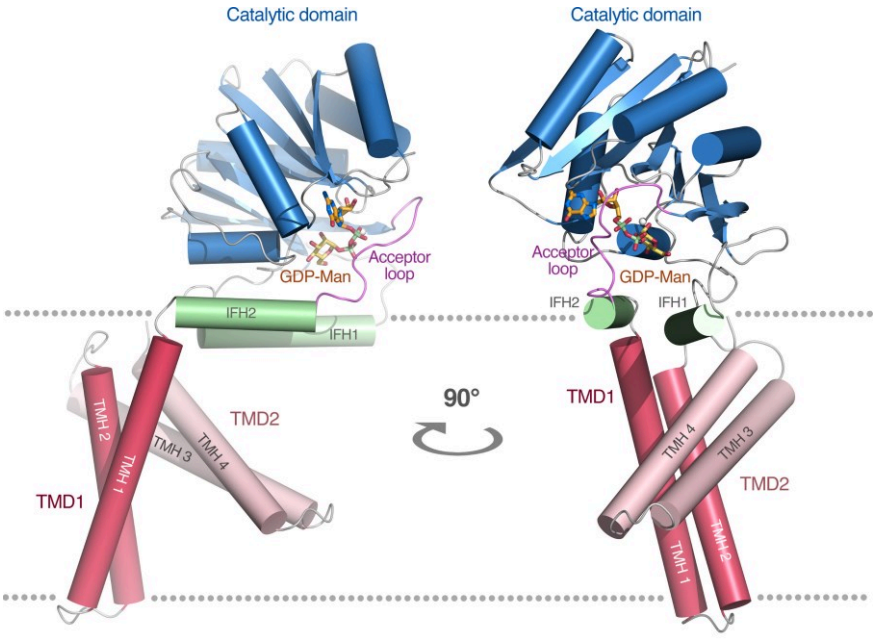


Figure 9: Overall structure of *Pf*DPMS in complex with GDP-Man and Mn^{2+}. The acceptor loop is in a closed conformation (purple coil).

By crystallizing a reaction mixture incubated at 60°C, a post-catalysis active-site state with bound Dol-P-Man was captured. The dolichyl chain is located between IFH1 and IFH2 such that it can interact with hydrophobic residues in both IF helices, and the acceptor's phosphate group is coordinated by a serine (Ser135) [82]. Interaction of the dolichyl chain with the IF helices causes IFH2 to undergo a conformational change and shift away from IFH1, which leads to opening of the A-loop and active site. In this state the “back door” is open, which allows the hydrolyzed donor to leave and allows binding of a new GDP-Man molecule. In the post-catalysis complex, GDP remains bound to the enzyme in the presence of Dol-P-Man but since the structure was produced from a reaction mixture it is expected that the experimental data contain information from several different structural states (Figure 10).

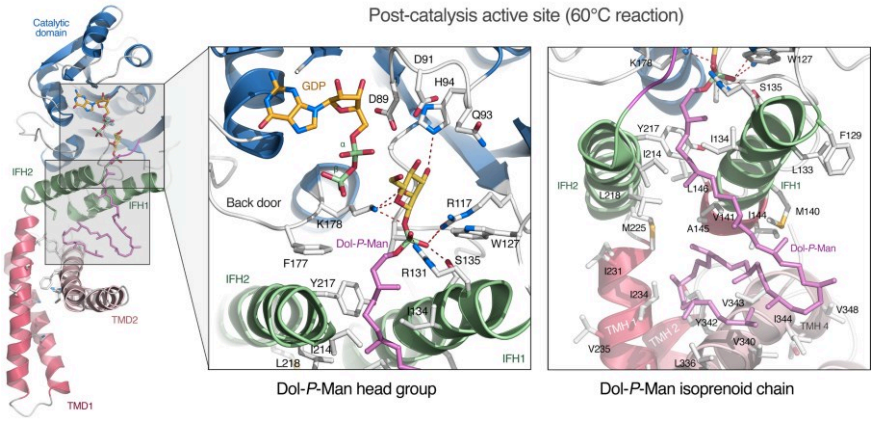


Figure 10: Structure of the *Pf*DPMS in complex with GDP and Dol55-P-Man. **Left inset (active site):** the active site of wild-type *Pf*DPMS binds the Dol-P-Man head group while the metal ion has departed and the GDP diphosphate group moved towards the "back door" to accommodate the Dol55-P-Man. **Right inset (TM domain):** the Dol-P-Man isoprenoid chain rests on TMH4 in the TM domain. The acceptor loop and Phe177 are in "open" conformation.

2.1.2 Insights in catalytic mechanism

Outline of the catalytic mechanism for biosynthesis of Dol-P-Man could be based on the obtained crystal structures (Figure 11).

First, before the acceptor enters, GDP-Man and the metal ion bind to the active site. The back door (Phe177) and front door (A-loop) are closed, preventing donor or metal to leave the active site. In the second step, IFH1 and IFH2 bind the first two isoprene units of Dol55-*P* and the phosphate group is coordinated by active site residue Ser135. This will trigger a conformational change of the A-loop into an open state, and Phe177 is opening the back door. When both donor and acceptor substrate are in place, the activated phosphate group of Dol55-P will perform a nucleophile attack at the α-mannosyl C1 [83].

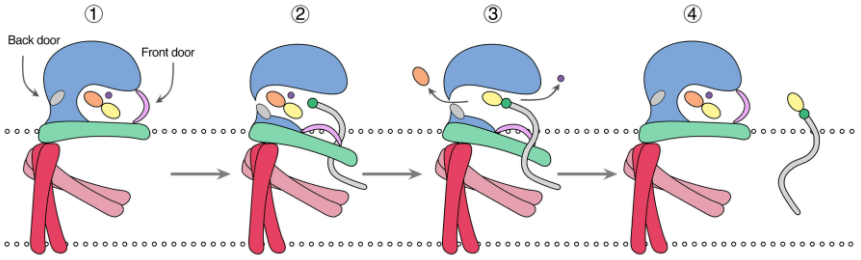


Figure 11: Proposed catalytic mechanism for biosynthesis of Dol-P-Man in schematic representation of the conformational changes: **Step 1:** GDP-Man and metal ion binding in the active site. **Step 2:** Dol-P acceptor binding leads to IFH2 movement and opens "back door" and the acceptor loop. **Step 3:** glycosyl transfer following release of GDP and metal ion. **Step 4:** binding of new GDP-Man and metal ion, release of Dol-P-Man, closure of back door and acceptor loop. Color scheme: Catalytic domain, blue; IFH2, green; TM domain, red; acceptor loop, pink; GDP-Man, orange and yellow ovals; metal ion, purple circle; Phe177, gray oval; Dol-P isoprenoid chain, gray; and Dol-P phosphate group, green circle.

In the last step, a new bond is formed, which with respect to the substrate has an inverted configuration. Once the Dol55-P-Man product has formed, GDP can leave through the back door while the metal ion escapes through the front door creating space for the Dol55-*P*-Man, which eventually will be released. The characterization of a truncation variant of *Pf*DPMS where the TM domain (residues 230–352) was deleted, revealed that the TM domain is not required for catalytic activity, which raised the question of the function of the TM domain.

2.1.3 Indication of Dol-P-Man translocation

The elusive role of the TM domain needed further investigation. From the mammalian DPM1/DPM2/DPM3 complex, it was known that DPM2 enhances the activity of DPM1 and DPM3 is needed to increase binding of DPM1 to the ER membrane. Nevertheless, the exact function of DPM2 could not be determined [52]. Early studies on *Saccharomyces cerevisiae* DPMS suggested that the enzyme has the ability to both synthesize and translocate Dol-*P*-Man, but later studies could not confirm this observation [84]. Furthermore, researchers have been trying to identify the protein responsible for Dol-*P*-Man translocation across the ER membrane but could only establish that the mechanism is protein-assisted and energy-independent [50, 77].

Based on 3D structure of *Pf*DPMS obtained by crystallizing a reaction mixture incubated at low temperature (25°C), the experimental data revealed a mixture of two new reaction states. Firstly, there are traces of GDP-Man bound in the active site in a productive orientation, i.e. exposing the C1 *α*-mannosyl for a nucleophilic attack from the Dol-*P* phosphate group, which is coordinated in the diphosphate-binding site. With both donor and acceptor bound, the A-loop was present in a disordered open state. Secondly, there were traces of a flipped Dol-*P*-Man molecule with the mannosyl group bound in a polar pocket between TMD1 and TMD2, indicating that the TM domain has a binding site for the mannosyl head group. The ability of one protein to perform at least two distinct biological functions is known as moonlighting [85].

2.1.4 Designing a translocation assay

Given the interesting structural results it was exciting to set out to test the hypothesis whether *Pf*DPMS is translocating its own product. However, to obtain biochemical evidence for Dol-P-Man translocation would prove to be extremely challenging. First, a system that can separate "flipped" and "non-flipped" Dol-P-Man was required, and therefore the use of proteoliposomes (PRLs) was logical. *Pf*DPMS incorporated in the lipid bilayer of PRLs can be supplied with donor substrate and metal ion from the surrounding aqueous phase, and any flipped Dol-P-Man would display its mannosyl headgroup to the interior of the PRLs. To prove the *anti*-hypothesis, i.e. that Dol-P-Man is *not* translocated by *Pf*DPMS, required that the mannosyl group could be labeled from the outside under conditions that would not disrupt the PRLs. The sensitivity of PRLs has been proven to be problematic. They are unstable under harsher conditions, e.g. elevated temperature, addition of surfactants, e.g. cholate, and acids [86]. Therefore, an appropriate method to label Dol-P-Man was to phosphorylate the lipid-bound mannosyl at the C6 position using hexokinase. This resulted in a coupled glycosyltransferase-hexokinase assay. *Pf*DPMS incorporated into phosphatidylcholine (PC) PRLs together with Dol-P was first allowed to catalyze the GT reaction (Figure 12, I), followed by the hexokinase reaction (Figure 12, II). After completing both reactions, the polar and apolar phases were separated by chloroform extraction (Figure 12, III). The apolar phase, which contains PC lipids, Dol-P-Man and Dol-P-Man6P, was subjected to mild acid hydrolysis to release Man or Man6P (Figure 12, IV).

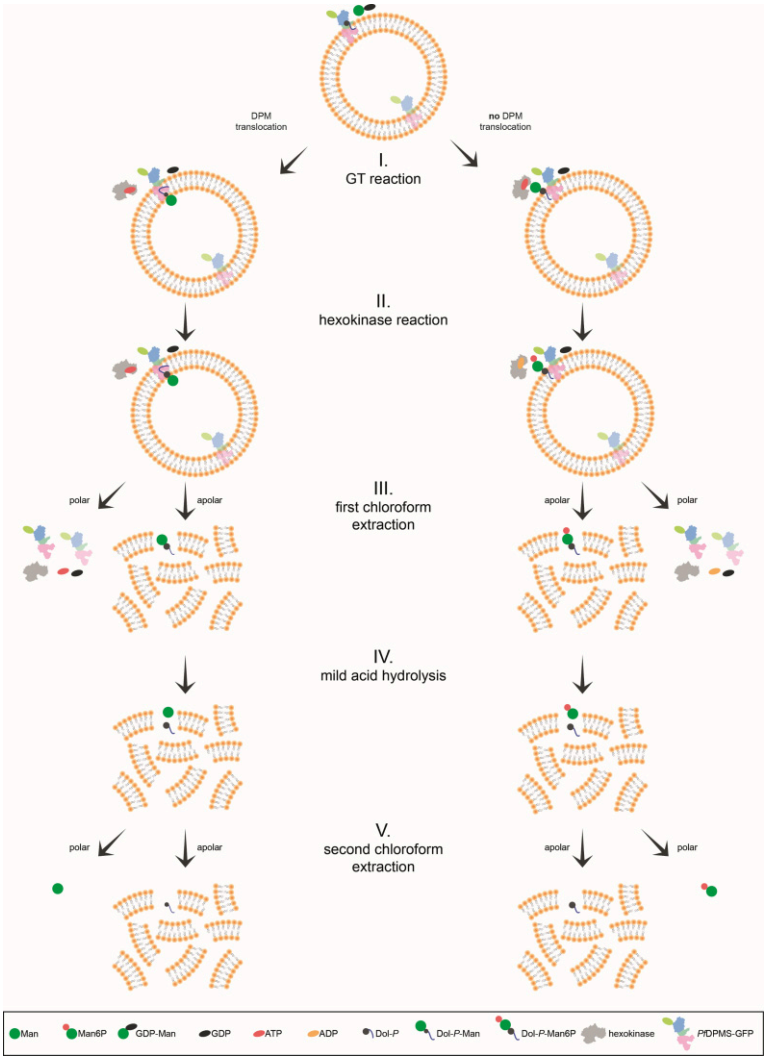


Figure 12: Coupled glycosyltransferase-hexokinase PRL-based assay designed to identify translocation activity of *Pf*DPMS. **Step I)** GT reaction: *Pf*DPMS-GFP catalyzes the mannosyl transfer. If translocation occurs, the Dol-P-Man product is flipped across the bilayer to expose its mannosyl headgroup to the PRL interior. **Step II)** Hexokinase reaction: mannose exposed to the outside of the PRLs will be phosphorylated by hexokinase. **Step III)** First chloroform extraction to retrieve apolar phase containing Dol-P-Man and/or Dol-P-Man6P (also retains detergent and PC). **Step IV)** Mild acid hydrolysis: hydrochloric acid is used to release Man or Man6P from Dol-P-Man or Dol-P-Man6P. **Step V)** Second chloroform extraction to capture released Man or Man6P.

In the final step the released Man or Man6P were captured in the polar phase after a second chloroform extraction (Figure 12, V).

The samples retrieved after the second extraction were analyzed by thin-layer chromatography, but no phosphorylation of Dol-P-Man could be detected. Provided that Dol-P-Man bound in PRLs is as accessible to hexokinase as it is in protein detergent complexes (PDCs), it appears that the synthesized glycolipid has been translocated across the liposome membrane to the inner face where it is inaccessible for hexokinase.

2.1.5 Structural comparison with known enzymes

Of the GTs that use lipids as substrate only two crystal structures exist, both of which only have nucleotide bound. The first is *Synechocystis* GtrB (PDB code: 5EKP), also a GT2 member, that catalyzes glucose transfer from UDP-Glc to undecaprenolphosphate (Und-P) [87]. GtrB shows a similar topology for the catalytic domain as *Pf*DPMS, but the IF helices and the TM domain are differently orientated relative to the catalytic domain, which enable formation of a homotetramer. The second is the aminoarabinose transferase ArnT from *Cupriavidus metallidurans* (PDB code: 5F15)[88], which in contrast to GtrB and *Pf*DPMS is a member of the GT83 family, and features the GT-C fold. Its catalytic domain is associated with three IF helices connected to a TM domain containing 13 TM helices. The IF helices seem to play an equally important role in ArnT as in *Pf*DPMS by binding the unloaded donor (Und-P) and triggering conformational changes that affect the A-loop. Based on the structural similarities of the catalytic domain, it can be assumed the *Pf*DPMS and GtrB share a similar catalytic mechanism, while ArnT seems to use a different mechanism [87, 88].

Comparing the TM domain of *Pf*DPMS with relevant ATP-independent membrane transporters leads to the GtrA family with members that participate in the glycosylation of cell-surface glycoconjugates, and contain 2 to 4 TM helices [89]. Four members (*Sfx*GtrA from *Shigella flexneri* bacteriophage X [89], ArnE/ArnF from *E. coli* [90], GtcA from *Listeria monocytogenes* serovar 1/a and *Bacillus subtilis* 168 [91], and Rv3789 from *Mycobacterium tuberculosis* [92]) of this family show low sequence identity but a membrane topology that is similar to that of *Pf*DPMS.

2.2 Crenarchaeal mannosyltransferase involved in *N*-glycan biosynthesis (PAPER III)

N-glycans are assembled on membrane-bound lipid carrier in all three domains of life [71-73]. The eukaryotic precursors, LLO, are built on the lipid carrier dolichyldiphosphate (Dol-PP), while archaea use Dol-P or Dol-PP, and bacteria use Und-P (also known as bactoprenolphosphate). In eukaryotes, the heptasaccharide core $(GlcNAc)_2$-$(Man)_5$ is generated at the cytoplasmic side of the ER membrane, and after translocation, further glycosylated in the lumal side of the ER to yield a tetradecasaccharide, $(GlcNAc)_2$-$(Man)_9$-$(Glc)_3$. The enzyme OST transfers the LLO "en-bloc" to a nascent protein, and further maturation steps will follow in the Golgi apparatus. In archaea, LLO synthesis is taking place at the cytoplasmic face of the plasma membrane but might be further modified on the extracellular side before being transferred to a protein [93].

Archaea do not share a common *N*-glycan core structure as in eukaryotes, and display a much wider variety of glycan structures [94, 95]. Interestingly, the biantennary high-mannose *N*-glycan (HMG) of the crenarchaeon *Pyrobaculum calidifontis* shows high similarity to eukaryotic HMG with respect to the sugar composition, including a modified chitobiose-like core. More specifically, the HMG of *P. calidifontis* comprises 11 sugars units, $Glc(NAc)_2$-$GlcA(NAc)_2$-$(Man)_9$, where the first mannose is α-1,4-linked to the chitobiose-like core, followed by branching into α-1,3- and α-1,6-mannosyl-linked antennae. On the α-1,6-mannosyl branch, follow four α-1,2-linked mannose units and a terminal α-1,3-linked mannose. The α-1,3-mannosyl branch has only one α-1,2-linked mannose added (Figure 13) [96]. The *P. calidifontis* is of particular interest since the crenarchaeota belong to the TACK superphylum [97], which is considered a possible origin of eukaryotes [98].

In PAPER III, the structure of the membrane glycosyltransferase *Pc*ManGT from *P. calidifontis* is discussed in relation to its potential role in *N*-glycan biosynthesis. The crystal structure of *Pc*ManGT was determined in complex with the donor substrate GDP-Man and metal ion. The results offer the first structural and biochemical characterization of a GT of TACK origin.

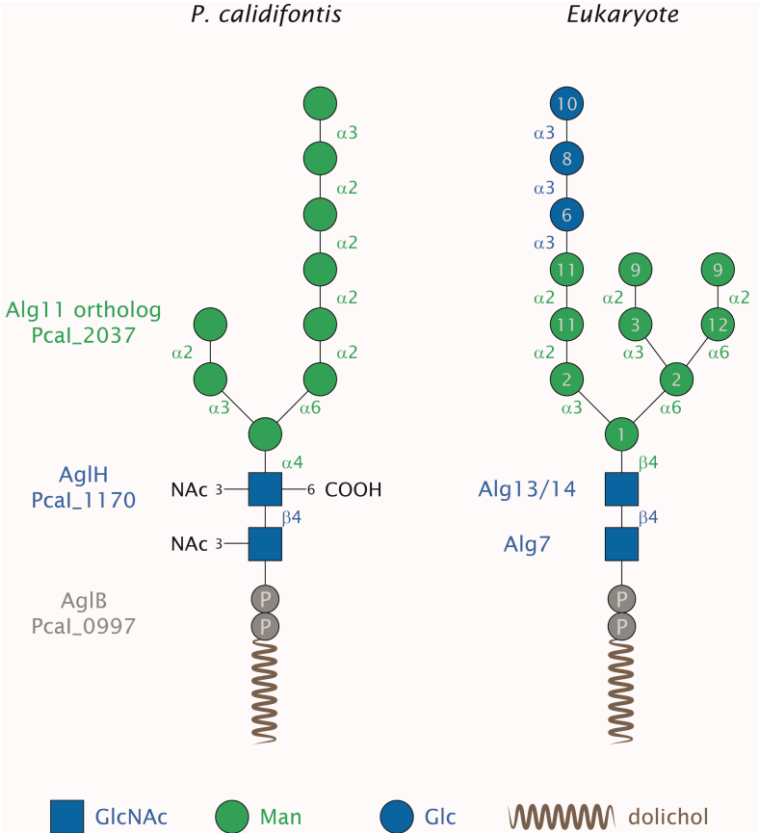


Figure 13: N-glycan structure. Comparison of the 11-residue biantennary high-mannose N-glycan of *P. calidifontis* Dol-PP-Glc$(NAc)_2$-GlcA$(NAc)_2$-$(Man)_9$ and the eukaryotic 14-residue counterpart Dol-PP-$(GlcNAc)_2$-$(Man)_9$-$(Glc)_3$. The picture was adapted from [96].

2.2.1 Overall structure and donor binding of *Pc*ManGT

The catalytic domain of *Pc*ManGT shows a GT-A fold, which is typically for members of the GT2 family. Two membrane-IF helices (IFH1 and IFH2) are connecting the catalytic domain with the TM domain, which consists of three antiparallel TM helices (TMH1, TMH2 and TMH3). TMH3 is midway disrupted by a 90°-kink and forms a substratum for IFH1 and IFH2, and therefore referred to as sub-IFH3 (Figure 14).

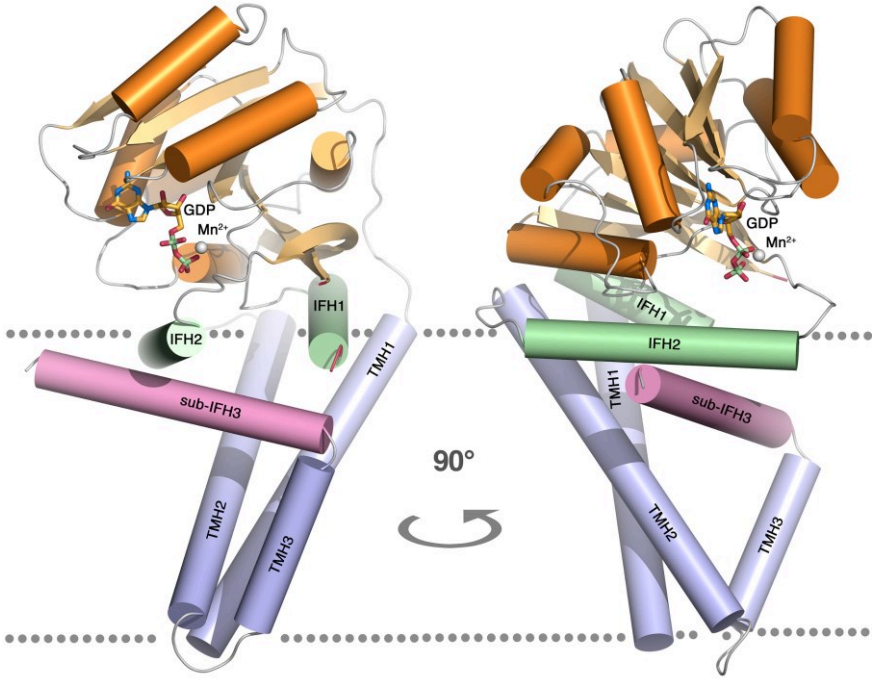


Figure 14: Overall fold of *Pc*ManGT in complex with GDP and Mn^{2+} shown in two views. The position of the disordered acceptor loop, which is protruding from IFH1, is indicated as red coil.

In absence of any acceptor substrate, GDP-Man resides in the active site pocket formed by the C-terminal ends of the parallel β-strands (β1-β4), where it interacts with the β-α loop regions. The guanosine ring is interacting with Ser82 and Arg112, and the ribose moiety is coordinated by hydrogen bonds with the backbone amides of Pro50 and Tyr52, as well as the side chain of Glu54. The DXD motif is defined by Asp135 and Asp137, where Asp137 is coordinating the diphosphate group *via* the Mn^{2+} ion while Asp135 is stacking with the ribose. The electron density for the mannosyl group of GDP-Man is quite featureless, possibly due to motion, and therefore the precise interactions with the active-site residues is somewhat speculative, although Lys115, Asp135, Gly220, Arg261 and His264 are all within in suitable distance to form hydrogen bonds.

2.2.2 Synthesis capabilities of *Pc*ManGT

Determination of a GT's biochemical function requires the identification of donor and acceptor substrate. It is not even certain that knowing the substrate specificity would conclusively reveal the true biological function, and genetic manipulation such as gene knock-out experiments combined with analysis of *N*-glycan composition would probably be needed. Unfortunately, archaea are difficult to manipulate genetically, including *P. calidifontis* [99, 100].

In the case of *Pc*ManGT, it could be shown by isothermal titration calorimetry (ITC) and structure determination that GDP-Man is the preferred donor substrate (Figure 15 A)., and that the enzyme prefers Mn^{2+} as metal cofactor instead of Mg^{2+} and Ca^{2+}. Since the precise identity of the dolichylphosphate sugar acceptor is unknown, the hydrolytic side reaction had to be used to study enzyme activity. The aspartate residues in the DXD motif were replaced by alanine, and the mutant was shown to be unable to hydrolyze the donor substrate, which proved the crucial role of Asp135 and Asp137 for donor and metal binding.

Prior to testing suitable acceptor substrates, the effect of lipids on protein stability was tested. *Pc*ManGT is significantly stabilized by Dol55-P, and a similar effect could be observed for Dol95-P and anionic lipids such as phosphatidylserine (PS) and phosphatidylglycerol (PG). Only weak stabilization could be achieved by adding neutral lipids like phosphatidylcholine (PC) and phosphatidylethanolamine (PE). To verify the observed effect of lipids on *Pc*ManGT, three additional proteins were tested for stabilization by lipids: *i*) the Dol-P-dependent membrane enzyme *Pf*DPMS showed a similar but weaker stabilization by Dol-P; *ii*) a prokaryotic transporter of the unrelated MFS superfamily was weakly stabilized by lipids but not specifically by Dol-P; and *iii*) the lipid-independent soluble protein ArnA from *E. coli* showed no stabilization by any lipid (Figure 15 B).

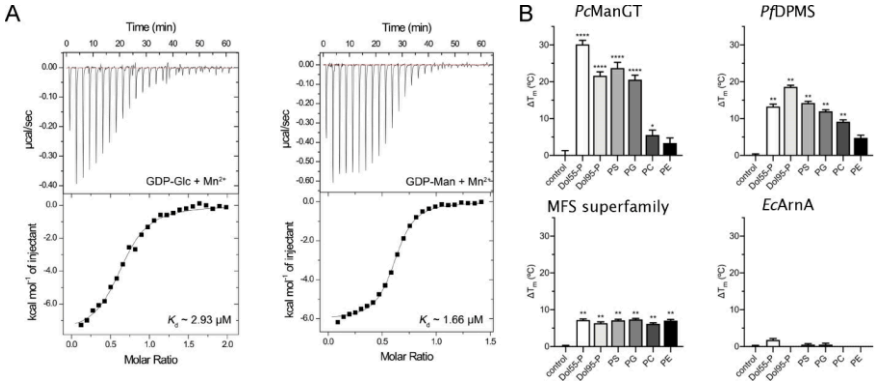


Figure 15: Interaction of donor and potential acceptor substrates or lipids with *Pc*ManGT. A) ITC for GDP-sugar donors. Dissociation constants derived from ITC for binding of different GDP sugars in the presence of Mn^{2+}. B) Thermal stabilization assay. The effect of lipids on the thermal stability of *Pc*ManGT, *Pf*DPMS, a prokaryotic transporter belonging to the MFS superfamily, and EcArnA, a soluble protein from E. coli. The relative stabilization to thermal unfolding by the addition of lipids is expressed as the change in melting temperature (ΔT_m). Used lipids were anionic dolichylmonophosphates (Dol55-P, Dol95-P); anionic phospholipids (PG, PS); and neutral phospholipids (PC, PE).

Based on the strong relative stabilization by Dol-P, it is expected that *P. calidifontis* uses an archaeal Dol-PP as lipid carrier for LLO synthesis. Since *P. calidifontis* uses Dol-PP as carrier, the natural acceptor substrate might be a Dol-PP-linked sugar [96, 101]. Dolichylphosphate-linked glycoconjugates are not commercially available to be tested as acceptors, and therefore only soluble *N*-glycan intermediates could be used, i.e. 2*α*-mannobiose, 3α-mannobiose, 6*α*-mannobiose, 3*α*,6*α*-mannotriose, and chitobiose. HPAEC-PAD was used to analyze possible products from mannosyltransfer to any of the soluble glycan intermediates, but only products originating from GDP-Man hydrolysis were observed, indicating that either the enzyme cannot transfer to any of the soluble substrates tested, or which is more likely, the enzyme requires that the acceptor sugar is attached to Dol-PP.

2.2.3 Comparison with relevant enzymes

The 3D structure of *Pc*ManGT is distinct from already determined structures. The structure most similar to full-length *Pc*ManGT is the catalytic subunit (BcsA) of cellulose synthase from *Rhodobacter spheroides* (PDB code: 4HG6) [102]. If only the catalytic domain of *Pc*ManGT is used for structural matching, chondroitin synthase (PDB code: 2Z86) is the closest match [103].

Bacterial cellulase synthase (Bcs) is a protein complex consisting of at least three subunits: BcsA the intracellular catalytic subunit, BcsB a periplasmic protein that may guides the polymer across the periplasm and BcsC a periplasmic domain that most likely facilitates complex assembly [104]. The *R. spheroides* BcsA consists of three principal parts: *i*) TM domain, that contains four N-terminal TM helices (TMH1-4) and four C-terminal TM helices (TMH5-8) that form the TM domain, while TMH1-2 coil with the C-terminal helix of the BcsB and TMH3-8 form a narrow transmembrane pore guiding the synthesized cellulose chain to the periplasm, *ii*) GT domain, an intracellular loop adopting a GT-A fold, located between TMH4 and TMH5; *iii*) PilZ domain, intracellular C-terminal β-barrel that is connect to the catalytical domain by an extensive α-helical region [102, 104].

For comparison with *Pc*ManGT only the extended catalytical domain of BcsA (TMH3-8 and GT domain) is of interest. *Pc*ManGT and *Rs*BcsA share a similar triangular arrangement of IFH1, IFH2 and sub-IFH3. The three TM helices of *Pc*ManGT arrange to form one half of the six TM helices in *Rs*BcsA that form the cellulose-binding channel [102]. That *Pc*ManGT lacks the sequence pattern typical for processive GTs like cellulose synthase (D,D,D,D,Q(Q/R)XRW), and the presence of a "half-channel" TM domain support a hypothesis where *Pc*ManGT catalyzes a mannosyl-transfer to a Dol-PP-linked acceptor sugar during *N*-glycan biosynthesis in *P. calidifontis*. The A-loop in *Pc*ManGT is identified as an extensive connection between β5 and IFH1. This loop corresponds to a similar loop in *Rs*BcsA that acts as a lid over a triangular opening between IFH1, IFH2 and the sub-IFH3. Although, the A-loop is in a different location compared with *Pf*DPMS, it seems clear that the IF helices play an important function in regulating the binding of lipid-linked acceptors in transmembrane GTs (Figure 16).

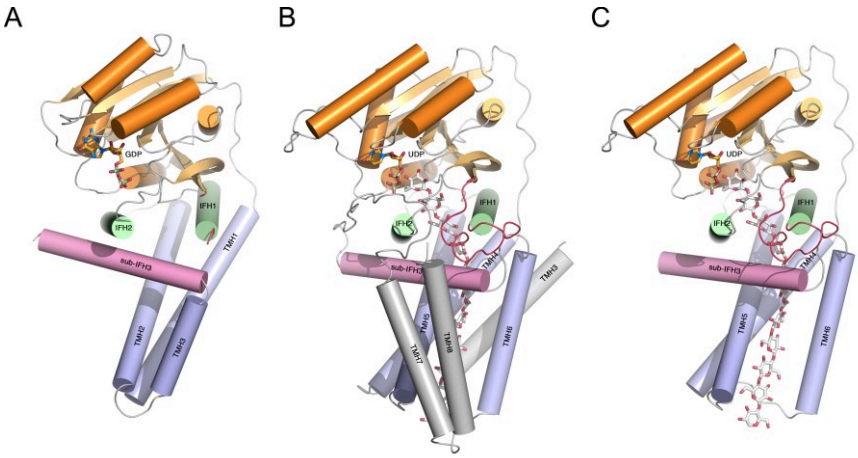


Figure 16: Comparison of the 3-D fold of *Pc*ManGT and BcsA. A) *Pc*ManGT as Figure 14, B) *R. spheroides* BcsA cellulose synthase unit (PDB code: 4HG6, residues 62-575) showing the six-TM domain that forms a cellulose-binding channel [102]. The TMH3/7/8 that are not present in *Pc*ManGT are colored gray. C) BcsA cellulose synthase unit (only TMH4-6) that are shared with *Pc*ManGT (PDB code: 4HG6, residues 96-496).

The recently reported cryo-electron microscopy structure of *Saccharomyces cerevisiae* Alg6 [36], an inverting glucosyltransferase using Dol-P-Glc as donor substrate, revealed a GT-C fold. Despite obvious differences, both *Pc*ManGT and Alg6 are Dol-P-dependent GTs. When docking a modeled Dol-PP chain in the TM domain of *Pc*ManGT, a similar kinked conformation of the isoprenoid chain as that seen in Alg6 seemed to give the best result. By further adding the chitobiose-like core to the docked Dol-PP molecule shows that an acceptor containing two sugar units can be accommodated in the active site, depending on the precise conformation of the A-loop for which there is no interpretable electron density in the experimental data.

2.2.4 Proposed function in *N*-glycan synthesis

The HMG of *P. calidifontis* consists exclusively of *α*-linked mannosyl residues. A glycosidic bond in *α*-configuration can be synthesized either by a retaining GT in the cytoplasm using GDP-*α*-Man as donor, or by a non-cytoplasmic inverting GT using Dol-P-*β*-linked mannose as donor substrate. Based on the sequence, the CAZy database has classified *Pc*ManGT as an inverting GT2 [25], which would imply the stereochemical anomeric inversion from the *α*-linked mannose in GDP-Man to a *β*-linked mannosyl residue in the resulting product, which is at odds with the HMG structure. It is also possible that *Pc*ManGT has been misclassified in the CAZy database and it instead is a retaining GT2-like enzyme. The willingness of *Pc*ManGT to accept nucleophilic water for donor hydrolysis is a trait typically observed for retaining GTs.

The 3D structure and biochemical data argue for *Pc*ManGT being a Leloir GT that transfers a mannosyl unit from GDP-Man to a Dol-PP-linked glycan intermediate containing two sugar units, which would mean that the acceptor substrate would be Dol-PP-Glc(NAc)$_2$-GlcA(NAc)$_2$, and the product Dol-PP-Glc(NAc)$_2$-GlcA(NAc)$_2$-Man. It was also noted during this study that the gene coding for *Pc*ManGT (Pcal_0472) is located in a large gene cluster in the in the *P. calidifontis* genome that contains 14 genes, which led to the suggestion that this cluster is a possible protein *N*-glycosylation (PNG) gene cluster. Of the 14 genes in the cluster, six genes code for GTs. Three of these GT genes show low but distinct sequence similarity to the sequences of the well-established eukaryotic *N*-glycan biosynthetic enzymes α-1,3/1,6-mannosyltransferase Alg2 (Pcal_0473, Pcal_0483), and Alg7 (Pcal_0478) that transfers the first GlcNAc unit to the Dol-PP carrier [96]. Based on the docking results, although speculative, *Pc*ManGT could be a *P. calidifontis* counterpart to Alg1, i.e. the GT that adds the first mannosyl unit to the chitobiose-like *N*-glycan core.

2.3 Host *N*-glycan degradation by *Cutibacterium acnes* (PAPER IV)

Considering that more than two-thirds of human proteins are glycosylated [105] it is not surprising that commensal and pathogenic bacteria have evolved pathways to degrade protein-linked glycans [106]. This capacity would serve as a nutritional source for bacteria [107], but also enable them to modulate the immune response of the host [108-110]. The gram-positive commensal bacteria *Cutibacterium acnes* is naturally occurring member of the human skin microbiota and can under certain circumstances cause a pathogenic condition, acne vulgaris [111-113].

In PAPER IV, the genome of *C. acnes* was successfully searched for genes that encode enzymes with the potential ability to degrade eukaryotic *N*-glycans. As a result, a novel *N*-glycan-degradation pathway could be proposed for *C. acnes*, and an enzyme with a role in the cytoplasmic pathway was studied and characterized in detail. Based on the CAZy classification, this gene is annotated as a β-mannosidase belonging to family GH5 subfamily 18. The biochemical and structural investigation revealed that the enzyme is a functional, dimeric exo-β-1,4-mannosidase that catalyzes the hydrolysis of the β-1,4-glucosidic bond between the second GlcNAc and the first mannose residue of the canonical eukaryotic *N*-glycan core. The results represent the first identification of a host *N*-glycan degradation pathway by the, often notorious, acne bacterium. Additionally, the study represented the first detailed structural and functional characterization of a GH belonging to the GH5_18 subfamily.

2.3.1 Discovery of an *N*-glycan degradation pathway in *C. acnes*

The genome of *C. acnes* strain 266 (phylogenetic group IA1) was searched for genes that code for CAZymes. With the help of operon-prediction algorithms and extensive bioinformatics of individual genes, an operon for cytoplasmic degradation of the eukaryotic *N*-glycan $(GlcNAc)_2$-$(Man)_3$ core, was identified in the genome of *C. acnes* (Figure 17) [114].

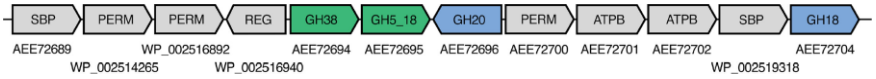


Figure 17: Proposed *C. acnes* 266 *N*-glycan-processing locus 1. GH genes predicted as mannosidases are colored green, and GH genes with predicted *N*-acetylhexosaminidase activity are blue. Other associated genes are colored light gray and includes predicted sugar ABC-transporter substrate-binding protein (SBP), sugar ABC-transporter permease (PERM), transcriptional-regulator gene (REG), and ABC-transporter ATP-binding protein (ATPB). Accession numbers (GenBank, or RefSeq when GenBank was not available) are shown below each gene.

The predicted operon contains four GH genes: a GH38 (putative exo-α-1,3- or exo-α-1,6-mannosidase; EC 3.2.1.24), a GH5 (subfamily 18, predicted β-mannosidase; EC 3.2.1.25), a GH20 (predicted exo-β-1,4-*N*-acetylhexoaminidase; EC.3.2.1.52) and a GH18 (putative endo-β-*N*-acetylglucosaminidase; EC 3.2.1.96), as well as relevant genes for sugar ABC transporter permeases and ABC transporter substrate-binding proteins. The three genes coding for GH38, GH5_18 and GH20 are lacking a secretion signal and would be sufficient to hydrolyze all glycosidic bonds present in the $(GlcNAc)_2$-$(Man)_3$ core to free monosaccharides in the cytoplasm. Additionally, a GH29 (predicted exo-*α*-1,6-fucosidase; EC 3.2.1.51) without secretion signal was found outside the predicted operon, however it would provide the required activity to remove fucose from the potentially fucosylated chitobiose core (Figure 18A).

Trimming of the *N*-glycans involves a sialidase (EC 3.2.1.18), *β*-1,4-galactosidase (EC 3.2.1.23) and *β*-1,2-N-acetyl-hexosaminidase (EC 3.2.1.52). In the genome of *C. acnes* gene candidates were identified coding for enzymes with the required activities. The sialidase activity could be performed by a GH33 gene, which corresponds to the catalytic sialidase domain of *Micromonospora viridifaciens* NanH [115]. The GH3 gene encodes a putative *β*-1,2-*N*-acetyl-hexosaminidase with sequence

similarity to *Bacillus subtilis* strain 168 NagZ [116]. The third required activity is coded by an annotated gene fragment showing sufficient similarity to a GH35 β-1,4-galactosidase [117]. The three genes coding for GH3, GH33 and GH35 all contain a secretion signal, but lack a typical C-terminal cell-wall anchoring LPxTG motif. However, the LPxTG motif does not seem to be an essential feature for secreted enzymes as it was shown for BgaC from *Streptococcus pneumoniae* [118]. Following *N*-glycan trimming, a suitable ABC transporter would be able to translocate the fucosylated *N*-glycan core to the cytoplasm of the bacterial cell (Figure 18 B).

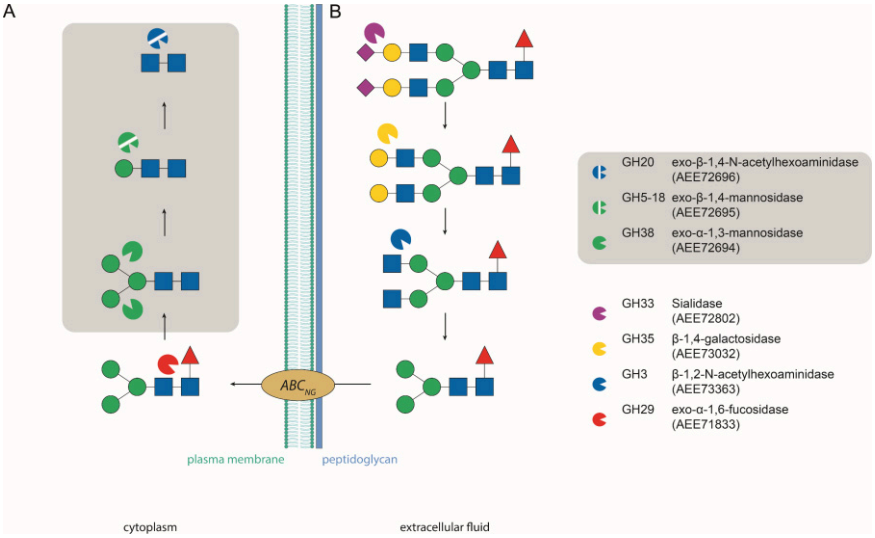


Figure 18: Proposed degradation pathway for host *N*-glycans by *C. acnes*. **A)** Proposed cytoplasmic pathway involving the enzymes GH38, GH5_18 and GH20, and **B)** a hypothetical extracellular pathway. Symbols: blue square: *N*-acetylglucosamine; green circle, mannose; yellow circle, galactose; purple diamond, sialic acid; red triangle, fucose.

For the human pathogen *Streptococcus pyogenes* a similar pathway for *N*-glycan degradation was suggested [119]. Although, *Sp*GH38 is a putative exo-α-1,3-mannosidase, *Sp*GH125 would serve to hydrolyze the α-1,6-linked mannose [120]. In *C. acnes*, a second locus was identified having a GH38/GH125 gene pair in addition to genes coding for GH20 and GH4. Despite lacking the gene coding for a β-mannosidase activity,

this second locus could implicate that *C. acnes* possesses alternative pathways for degradation of *N*-glycans.

2.3.2 Structural investigation of exo-*β*-mannosidase from *C. acnes*

The crystal structure of the exo-*β*-1,4-mannosidase from *C. acnes* (*Ca*Man5_18) was determined without substrate or product bound in the active site, although many attempts had been made to generate an enzyme complex with mannose. Members of the GH5 family belong to the GH-A clan with the canonical $(\beta/\alpha)_8$ TIM-barrel fold. GH5 members use a retaining reaction mechanism that requires two catalytic glutamate residues. These active-site residues are buried in an active-site pocket, approximately 12 Å below the protein surface (Figure 19). Interestingly, GHs with a nearly identical overall structure still show very low sequence identity (10-19%). With the aim to capture the expected hydrolysis product (Man) in the active site, both catalytic residues (Glu140 and Glu259) were mutated to alanine to produce an inactive variant for co-crystallization experiment. However, no crystals could be obtained for this mutant. Instead, a $(GlcNAc)_2$-Man molecule was manually docked in the active site to evaluate potential interaction in the active site. To avoid clashes with amino acid residues, the subsite -1 had to be adapted based on a corresponding mannose conformation present in the structure of *Rhizomucor miehei* Man5 (PDB codes: 4NRR and 4NRS) [121, 122].

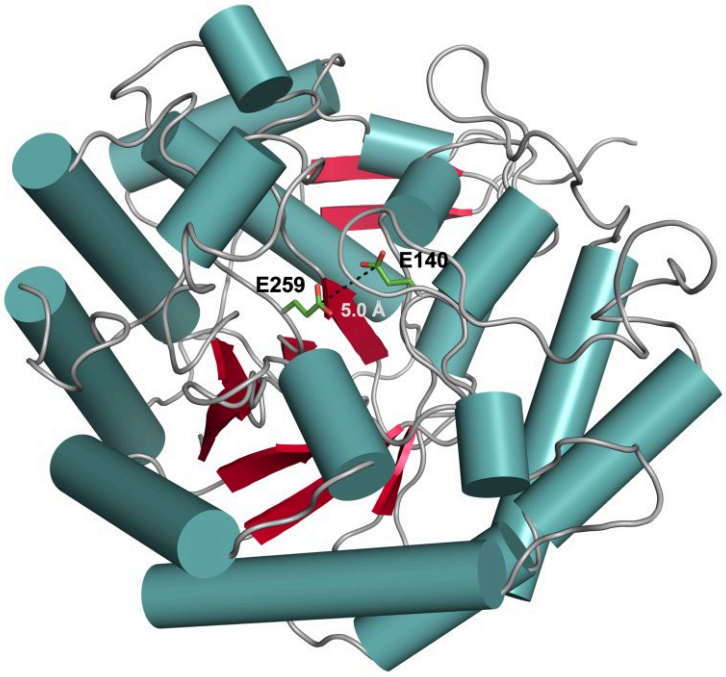


Figure 19: Ribbon representation of the *Ca*Man5_18 subunit structure. The subunit structure of *Ca*Man5_18 features a $(\beta/\alpha)_8$ TIM-barrel fold typical for members of the GH5 family. The catalytic acid/base Glu140 and the nucleophile Glu259 are represented as stick models.

The experimentally determined 3D structure of *Ca*Man5_18 revealed a homodimer. At the dimer interface, several loops (LP1-LP5) serve to stabilize the dimeric state, where LP2 and LP3 are of particular interest because of their involvement in the active site. That *Ca*Man5_18 is a functional homodimer was further supported by analysis of the interface, size-exclusion chromatography and protein unfolding experiments as a function of temperature.

2.3.3 Substrate specificity of *Ca*Man5_18

Initial substrate screening using 4-nitrophenyl β-D-glycopyranosides (*p*NP-sugars) returned detectable activity only towards *p*NP-βMan. Further testing of more complex substrates like mannan or glucomannan did not show any activity. More detailed product analysis was performed using TLC and MALDI-TOF mass spectrometry with manno-oligosaccharides and $(GlcNAc)_2$-Man as substrates. TLC results showed that short mannooligosaccharides were favored over longer ones. Since mannose was the final product, *Ca*Man5_18 could be assigned as an exo-acting GH. The enzyme preferred $(GlcNAc)_2$-Man over mannooligosaccharides as judged by the efficiency of hydrolysis.

2.4 Bacterial *β*-1,3-glucanases derived from moose rumen microbiome (PAPER V)

β-glucans and *β*-glucan-metabolizing enzymes are gaining increasing interest for medical and biotechnological applications, mainly because *β*-glucans are known to support human and animal health, but also because of the potential usefulness of *β*-glucan-degrading enzymes for biomass conversion to generate sustainable value-added products like biofuel and biogas [123]. Currently, algal biomass is being prospected as a new and sustainable feedstock for third-generation biofuels. The principal *β*-glucan in brown algae is laminarin, a *β*-1,3-linked glucopolysaccharide with branches containing *β*-1,6-linked glucose. The high glucose content and the lack of lignin make it a suitable candidate for saccharification and fermentation to produce bioethanol [123]. Naturally, *β*-1,3-glucans are metabolized by *β*-glucanases, which catalyze the hydrolysis of *β*-1,3-linked glucose units [64].

In PAPER V, a GH5 gene from a metagenome-assembled genome originating from a bacterium inhabiting the moose rumen was characterized structurally and biochemically. The gene has been assigned to CAZy GH5 subfamily 44. The expressed gene product, *mrb*Exg5, was found to be an exo-*β*-1,3-glucanase with a homodimeric crystal structure similar to yeast exo-*β*-1,3-glucanases.

2.4.1 Structural features of *mrb*Exg5

The overall 3D structure of *mrb*Exg5 shows the $(\beta/\alpha)_8$ TIM-barrel typical of GH5 members. As expected, the active site contains two characteristic glutamate residues - Glu167 (acid/base catalyst) and Glu295 (nucleophile). A highly flexible loop between strand *β*7 and helix *α*7 lacked interpretable electron density, and attempts to capture complexes of wild-type *mrb*Exg5 with *β*-1,3-linked oligosaccharides bound in the active site failed. Further attempts were made using the inactive mutant E167Q/E295Q, but without success. The binding of a polyethylene glycol molecule in the active site that blocks the entrance is likely to explain the difficulties obtaining a crystal complex. Based on modeling, a bound substrate molecule in the active site is likely to form interactions with the extensive *β*7-*α*7 loop like a similar positioned loop in yeast enzymes closes the active site.

*mrb*Exg5 forms a homodimer in the crystal, and according to analysis of the dimer interface, the homodimer can be stabilized by 14 potential salt bridges, indicating that it is likely to be an authentic dimer (Figure 20). Additional experiments (chemical crosslinking, SEC and biphasic melting-temperature curve) were used to confirm that *mrb*Exg5 is a functional homodimer in solution.

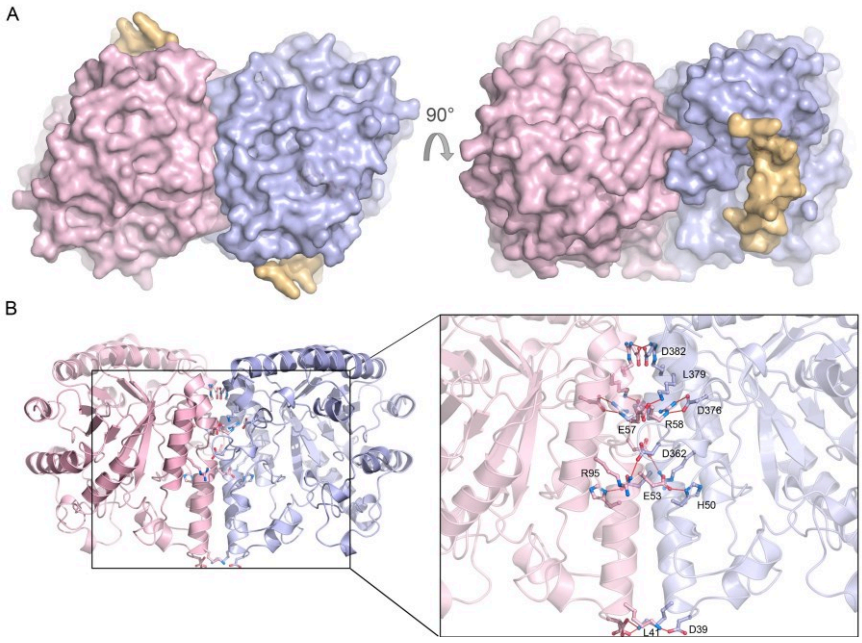


Figure 20: Dimerization of *mrb*Exg5. **A)** Surface representation of the *mrb*Exg5 dimer with a 2-fold symmetry relating the two monomers (pink and blue). The orange surface represents the active-site loop (residues 305-325) that is missing in the crystallographic model. Here, it has been roughly modeled to show its presumed location as a lid over the active site. **B)** Ribbon drawing showing the overall dimer with stabilizing salt links at the dimer interface to the left, and a zoomed in view of dimer interface to the right. The intersubunit ionic interactions include the following pairs from each subunit: Asp39-Lys41, His50-Glu53, Glu57-Asp376, Arg58-Lys379 and Arg62-Asp382.

2.4.2 Catalytic specificity of *mrb*Exg5

To determine the substrate specificity of *mrb*Exg5, different *p*NP-sugars were tested. The highest activity was observed with *p*NP-*β*Glc, while *p*NP-*α*Fuc and *p*NP-*β*Gal performed poorly as substrates, and no activity was detected for *p*NP-*β*Cel and *p*NP-*β*Man. Testing oligosaccharides and polysaccharides showed that only substrates containing *β*-1,3 linkages (laminarioligosaccharides L2-L5 and laminarin) and *β*-1,3-1,4 linkages (3^2- *β*-D-glucosyl cellobiose) were hydrolyzed.

For the laminarioligosaccharides tested, the specificity constant k_{cat}/K_m increased with increasing degree of polymerization (DP) up to DP 5 (L5, laminaripentaose), but *p*NP-*β*Glc still performed better as judged by the 7-fold higher k_{cat}/K_m ratio compared with L5.

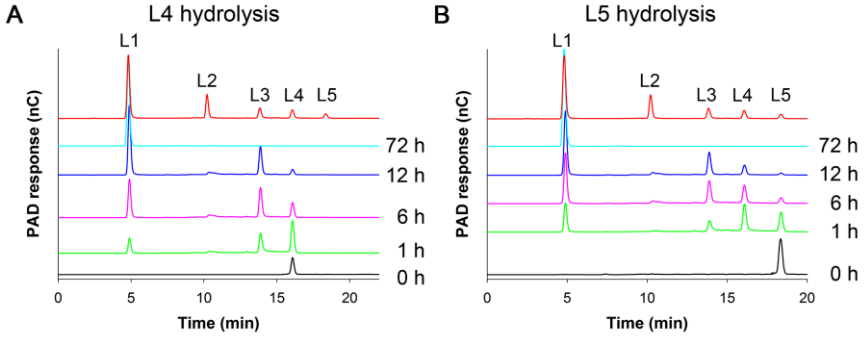


Figure 21: Time-dependent product analysis for hydrolysis of laminaritetraose (L4) and laminaripentaose (L5) as substrate. HPAEC-PAD product analysis for hydrolysis of L4 (A) and L5 (B). The red curve represents the laminarin standards with DP 1-5.

The time-dependent hydrolysis of the L2-L5 and laminarin was monitored using HPAEC-PAD. Complete conversion of L4 and L5 to L1 was achieved after 72h incubation at a temperature of 35°C (pH 6.0). During the initial phase, only L3 and L1, but no L2, could be detected. This observation is consistent with exo-activity (Figure 21). For hydrolysis of L2 and L3, an initial accumulation of L1 was detected, but after a longer incubation time, products of higher DP appeared indicating transglycosylation activity. Similar behavior has been reported for the related yeast enzymes Exg from *Candida albicans* [124] and Exg1 from *Saccharomyces cerevisiae* [125].

When incubating *mrb*Exg5 with laminarin during an extended period of time, a more complex product pattern emerged with initial appearance of L1, L2 and an unidentified oligosaccharide, and appearance of laminarioligosaccharides with DP 3-5 after 24 h.

2.4.3 Comparison with yeast *β*-1,3-glucanases

The search for structural homologs resulted in two enzymes with approximate sequence identities of 30%: the exo-*β*-1,3-glucanase from *Saccharomyces cerevisiae* (*Sc*Exg1, PDB code: 1H4P) [125] and the exo-*β*-1,3-glucanase from *Candida albicans* (*Ca*Exg, PDB codes: 2PB1 and 3N9K) [124].

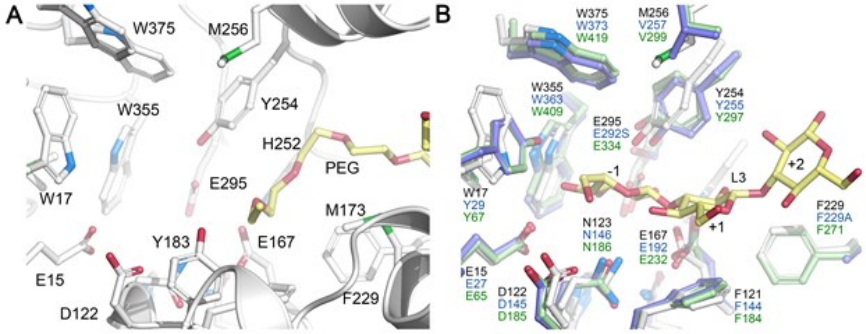


Figure 22: The active site of *mrb*Exg5. A) Active site chains of *mrb*Exg5 (white) and the coordinated PEG molecule (yellow). B) Overlay of the active sites of *mrb*Exg5 (white), with the F229A/E295S mutant of *Ca*Exg (blue) with bound laminaritriose in yellow, and *Sc*Exg1 in green.

A structural alignment of *mrb*Exg5 with *Sc*Exg1 and *Ca*Exg revealed similar active sites in all three structures (Figure 22). Laminaritriose was modeled in the active site of *mrb*Exg5 based on the *Ca*Exg complex, which allowed prediction of possible active-site interactions with laminaritriose in three subsites, -1, +1 and +2. Based on the kinetic parameters, natural substrates with higher DP appear more favorable, and modeling suggests that additional binding sites beyond +2 may be formed. Unfortunately, the lack of structural information for the flexible active-site loop prevented a more detailed analysis of possible subsite interactions.

3 Concluding remarks and outlook

The aim of this thesis was to provide insight in the biosynthesis and degradation of *N*-glycans as well as the breakdown of carbohydrates that are a suitable source for production of sustainable biofuel.

Protein glycosylation is achieved by complex pathways involving extensive and diverse sets of glycosyltransferases. Defects in these pathways can lead to severe diseases in humans, i.e. congenital disorders of glycosylation (CDG). It is therefore important to investigate the function and structure of the enzymes responsible for specific steps of the pathways that involves defective enzyme performance. DPMSs are coupled to several CDGs, but understanding of the underlying molecular mechanisms of action of this group of enzymes has been lacking, including the events of glycolipid membrane translocation.

This thesis presents the first determined three-dimensional structure of a DPMS, from the archaeal source *Pyrococcus furiosus*. The results enabled rationalization of structure-function relationships concerning Dol-P-Man biosynthesis, and also presented unexpected new information that indicated a moonlighting function of *P. furiosus* DPMS that would include both the synthesis and translocation of Dol-P-Man across the membrane. Initial results from a customized proteoliposome assay offered preliminary support for the translocation hypothesis, but further studies are required to fully establish a coupled synthase-flippase function for *P. furiosus* DPMS.

Furthermore, this thesis identified and characterized a mannosyltransferase from *Pyrobaculum calidifontis* (*Pc*ManGT), and suggests that the enzyme is involved in the synthesis of the *P. calidifontis* high-mannose *N*-glycan, and acting as a possible archaeal counterpart of the eukaryotic Alg1. The gene coding for *Pc*ManGT is present in a large gene cluster in the *P. calidifontis* genome, which is hypothesized in the thesis to constitute a gene cluster for *N*-glycan biosynthesis. Future functional and structural characterization of additional genes in this cluster will shed further light on this hypothesis.

Commensal and pathogenic bacteria have evolved pathways to degrade *N*-glycans for nutrition, and in some cases to escape the immune response of the affected host. It is therefore important to understand the degradation pathways in order to investigate new and more effective ways

to tackle bacterial infections. In the thesis, an *N*-glycan degradation pathway was identified and described for the commensal bacterium *Cutibacterium acnes* (the bacterium that causes acne), and one enzyme in this pathway (*Ca*Man5_18) was characterized in detail. Future studies would include functional characterization of additional enzymes of this pathway to confirm the identity of the gene locus as an *N*-glycan degradation operon, as well as inhibitor-binding studies where specific steps in the pathway are selectively blocked to interfere with bacterial metabolism and pathogenicity.

Microbial carbohydrate-active enzymes (CAZymes) are interesting for the purpose of exploring new sustainable processes for production of biofuels and value-added carbohydrate-based products. Metagenomic databases contain a wealth of genomic information concerning potential CAZyme candidates that can be of use for the degradation of complex polysaccharides. In the thesis, the exo-β-1,3-glucanase *mrb*ExG5 produced by a bacterium thriving in the moose rumen was characterized structurally and biochemically. Based on substrate specificity towards laminarin, *mrb*ExG5 could be a suitable candidate for the industrial treatment of brown-algal biomass as a sustainable feedstock for biofuel production. Continued studies would include conducting small-scale experiments with *mrb*ExG5 using algal biomass as substrate.

4 Acknowledgements

First, I would like to thank my main supervisor **Christina Divne** for our journey together the last five years. You sparked my interest in structure biology and let me follow my curiosity to explore this fantastic research field. I am very thankful for your guidance and support throughout my PhD project, and giving me all the freedom I could ask for during this time! Your door was always open, and you encouraged me to travel wherever I wanted to go, so that could collect a lot of valuable experience during workshops, seminars and conferences.

I would also like to thank **Rosaria** for all the guidance in the beginning of my PhD project! Showing me how to deal with membrane proteins (if nothing helps, soap can be the last detergent to test) and how to spend the nights at synchrotron without falling asleep.

After moving with our lab (back) to AlbaNova I was lucky to share the office with **Maja** and **Stefan**. Thank you for easing the first time there, providing a relaxing office atmosphere and extending short weekends with a long coffee break on Monday mornings.

Thanks to my current office mates that amplified the interest in structure biology enormous in our department. **Dayanand**, thank you for always having my back in the lab and sharing the mood-elevating GH5-projects with me. **Markus**, thank you for educating us with your huge common knowledge about things that most people don't know, like abnormal behavior of otters and dolphins. Big thanks to **Valentina**, you are not just an average colleague, but also the best travel companion I could have ever asked for. Thank you for keeping me company during long days, and for the fun times outside the lab. Thank you all for each scientific and non-scientific discussions we had as a group. I am very grateful that all of you joined Christina's lab!

Thanks to the people at PSF for the constant support and input during my crystallization trials. And a big thanks to all the local contacts at the synchrotron (Bessy, Diamond, DESY, ESRF, MaxIV and Soleil) taking care of me and my crystals during the beamtimes. A special thanks to **Serena** and **Pierre** for your awesome support and the extra morning coffee!

Thanks to all past and present co-workers of DIB, for entertaining Kick-off days together and fantastic celebrations of other PhD students reaching the finish line. Thanks for everyone joining our movie nights and after work beers. It was great having you as colleagues and our fruitful discussions during our lunch and PhD seminar sessions. Special thanks to **Teun** for always being available for endless coffee breaks to brainstorm, vent or just discuss about everything else!

Thanks to all the colleagues of Glycoscience for creating a nice working atmosphere and discussions during lunch breaks about everything and anything.

Thank you, **Elias Arnér** for giving me the opportunity to come to Stockholm and for laying the foundation to stay in Sweden to pursue my PhD studies.

I am very thankful for all my friends in Stockholm, **Aida** & **Lucio**, **Åsa** & **Hossein**, **Elle** & **Mattias, Johan**, **Lisa** & **Manuel**, **Mattias H.**, **Michał**, **Robert, Phillia** & **Simon**, **Steffi** & **Tobias**, **Tomas L., Tomas M.**, **Katrin** & **Moritz**, **Karen** & **Juan**, **Magali**, **Nacho**, **Antonina** & **Mohamed**, **Jessie** & **Axel**, **Julen**, **Médoune**, **Martin**, **Viktoria**, **Caterina** & **Serge**, **Laura** & **Michele**, **Jorge**, **Teresa** & **Mino**, you are one of the reasons why living in Stockholm makes me feel at home. Spending time with you and discovering Sweden together gave me the energy that I needed in the last five years!

I am thankful for my football group and our marvelous Korpen team HTS, **Axel, Joep**, **Kipat**, **Leo**, **Michael**, **Teun**, **Tim**, **Tobias** for letting me have a break from everyday work troubles and just having fun on the pitch!

Ich möchte mich auch bei meiner Familie, **Mama**, **Papa**, **Felix** & **Lea**, **Elinor**, **Jens-Peter** und **Johanna** & **Matthias** bedanken für die bedingungslose Unterstützung in all den Jahren! Vielen Dank an meine Eltern, ihr habt schon früh mein Interesse an den Naturwissenschaften gefördert und danke für euer Bemühen zu verstehen was ich den ganzen Tag so auf Arbeit treibe. Ohne euren Zuspruch wären manche Dinge wohl anders verlaufen. Vielen Dank euch allen, für all die Sommer, die wir in Schweden oder sonst wo in der Welt zusammen verbracht haben, um mal Luft zu holen vom Alltag.

Auch möchte ich mich bei meinen Freunden aus der Heimat, **Leonie** & **Daniel**, **Marie** & **Christoph**, **Nichole** & **Tony**, **Sarah** & **Jan, Susi** & **Basti**, **Ulli** & **Peter**, **Falko**, **Martin** bedanken für die zahlreichen Weihnachtsmarktbesuche und dass manche sich sogar auf den Weg nach Schweden gemacht habt um uns zu besuchen!

Finally, **Theresa** it is difficult to express in words how lucky I am to have you at my side, you are the love of my life! Without you I would never have managed to get there where I am now. Thank you for taking care of me in the stressful times during the last month and for pushing me to do things that are not my strong suite. Thank you for exploring the world with me together and for suffering by my side during our hikes no matter how much it rained, how steep the mountains were and how far the next camp site was!

5 References

1. Binkley, R.W., *Modern carbohydrate chemistry*. 1988: New York: Marcel Dekker 1988.
2. Ionescu, A.R., et al., *Conformational pathways of saturated six-membered rings. A static and dynamical density functional study.* J Phys Chem A, 2005. **109**(36): p. 8096-105.
3. Ma, B.Y., H.F. Schaefer, and N.L. Allinger, *Theoretical studies of the potential energy surfaces and compositions of the D-aldo- and D-ketohexoses.* Journal of the American Chemical Society, 1998. **120**(14): p. 3411-3422.
4. *Biochemistry 3rd edition*. 2000: Addison Wesley Longman.
5. Jennewein, M.F. and G. Alter, *The Immunoregulatory Roles of Antibody Glycosylation.* Trends Immunol, 2017. **38**(5): p. 358-372.
6. Franklin, E.C., *Structure and Function of Immunoglobulins.* Acta Endocrinologica, 1975. **78**: p. 77-95.
7. Jung, D. and F.W. Alt, *Unraveling V(D)J recombination; insights into gene regulation.* Cell, 2004. **116**(2): p. 299-311.
8. Limoli, D.H., C.J. Jones, and D.J. Wozniak, *Bacterial Extracellular Polysaccharides in Biofilm Formation and Function.* Microbiol Spectr, 2015. **3**(3).
9. Reily, C., et al., *Glycosylation in health and disease.* Nat Rev Nephrol, 2019. **15**(6): p. 346-366.
10. Kandiba, L., et al., *Diversity in prokaryotic glycosylation: an archaeal-derived N-linked glycan contains legionaminic acid.* Mol Microbiol, 2012. **84**(3): p. 578-93.
11. Bieberich, E., *Synthesis, Processing, and Function of N-glycans in N-glycoproteins.* Adv Neurobiol, 2014. **9**: p. 47-70.
12. Stanley, P., N. Taniguchi, and M. Aebi, *N-Glycans, in Essentials of Glycobiology*, in *Essentials of Glycobiology*, rd, et al., Editors. 2015: Cold Spring Harbor (NY). p. 99-111.
13. Moremen, K.W. and M. Molinari, *N-linked glycan recognition and processing: the molecular basis of endoplasmic reticulum quality control.* Curr Opin Struct Biol, 2006. **16**(5): p. 592-9.
14. Vembar, S.S. and J.L. Brodsky, *One step at a time: endoplasmic reticulum-associated degradation.* Nat Rev Mol Cell Biol, 2008. **9**(12): p. 944-57.
15. Nagae, M., et al., *3D Structure and Function of Glycosyltransferases Involved in N-glycan Maturation.* Int J Mol Sci, 2020. **21**(2).
16. Ng, B.G. and H.H. Freeze, *Perspectives on Glycosylation and Its Congenital Disorders.* Trends Genet, 2018. **34**(6): p. 466-476.
17. Verheijen, J., et al., *Therapeutic approaches in Congenital Disorders of Glycosylation (CDG) involving N-linked glycosylation: an update.* Genet Med, 2020. **22**(2): p. 268-279.
18. Oliveira-Ferrer, L., K. Legler, and K. Milde-Langosch, *Role of protein glycosylation in cancer metastasis.* Semin Cancer Biol, 2017. **44**: p. 141-152.
19. Pauly, M., et al., *Hemicellulose biosynthesis.* Planta, 2013. **238**(4): p. 627-42.
20. Usoltseva, R.V., et al., *Laminarans and 1,3-beta-D-glucanases.* Int J Biol Macromol, 2020.
21. Cifuente, J.O., et al., *Structural basis of glycogen metabolism in bacteria.* Biochem J, 2019. **476**(14): p. 2059-2092.

22. Kang, Q., et al., *Bioethanol from lignocellulosic biomass: current findings determine research priorities.* ScientificWorldJournal, 2014. **2014**: p. 298153.
23. Hallac, B.B. and A.J. Ragauskas, *Analyzing cellulose degree of polymerization and its relevancy to cellulosic ethanol.* Biofuels Bioproducts & Biorefining-Biofpr, 2011. **5**(2): p. 215-225.
24. Malgas, S., et al., *A mini review of xylanolytic enzymes with regards to their synergistic interactions during hetero-xylan degradation.* World J Microbiol Biotechnol, 2019. **35**(12): p. 187.
25. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013.* Nucleic Acids Res, 2014. **42**(Database issue): p. D490-5.
26. Henrissat, B., *A Classification of Glycosyl Hydrolases Based on Amino-Acid-Sequence Similarities.* Biochemical Journal, 1991. **280**: p. 309-316.
27. International Union of Biochemistry and Molecular Biology. Nomenclature Committee. and E.C. Webb, *Enzyme nomenclature 1992 : recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes*. 1992, San Diego: Published for the International Union of Biochemistry and Molecular Biology by Academic Press. xiii, 862 p.
28. Ardevol, A. and C. Rovira, *Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations.* J Am Chem Soc, 2015. **137**(24): p. 7528-47.
29. Davies, G. and B. Henrissat, *Structures and mechanisms of glycosyl hydrolases.* Structure, 1995. **3**(9): p. 853-9.
30. Henrissat, B. and G. Davies, *Structural and sequence-based classification of glycoside hydrolases.* Curr Opin Struct Biol, 1997. **7**(5): p. 637-44.
31. Gloster, T.M., *Advances in understanding glycosyltransferases from a structural perspective.* Curr Opin Struct Biol, 2014. **28**: p. 131-41.
32. Hurtado-Guerrero, R. and G.J. Davies, *Recent structural and mechanistic insights into post-translational enzymatic glycosylation.* Current Opinion in Chemical Biology, 2012. **16**(5-6): p. 479-487.
33. Albuquerque-Wendt, A., et al., *Membrane Topological Model of Glycosyltransferases of the GT-C Superfamily.* Int J Mol Sci, 2019. **20**(19).
34. Ardiccioni, C., et al., *Structure of the polyisoprenyl-phosphate glycosyltransferase GtrB and insights into the mechanism of catalysis.* Nat Commun, 2016. **7**: p. 10175.
35. Morera, S., et al., *High resolution crystal structures of T4 phage beta-glucosyltransferase: induced fit and effect of substrate and metal binding.* J Mol Biol, 2001. **311**(3): p. 569-77.
36. Bloch, J.S., et al., *Structure and mechanism of the ER-based glucosyltransferase ALG6.* Nature, 2020. **579**(7799): p. 443-447.
37. Sinnott, M.J., *Catalytic mechanism of enzymic glycosyl transfer.* Chemical Reviews, 1990. **90**(7): p. 1171-1202.
38. Lairson, L.L., et al., *Glycosyltransferases: structures, functions, and mechanisms.* Annu Rev Biochem, 2008. **77**: p. 521-55.
39. Palcic, M.M., *Glycosyltransferases as biocatalysts.* Curr Opin Chem Biol, 2011. **15**(2): p. 226-33.

40. Liang, D.M., et al., *Glycosyltransferases: mechanisms and applications in natural product development.* Chemical Society Reviews, 2015. **44**(22): p. 8350-8374.
41. Unligil, U.M. and J.M. Rini, *Glycosyltransferase structure and mechanism.* Curr Opin Struct Biol, 2000. **10**(5): p. 510-7.
42. Ardevol, A., et al., *The reaction mechanism of retaining glycosyltransferases.* Biochem Soc Trans, 2016. **44**(1): p. 51-60.
43. Koshland, D.E., Jr., *Stereochemistry and the mechanism of enzymatic reactions.* Biol Rev, 1953. **28**: p. 416-433.
44. McArthur, J.B. and X. Chen, *Glycosyltransferase engineering for carbohydrate synthesis.* Biochem Soc Trans, 2016. **44**(1): p. 129-42.
45. Li, S.T., et al., *Alternative routes for synthesis of N-linked glycans by Alg2 mannosyltransferase.* FASEB J, 2018. **32**(5): p. 2492-2506.
46. Lussier, M., A.M. Sdicu, and H. Bussey, *The KTR and MNN1 mannosyltransferase families of Saccharomyces cerevisiae.* Biochim Biophys Acta, 1999. **1426**(2): p. 323-34.
47. Orlean, P., C. Albright, and P.W. Robbins, *Cloning and Sequencing of the Yeast Gene for Dolichol Phosphate Mannose Synthase, an Essential Protein.* Journal of Biological Chemistry, 1988. **263**(33): p. 17499-17507.
48. Lommel, M. and S. Strahl, *Protein O-mannosylation: Conserved from bacteria to humans.* Glycobiology, 2009. **19**(8): p. 816-828.
49. Doucey, M.A., et al., *Protein C-mannosylation is enzyme-catalysed and uses dolichyl-phosahate-mannose as a precursor.* Molecular Biology of the Cell, 1998. **9**(2): p. 291-300.
50. Rush, J.S., *Role of Flippases in Protein Glycosylation in the Endoplasmic Reticulum.* Lipid Insights, 2015. **8**(Suppl 1): p. 45-53.
51. Orlean, P., *Dolichol Phosphate Mannose Synthase Is Required Invivo for Glycosyl Phosphatidylinositol Membrane Anchoring, O-Mannosylation, and N-Glycosylation of Protein in Saccharomyces-Cerevisiae.* Molecular and Cellular Biology, 1990. **10**(11): p. 5796-5805.
52. Maeda, Y. and T. Kinoshita, *Dolichol-phosphate mannose synthase: structure, function and regulation.* Biochim Biophys Acta, 2008. **1780**(6): p. 861-8.
53. Council, C.E., et al., *Enzymatic glycosylation involving fluorinated carbohydrates.* Org Biomol Chem, 2020. **18**(18): p. 3423-3451.
54. Coines, J., L. Raich, and C. Rovira, *Modeling catalytic reaction mechanisms in glycoside hydrolases.* Curr Opin Chem Biol, 2019. **53**: p. 183-191.
55. Costa, D.A.L. and E.X.F. Filho, *Microbial beta-mannosidases and their industrial applications.* Appl Microbiol Biotechnol, 2019. **103**(2): p. 535-547.
56. Moreira, L.R.S. and E.X.F. Filho, *An overview of mannan structure and mannan-degrading enzyme systems.* Applied Microbiology and Biotechnology, 2008. **79**(2): p. 165-178.
57. Yamabhai, M., et al., *Mannan biotechnology: from biofuels to health.* Critical Reviews in Biotechnology, 2016. **36**(1): p. 32-42.
58. Couturier, M., et al., *Structural and biochemical analyses of glycoside hydrolase families 5 and 26 beta-(1,4)-mannanases from Podospora anserina reveal differences upon manno-oligosaccharide catalysis.* J Biol Chem, 2013. **288**(20): p. 14624-35.

59. Herve, C., et al., *Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects.* Proc Natl Acad Sci U S A, 2010. **107**(34): p. 15293-8.
60. Rovira, C., et al., *Mannosidase mechanism: at the intersection of conformation and catalysis.* Curr Opin Struct Biol, 2020. **62**: p. 79-92.
61. Synytsya, A. and M. Novak, *Structural analysis of glucans.* Ann Transl Med, 2014. **2**(2): p. 17.
62. Keller, M.B., et al., *Activity of fungal beta-glucosidases on cellulose.* Biotechnol Biofuels, 2020. **13**: p. 121.
63. Linton, S.M., *Review: The structure and function of cellulase (endo-beta-1,4-glucanase) and hemicellulase (beta-1,3-glucanase and endo-beta-1,4-mannase) enzymes in invertebrates that consume materials ranging from microbes, algae to leaf litter.* Comp Biochem Physiol B Biochem Mol Biol, 2020. **240**: p. 110354.
64. Ketudat Cairns, J.R. and A. Esen, *beta-Glucosidases.* Cell Mol Life Sci, 2010. **67**(20): p. 3389-405.
65. Hamilton, B.S., et al., *A library of chemically defined human N-glycans synthesized from microbial oligosaccharide precursors.* Scientific Reports, 2017. **7**.
66. Chao, Q., et al., *Recent Progress in Chemo-Enzymatic Methods for the Synthesis of N-Glycans.* Frontiers in Chemistry, 2020. **8**.
67. Khoo, C.G., et al., *Algae biorefinery: Review on a broad spectrum of downstream processes and products.* Bioresour Technol, 2019. **292**: p. 121964.
68. Zargarzadeh, M., et al., *Biomedical applications of laminarin.* Carbohydr Polym, 2020. **232**: p. 115774.
69. Dobrincic, A., et al., *Advanced Technologies for the Extraction of Marine Brown Algal Polysaccharides.* Mar Drugs, 2020. **18**(3).
70. Blanch, H.W., *Bioprocessing for biofuels.* Curr Opin Biotechnol, 2012. **23**(3): p. 390-5.
71. Moremen, K.W., M. Tiemeyer, and A.V. Nairn, *Vertebrate protein glycosylation: diversity, synthesis and function.* Nat Rev Mol Cell Biol, 2012. **13**(7): p. 448-62.
72. Eichler, J., *Extreme sweetness: protein glycosylation in archaea.* Nat Rev Microbiol, 2013. **11**(3): p. 151-6.
73. Nothaft, H. and C.M. Szymanski, *Protein glycosylation in bacteria: sweeter than ever.* Nat Rev Microbiol, 2010. **8**(11): p. 765-78.
74. Aebi, M., *N-linked protein glycosylation in the ER.* Biochim Biophys Acta, 2013. **1833**(11): p. 2430-7.
75. Jarrell, K.F., et al., *N-Linked Glycosylation in Archaea: a Structural, Functional, and Genetic Analysis.* Microbiology and Molecular Biology Reviews, 2014. **78**(2): p. 304-341.
76. Hirschberg, C.B. and M.D. Snider, *Topography of Glycosylation in the Rough Endoplasmic-Reticulum and Golgi-Apparatus.* Annual Review of Biochemistry, 1987. **56**: p. 63-87.
77. Sanyal, S. and A.K. Menon, *Stereoselective transbilayer translocation of mannosyl phosphoryl dolichol by an endoplasmic reticulum flippase.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(25): p. 11289-11294.
78. Nagata, S., T. Sakuragi, and K. Segawa, *Flippase and scramblase for phosphatidylserine exposure.* Curr Opin Immunol, 2020. **62**: p. 31-38.

79. Montigny, C., et al., *On the molecular mechanism of flippase- and scramblase-mediated phospholipid transport.* Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids, 2016. **1861**(8): p. 767-783.
80. Colussi, P.A., et al., *Human and Saccharomyces cerevisiae dolichol phosphate mannose synthases represent two classes of the enzyme, but both function in Schizosaccharomyces pombe.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(15): p. 7873-7878.
81. Breton, C., et al., *Structures and mechanisms of glycosyltransferases.* Glycobiology, 2006. **16**(2): p. 29R-37R.
82. Banerjee, D.K., et al., *In vitro phosphorylation by cAMP-dependent protein kinase up-regulates recombinant Saccharomyces cerevisiae mannosylphosphodolichol synthase.* Journal of Biological Chemistry, 2005. **280**(6): p. 4174-4181.
83. Herscovics, A., C.D. Warren, and R.W. Jeanloz, *Anomeric Configuration of Dolichyl D-Mannosyl Phosphate Formed in Calf Pancreas Microsomes.* Journal of Biological Chemistry, 1975. **250**(20): p. 8079-8084.
84. Schutzbach, J.S. and J.W. Zimmerman, *Yeast Dolichyl-Phosphomannose Synthase - Reconstitution of Enzyme-Activity with Phospholipids.* Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire, 1992. **70**(6): p. 460-465.
85. Jeffery, C.J., *An introduction to protein moonlighting.* Biochemical Society Transactions, 2014. **42**: p. 1679-1683.
86. Chang, E.L., *Unusual thermal stability of liposomes made from bipolar tetraether lipids.* Biochem Biophys Res Commun, 1994. **202**(2): p. 673-9.
87. Ardiccioni, C., et al., *Structure of the polyisoprenyl-phosphate glycosyltransferase GtrB and insights into the mechanism of catalysis.* Nature Communications, 2016. **7**.
88. Petrou, V.I., et al., *Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation.* Science, 2016. **351**(6273): p. 608-612.
89. Guan, S., D.A. Bastin, and N.K. Verma, *Functional analysis of the O antigen glucosylation gene cluster of Shigella flexneri bacteriophage SfX.* Microbiology-Uk, 1999. **145**: p. 1263-1273.
90. Yan, A.X., Z.Q. Guan, and C.R.H. Raetz, *An undecaprenyl phosphate-aminoarabinose flippase required for polymyxin resistance in Escherichia coli.* Journal of Biological Chemistry, 2007. **282**(49): p. 36077-36089.
91. Rismondo, J., et al., *GtcA is required for LTA glycosylation in Listeria monocytogenes serovar 1/2a and Bacillus subtilis.* Cell Surf, 2020. **6**: p. 100038.
92. Larrouy-Maumus, G., et al., *A Small Multidrug Resistance-like Transporter Involved in the Arabinosylation of Arabinogalactan and Lipoarabinomannan in Mycobacteria.* Journal of Biological Chemistry, 2012. **287**(47).
93. Larkin, A. and B. Imperiali, *The expanding horizons of asparagine-linked glycosylation.* Biochemistry, 2011. **50**(21): p. 4411-26.
94. Lombard, J., *The multiple evolutionary origins of the eukaryotic N-glycosylation pathway.* Biology Direct, 2016. **11**.
95. Meyer, B.H. and S.V. Albers, *Hot and sweet: protein glycosylation in Crenarchaeota.* Biochemical Society Transactions, 2013. **41**: p. 384-392.
96. Fujinami, D., Y. Taguchi, and D. Kohda, *Asn-linked oligosaccharide chain of a crenarchaeon, Pyrobaculum calidifontis, is reminiscent of the eukaryotic high-mannose-type glycan.* Glycobiology, 2017. **27**(8): p. 701-712.

97. Guy, L. and T.J.G. Ettema, *The archaeal 'TACK' superphylum and the origin of eukaryotes.* Trends in Microbiology, 2011. **19**(12): p. 580-587.
98. Koonin, E.V. and N. Yutin, *The Dispersed Archaeal Eukaryome and the Complex Archaeal Ancestor of Eukaryotes.* Cold Spring Harbor Perspectives in Biology, 2014. **6**(4).
99. Schleper, C., G. Jurgens, and M. Jonuscheit, *Genomic studies of uncultivated archaea.* Nat Rev Microbiol, 2005. **3**(6): p. 479-88.
100. Bernick, D.L., et al., *Comparative genomic and transcriptional analyses of CRISPR systems across the genus Pyrobaculum.* Frontiers in Microbiology, 2012. **3**.
101. Taguchi, Y., D. Fujinami, and D. Kohda, *Comparative Analysis of Archaeal Lipid-linked Oligosaccharides That Serve as Oligosaccharide Donors for Asn Glycosylation.* Journal of Biological Chemistry, 2016. **291**(21): p. 11042-11054.
102. Morgan, J.L.W., J. Strumillo, and J. Zimmer, *Crystallographic snapshot of cellulose synthesis and membrane translocation.* Nature, 2013. **493**(7431): p. 181-U192.
103. Osawa, T., et al., *Crystal structure of chondroitin polymerase from Escherichia coli K4.* Biochem Biophys Res Commun, 2009. **378**(1): p. 10-4.
104. Omadjela, O., et al., *BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(44): p. 17856-17861.
105. Apweiler, R., H. Hermjakob, and N. Sharon, *On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database.* Biochimica Et Biophysica Acta-General Subjects, 1999. **1473**(1): p. 4-8.
106. Marcobal, A., et al., *A refined palate: bacterial consumption of host glycans in the gut.* Glycobiology, 2013. **23**(9): p. 1038-46.
107. Garbe, J. and M. Collin, *Bacterial hydrolysis of host glycoproteins - powerful protein modification and efficient nutrient acquisition.* J Innate Immun, 2012. **4**(2): p. 121-31.
108. Byers, H.L., et al., *Sequential deglycosylation and utilization of the N-linked, complex-type glycans of human alpha1-acid glycoprotein mediates growth of Streptococcus oralis.* Glycobiology, 1999. **9**(5): p. 469-79.
109. Cao, Y., E.R. Rocha, and C.J. Smith, *Efficient utilization of complex N-linked glycans is a selective advantage for Bacteroides fragilis in extraintestinal infections.* Proc Natl Acad Sci U S A, 2014. **111**(35): p. 12901-6.
110. Collin, M. and A. Olsen, *Effect of SpeB and EndoS from Streptococcus pyogenes on human immunoglobulins.* Infect Immun, 2001. **69**(11): p. 7187-9.
111. Achermann, Y., et al., *Propionibacterium acnes: from commensal to opportunistic biofilm-associated implant pathogen.* Clin Microbiol Rev, 2014. **27**(3): p. 419-40.
112. Grice, E.A. and J.A. Segre, *The skin microbiome.* Nat Rev Microbiol, 2011. **9**(4): p. 244-53.
113. Lichtenberger, R., et al., *Genetic architecture of acne vulgaris.* J Eur Acad Dermatol Venereol, 2017. **31**(12): p. 1978-1990.
114. Price, M.N., et al., *A novel method for accurate operon predictions in all sequenced prokaryotes.* Nucleic Acids Res, 2005. **33**(3): p. 880-92.

115. Gaskell, A., S. Crennell, and G. Taylor, *The three domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll.* Structure, 1995. **3**(11): p. 1197-205.
116. Litzinger, S., et al., *Structural and kinetic analysis of Bacillus subtilis N-acetylglucosaminidase reveals a unique Asp-His dyad mechanism.* J Biol Chem, 2010. **285**(46): p. 35675-84.
117. Rojas, A.L., et al., *Crystal structures of beta-galactosidase from Penicillium sp. and its complex with galactose.* J Mol Biol, 2004. **343**(5): p. 1281-92.
118. Jeong, J.K., et al., *Characterization of the Streptococcus pneumoniae BgaC protein as a novel surface beta-galactosidase with specific hydrolysis activity for the Galbeta1-3GlcNAc moiety of oligosaccharides.* J Bacteriol, 2009. **191**(9): p. 3011-23.
119. Suits, M.D., et al., *Structure and kinetic investigation of Streptococcus pyogenes family GH38 alpha-mannosidase.* PLoS One, 2010. **5**(2): p. e9006.
120. Robb, M., et al., *Molecular Characterization of N-glycan Degradation and Transport in Streptococcus pneumoniae and Its Contribution to Virulence*. PLoS Pathog, 2017. **13**(1): p. e1006090.
121. Leaver-Fay, A., et al., *Computationally Designed Bispecific Antibodies using Negative State Repertoires.* Structure, 2016. **24**(4): p. 641-651.
122. Zhou, P., et al., *Structural insights into the substrate specificity and transglycosylation activity of a fungal glycoside hydrolase family 5 beta-mannosidase.* Acta Crystallogr D Biol Crystallogr, 2014. **70**(Pt 11): p. 2970-82.
123. Wei, N., J. Quarterman, and Y.S. Jin, *Marine macroalgae: an untapped resource for producing fuels and chemicals.* Trends in Biotechnology, 2013. **31**(2): p. 70-77.
124. Stubbs, H.J., et al., *Hydrolase and transferase activities of the beta-1,3-exoglucanase of Candida albicans.* European Journal of Biochemistry, 1999. **263**(3): p. 889-895.
125. Suzuki, K., et al., *Characterization of recombinant yeast exo-beta-1,3-glucanase (Exg 1p) expressed in Escherichia coli cells.* Bioscience Biotechnology and Biochemistry, 2001. **65**(6): p. 1310-1314.