Chemoenzymatic Synthesis of Xylan and Glucoxylan Oligosaccharides for the Study of Xylan Biosynthetic Enzymes

Inaugural-Dissertation

to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by

Ignacio José Álvarez-Martínez

Berlin, 2023

The work in this dissertation was done between February 2018 and January 2023. First, until February 2020, at the Department of Biomolecular Systems of the Max Planck Institute of Colloids and Interfaces (MPIKG) in Potsdam, Germany. Later, between February 2020 and August 2022, at the Institute of Organic Chemistry of the University of Natural Resources and Life Sciences of Vienna (BOKU) in Vienna, Austria. The work presented in this dissertation has been supervised by Prof. Dr. Fabian Pfrengle.

1st reviewer:	Prof. Dr. Fabian Pfrengle
2nd reviewer:	Prof. Dr. Christoph A. Schalley
Date of defence:	10 May 2023

Acknowledgements

I would like to express my gratitude to Prof. Dr. Fabian Pfrengle for the ideas and guidance over these years of doctoral research and for granting us the opportunity of doing research at two different locations with different dynamics and input as we moved from the MPIKG in Potsdam to the BOKU in Vienna. I would also like to thank Prof. Dr. Christoph A. Schalley for kindly agreeing on reviewing this thesis.

I would like to thank all the current and former members of the group, with special mention to Colin and Deborah for the research that we had the chance to do together. I would like to also thank Max, Uwe, Manish, Agata and Sebastian for the enriching discussions and for contributing to the good environment in the lab at the MPIKG in Potsdam. I would also like to thank all the group members incorporated after our arrival at the BOKU in Vienna for all the specialized knowledge and for showing me the importance of a good environment for making science. Special mention to Matteo, Irene, Daniele, Alessandro, Angela, and Sara for the chemistry and all the time that we spent out of the lab; to Nino and Charlie, for the good dynamics, the advice we gave to each other and the few but very appreciated times out of the lab, and to Katharina, Nivedya, and Remya for all the fresh air that they have brought to the group.

I would also like to mention the good dynamics at the department of Biomolecular Systems at the MPIKG in Potsdam. Thank you to all my colleagues there for the good chemistry and for showing me that science relies strongly on collaboration. Special mention for the good scientific discussions, the everyday sharing, the board games and many more activities to Felix and Elena, and to all the other great colleagues for the science we could discuss in and out of Chezbriel.

I would like to thank my family, my parents Emilia and Ignacio, my brother Mario, and my sister-in-law Carla for all their support over these years, for the patience and the good guidance.

Lastly, I would like to thank the many excellent researchers and teachers in Valencia, Copenhagen and Prague with a contagious energy and enthusiasm with whom I have had the opportunity to meet and work before this thesis, who share a great appreciation for good science and that have also contributed to who I am today.

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me. Intellectual property of other authors has been marked accordingly. I also declare that I have not applied for an examination procedure at any other institution and that I have not submitted the dissertation in this or any other form to any other faculty as a dissertation.

Table of Contents

Summary

Glycans are essential molecules for living beings and, in land plants, they play an important structural role as part of the plant cell wall. A major component of the cell wall of many plants is the hemicellulose xylan, which is, after cellulose, the second most abundant glycan in the plant biomass.

Xylan polysaccharides are composed of a linear 1,4-β-linked backbone of xylosyl residues and are decorated with diverse substituents such as glucuronic acid and arabinosyl residues. Only recently, the cohesive role that xylan polysaccharides play between the other major components of the secondary cell wall, cellulose and lignin, is being recognised. Developing synthetic methodologies to produce xylan oligosaccharides that can serve as substrates for enzymatic studies is important for unravelling some of the gaps in the knowledge of its biosynthesis.

Another related xylose-containing polysaccharide was recently revealed in the cell walls of barley. The newly discovered hemicellulose glucoxylan is a linear 1,4-β-linked glycan composed of alternating xylosyl and glucosyl residues. This glycan was previously identified only in the cell wall of a sea lettuce species (a type of green algae). In barley, it is synthesised by genes in the monocot-specific *CslF* subfamily and is believed to play a structural role in the plant cell wall.

In Chapter 2 of this thesis, a chemoenzymatic synthesis of xylan dodecasaccharides for the study of xylan-modifying enzymes is described. The synthesis of these materials was designed around two central ideas, (a) the implementation of a divergent-convergent iterative synthetic strategy, and (b) the use of a glycosynthase for the glycosylation reactions. Because of the limited solubility of long unprotected glycans in water, two parallel syntheses were performed, one to obtain xylan oligosaccharides with a methyl group at the 3-OH of the reducing-end xylosyl residue and one to obtain xylan oligosaccharides without the methylation. Besides unprotected xylosyl acceptors, the xylan glycosynthase (XynAE265G) employed in these syntheses requires unprotected α-xylosyl fluorides as donors. These donors were equipped with a THP group at the 4-OH of the non-reducing end xylosyl residue

to prevent self-condensation of these molecules. The enzymatic glycosylation reactions with this donor gave exclusively the desired 1,4-β-linked glycosyl products, and no sugar-based side-products were detected.

The xylans oligosaccharides equipped with a methyl group showed superior solubility in comparison to the unprotected ones, so the methyl-substituted dodecasaccharide was used for biosynthetic studies using xylan-modifying enzymes. Treatment of this substrate with glucuronosyltransferase *At*GUX3 was found to install a single GlcA substituent at the substrate. The data obtained by MS/MS analysis of the reaction products is compatible with the substituent being installed at one of the two central residues of the substrate, which is in agreement with previously reported *in vivo* studies.

In Chapter 3, a small library of glucoxylan oligosaccharides was produce by chemical synthesis. These molecules have the same structural features as those reported in the literature to be synthesised by the two barley glycosyltransferases *Hv*CslF3 and *Hv*CslF10. The synthesis of these target molecules, ranging from di- to hexasaccharides, was designed to maximize the number of convergent and divergent steps in order to minimize the number of required synthetic transformations. As a temporary protecting group for chain elongation, a TBS group was used with great success to protect the 4-OH glycosylation site. The protected glucoxylan molecules in this small library are equipped with an anomeric azidopentyl linker, which may be used after reduction to the aminopentyl linker, for immobilization of these molecules as microarrays for biosynthetic studies and the characterization of antibodies.

Zusammenfassung

Glykane sind essenzielle Moleküle für Lebewesen und spielen bei Landpflanzen eine wichtige strukturelle Rolle als Teil der pflanzlichen Zellwand. Ein Hauptbestandteil der Zellwand vieler Pflanzen ist die Hemicellulose Xylan, die nach der Cellulose das zweithäufigste Glykan in der pflanzlichen Biomasse ist.

Xylan-Polysaccharide bestehen aus einem linearen 1,4-β-verknüpften Grundgerüst aus Xylosyl-Resten und sind mit verschiedenen Substituenten wie Glucuronsäure und Arabinosyl-Resten dekoriert. Erst in jüngster Zeit wird die kohäsive Rolle erkannt, die Xylan-Polysaccharide zwischen den anderen Hauptbestandteilen der sekundären Zellwand, Cellulose und Lignin, spielen. Die Entwicklung synthetischer Methoden zur Herstellung von Xylan-Oligosacchariden, die als Substrate für enzymatische Untersuchungen dienen können, ist wichtig, um einige Lücken im Wissen über ihre Biosynthese zu schließen.

Ein weiteres verwandtes xylosehaltiges Polysaccharid wurde kürzlich in den Zellwänden von Gerste entdeckt. Die neu entdeckte Hemicellulose Glucoxylan ist ein lineares 1,4-β-verknüpftes Glykan, das aus abwechselnden Xylosyl- und Glucosylresten besteht. Dieses Glykan wurde bisher nur in der Zellwand einer Seesalatart (einer Grünalgenart) nachgewiesen. In Gerste wird es von Genen der monokotylen CslF-Unterfamilie synthetisiert und spielt vermutlich eine strukturelle Rolle in der Pflanzenzellwand.

In Kapitel 2 dieser Arbeit wird eine chemoenzymatische Synthese von Xylan-Dodekasacchariden für die Untersuchung von Xylan-modifizierenden Enzymen beschrieben. Die Synthese dieser Materialien basiert auf zwei zentralen Ideen: (a) die Umsetzung einer divergent-konvergenten, iterativen Synthesestrategie und (b) die Verwendung einer Glycosynthase für die Glycosylierungsreaktionen. Aufgrund der begrenzten Löslichkeit langer ungeschützter Glykane in Wasser wurden zwei parallele Synthesen durchgeführt, eine zur Gewinnung von Xylan-Oligosacchariden mit einer Methylgruppe am 3-OH des Xylosyl-Restes am reduzierenden Ende und eine zur Gewinnung von Xylan-Oligosacchariden ohne die Methylierung. Neben ungeschützten Xylosyl-Akzeptoren benötigt die in diesen Synthesen eingesetzte Xylan-Glycosynthase (XynAE265G) ungeschützte α-Xylosylfluoride als Donatoren.

Diese Donatoren wurden mit einer THP-Gruppe am 4-OH des nicht reduzierenden Xylosylrestes versehen, um eine Selbstkondensation dieser Moleküle zu verhindern. Die enzymatischen Glykosylierungsreaktionen mit diesem Donor ergaben ausschließlich die gewünschten 1,4-β-verknüpften Glykosylprodukte, und es wurden keine zuckerbasierten Nebenprodukte nachgewiesen.

Die mit einer Methylgruppe versehenen Xylan-Oligosaccharide zeigten eine bessere Löslichkeit als die ungeschützten, so dass das methylsubstituierte Dodecasaccharid für Biosynthese-Studien mit Xylan-modifizierenden Enzymen verwendet wurde. Bei der Behandlung dieses Substrats mit der Glucuronosyltransferase AtGUX3 wurde festgestellt, dass ein einzelner GlcA-Substituent am Substrat installiert wird. Die durch MS/MS-Analyse der Reaktionsprodukte erhaltenen Daten sind mit der Installation des Substituenten an einem der beiden zentralen Reste des Substrats vereinbar, was mit zuvor berichteten In-vivo-Studien übereinstimmt.

In Kapitel 3 wurde eine kleine Bibliothek von Glucoxylan-Oligosacchariden durch chemische Synthese hergestellt. Diese Moleküle weisen die gleichen strukturellen Merkmale auf, wie sie in der Literatur für die Synthese durch die beiden Gersten-Glykosyltransferasen HvCslF3 und HvCslF10 beschrieben werden. Die Synthese dieser Zielmoleküle, die von Di- bis Hexasacchariden reichen, wurde so konzipiert, dass die Anzahl der konvergenten und divergenten Schritte maximiert wurde, um die Anzahl der erforderlichen synthetischen Transformationen zu minimieren. Als temporäre Schutzgruppe für die Kettenverlängerung wurde mit großem Erfolg eine TBS-Gruppe zum Schutz der 4-OH-Glykosylierungsstelle eingesetzt. Die geschützten Glucoxylanmoleküle in dieser kleinen Bibliothek sind mit einem anomeren Azidopentyl-Linker ausgestattet, der nach Reduktion zum Aminopentyl-Linker für die Immobilisierung dieser Moleküle als Mikroarrays für Biosynthesestudien und die Charakterisierung von Antikörpern verwendet werden kann.

List of Abbreviations

Ac	acetyl
AEC	Anion-exchange chromatography
AGA	automated glycan assembly
All	allyl
Ara	arabinosyl *or* arabinose
AX	arabinoxylan
AXY	altered xyloglucan
At	*Arabidopsis thaliana*
BB	building block
Bn	benzyl
Bu	butyl
Bz	benzoyl
DP	degree of polymerization
DUF	domain of unknown function
cat.	catalysis
CAZy	carbohydrate-active enzyme
COSY	correlation spectroscopy
CSA	camphorsulfonic acid
CSC	cellulose synthase complex
d	doublet
DAST	diethylaminosulfur trifluoride
DCM	dichloromethane
dd	doublet of doublets
ddd	doublet of doublet of doublets
DHP	3,4-dihydropyran
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DP	degree of polymerization
E	glutamic acid
ϵ	reaction yield
equiv	equivalent
ESI	electrospray ionization
Et	ethyl
f	furanosyl *or* furanose
Fuc	fucosyl *or* fucose
G	glycine
Gal	galactosyl *or* galactose
GATL	galacturonosyltransferase-like

GAUT galacturonosyltransferase

GAX glucuronoarabinoxylan

GH glycoside hydrolase *or* glycosidase

Glc glucosyl or glucose

GlcA glucuronic acid

Glu glutamic acid

Gly glycine

GSD ground state destabilization

GT glycosyltransferase

GUX xylan glucuronyltransferase (glucuronic acid substitution of xylan)

GX glucuronoxylan

GXM glucuronoxylan methyltransferase

HEK293 cells Human embryonic kidney 293

HSQC heteronuclear single quantum coherence

Hz Hertz

IDC iterative divergent–convergent

IRX irregular xylem

J coupling constant

LC liquid chromatography

m multiplet

MALDI matrix-assisted laser desorption/ionization

Me methyl

MS mass spectrometry

MS/MS tandem mass spectrometry

n number of moles

Nap (2-naphthyl)methyl

NIS *N*-iodosuccinimide

NMP nucleoside monophosphate

Os *Oryza sativa* (rice)

o/n overnight

p pyranosyl *or* pyranose

PCW primary cell wall

PE petroleum ether

PG protecting group

Ph phenyl

PMB *p*-methoxybenzyl

ppm part per million

Pr	propyl
Pv	*Panicum virgatum* (switchgrass)
Py	pyridine
Ptt	*Populus tremulata x tremuloides*
q	quartet
RT	room temperature
RWA	reduced wall acetylation
s	singlet
sat.	saturated
SEC	size exclusion chromatography
SCW	secondary cell wall
SPE	Solid-phase extraction
t	triplet
TBL	trichome birefringence-like
TBR	trichome birefringence
TBDPS	*tert*-butyldiphenylsilyl
TBS	*tert*-butyldimethylsilyl
*t*Bu	*tert*-butyl
TCAA	trichloroisocyanuric acid
TES	triethylsilyl
Tf	trifluoromethylsulfonyl (triflyl)
TFA	trifluoroacetic acid
TIPS	triisopropylsilyl
THF	tetrahydrofuran
THP	tetrahydropyranyl
TLC	thin layer chromatography
TMS	trimethylsilyl
TOF	time of flight
Tol	*p*-tolyl
Tr	triphenylmethyl (trityl)
TS	transition state
TSS	transition state stabilization
XyG	xyloglucan
XAT	xylan arabinosyltransferase
XOS	xylo-oligosaccharide
XSC	xylan synthase complex
Xyl	xylosyl *or* xylose

Chapter 1
General Introduction

1.1. Glycochemistry: from Mono- to Polysaccharides

1.1.1. Nomenclature and Structure of Glycans

The term “glycan” is defined by the IUPAC as a synonym of polysaccharide, although this term is called to slowly replace the more general and less scientifically accurate term “carbohydrate”. This family of polyhydroxylated carbonyl compounds has been historically defined by the formula $C_m(H_2O)_n$, however, nowadays the diversity of compounds

encompassed by this family extends way beyond the formula. Because compounds like glucose or sucrose are sweet, terms like “sugars” or “saccharides” as well as the prefix *glyc*- are often used in the field.[1]

In glycochemistry,[2] traditionally known as carbohydrate chemistry, the basic unit is the monosaccharide, and bigger structures are built by combination of them through glycosidic linkages, e.g., di-, tri-, tetra-…, oligo-, polysaccharides. The most basic monosaccharides are polyhydroxylated aldehydes (aldoses) or 2-ketones (ketoses) with three to nine carbon atoms; in nature, the most common examples of monosaccharides are with five (pentoses) and six (hexoses) carbon atoms. Modifications at the hydroxy groups, e.g., substitution of the H atom by another functional group (acetyl, benzyl, etc.) or replacement of the hydroxyl group itself by another group (e.g., NH_2, NHAc…), do not change drastically the reactivity of the molecule and therefore these compounds are still regarded as part of the glycan family. Other modifications such as reduction of the carbonyl group to a hydroxyl group have a greater impact on their chemistry and, thus, compounds like polyols (e.g., xylitol, inositol) are not considered to belong to the glycan family, but they are somewhat related.

The complexity of monosaccharides, or glycans in general, results to a great extent from the number of stereogenic carbon atoms that these molecules have, which can be represented in a Fischer projection (Figure 1.1). In a pentose like xylose, there are three stereogenic carbon atoms in the open-chain form of the molecule and one more in the cyclic form (Figure 1.1A). The relative configuration of all these stereogenic carbon atoms in monosaccharides is given by a prefix: for aldohexoses, gluc-, mann-, galact-, id-, etc.; for aldopentoses, xyl-, arabin-, etc.[1] Each prefix describes two enantiomeric forms of the same relative configuration, which can be further differentiated by the letters L and D (in small caps) (Figure 1.1A).[3] Upon ring closure, the additional stereocentre formed at the anomeric carbon atom can be then described by the Greek letters α and β, which are assigned by comparison of the anomeric carbon configuration with the configuration of the furthest stereocentre from the anomeric carbon (the reference atom).[3] The ring is formed by intramolecular attack of one of the hydroxyl groups on the carbonylic carbon atom, resulting in the

formation of a hemiacetal. If the H atom on the exocyclic oxygen is replaced by an organic fragment, the group is called acetal. The ring can be either a five- or a six-membered ring. Incorporation of the ring size into the nomenclature is done by the prefixes *furan*- (5-membered ring) and *pyran*- (six-membered ring) (Figure 1.1), due to their structural similarities with the heterocyclic compounds of the same name. Whether a five- or six-membered ring predominates in a monosaccharide depends on the compound structure and on the given physical conditions. Unlike in heterocyclic chemistry, the atom numbered as *one* within a monosaccharide is the anomeric carbon atom. As an example for the nomenclature of monosaccharides, the molecule in Figure 1.2A is named β-D-xylopyranose.

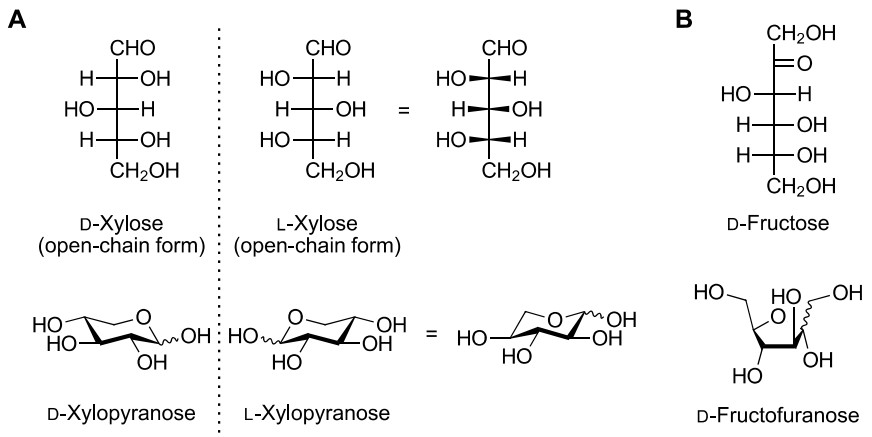


Figure 1.1. A. The two enantiomers of xylose (an aldopentose), in open-chain form (Fischer projection) and as a six-membered ring (chair conformation). **B.** D-Fructose (a ketohexose) in open-chain and five-membered ring form.

Beyond the configuration of the stereogenic carbon atoms, every monosaccharide exists predominantly in a certain conformation. The most common ones in pyranoses are the chair conformations $^4C_1$ and $^1C_4$ (Figure 1.2A), but others like the twist-boat conformation can also predominate under certain conditions, e.g., with very bulky substituents.[4]

The prefix *glyc*- or *glyco*- is often used to indicate a relationship with the glycan family. Glycoproteins or glycolipids are glycoconjugates consisting of a glycan (of indeterminate length) and a protein or a lipid. The prefix *glyc*- can often be substituted by the prefix of

the specific sugar configuration, e.g., α-D-mannosylation, galactolipic, etc. In some cases, the suffix *-an* can be added to a configuration prefix to denote a homoglycan, which is a glycan composed of only one type of monosaccharide, e.g., glucan. Glycans made of different monosaccharides, whether as a part of the backbone chain or as a side-chain substituents, are referred to as heteroglycans (Figure 1.2B).[3] In practice, names such as xylan or mannan, among others, are used to refer to heteroglycans with a backbone composed of a single type of glycosyl residue, in this case, xylosyl and mannosyl residues, respectively.

The two extremes of a glycan are referred to as the non-reducing end and the reducing end (Figure 1.2B). The former is the first ring with a sugar attached to the anomeric carbon (usually on the lefthand side in a structural representation) and the latter is the sugar ring that has a hemiacetal, or a non-sugar residue attached to the anomeric carbon (usually on the righthand side). The name originates from the ability of an aldehyde to act as a reducing agent, becoming itself oxidized to a carboxylic acid. Each monosaccharide that forms part of a larger structure can be referred to as a sugar residue, which is reminiscent of the amino acid residues in protein terminology.

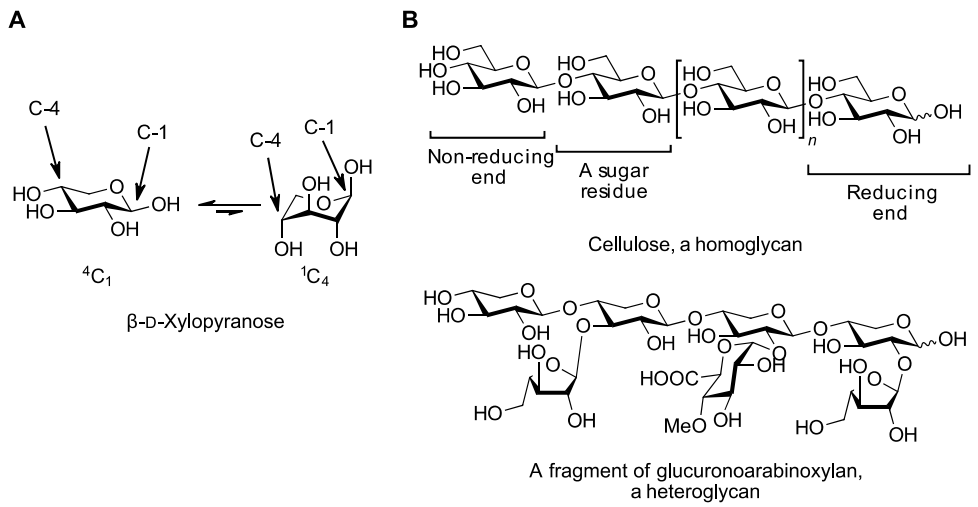


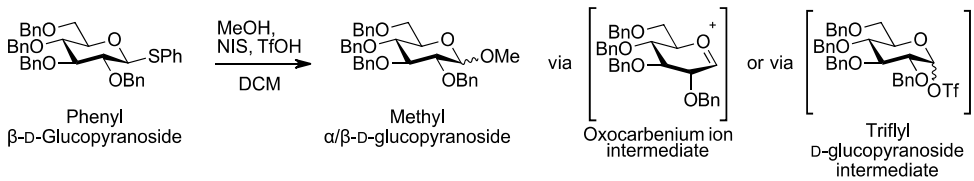
Figure 1.2. A. β-D-xylopyranose in two possible chair conformations. **B.** A homoglycan and a heteroglycan.

1.1.2. The Glycosylation Reaction

The hallmark of glycochemistry is the glycosylation reaction. By this reaction, a C-X bond (X = heteroatom) is formed between the anomeric carbon atom of a glycosyl group (the *glycosyl donor*) and another molecule or anion. If the nucleophile is a sugar molecule, it can be referred to as *glycosyl acceptor*. If the nucleophile is a non-sugar molecule, the glycosylation product is a *glycoside*, with the sugar residue known as the *glycone* and the non-sugar residue as the *aglycon*. Depending on the type of heteroatom acting as a nucleophile, the reactions can be classified as an *O*-glycosylation, *N*-glycosylation, *S*-glycosylation, etc. The term "*C*-glycosylation" is also sometimes found in the literature, however, these reactions form a homoatomic C-C bond, which lacks a polar component in the bond, and therefore the properties of these reaction products can differ from the other types of glycosylation products. In the laboratory, there is a plethora of possibilities to carry out such a transformation. For instance, in the present of an acceptor such as a molecule with a free hydroxyl group, treatment of a phenyl thioglycoside with NIS/TfOH in DCM at low temperatures or treatment of a glycosyl trichloroacetimidate with TMSOTf in DCM at low temperatures are among the most popular procedures to perform a chemical glycosylation reactions. In the first example, NIS acts as a promoter and TfOH as a catalyst, and in the second case, TMSOTf acts as a promoter. In living organisms, this transformation is usually catalysed by glycosyltransferases (GTs) using sugar nucleotides and other glycans (or other biomolecules) as glycosyl substrates.

The mechanism of a chemical glycosylation reaction in solution has been traditionally described by an S_N1 mechanism with the formation of a free oxocarbenium ion intermediate in solution (Scheme 1.1). However, more recent studies point towards the existence of covalent or contact ion pair (CIP) intermediate species predominating in solution after the leaving group departure. This species would be formed between the pyranosyl oxocarbenium ion and the counterion of the catalyst or the promoter, often a triflate ion. Under this scenario, the Curtin-Hammett principle can also play a role, what difficult the development of a stereochemical model for the prediction of glycosylation reactions. The measurement of bimolecular kinetics in glycosylation reactions or the limited number of observed

oxocarbenium ions in solution, which has been done in superacid media, are some of the evidence supporting a S_N2 reaction mechanism.[2,5]



Scheme 1.1. Example of a stereotypical chemical glycosylation in DCM between a phenyl thioglycoside and methanol using NIS and TfOH as the activation system.

The glycosylation reactions catalysed by enzymes often follow a different type of mechanism.[6] In this case, the stereoselectivity is provided, not only by the interaction between the substrate and the nucleophile, but also by the interactions between the enzyme's active site and both the substrate and the nucleophile.

1.2. Enzymes

1.2.1. Introduction and Historic View

Enzymes are protein-based biomolecules that catalyse chemical reactions. They are essential to living beings and take part in a vast number of chemical processes within the cell. Together with the protein units of the enzyme, often non-protein small molecules and/or metallic ions can be found; these complementary elements are known as cofactors. There is a number of RNA molecules, which are not proteins, that are also known to act on their own as catalysts; these are however known as ribozymes.

The beginnings of biochemistry are arguably the beginnings of the study of enzymes. At the end of the 19th century, Eduard Buchner found that extracts from yeast can catalyse the transformation of sugar into alcohol in the absence of living cells. Three decades later, in 1926, the first crystallization of an enzyme (urease) was done by James Summer, who also discovered that urease was purely a protein. With the subsequent crystallizations of a few

more enzymes by John Northrop and Moses Kunitz in the decade of 1930, the idea of enzymes being proteins was finally established.[7] All these discoveries precipitated the end of the old theory of vitalism among the great majority of scientists.

As early as in 1930, J. B. S. Haldane hypothesised in his book "Enzymes" that weak interactions between enzyme and substrate are responsible for the rate enhancement produced by enzymes: "the key does not fit the lock perfectly but exercises a certain strain on it".[7] Sixteen years later, in 1946, Pauling suggested that the active region of the enzyme is closely complementary to the activated complex rather than to the substrate.[7,8] From this first hypothesis to the current state of mechanistic enzymology, a lot of progress has been made.

1.2.2. Origins of the Catalytic Power of Enzymes

It is apparent that naturally evolved enzymes catalyse reactions by lowering the energy of the transition state, and not by ground state destabilization (GSD), as it has often been suggested.[9] The key idea behind the catalytic power of enzymes seems to be the polar preorganization of the enzyme's active site, which leads to the reduction of the energy of the transition state. The interactions that make this energy reduction possible seems to be of electrostatic nature.[10] However, there is a general agreement that all the precise details about the origin of the enormous catalytic power of enzymes remains unclear.[11,12]

The concept of polar preorganization of the enzyme's active site has been often misunderstood. The idea is that, during a non-catalysed reaction, there is an energy penalty paid due to the reorganization of the water molecules during the TS formation. Since the enzyme's active site is already oriented to accommodate the developing charges during the TS formation, this reorganization energy penalty is much smaller and therefore the overall energy barrier for the reaction is lower.[10,13] The interaction energy between the TS and its environment seems not to vary much between the reaction within the active site and the reaction in aqueous solution, but this energy is not equal to the amount of energy required to move from the reactant state to the TS.[13]

Over the years, many alternative hypotheses have been formulated to account for the catalytic power of enzymes. Concepts like entropic effects or the so-called near attack conformations are recurrent in the literature of the last decades and have been the source of heated debates. However, these ideas do not seem to agree with the available experimental/computational data. Even though there seem to be examples of artificial enzymes that catalyse reactions by GSD,[9] naturally evolved enzymes seem to catalyse reactions only by transition state stabilization (TSS).[9,10]

In the recent years, there has also been a lot of debate over the influence of protein dynamics on the catalytic power of enzymes. Protein dynamics does not take into consideration stochastic motions; only those motions that are non-statistical, which are not in thermal equilibrium with the environment, are to be considered.[11] Warshel and colleagues argue against the idea of protein dynamics contributing to catalysis, stating that "*there has not yet been any study that consistently established a connection between an enzyme's conformational dynamics and a significant increase in the catalytic contribution of the chemical step*" and that neither is there consistent theoretical studies supporting the idea.[11] However, several authors agree with the notion that the enzyme beyond the active site is more than a passive structure.[12,14]

1.2.3. Enzyme Classification

Enzymes are usually classified according to their function, i.e., the reaction that they catalyse.[15] For that, a numerical classification scheme for enzymes was developed, the Enzyme Commission number (EC number). With this system, every different enzymatic function is labelled with numerical code, which consists of the letters EC followed by four numbers separated by periods, e.g., EC 3.2.1.4. Every number from left to right gives a more refined description of the type of function represented by the EC number. For instance, in EC 3.2.1.4, the 3 in the first position corresponds to hydrolases; the 2 in the second position, to glycosylases (which is a type of hydrolase); the 4 in the third position, to glycosidases (which is a type of glycosylase); and in this way the function is progressively getting described more specifically. The limitation of this system is that it does not have significant

predictive value, as it does not reflect the structural and mechanistic characteristics of the enzymes that they are assigned to; it only describes a given function.[16] On top of that, a given enzyme can catalyse more than one reaction, and therefore more than one EC number will be assigned to that enzyme, e.g., some glycosidases (EC 3.2.1.*) can also act as glycosyltransferases (EC 2.4.*.*) because both reactions transfer a glycosyl group and therefore can mechanistically be very similar.[17]

Enzymes can also be classified in protein families according to their amino acid sequence.[18] The advantage of this classification system is that the general folds of the enzyme, the active site topology and the mechanisms of the catalysed reactions are often tightly interconnected with the amino acid sequence[16,19] and, so, the system has a certain predictive power.[20] A system for the classification of carbohydrate-active enzymes based on sequence similarities was developed by Henrissat and colleagues[18] and initially included the first thirty-five families of glycoside hydrolases. After several updates, the authors launched the Carbohydrate-Active Enzymes (CAZy) Database.[21] Thus, this online database contains the latest data regarding the classification of carbohydrate-active enzymes and related proteins according to their amino acid sequence.[16]

Some protein families are further classified as clans. Within a clan, families are assumed to have a common ancestry. In addition, they generally share the tertiary structure of the enzymes and the catalytic centres at the active sites as well as the reaction mechanism of the catalysed transformations.[16,22]

The classification system based on enzymatic function (EC number) and the one based on the amino acid sequence (available for carbohydrate-active enzymes on the CAZy database) are complementary. A given enzymatic function can often be found in several protein families, which is an example of convergent evolution. Simultaneously, in a given family, often different functions can be found (polyspecific family[18]), which is, in turn, an example of divergent evolution.[16,20] As an example, the enzymatic function represented by EC 3.2.1.4 (hydrolysis of 1,4-β-D-glucosidic linkages) can be found assigned to fourteen different GH families; and, in contrast, family GH1 has more than twenty-two EC number

assigned.[23] In contrast to a polyspecific family, a family that has only one enzymatic function is known as monospecific.[18]

As of today, there are seven different classes of proteins classified according to their function (Table 1.1).[24]

Class	Function	EC Number
Oxidoreductases	Redox reactions; transfer of H, O atoms or electrons from one substrate to another.	EC 1
Transferases	Transfer of a functional group from one substrate to another (glycosyl, methyl, acyl, amino, phosphate group…).	EC 2
Hydrolases	Hydrolysis of bonds, leading to the formation of two products.	EC 3
Lyases	Addition or removal of functional groups in a non-hydrolytic fashion (C-C, C-N, C-O or C-S may be cleaved/formed).	EC 4
Isomerases	Isomerization; intramolecular rearrangement of atoms.	EC 5
Ligases	Formation of a bond between two substrates with consumption of ATP (C-C, C-N, C-O, C-S).	EC 6
Translocases[25]	Movement of molecules or ions across membranes or separation of them within a membrane.	EC 7

Table 1.1. Classes of enzymes according to their function.

Carbohydrate-active enzymes (CAZy)

Among all existing enzymes, there is a number of them involved in the biosynthesis, modification and breakdown of glycans, which are known as carbohydrate-active enzymes.

The CAZy database[21,26] classifies the carbohydrate-active enzymes according to their amino acid sequence. As of today, it covers 5 different groups of families plus one associated (non-catalytic) group. The most common functions within the families are:

- Glycoside Hydrolase families (GHs): hydrolysis and/or rearrangement of glycosidic bonds
- Glycosyltransferase families (GTs): transfer of glycosides; formation of glycosidic bonds.
- Polysaccharide Lyase families (PLs): non-hydrolytic cleavage of glycosidic bonds.
- Carbohydrate Esterase families (CEs): hydrolysis of esters.
- Auxiliary Activities families (AAs): redox enzymes that act in conjunction with carbohydrate-active enzymes.
- Carbohydrate-Binding Modules (CBMs): non-catalytic binding.

These groups of families are sometimes also regarded as classes;[21] however, they should not be confused with the classes associated with function described earlier.

The group of Auxiliary Activities (AAs) has been the last one to be incorporated into the CAZy database. This group includes lytic polysaccharide mono-oxygenases (incl. the recently reclassified families previously known as CBM33 and GH61) as well as enzymes involved in the degradation of lignin. Even though lignin is not a glycan, this aromatic polymer is found, without variation, together with glycans in the plant cell wall.[27]

Carbohydrate-active enzymes involved in the formation and cleavage of *O*-glycosidic bonds

Glycosyltransferases, glycosidases (for C-O and C-S glycosidic linkages) and lyases constitute the machinery that nature employs for the formation and cleavage of glycosidic bonds.

The glycoside hydrolase families (GHs) consist predominantly of *O*-glycoside hydrolases (among EC 3.2.1.*), whose function is to catalyse the hydrolysis of *O*-glycosidic bonds. However, because the classification in protein families is sequence-based, there is a significant number of enzymes found within the GH families with other functions, e.g, hexosyltransferases, *S*-glycoside hydrolases (e.g., in GH1), phosphorylases (e.g., in GH13), lyases (e.g., in GH31), isomerases (in GH13), etc.[23] On top of that, some enzymes within the GH families can act as catalysts for both hydrolysis and transglycosylation reactions, which makes them

part of two different classes of enzymes at the same time. Both the hydrolysis of a glycosidic bond and the transglycosylation reaction are glycosyl transfer reactions that differentiate only in the kind of acceptor being glycosylated (water vs a glycan or biomolecule), and therefore, they can occur following very similar reaction mechanisms.[17] In addition to the sequence-based classification, some of the GH families are further grouped in eighteen different clans (GH-A to GH-R) according to both sequence and fold.[23]

The glycosyltransferase families (GTs) consist almost exclusively of glycosyltransferases,[28] which catalyse the transfer of a glycosyl group from a phosphate donor to a glycosyl acceptor or another biomolecule. The glycosyl donors employed by these enzymes are most frequently nucleotide diphosphate sugars, e.g., uridine diphosphate xylose (UDP-xylose), guanosine diphosphate mannose (GDP-mannose), etc.; however, nucleotide monophosphates, e.g., cytidine monophosphate *N*-acetylneuraminic acid (CMP-sialic acid) are somewhat common, too. Glycosyl transferases that require nucleotide sugars are often known as Leloir enzymes in honour of the Argentinian physician and biochemist L. F. Leloir. Lipid phosphate or pyrophosphate sugars, e.g., dolichol pyrophosphate mannose, are another example of glycosyl donors.[6] Unsubstituted phosphate or pyrophosphate sugars, e.g., glucose 1-phosphate, are also used as substrates by some GTs, and as a result, they are known as phosphorylases or pyrophosphorylases. Phosphorylases catalyse reversibly the formation of sugar 1-phosphates;[29,30] and there are examples of them not only in the GT families (e.g., GT4)[28] but also in the GH families (e.g., GH94)[23]. As acceptors, glycosyltransferases generally use other sugars, but also proteins, lipids, nucleic acids, or other small molecules can serve as acceptors as well.[6] GTs, on one hand, are classified according to their fold in three superfamilies: GT-A, GT-B, and GT-C; but there are also examples of GTs adopting other folds.[6,31] On the other hand, they are classified in clans according to both fold and activity: Clan I (GT-A fold; inverting), clan II (GT-B fold; inverting), clan III (GT-A fold; retaining), clan IV (GT-B fold; retaining).[32]

Endo- and *exo*-glycosidases

Glycoside hydrolases can also be classified according to the position in the glycan where the glycosidic bond is cleaved. *Exo*-glycoside hydrolases will hydrolyse only the last

glycosidic bond, generally, at the non-reducing end of a glycan.[17] *Endo*-glycoside hydrolases, however, will catalyse the hydrolysis of a glycosidic bond within a glycan chain.

As usual with many binary classifications, there are cases that do not fit any of the two categories. There are examples of *exo*-glycosidases with some degree of *endo*-glycosidase activity; or, in addition, there are endoglucanases that perform a random, initial attack followed by consecutive hydrolytic cleavages at one of the ends of the cellulose molecule before finally releasing the molecule.[33] The ability of an enzyme to perform several catalytic cycles before releasing the substrate is known as processivity.

1.2.4. Mechanism of GH- and GT-Catalysed Reactions

Reaction mechanism of glycoside hydrolase (GHs)

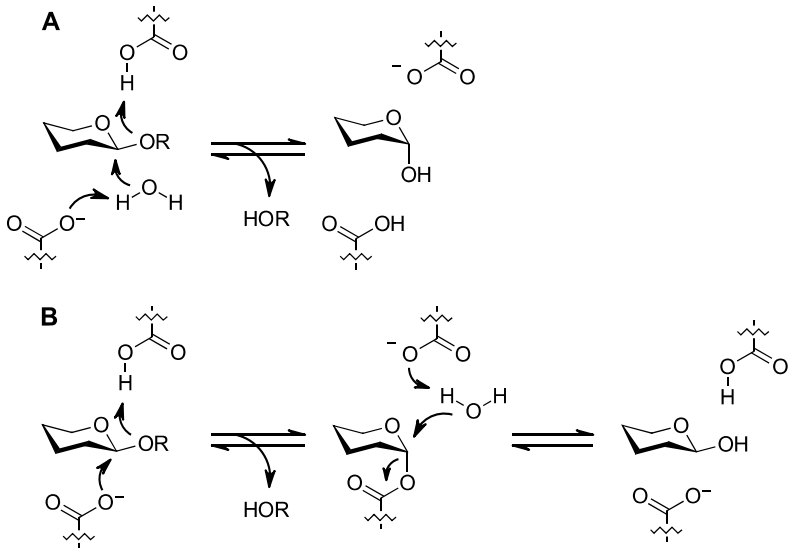
Glycoside hydrolases can also be classified into retaining and inverting enzymes depending on whether or not the product of the enzymatic reaction has a different configuration at the anomeric centre than the starting material. As it was initially proposed by Koshland,[34] both inverting and retaining enzymes have a classical reaction mechanism associated with them: single-displacement mechanism for inverting enzymes and double-displacement mechanism for retaining enzymes.[35] However, there are many examples of enzymes following variations of them.[36]

The classical single-displacement mechanism of inverting enzymes begins with the formation of a non-covalent enzyme-substrate complex (Scheme 1.2A). Then, with the assistance of a base catalyst, a nucleophilic attack by a water molecule at the anomeric centre will take place. That will induce the departure of the leaving group with assistance given by the proton donor catalytic centre. As a result, inversion of the configuration at the anomeric carbon atom has occurred.[37] This displacement step is concerted and resembles a classical S_N2 mechanism, with the additional participation of the catalytic centres of the active site. If the formation and cleavage of bonds in the TS has occurred to the same extent, it would be, in addition, a synchronous mechanism.

The classical double-displacement mechanism of retaining enzymes starts also with the formation of a non-covalent enzyme-substrate complex (Scheme 1.2B). Then, attack of the nucleophilic centre will induce the departure of the leaving group, which is assisted by a proton donor catalytic centre. As a result, a covalent glycosyl-enzyme species is formed.[38] This intermediate can also be of purely electrostatic nature, without the covalent bond between the enzyme and the glycoside.[39] With the assistance of the acid/base catalytic centre, a nucleophilic attack by a water molecule at the anomeric carbon will form the new glycosidic bond.[37] With the two inversions in the configuration of the anomeric carbon atom over the course of the whole transformation, the configuration is overall retained.

Aside from the catalytic centres, the active site provides the substrate with a range of electrostatic interactions that stabilize the developing charges during the TS formation. This stabilization lowers the energy of the TS, and thus enhancing the rate of the reaction.

Generally, the catalytic centres in the active sites of the enzymes in the glycoside hydrolase families are two carboxylic acids from glutamic or aspartic acid residues. Depending on the reaction mechanism, one can act as a general base catalyst or nucleophile/leaving group and the other as a general acid or general acid/base catalyst.[37,40] However, there are several examples of other functional groups or chemical species performing those catalytic roles.[36] As nucleophilic catalytic centre, there can be found phosphate ions (GH-Q, e.g., GH94), Zn associated with a cysteine residue (GH-P, e.g., GH127), the carbonyl oxygen atom from the C-2 acetamido group of the reaction substrate (GH-K, e.g., GH18), a tyrosine residue assisting a glutamic acid residue (GH-E, e.g., GH33), etc.[23] Similarly, a histidine residue (GH107) has been found to act as a proton donor catalytic centre.[23]



Scheme 1.2. A. Inversion by a single displacement mechanism. **B.** Retention by a double displacement mechanism.

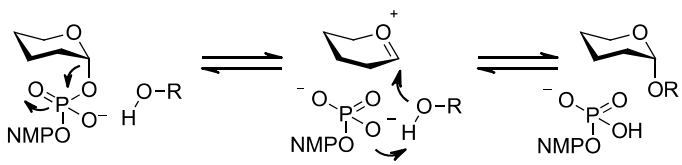
Reaction mechanism of glycosyltransferases (GTs)

Glycosyltransferases, like glycoside hydrolases, either retain or invert the stereochemistry of the anomeric centre at the glycosyl donor during the glycosyl-transfer reaction. Most inverting glycosyltransferases follow a single-displacement reaction mechanism similar to that found in inverting glycoside hydrolases.[6] However, most retaining glycosyltransferases, unlike retaining glycoside hydrolases, seem to follow a S_Ni-type (internal-nucleophilic-substitution-type) mechanism where the nucleophilic attack happens on the same side as where the leaving group departs (Scheme 1.3).[41–43] So, upon this first step, a short-lived ion pair intermediate will have been formed, which will be attacked by the nucleophilic acceptor. The existence of this mechanism is supported by the crystal structures of the enzyme-substrate tertiary complexes formed prior to the chemical reaction as well as by the absence of protein residues capable of forming a covalent glycosyl-enzyme intermediate.[41,44,45] A pentacoordinate transition state instead of the aforementioned short-lived intermediate would be very unlikely,[46] as simultaneous bond breaking and bond formation

on the same face is hardly depictable as a more energetically favourable scenario than a stepwise substitution through an intermediate.

The term S_Ni describes the mechanism of reactions such as the thermal decomposition of species such as alkyl sulfonyl chlorides or alkyl chloroformates.[47] However, this term applied to the mechanisms of reactions catalysed by GTs is somewhat unfortunate since, in this case, the acceptor is an external nucleophile and not an internal one.[6]

The classic double-displacement mechanism assigned to most inverting glycoside hydrolases is however still a viable option for the family GT6; this family possesses a nucleophilic residue in a suitable location to form the glycosyl-enzyme intermediate characteristic of the double-displacement reaction mechanism.[46,48]



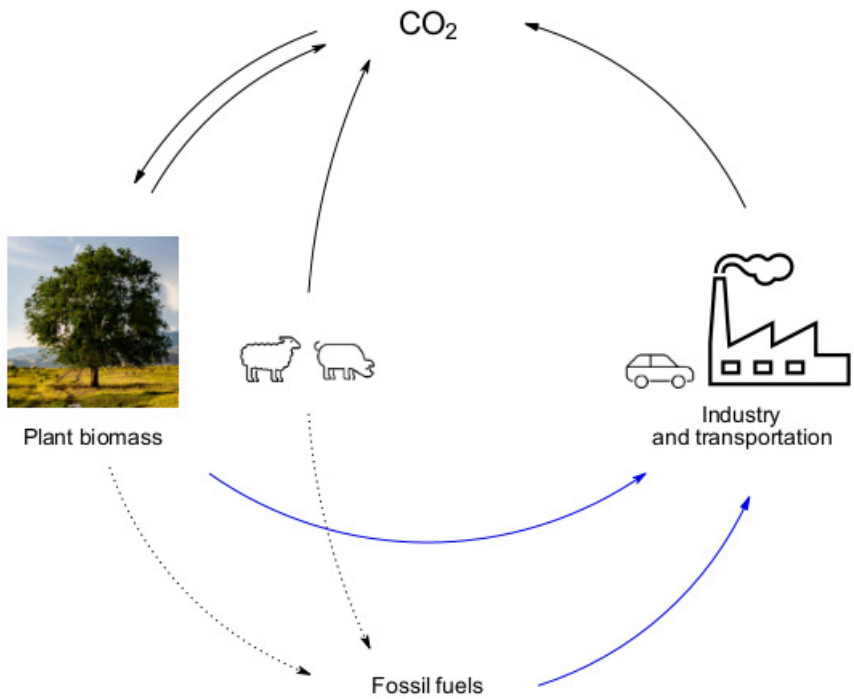
Scheme 1.3. S_Ni-type mechanism proposed for most retaining glycosyltransferases; NMP = nucleoside monophosphate.

1.3. Plants: Importance and Classification

1.3.1. The Importance of Plant Biomass

Plant biomass, incl. marine plants, accounts for 80% of world's biomass, and land plants are the dominating group.[49] Since the beginning of time, plant biomass has been used by humans as a source of energy to provide heat for survival and comfort, for building tools and houses, for cooking food, etc. With the industrial revolution in the 18th and 19th centuries, the use of wood[50] as a source of energy was progressively replaced by fossil fuels; for instance, in the iron industry, coal was introduced since its production requires less labour per heat unit. In the 19th century, the commercial exploitation of fossil fuels began.

Unlike plant biomass if harvested sustainably, fossil fuels cannot be replaced at a pace to keep up with consumption and therefore are considered a non-renewable source of energy. In addition, since fossil fuels are a source of carbon, burning these materials release carbon dioxide into the atmosphere. The increment of this greenhouse gas is rising the average global temperature, which can alter the climate and disrupt the fragile equilibrium on which Earth's life as known today depends.



Scheme 1.4. Simplified carbon cycle.

Instead of satisfying the increasing energy demand with fossil fuels, a more sustainable alternative is to obtain that energy directly from plant lignocellulosic biomass. Even though combustion of plant products equally releases CO_2 into the atmosphere (Scheme 1.4), the plant uses atmospheric CO_2 as a source of carbon during growth and therefore this released CO_2 will not alter the atmospheric composition as long as the amount of plant biomass on Earth remains constant.[51] The knowledge on the glycan composition of plants and glycan

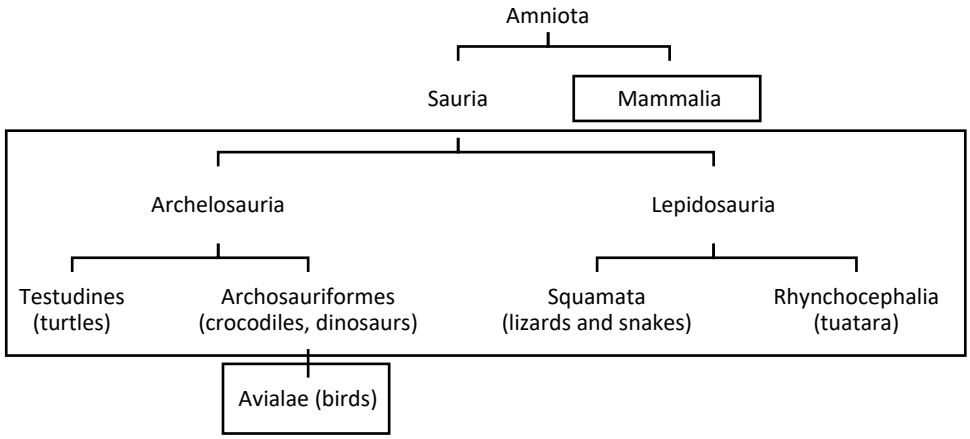
biosynthesis is essential both for producing modified plants that can satisfy industrial needs and for designing efficient industrial processes to transform lignocellulosic biomass into biofuels.[52] These second-generation biofuels take advantage of the existing energy infrastructure without the irreversible effects of fossil fuels.

1.3.2. Classification of Organisms: Taxonomy and Cladistics

Taxonomy, in general terms, is the science of classification. It originated from a desire to classify living and extinct organisms but can be applied to any system that comprises things or concepts. The schemes of classification are usually hierarchical, and the classification unit is the taxon (pl. taxa). In the early days, classification of organisms was based exclusively on morphological characteristics. However, with the development of DNA sequence analysis in the late 1970s, evolutionary relationships are built predominantly upon genetic data. The Sanger sequencing[53] has dominated the field of DNA sequencing for four decades until it has been progressively replaced by massive parallel sequencing. Morphological studies, however, are still a necessary pillar in phylogenetics since DNA data is not always available, e.g., fossils.[54] The collected phylogenetic data can then be presented in form of a tree, which would show the evolutionary history of the included organisms and their phylogenetic relationships.

In a phylogenetic tree, a clade or a monophyletic group is a group of species that includes a common ancestor and all its linear descendants. Mammals or birds are examples of clades (Scheme 1.5). In contrast, a paraphyletic group is a group that includes the common ancestor and its linear descendants except for one or more monophyletic subgroups. If there are several subgroups missing, the group can be regarded as polyparaphyletic. For instance, sauria is the clade that includes all modern reptiles plus extinct relatives and births; modern reptiles plus extinct relatives by themselves are paraphyletic with respect to birds (Scheme 1.5), which are a clade themselves. Lastly, a polyphyletic group is formed by groups of organisms that come from different ancestral sources, and which usually share a common characteristic as a result of convergent evolution. A polyphyletic group is not considered a

taxon by modern systematics. Warm-blooded animals form a polyphyletic group encompassed by mammals and birds (Scheme 1.5).



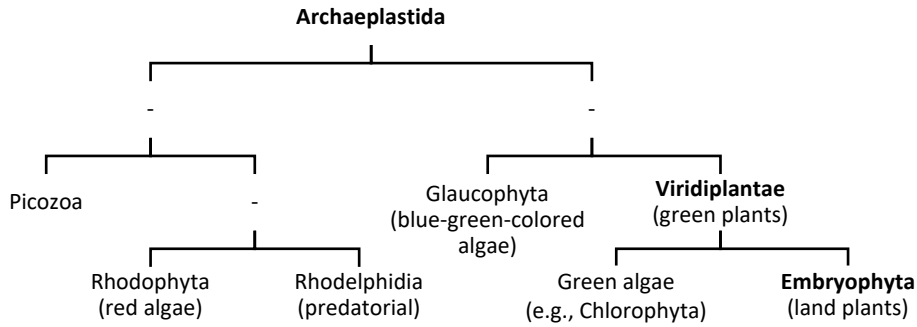
Scheme 1.5. Simplification of the phylogenetic relationships between modern reptiles, birds, and mammals**.**

1.3.3. The Problem with the Definition of "Plant"

Everyday language is often shaped by the macroscopic properties of the things that we get in contact with. Cladistics, however, analyses not only morphologic characteristics (both macro- and microscopic), but also genetic data to establish evolutionary relationships. This can sometimes give rise to incompatibilities where the terms proposed by cladistics do not satisfy the needs of everyday language, and vice versa. An example of this is the term "plant". A lay person may suggest that land plants, green algae and red algae are plants, but not the non-photosynthetic, unicellular organisms that are sometimes present in water. Cladistics, however, does not use specifically the term "plant" and establishes that red algae are more closely related to other organisms that might not resemble the everyday idea of a plant than they are to green algae or land plants (Scheme 1.6).

In the general scientific literature, the term "plant" is used in agreement with the relationships proposed by cladistics in a broader or narrower sense depending on the

characteristics being discussed. Plantae *sensu strictissimo* (strictest sense) refers to plants as a synonym of land plant (Embryophyta). Plantae *sensu stricto* (strict sense) refers to plants as a synonym of green plant (Viridiplantae) which comprises green algae and land plants but leaves out red algae and related organisms. Lastly, plantae *sensu lato* (broad sense) refers to plants as a synonym of the clade Archaeplastida, which includes red and green algae, land plants and other organisms, including non-photosynthetic, single-celled organisms as those in the genus *Rhodelphis*.



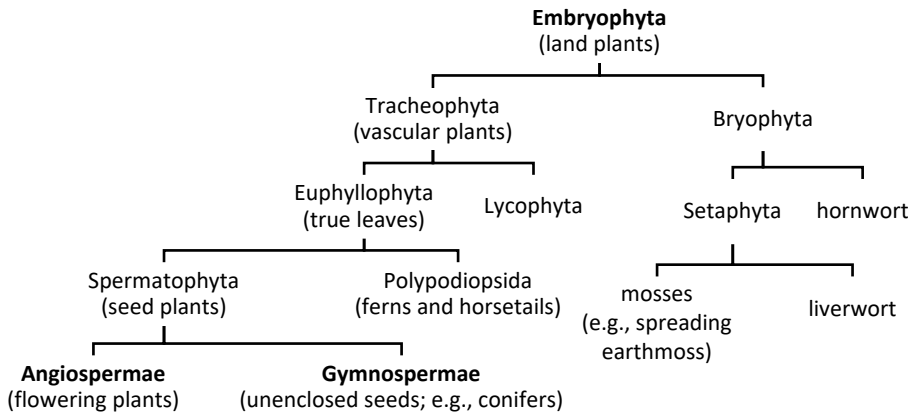
Scheme 1.6. Phylogenetic position of Viridiplantae and Embryophyta within Archaeplastida.[55] Green algae are paraphyletic with respect to land plants.

1.3.4. Land Plants and their Classification

Land plants, also called embryophytes or, collectively, Embryophyta, are multicellular eukaryotes with complex reproductive organs, most of which make 100% of the biomolecules that they need for growth and development. Glycans, lipids, protein, vitamins, etc. are synthesised within the plant from water, sunlight, carbon dioxide, oxygen, and minerals; therefore, most land plants are photoautotrophs.[56] Land plants are divided into two clades: Vascular plants (tracheophytes) and non-vascular plants (bryophytes) (Scheme 1.7).

Vascular plants

Vascular plants, also known as tracheophytes, are the most diverse and abundant type of land plant; only 7% of land plant species are non-vascular (bryophytes). The tracheophyte clade is characterized by having vascular tissue: xylem and phloem tissues. Through the xylem, water and minerals are transported from the roots to the stems and leaves; through the phloem, photosynthesis products are transported from the leaves to the rest of the plant.

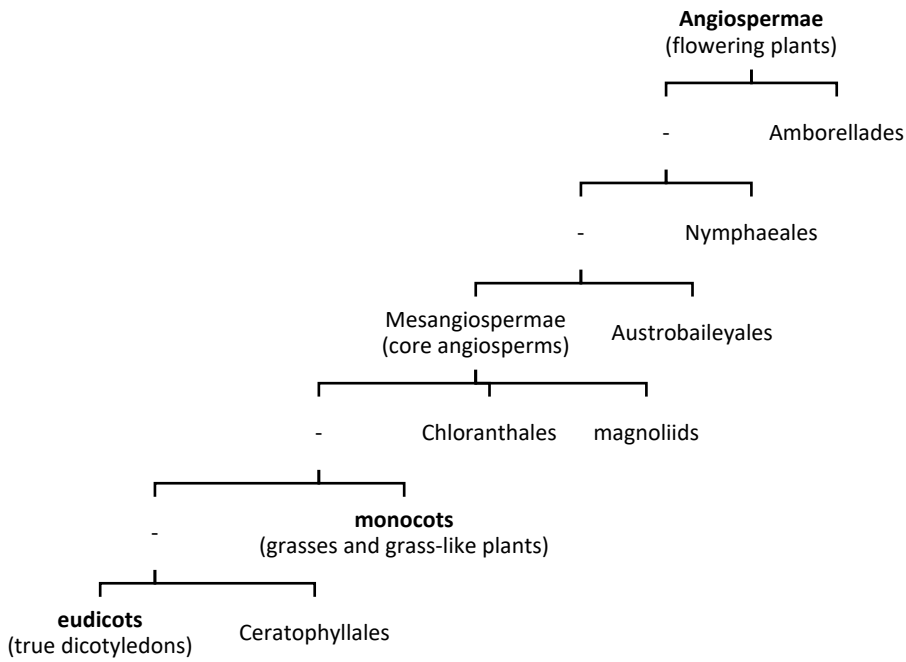


Scheme 1.7. Phylogenetic position of the angiosperms within the living land plants**.** This phylogenetic tree is done in agreement with the consensus for the classification of land plants;[57] in agreement with the Pteridophyte Phylogeny Group[58] for the term "polypodiopsida" to refer to ferns and horsetails. The exact relationships between liverworts, mosses, hornworts, and vascular plants are however still not fully clear.[59]

Angiosperms and gymnosperms

Seed-bearing plants (spermatophytes) form five extant divisions (Scheme 1.7). The clade of Angiosperms (flowering plants) and four other divisions traditionally known as gymnosperms (Cycadophyta, the cycads; Ginkgophyta, the ginkgo; Pinophyta, the conifers; Gnetophyta, the gnetophytes). The living members among gymnosperms are known as Acrogymnospermae, and the largest and most widely distributed group is the conifers.[60]

Angiosperm species comprise more than 90% of all land plant species.[61] Traditionally they have been divided into two groups (Scheme 1.8): monocots (grasses and grass-like plants) and dicots, depending on whether their seeds have one or two embryonic leaves (cotyledons), respectively. While monocots form a monophyletic group, dicots do not; they form a paraphyletic group with respect to monocots. The major clade within dicots is eudicots (Eudicotidae), which alone comprises 75% of all flowering plant species.[62]

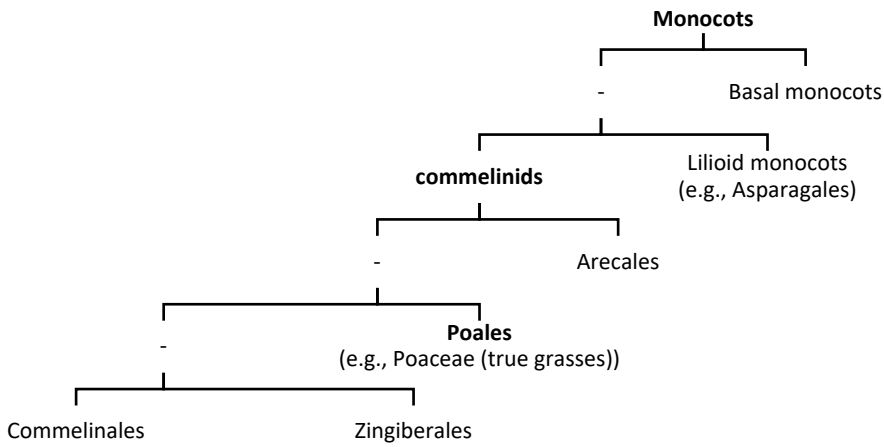


Scheme 1.8. Phylogenetic position of the monocots within the angiosperms in APG IV (2016).[63]

Monocots and eudicots

Monocots are the clade of grass and grass-like species of flowering plants. They are characterized by having one cotyledon (embryonic leaf), unlike all other flowering plants, which have two. This clade is divided between alismatid or basal monocots (Acorales and Alismatales), lilioid monocots (Petrosaviales, Dioscoreales, Pandanales, Liliales and Asparagales) and commelinids (Arecales, Poales, Zingiberales and Commelinales) (Scheme 1.9). Examples of monocots are the genus *Allium* (onion, garlic, leek…) and several crops in the

commelinids clade, like bananas, coconuts, pineapples, etc. Commelinids also contains the family of true grasses (Poaceae), which comprises cereal grains, bamboos, and several grassland species; examples of cereal crops are rice, wheat, maize, barley, oat, etc, with many of these belonging to the Pooideae subfamily. Commelinids are known for containing ferulic acid in their primary cell walls.



Scheme 1.9. Phylogenetic position of the Poales clade within the monocots in APG IV (2016).[63] Basal monocots are paraphyletic with respect to the rest of monocots, and lilioid monocots are too with respect to commelinids.

Eudicots, which are also known as true dicotyledons, are characterized by their tricolpate pollen structure. Within this clade, many economically important families can be found, e.g., Fabaceae (beans, peas, lentils), Fagaceae (beeches, chestnuts, oaks)**,** Brassicaceae (mustard, radish, cabbage, the model organism species *Arabidopsis thaliana*), etc. The model organism species *Populus trichocarpa,* in the genus *Populus,* also belongs within eudicots; in 2006, this species became the first tree species to have the genome sequenced.[64]

1.4. Glycans in the Plant Cell Wall

1.4.1. General Cell Wall Composition and Classification

Cell walls are present in most prokaryotes and eukaryotes, including plants and fungi, but are absent in animal cells, which are only surrounded by the cell membrane. In plants, cell walls provide the rigid network that gives the plant its structure, which prevents cytolysis and allows the cell to develop an internal pressure between 1 and 30 bar.[65] Cell walls are also the first line of defence against pathogens.[66] In addition to its rigid, shield-like characteristics, cell walls need to be metabolically active and act as a filter between the cell and its surroundings. Until cell expansion ceases, cell walls also need to be flexible.

The composition and structure of plant cell walls vary to some extent between species, types of cells, and stage of cell development. Generally, plant cell walls are made of cellulose microfibrils, tethered by cross-linking glycans and other components that, among other things, can provide additional stability and mechanical support. Plant cell walls are usually classified as primary (PCW) and secondary cell walls (SCW) according to the phase of cell development at which they form.[67]

The PCW develops during cell expansion and is deposited between the outermost layer (the middle lamella) and the innermost layer (the cell membrane). SCWs form after cell expansion has ceased and only in cell types that require additional robustness, e.g., water conductive and mechanically supportive tissues in certain plants. The SCW is deposited between the PCW and the innermost layer, the cell membrane.[68] If a tertiary cell wall is to exist, it should develop in another phase, e.g., by post-mortem deposition.[67] This classification correlates to a certain extent with the wall glycan composition. In addition, PCWs are characterized by containing variables amounts of pectic glycans while SCWs are instead lignified. As it is often the case, exceptions to this rule exist[69] because models are a good tool to facilitate communication and didactics[70] but very rarely can reflect the actual diversity.

Historically, plant cell wall glycans have been classified in three groups according to their extractability. Pectins,[71] which plays primarily a structural role in the PCW, can be extracted with dilute acidic solutions or calcium chelators; hemicelluloses[66] require strong alkaline treatment, and cellulose remains unextracted under those treatments.[72] However, this classification is rather vague. Chemically, pectins are very complex molecules, often made up of a galacturonan backbone chain (α-1,4-linked galacturonic acids), but at least seventeen different monosaccharide residues can form part of it.[73] Galactans, arabinans, arabinogalactans or callose ((1,3)-β-glucan) are often considered in this group too.[66] Hemicelluloses are xylans, xyloglucans, mannans, glucomannans and 1,3;1,4-β-D-glucan; they are non-cellulosic, non-pectic, 1,4-β-linked xyloses, glucoses, or mannoses.[66] Lastly, cellulose consists of unsubstituted, unramified chains of 1,4-β-linked D-glucan.

1.4.2. The Primary and Secondary Cell Walls

PCWs[74] are present in all plant cells, are relatively thin, highly hydrated, and are usually almost exclusively made of glycans. Unlike SCWs, PCWs form during cell expansion and therefore need to be extensible and incorporative to allow for morphogenesis while still being mechanically strong to resist turgor pressure.[68]

PCWs are classified stereotypically according to their composition as type I and type II.[75] Commelinid monocots possess a PCW type II and all the rest of flowering plants (dicots and non-commelinid monocots) have a PCW type I. Even though this classification can be useful, there are many exceptions that limit its applicability.[66]

SCWs[69] form only in tissues that requires additional mechanical support, e.g., tracheary elements (vessels and tracheids) and fibres. However, in plants like trees, they constitute most of the plant mass.[69] SCWs are generally composed of cellulose microfibrils made of 18 glucan chains[76,77] and hemicelluloses (xylans or mannans predominate) embedded in a matrix of lignin.

Lignin is a polyphenolic material characteristic of SCWs; PCWs are rarely lignified. The lignification, the deposition of lignin, is the final step of the SCW synthesis[69] and it improves

the wall resistance to compressive forces as well as limits the number of small molecules that can pass through.[78]

SCWs are diverse and play different specific structural roles within the plant. Often, they can be found subdivided in sublayers (S1, S2, S3) that have different cellulose microfibrils angles.[54] These layers can also sometimes be modified as a response to external factors. For instance, when an external force induces a severe curvature in a stem of a woody plant, a type of reaction wood forms to compensate for that change. In softwood (the wood of gymnosperms), the tissue is called compression wood and forms in the lower part of the title axis. In hardwood (the wood of non-monocot angiosperms), the tissue, which forms in the upper part of title axis, is called tension wood and is characterized by a gelatinous layer (G-layer) that forms in the fibres and allow for their contraction during maturation,[67] which may include a final step of lignification.[54] This example is just one of many that shows how dynamic SCWs are and the role that they play in the plant.

1.4.3. Hemicelluloses

Hemicelluloses[66] are ubiquitous to both PCWs and SCWs and can be defined as non-cellulosic, non-pectic, 1,4-β-linked glycans (e.g., xylans, xyloglucans, mannans, glucomannans, or 1,3;1,4-β-D-glucan). These glycans act as tethering agents in the wall by cross-linking cellulose microfibrils, and, in some cases, lignin, providing additional mechanical strength to the cell wall structure. For that reason, the term "cross-linking glycans" has been proposed to replace[79,80] the traditional term "hemicelluloses" (*hemi-* means *half*) that is used since these glycans were believed to be chemically and structurally related to cellulose.[81] However, it is not evident that all hemicelluloses are cross-liking glycans.

There are two mechanisms by which cross-linking glycans, e.g., xyloglucan (XyG) in the PCW of non-commelinid flowering plant, can tether cellulose microfibrils. One, by getting trapped during cellulose crystallization; and two, by hydrogen bond formation with the cellulose chain.

Without including the diversity of cells within a plant and consequently the diversity of PCW and SCW in the plant, there are two factors affecting cell wall chemical composition: type of plant and whether it is a PCW or a SCW. Scheller and Ulvskov[66] summarized the chemical composition of plant cell walls (Table 1.2; Figure 1.3A) for three types of plants: dicots, grasses (sometimes as representatives of commelinid monocots) and conifers (as the most important clade within gymnosperms).

Traditionally, XyG has been associated with the PCW and xylan polysaccharides with the SCW since those are the predominant glycans in the respective cell walls of dicots. In grasses, however, xylans are the main hemicellulose in both the PCW and the SCW. The composition of grass cell walls is considerably different to other angiosperms, as they contain much less XyG, and more xylan and 1,3;1,4-β-glucan, which is a clear evolutionary shift in cell wall composition.[82] SCWs in non-grass monocots are more similar to SCWs in eudicots than to SCWs in grass monocots.[83] In conifers, mannans are usually the most abundant hemicelluloses in the SCW (Table 1.2).

% (w/w) of glycans in the plant cell wall						
Glycans	**Dicots**		**Grasses**		**Conifers**	
	PCW	**SCW**	**PCW**	**SCW**	**PCW**	**SCW**
Xyloglucan (XyG)	20-25	Min.	2-5	Min.	10	–
Glucuronoxylan (GX)	–	20-30	–	–	–	–
Glucuronoarabinoxylan (GAX)	5	–	20-40	40-50	2	5-15
(Gluco)mannan	3-5	2-5	2	0-5	–	–
Galactoglucomannan	–	0-3	–	–	+	10-30
1,3;1,4-glucan	Abs.	Abs.	2-15	Min.	Abs.	Abs.

Table 1.2.[66] % (w/w) of glycans in the cell walls of different types of plants. Min, minor; Abs., absent; –, absent or minor; +, present but quantitative data not available.

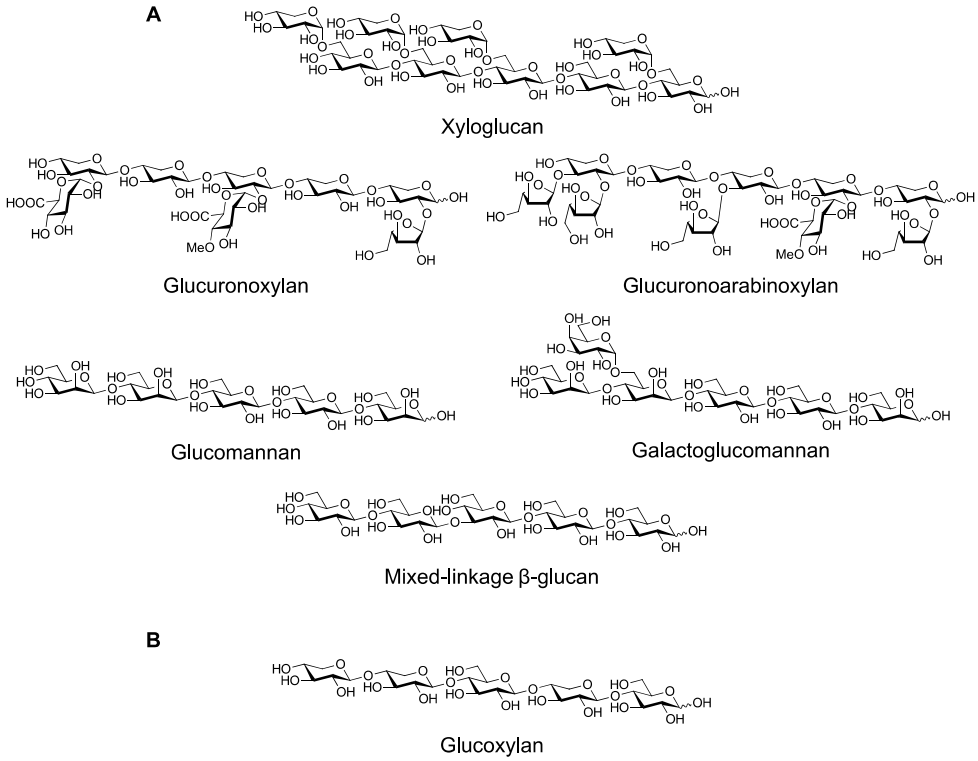


Figure 1.3. A. Representative fragments of the hemicelluloses from Table 1.2. **B.** A fragment of the newly discovered 1,4-β-linked glucoxylan.

Recently, a new type of glycan has been found in land plants, a linear glucoxylan (Figure 1.3B).[84] This polysaccharide, made of a backbone chain of β-1,4-linked xylosyl and glucosyl residues, which can be considered a hemicellulose, was first isolated[85] from the cell walls of *Ulva rigida*, a species of green algae, in the division of Chlorophyta, belonging to the genus of sea lettuces. Now it has also been found in land plants, precisely in barley.

1.4.4. Xylan Polysaccharides. Origin and General Structure

Xylan polysaccharides are characteristic of angiosperm SCWs, which are present in tissues that requires additional mechanical strength, e.g., conductive vessels and fibres. In angiosperms, xylan polysaccharides can constitute up to 35% of hardwood.[86] Plants with

inoperative genes, i.e., knockout genes, that are essential for xylan synthesis often display collapsed xylem vessels. However, this glycan precedes plant vascularisation.

The most common xylan backbone structure and the only one found in vascular plants (incl. in the ispaghula seed husk[87,88]) is comprised of 1,4-β-linked D-xylopyranosyl residues.[89] However, the term xylan technically refers to any cell wall glycan that has a backbone of xylosyl residues, even though only those containing 1,4-β-Xyl*p* linkages as part of the backbone are considered hemicelluloses.[66] Both 1,3-β-D-xylan and 1,3;1,4-β-D-xylan backbones have also been found in chlorophytes (a type of green algae) and rhodophytes (red algae).[90] Xylan polysaccharides within vascular plants are always decorated with either glycosyl groups (e.g., arabinosyl, glucuronosyl groups, etc.) or acyl groups (e.g., acetyl groups).[91] Examples of linear unsubstituted xylan polysaccharides, i.e., homoxylans, have been reported in chlorophytes and red algae, although they do not seem to occur in land plants.[90]

1.4.5. Xylan Decoration in Land Plants

Xylans	**Dicots**	**Grasses (Poaceae)**	**Conifers**
Sequence 1	✓	–	✓
Glucuronic acid (Glc*p*A)	✓	✓	✓
4-*O*-Me-Glc*p*A	✓	✓	✓
Ara*f* (terminal)	–	✓	✓
2-*O*-linked Ara*f*	–	✓	–
Acetyl esters	✓	✓	–
Feruloyl/*p*-coumaroyl esters	–	✓	–

Table 1.3.[92] Presence of several structural features on xylan polysaccharides of different types of land plants. Dicots are represented by poplar (*Populus trichocarpa)*, *Arabidopsis thaliana,* and psyllium; grasses, by switchgrass, the genus *Miscanthus*, corn, rice and *Brachypodium distachyon*; conifers, by the genus *Pinus*.

Xylan can be grouped according to the presence of three different substituents: α-L-arabinosyl (Ara*f*), (4-*O*-methyl)glucuronic acid ((Me)Glc*p*A) and acetyl (Ac) groups. As a very general rule, xylans from eudicots and early-branching angiosperms (traditionally known as dicots) do not generally contain Ara*f* substituents (AcGX); xylans from conifers and most other gymnosperms are not acetylated (GAX); and, within commelinids (incl. grasses), all substituents tend to be present on the xylan chain (AcGAX);[83] however, in monocots, xylans are rather structurally diverse.[83]

In eudicots and early branching angiosperms, xylan polysaccharides are found as AcGX (Table 1.3), with a (Me)Glc*p*A substituent generally every six, eight, ten or twelve xylosyl residues.[93] Acetyl groups are present every other residue, with all hydroxyl groups being available for acetylation, although *O*-3 positions of xylosyl residues containing (Me)Glc*p*A are virtually always acetylated.[94] These glycans have virtually no Ara*f* substituent,[95] and a length of approx. one hundred xylosyl residues.[92]

Interestingly, the substitution pattern of (Me)Glc*p*A in the SCWs of *Arabidopsis thaliana* has been found to be divided into two domains. The major domain, controlled by the glucuronyltransferase GUX1, is evenly decorated with a (Me)Glc*p*A substituent every eighth or tenth xylosyl residue. Acetyltransferase XOAT1 has been found to be essential for this domain to be formed.[96] In the minor domain, which is controlled by GUX2, (Me)Glc*p*A groups are however deposited every six, seven or eight residue and therefore without an even substitution pattern.[93] Different substitution patterns are believed to favour interactions with different components in the cell wall.

In coniferous, even though xylans are usually not the predominant hemicellulose, they are generally found as GAX (Table 1.3). These polysaccharides, which tend not to be acetylated, have two (Me)Glc*p*A substituent on *O*-2 and one Ara*f* on *O*-3 every twelve xylosyl residues.[95] Since Ara*f* substituents are found in small proportion, these glycans are also referred to as AGX in the literature.[83,92] Interestingly, in Gnetophyta, another clade in Gymnosperms, xylans have structural features typical of dicots: *O*-acetylation, no Ara*f* substituents, and low amounts of (Me)Glc*p*A.[95]

In monocots, xylan polysaccharides show extensive structural diversity.[83] Within commelinids, AcGAX is the characteristic xylan.[83] In the SCWs of grasses, the AcGAXs are predominantly substituted on *O*-3 with Ara*f*; these sidechains are often further substituted with Ara*f* or β-D-Xyl*p*. In the PCWs, however, Ara*f* substituents are attached to *O*-2, *O*-3 or to both. In the cell walls of starchy cereal grains, neutral AcAX can be found. Another example of the diversity of grass xylans is the presence of α-D-Gal*p* or α-L-Gal*p* in corn bran or corn fibre. Acetylation in grasses occurs to a lesser degree than in dicots, but Ara*f* groups are also sometimes acetylated on *O*-2. The presence of feruloyl or *p*-coumaroyl esters is a hallmark of xylan in grasses. These substituents are attached to Ara*f* groups on *O*-5 in both AcGAX and AcXG and can dimerize or trimerize xylan chains or crosslink them with lignin. This fact suggests that of feruloyl or *p*-coumaroyl groups might be the initiation point for lignification of cell walls in grasses.[92]

Methylation of Glc*p*A has also been suggested to be a key characteristic of tissues that need additional hydrophobic properties, since it is more abundant in vessels than in fibres in *Arabidopsis thaliana* and it seems to be only present in vascular plants.[92]

Sequence 1 is a reducing-end tetrasaccharide sequence found in xylan polysaccharides of eudicots, basal angiosperms, and gymnosperms (Table 1.3). In monocots, it has been detected in at least one species per order, including Poales, although not in the grass family (Poaceae).[83] Its function is still unknown, but it is believed to act as a primer to initiate xylan backbone elongation.[97.] It has the following sequence of sugar residues: →4)-β-D-Xyl*p*-(1→3)-α-L-Rha*p*-(1→2)-α-D-Gal*p*A-(1→4)-D-Xyl*p.*[98]

1.4.6. Xylan–Cellulose Interactions and the Importance of Domains

Xylan polysaccharides exist generally in two major conformations: a twofold and a threefold helical screw conformation. In solution or in the cell wall matrix, a threefold screw conformation seems to dominate while a twofold helical screw conformation is adopted for the adsorption onto cellulose or other glycans, favouring the interaction.[99]

In eudicots, early-branching angiosperms and gymnosperms, the spacing between substituents seems to be strictly controlled[95] and an even pattern where only every other xylosyl residue bears substitution is present in large parts of the xylan backbone. This pattern facilitates the xylan chain to adopt a twofold helical screw conformation which has all the substituents on the same face. This conformation has therefore a free face that enables the interaction with the hydrophilic surfaces of cellulose through hydrogen bonding.[96] The overall process of adsorption of a hemicellulose like xylan onto cellulose in water seems to be, however, entropy driven.[100]

Domains in xylan polysaccharides with other substitution patterns have also been identified in conifers[101] and eudicots.[93] These domains in xylan and in other glycans determine the glycan conformation leading to different types of interactions and thus different roles in the cell wall. The biosynthetic machinery seems to regulate these domains to control the function of the cell wall components.[96] Recently, much of the complexity of the molecular architecture of softwood has been finely depicted in an elaborated model by Terret and colleagues.[102]

In grasses, even substitution patterns at the xylan chains are of low abundance, and, because of that, the xylan-cellulose interaction seems to happen predominantly between the threefold helical screw conformation of xylan and amorphous cellulose. This interaction would be based on Van der Waals forces and to a lesser extent on hydrogen bonding and seems to be of smaller importance for the mechanical properties of SCW in grasses than the covalent xylan-lignin linkages based on feruloyl or *p*-coumaroyl substituents.[103] In agreement with that, the predominant interaction that maintain the cohesion of cellulose (and probably of other glycans) seems to be also London dispersion forces, even though hydrogen bonding plays a role as well.[104]

Between xylan and lignin in grasses, there seem to be an additional interaction that might be of importance. Xylan polysaccharides with a threefold or a distorted twofold screw conformation seem to also interact with lignin. This interaction is also influenced by the extent of lignin substitution since more highly methylated syringyl units of lignin seem to interact more strongly with xylan.[105] Because of the interactions with both lignin and cellulose,

xylan polysaccharides play a central cohesive role in the structure of SCWs in grasses, where this glycan acts as an adhesive between the other major components of the wall.

The thermodynamics of glycan aggregation

The forces that dominate the interaction between glycans seem to be London dispersion forces and, to a lesser extent, hydrogen bonding.[100,104] However, the overall process of glycan aggregation in water or the adsorption of a hemicellulose like xylan onto cellulose in water seems to be an entropy-driven process.[100] These phenomena can be explained in the following terms.

The change in the enthalpy of the system upon glycan aggregation in water (or glycan adsorption) would be very small since the value of the water–glycan chain interactions that would have been broken upon aggregation would be similar as the value of the glycan chain–glycan chain interactions formed at the aggregation event.

Before glycan aggregation, however, because the electrostatic interactions at the interface between the hydrophobic regions of the glycan and the water molecules are very weak, the water molecules at the interface would arrange themselves in an ice-like structure in order to maintain their overall electrostatic stabilization that water molecules experience with each other. If the electrostatic stabilization is maintained with fewer interactions, these interactions must be of stronger value and that would cause a limitation in the degrees of movement of those water molecules. This phenomenon is an example of the enthalpy–entropy compensation effect.[106]

Upon glycan aggregation, the hydrophobic regions, i.e., the non-polar surfaces, of the glycan chains would get in contact with each other and that would cause that the water molecules would no longer be in contact with those non-polar regions. This event would precipitate the dismantling of the ice-like structure that the water molecules adopted and, therefore, an increase in their entropy. This favourable change in entropy caused by the increase in the number of degrees of freedom of the water molecules would be of bigger value than the entropic penalty paid by immobilizing the glycan chains into the aggregate

or onto a molecular surface. Therefore, the glycan aggregation or adsorption event would be a thermodynamically favourable, entropy-driven process.[106]

The tendency of non-polar molecules or the non-polar surfaces of a molecule to aggregate in a solvent like water is known as the hydrophobic effect.[106]

1.4.7. Brief Introduction to Functional Genomics

Functional genomics is a field within molecular biology that focusses on identifying the functions and interactions of genes and proteins. Traditionally, these studies began with the identification of a phenotype in a group of organisms, and, from that point, efforts were put into discovering the gene that encoded the protein responsible for the atypical phenotype (forward genetics). Nowadays, with many organisms having their genome sequenced, the studies are often designed in the reverse order (reverse genetics): inducing a mutation (often to make the gene inoperative, i.e., a gene knockout) on the gene of interest and analysing the phenotypic effects. With this, information about the function of the protein encoded in the gene of interest is regularly obtained. Often, the same phenotype is observed with the knockout of different individual genes, which indicates to a certain extent that those genes, and therefore their encoded proteins, have related functions. The functional characterization of a protein is often a laborious work that reaches a milestone with the protein performing its function *in vitro*.

Another way of identifying potential gene functions is by sequence screening and identification of similar sequences in different genes; if the specific function of these sequences or domains is unknown, they are named DUF (domains of unknown function), e.g., DUF231, DUF579, etc. Identical or similar sequences in key sections of a protein lead to proteins with similar or related function.

Gene/protein nomenclature

Gene symbols are generally written in italicized uppercase (e.g., *IRX9*); the respective encoded protein will keep the symbol but will not be italicized (e.g., IRX9); mutant organisms

are named after the mutated gene, in italicized lowercase (e.g., *irx9* mutant).[107] Often a gene and the respective encoded protein under study (of unknown function) are named after the phenotype observed in the knockout organism, i.e., the mutant phenotype; examples in the biosynthesis of the plant cell wall are *IRREGULAR XYLEM, IRX; REDUCED WALL ACETYLATION, RWA; TRICHOME BIREFRINGENCE-LIKE, TBL,* etc. Several genes associated with the same mutant phenotype are often assigned the same name with different numbers, e.g., *IRX9*, *IRX10*, etc. If two genes in the same genome are homologs, they will be differentiated by the suffix -like or the letter L, e.g., *IRX9* and *IRX9L*. To differentiate genes or proteins between species a prefix in italics is often used, e.g., *At*IRX9L for an enzyme in *Arabidopsis thaliana*.

Gene homology is identified by statistical significance similarity, which reflects common ancestry;[108] sometimes percentage of identity is used as an indicator, but there are better indicators.[108] Gene homologs can be either paralogs or orthologs depending on whether they are the result of gene duplication or gene speciation, respectively; homologs in the same genome will always be paralogs.[109]

Model organisms

Genetic studies are often conducted on model organisms, which are species that have a set of characteristics that make research on them easier. In addition, working on the same species allows for cooperation between research groups which enables faster progress within a field (the so-called synergic effects). Discoveries made on model organisms, very often produce insight into related species.

Arabidopsis thaliana[110] is a small herbaceous flowering plant species in the eudicot clade, native to Eurasia and Africa. It is arguably the most important model organism in plant biology and much of the research conducted in plants begins with this species. Its popularity lies in being a eudicot, having a small genome and a short generation time as well as being self-fertile and easy to grow. Much of the information produced from research with this species is collected by the Arabidopsis Information Resource (TAIR).[111] This community

resource based on a non-profit corporation maintains an online database of genetic and molecular biology data for *Arabidopsis thaliana*.

For genetic studies conducted specifically on hardwood or eudicot trees, *Populus trichocarpa* is often the model organism of choice. In grasses, different crops of great economic value for food or energy e.g., rice, switchgrass, etc. are often chosen for research.

1.4.8. Xylan Biosynthesis

The plant machinery involved in the biosynthesis of xylan polysaccharides has been completely unknown until mid-2000s when the first enzyme candidates in *Arabidopsis thaliana* were proposed. In 1997, identification of the first genes within the *IRREGULAR XYLEM* (*IRX*) gene family was done.[112] In 2005, a landmark in the field of xylan biosynthesis was achieved with the identification of several novel genes in *Arabidopsis thaliana,*[113] some of which have been found to be involved in xylan biosynthesis. In the following years, more candidates were proposed[114,115] and the function of many of those in xylan biosynthesis were identified.

The biosynthesis of xylans[97] happens in the Golgi apparatus and, from there, xylans are secreted into the cell wall by vesicles. Xylan biosynthesis in eudicots can be divided between the enzymes involved in β-1,4-xylan backbone elongation, enzymes involved in the synthesis of the sequence 1, and enzymes involved in the introduction of substituents, such as acetyl and Glc*p*A substituents, and methylation of the Glc*p*A substituents.

Xylan backbone elongation

Xylan backbone synthesis in eudicots seems to begin with Sequence 1 acting as a primer and continue by elongation from the nonreducing end of the xylan chain. From genetic studies in *Arabidopsis thaliana*, it was found that xylan backbone synthesis (Table 1.4) involves a group of proteins from family GT47, IRX10/IRX10L, and two groups from family GT43, namely IRX9/IRX9L and IRX14/IRX14L.[98,116–120] In asparagus[121] and wheat,[122] IRX9, IRX10, and IRX14 have been shown to form a xylan synthase complex (XSC) where IRX10

would perform the catalytic function, while IRX9 and IRX14 are believed to play a structural role.[123] The role of IRX10/IRX10L is supported by the fact that only those were shown to have activity *in vitro*[124,125] and IRX10 lacks a transmembrane domain.[69] Because IRX14 has a DXD motif, it might play a role in substrate binding[123] since these motifs are known to coordinate by a divalent cation, e.g., Mn^{2+}, with phosphate groups of nucleotide donors.[126] Interestingly, the DXD motif is lacking in IRX9,[127] while is conserved in IRX9L.[125] While the pairs of paralogs mentioned, e.g., IRX9 and IRX9L, seem to be functionally redundant,[128] they might be active at different locations, e.g., PCW and SCW.[129]

Xylan backbone elongation

Enzyme name	Gene code (AGI code)	Function	Family	Notes
IRX10	At1g27440	Catalytic	GT47	First appearance[113] Characterization[124]
IRX10L / XYS1	At5g61840			First appearance[117,119] Characterization (as XYS1)[125] Claiming characterization[117]
IRX9	At2g37090	Probably structural	GT43	First appearance[113] Characterization[98] Function suggested[116]
IRX9L	At1g27600			First appearance[120]
IRX14	At4g36890			First appearance[114] Function suggested[116]
IRX14L	At5g67230			First appearance[120]

Table 1.4. Enzymes candidates involved in xylan backbone elongation.

GlcA installation on the xylan backbone

Xylan glucuronyltransferases (GUXs)				
Enzyme name	**Gene code (AGI code)**	**Function**	**Family**	**Notes**
GUX1	At3g18660	Addition of Glc*p*A; major domain in SCW[93]	GT8	In young stems and roots, *gux1gux2* mutants have unchanged amounts of PUX_5, but highly reduced amounts of $^{[m]}UX_4$.[129]
GUX2	At4g33330	Addition of Glc*p*A; minor domain in SCW[93]		
GUX3	At1g77130	Addition of Glc*p*A. Solely responsible for (Me)Glc*p*A decoration of PCW specific xylan[129]		In young stems and roots, *gux3* mutants have undetectable amounts of PUX_5, but only a small reduction in $^{[m]}UX_4$.[129]
GUX4	At1g54940	Potentially, addition of Glc*p*A		In young stems and roots, *gux4gux5* mutants have no change in xylan structure[129]
GUX5	At1g08990			

Table 1.5. Putative xylan glucuronyltransferases (GUXs).

Installation of α-1,2-GlcA substituents on the xylan backbone (Table 1.5) is catalysed by xylan glucuronyl transferases (GUXs) which belong to family GT8 (GUXs). *At*GUX1 and *At*GUX2 have been shown to produce different substitution patterns leading to different domains in xylans polysaccharides *in vivo*.

GlcA methylation

Methylation of the 4-OH of Glc*p*A (Table 1.6) is catalysed by glucuronoxylan methyltransferases (GXMs) belonging to family DUF579.[130,131]

IRX15/IRX15L are also members of the family DUF579 and therefore a similar function could be expected (Table 1.6). *Irx15 irx15l* mutants show decrease in xylan content in

Arabidopsis,[132] but the methylation of GXs increases.[130] Thus, although they seem to be involved in xylan synthesis or deposition, they do not seem to act as xylan methyltransferases. They might either play a structural rather than catalytic role in the methylation of GXs or be involved in the methylation of other glycans such as fucose or xylose residues in pectins, which might affect xylan content indirectly.[128]

DUF579				
Enzyme name	**Gene code (AGI code)**	**Function**	**Family**	**Notes**
GXM1	At1g09610	Methylation of 4-OH in Glc*p*A substituents on xylan	DUF579	First DUF579[113]
GXM2	At4g09990			[130,131]
GXM3	At1g33800			
IRX15	At3g50220	Unknown function in xylan synthesis or deposition		[132,133]
IRX15L	At5g67210			

Table 1.6. DUF579 family in *Arabidopsis thaliana*.

Synthesis of Sequence 1

At least four GTs (Table 1.7) are involved in the synthesis of the tetrasaccharide primer with the structure →4)-β-D-Xyl*p*-(1→3)-α-L-Rha*p*-(1→2)-α-D-Gal*p*A-(1→4)-D-Xyl*p* (Figure 1.4). Two GTs are a pair of paralogs belonging to family GT47, FRA8/F8H, which are phylogenetically related to putative xylosyltransferases IRX10/IRX10L, and therefore they might also be xylosyltransferases.[97] The two other GTs are from family GT8, IRX8 might be a galacturonosyltransferase[134] and PARVUS might link the first xylose residue to the acceptor.[97]

It is worth noting that, *irx8* and *fra8* mutants of *Arabidopsis thaliana*, which are lacking enzymes involved in the synthesis of Sequence 1 and therefore are almost depleted of such a sequence, still produce a certain amount of xylans, without the reducing-end sequence.[98] Thus, Arabidopsis might have a mechanism for de novo synthesis of xylan or there might be small amounts of xylobiose or a related glycan for the initiation of the xylan chain biosynthesis since IRX10L requires an acceptor of at least two sugar residues for its activity.[125]

Synthesis of Sequence 1				
Enzyme name	**Gene code (AGI code)**	**Function**	**Family**	**Notes**
FRA8/ IRX7	At2g28110	Maybe xylosyltransferases, since they are phylogenetically related to IRX10/IRX10L[97]	GT47	First appearance[113] Function suggested[98]
F8H/ IRX7L	At5g22940			Homolog of FRA8
IRX8/ GAUT12	At5g54690	Potentially acting as a galacturonosyltransferase[134]	GT8	First appearance[113] Function suggested[98]
PARVUS/ GATL1	At1g19300	Unknown, but it might link the tetrasaccharide to the acceptor[97]		116,135

Table 1.7. Enzymes candidates involved in xylan Sequence 1 synthesis in *Arabidopsis thaliana*.

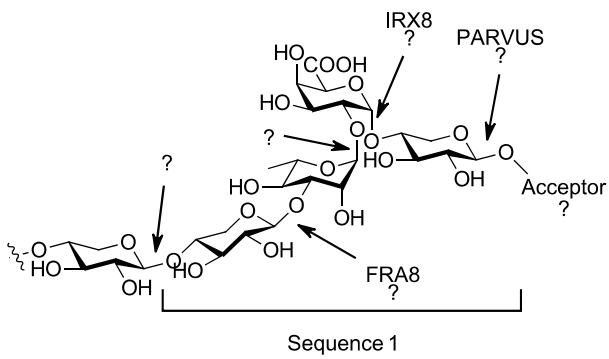


Figure 1.4. Sequence 1 and the enzyme candidates for its synthesis.[97]

Xylan acetylation

Acetylation – DUF231, RWA and AXY9

Enzyme name	Gene code (AGI code)	Function	Family	Notes
TBL29/ESK1/ XOAT1	At3g55990	2-*O*-monoacetylation[91]	DUF 231	Characterization (as XOAT1)[125]
TBL28/ XOAT2	At2g40150	[a 136]		Expressed preferentially in the stems, predominantly in both xylem cells and interfascicular fibre cells.[136]
TBL30/ XOAT3	At2g40160	2-*O*-monoacetylation and 3-*O*-monoacetylation.[136]		Expression is restricted to xylem cells.
TBL3/XOAT4	At5g01360	[a] with positional preference for 3-*O*-monoacetylation.[136]		[137]
TBL31/ XOAT5	At1g73140			
TBL32/ XOAT6	At3g11030	3-*O*-acetylation of 2-*O*-Glc*p*A-substituted residues.[136]		[138]
TBL33/ XOAT7	At2g40320			
TBL34/ XOAT8	At2g38320	[a] with positional preference for 3-*O*-monoacetylation.[136]		[139]
TBL35/ XOAT9	At5g01620	2,3-*O*-diacetylation.[136]		
AXY9	At3g03210	They might help transferring acetyl groups from donor to acetyltransferase.	-	[140]
RWAs 1/2/3/4	At5g46340 At3g06550 At2g34410 At1g29890	They might be Ac-donor transporters.	-	[141,142]

Table 1.8. Enzymes believe to be involved in xylan acetylation in *Arabidopsis thaliana*. [a]2-*O*-monoacetylation, 3-*O*-monoacetylation and 2,3-*O*-diacetylation.

Acetyl groups are an important motif in xylan polysaccharides and the suppression of enzymes involved in producing this decoration often causes collapsed vessels in the plant. In Arabidopsis, many of the proteins involved in the acetylation process (Table 1.8) belong to the family DUF231, which is also known as the TBL family, with 46 members in Arabidopsis, the TRICHOME BIREFRINGENCE (TBR) protein and 45 TBR-like (TBL) proteins.[115] All these proteins conserve a DUF231 domain and a TBL domain, which are associated with acetyltransferase activity with different substrates. Initially, these domains were believed to be plant-specific, but recently they have also been found in green algae.[143]

The DUF231 domain contains a DxxH motif (Asp–X–X–His) and the TBL domain, a GDS motif (Gly–Asp–Ser). The combination of the Ser from DxxH with the His and Asp from GDS forms the Ser-His-Asp catalytic triad, which is a well-known catalytic triad.[144] This motif has been found to be essential for the catalytic activity of these *O*-acetyltransferases.[91,136,145]

AXY9 (ALTERED XYLOGLUCAN) and RWAs (1-4) (REDUCE XYLAN ACETYLATION) are also believed to be involved in the acetylation of xylan. RWA potentially function as an acetyl-donor translocator[142] and AXY9 might act between a RWA and the acetyltransferase.[140] While TBL and AXY9 have a single transmembrane domain, RWA has ten.[146]

Xylan arabinosylation

There are several enzymes involved in the decoration of xylan in grasses with α-L-arabinosyl substituents. XATs, which belong to family GT61, are putative glycosyltransferases that mediate in the installation of the 3-*O*-Ara*f* substitutions (Table 1.9). GTs involved in the production of 2-O-Ara*f* or 2,3-O-Ara*f* substitutions have not yet been identified.[97]

Grass 3-O-*arabinosyltransferases*

Enzyme name	Gene code (AGI code)	Function	Family	Notes
XATs	-	Ara*f* substitution in grasses	GT61	[147,148]

Table 1.9. Putative grass 3-*O*-arabinosyltransferases.

1.5. Glycan Synthesis

1.5.1. Challenges of Glycan Synthesis

Glycans are complex molecules with many stereocentres that are synthesised, both in nature and in the laboratories, by forming glycosidic bonds between monosaccharides or larger building blocks (BB). Three aspects are key for the laboratory synthesis of glycans: (a) the choice of building blocks, which already include all stereocentres except for the one formed in the glycosylation reaction and, (b) and (c), the regioselectivity and stereoselectivity of the glycosylation reaction.

Building blocks with, for instance, the very abundant gluco-configuration are cheap and easy to make since a lot of methodologies have been developed for their synthesis and modification and the natural monosaccharides are abundant in nature. Less abundant sugar configurations, e.g., the so-called rare sugars, are much less investigated, and the monosaccharides are generally more expensive. Post-glycosylation modification of the sugar configuration is possible, but the number of efficient methodologies available is limited.

The regioselectivity and stereoselectivity of the glycosylation reactions depend on the reaction conditions chosen, incl. reagents, and on the structures of the building blocks. The regioselectivity is often controlled in purely chemical approaches by the use of protecting groups. The stereoselectivity, however, poses a bigger challenge since it is not only influenced by the structure of the BBs, i.e., protective group pattern (incl. leaving group) and sugar configuration, but also by the reaction conditions, e.g., solvent, temperature, reagents, etc. Unfortunately, in many cases, proper models to predict the stereochemical outcome of a glycosylation reaction are lacking and finding the conditions and the position and type of substituents to obtain the desired products in chemical approaches often implies some amount of trial and error.

In addition to the challenges in the individual reactions, there are several other key aspects to consider in the multi-step process of making glycans. Time is one of them, which is

influenced by the number of synthetic steps, reaction times, number of purification steps, etc. The economic costs are also important, which depends on the required time to make the glycans, but also on price and amounts of reagents, the efficiency of the reactions and the yields, etc.

1.5.2. Strategies for the Preparation of Glycans

Aside from the traditional approaches within chemical synthesis (e.g., tuning reaction conditions, application of convergent synthetic routs, one-pot procedures, etc.), there are further strategies to make glycan synthesis faster and cheaper. Automated solid-phase glycan synthesis and the use of enzymes are among the most promising strategies.

Automated glycan assembly (AGA)

Automated solid-phase synthesis of glycans (Automated Glycan Assembly; AGA) was first reported by Seeberger and colleagues in 2001.[149] Because automated solid-phase synthesis has revolutionized the field of peptide synthesis, there has been a great interest in applying the same principles to glycan assembly. However, because glycans are more complex molecules, this methodology cannot be routinely used to a comparable extent and its application is limited by the availability of suitable glycosylation strategies, which involves the choice of reactions conditions, BB structures, etc. Among the most challenging glycosylation reactions are β-mannosylation and α-sialylation reactions.[150] One of the important advantages of automated solid-phase glycan assembly (AGA) is that it enables the use of excess reagent to drive reactions to completion as any impurity can be easily washed away after each cycle. Once the synthesis is complete, the protected glycan is cleaved from the solid support and deprotected. Among the recent examples of AGA can be found a branched 151-mer mannan polysaccharide[151] or a library of arabinoxylan oligosaccharides synthesised previously in our group.[152,153]

Strategies for enzyme-mediated glycosylation reactions

The use of enzymes, such as glycosynthases or glycosyltransferases, is another promising strategy in glycan synthesis that is gaining in popularity in recent times. Particularly in industry, enzymes are regularly used in glycan-related transformations as they provide high control over the regio- and stereoselectivity of the reactions, enable the use of unprotected glycans in aqueous medium, and have low toxicity.[154] In the synthetic chemistry laboratories, their use is becoming more common, and, in recent years, there are many examples of enzyme-mediated syntheses of highly complex glycans;[155,156] however, it is far from being routinely stablished.

Glycosyltransferases are the tools that nature employs to catalyse glycosylation reactions. These enzymes have the potential to be heterologously expressed and used as reagents for glycan synthesis in the laboratory. However, finding a GT for a desired transformation with the desired substrate specificity may not always be possible and, since most of these enzymes are membrane bound, there might be difficulties in its expression and handling.[157] On top of that, the required nucleotide-phosphate glycosyl donors can sometimes be difficult to obtain, can be expensive and relatively unstable.

Strategies to use GHs as catalysts for glycosylation reactions

Glycoside hydrolases (together with lyases) form the machinery that nature employs for the cleavage of glycosidic bonds. However, because the hydrolysis and the transglycosylation reaction differ only in the acceptor being glycosylated, there is a number of enzymes in the GH families whose primary function is the catalysis of the transglycosylation reaction (transglycosylases). In addition, there are strategies to favour the transglycosylation reaction over hydrolysis (briefly reviewed below). Because GHs are more structurally diverse than GTs,[6] the chances of finding a GH with the desired substrate specificity are also potentially higher.

The transglycosylation/hydrolysis (T/H) ratio is an indicator of the ability of an enzyme to perform one transformation or the other.[17] One way to favour to some extent the transglycosylation reaction in enzymes with both functions is by increasing acceptor concentration.

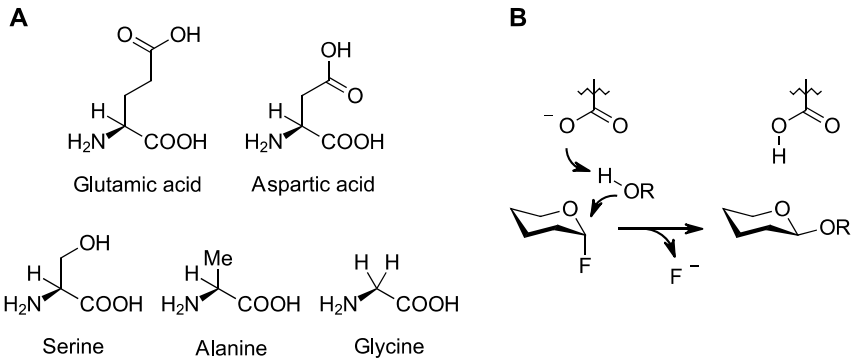
Once the glycosyl-enzyme intermediate is formed, higher concentrations of acceptor will increase the chances of the formation of the desired transglycosylation product.[17,158] Alternatively, a reactive donor can be used to increase the concentration of the glycosyl-enzyme intermediate formed from the glycosyl donor, and thus reducing the hydrolysis of the glycosyl product.[159] However, with the hydrolytic function of the enzyme remaining active, complex mixtures of products and low yields are often obtained.

A step further is to apply enzyme engineering. Inducing random or site-directed mutations in combination with direct evolution or *in silico* methods can increase the preference of an enzyme to catalyse the transglycosylation reaction over hydrolysis.[159] The main problem of this approach is that the enzyme, to a certain extent, still catalyse the hydrolysis of the desired glycosyl product.

Enzyme engineering can also be used to completely deactivate the enzyme's function to cleave glycosidic bonds and combine this modification with the use of unnatural activated glycosyl donors (e.g., fluoride donors) as reaction substrates. These artificial enzymes are known as glycosynthases and were first developed from retaining *exo*-glycoside hydrolases by the Withers group[160] and soon after from *endo*-glycoside hydrolases by the Planas group.[161] The advantage of *endo*-glycosidase-based glycosynthases is that they can accept glycosyl donors of more than one sugar residue. Eight years later, the first glycosynthase based on an inverting enzyme was also reported.[162]

To obtain glycosynthases, the nucleophilic centre of a glycoside hydrolase is modified by direct evolution. More precisely, the nucleophilic residue, often a glutamic or an aspartic acid, is replaced by a non-nucleophilic one, such as glycine (Scheme 1.10A). In glycosynthases derived from xylanases of family GH10, glycine has shown to induce higher glycosylation activities than other non-nucleophilic residues such as alanine or serine.[163] As a result of the modification in the catalytic centre, the enzyme lacks the ability to cleave glycosidic bonds and therefore to form the glycosyl-enzyme intermediate. However, the active site still maintains the topology that can stabilize the charges that develop during TS formation. So, once an activated donor enters the active site of the glycosynthase, the donor is accommodated in an orientation and conformation that resembles the intermediate formed

in the natural hydrolase. Given the high reactivity of the donor and the stabilising interactions that the enzyme provides, a glycosylation reaction with an incoming glycosyl acceptor will take place, often in a highly regio- and stereoselective manner, as would have happened with the natural enzyme (Scheme 1.10B).



Scheme 1.10. A. Common catalytic residues: Glutamic acid (Glu, E), aspartic acid (Asp, D); common non-nucleophilic residues: serine (Ser, S), alanine (Ala, A), glycine (Gly, G). **B.** Mechanism of a reaction mediated by a glycosynthase based on a retaining hydrolase.

1.6. References

(1) Seeberger, P. H. Monosaccharide Diversity. In *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press: New York, 2022.

(2) Crich, D. En Route to the Transformation of Glycoscience: A Chemist's Perspective on Internal and External Crossroads in Glycochemistry. *J Am Chem Soc* **2021**, *143*, 17–34. https://doi.org/10.1021/jacs.0c11106.

(3) McNaught, A. D. Nomenclature of Carbohydrates. *Pure & Appl. Chem.* **1996**, *68* (10), 1919–2008.

(4) Pedersen, C. M.; Nordstrøm, L. U.; Bols, M. "Super Armed" Glycosyl Donors: Conformational Arming of Thioglycosides by Silylation. *J Am Chem Soc* **2007**, *129* (29), 9222–9235. https://doi.org/10.1021/ja071955l.

(5) Santana, A. G.; Montalvillo-Jiménez, L.; Díaz-Casado, L.; Corzana, F.; Merino, P.; Cañada, F. J.; Jiménez-Osés, G.; Jiménez-Barbero, J.; Gómez, A. M.; Asensio, J. L. Dissecting the Essential Role of Anomeric β-Triflates in Glycosylation Reactions. *J Am Chem Soc* **2020**, *142* (28), 12501–12514. https://doi.org/10.1021/jacs.0c05525.

(6) Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. Glycosyl Transferases: Structures, Functions, and Mechanisms. *Annu Rev Biochem* **2008**, *77*, 521–555. https://doi.org/10.1146/annurev.biochem.76.061005.092322.

(7) Blow, D. So Do We Understand How Enzymes Work? *Structure* **2000**, *8* (4), R77–R81.

(8) Kraut, J. How Do Enzymes Work? *Science (1979)* **1988**, *242*, 533–540.

(9) Frushicheva, M. P.; Cao, J.; Chu, Z. T.; Warshel, A. Exploring Challenges in Rational Enzyme Design by Simulating the Catalysis in Artificial Kemp Eliminase. *Proc Natl Acad Sci U S A* **2010**, *107* (39), 16869–16874. https://doi.org/10.1073/pnas.1010381107/-/DCSupplemental.

(10) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem Rev* **2006**, *106* (8), 3210–3235. https://doi.org/10.1021/cr0503106.

(11) Warshel, A.; Bora, R. P. Perspective: Defining and Quantifying the Role of Dynamics in Enzyme Catalysis. *Journal of Chemical Physics* **2016**, *144* (18), 180901–180917. https://doi.org/10.1063/1.4947037.

(12) Nagel, Z. D.; Klinman, J. P. A 21st Century Revisionist's View at a Turning Point in Enzymology. *Nat Chem Biol* **2009**, *5* (8), 543–550. https://doi.org/10.1038/nchembio.204.

(13) Jindal, G.; Warshel, A. Misunderstanding the Preorganization Concept Can Lead to Confusions about the Origin of Enzyme Catalysis. *Proteins: Structure, Function and Bioinformatics* **2017**, *85* (12), 2157–2161. https://doi.org/10.1002/prot.25381.

(14) Markin, C. J.; Mokhtari, D. A.; Sunden, F.; Appel, M. J.; Akiva, E.; Longwell, S. A.; Sabatti, C.; Herschlag, D.; Fordyce, P. M. Revealing Enzyme Functional

Architecture via High-Throughput Microfluidic Enzyme Kinetics. *Science (1979)* **2021**, *373* (6553), 411. https://doi.org/10.1126/science.abf8761.

(15) Tipton, K.; McDonald, A. *A Brief Guide to Enzyme Nomenclature and Classification*. https://iubmb.org/wp-content/uploads/sites/10116/2018/11/A-Brief-Guide-to-Enzyme-Classification-and-Nomenclature-rev.pdf.

(16) Davies, G. J.; Sinnott, M. L. Sorting the Diverse: The Sequencebased Classifications of Carbohydrateactive Enzymes. *Biochem (Lond)* **2008**, *30* (4), 26–32.

(17) Bissaro, B.; Monsan, P.; Fauré, R.; O'Donohue, M. J. Glycosynthesis in a Water-world: New Insight into the Molecular Basis of Transglycosylation in Retaining Glycoside Hydrolases. *Biochemical Journal* **2015**, *467* (1), 17–35. https://doi.org/10.1042/BJ20141412.

(18) Henrissat, B. A Classification of Glycosyl Hydrolases Based on Amino Acid Sequence Similarities. *Biochemical Journal* **1991**, *280*, 309–316.

(19) Gebler, J.; Gilkes, N. R.; Claeyssens, M.; Wilson, D. B.; Beguin, P.; Wakarchuk, W. W.; Kilburn, D. G.; Miller, J. R. C.; Warren, R. A. J.; Withers, S. G. Stereoselective Hydrolysis Catalyzed by Related Beta-1,4-Glucanases and Beta-1,4-Xylanase. *The Journal of Biologogical Chemistry* **1992**, *267* (18), 12559–12561.

(20) Henrissat, B.; Daviest, G. Structural and Sequence-Based Classification of Glycoside Hydrolases. *Curr Opin Struct Biol* **1997**, *7*, 637–644.

(21) *Carbohydrate-Active Enzymes Database*. http://www.cazy.org/ (accessed 2022-07-18).

(22) Henrissat, B.; Bairoch, A. Updating the Sequence-Based Classification of Glycosyl Hydrolases. *Biochemical Journal* **1996**, *316*, 695–696.

(23) *Glycoside Hydrolase family classification*. http://www.cazy.org/Glycoside-Hydrolases.html (accessed 2022-07-18).

(24) McDonald, A. G.; Tipton, K. F. Enzyme Nomenclature and Classification: The State of the Art. *FEBS Journal* **2021**. https://doi.org/10.1111/febs.16274.

(25) Tipton, K. *Translocases (EC 7): A New EC Class*; 2018. https://iubmb.org/wp-content/uploads/sites/10116/2018/10/Translocases-EC-7.pdf (accessed 2022-07-11).

(26) Drula, E.; Garron, M. L.; Dogan, S.; Lombard, V.; Henrissat, B.; Terrapon, N. The Carbohydrate-Active Enzyme Database: Functions and Literature. *Nucleic Acids Res* **2022**, *50*, D571–D577. https://doi.org/10.1093/nar/gkab1045.

(27) Levasseur, A.; Drula, E.; Lombard, V.; Coutinho, P. M.; Henrissat, B. Expansion of the Enzymatic Repertoire of the CAZy Database to Integrate Auxiliary Redox Enzymes. *Biotechnol Biofuels* **2013**, *6* (41). https://doi.org/10.1186/1754-6834-6-41.

(28) *Glycosyltransferase family classification*. http://www.cazy.org/GlycosylTransferase-family (accessed 2022-07-18).

(29) Nakai, H.; Kitaoka, M.; Svensson, B.; Ohtsubo, K. Recent Development of Phosphorylases Possessing Large Potential for Oligosaccharide Synthesis. *Curr Opin Chem Biol* **2013**, *17* (2), 301–309. https://doi.org/10.1016/j.cbpa.2013.01.006.

(30) Awad, F. N. Glycoside Phosphorylases for Carbohydrate Synthesis: An Insight into the Diversity and Potentiality. *Biocatal Agric Biotechnol* **2021**, *31*. https://doi.org/10.1016/j.bcab.2020.101886.

(31) Taujale, R.; Zhou, Z.; Yeung, W.; Moremen, K. W.; Li, S.; Kannan, N. Mapping the Glycosyltransferase Fold Landscape Using Interpretable Deep Learning. *Nat Commun* **2021**, *12*, 5656. https://doi.org/10.1038/s41467-021-25975-9.

(32) Coutinho, P. M.; Deleury, E.; Davies, G. J.; Henrissat, B. An Evolving Hierarchical Family Classification for Glycosyltransferases. *J Mol Biol* **2003**, *328* (2), 307–317. https://doi.org/10.1016/S0022-2836(03)00307-3.

(33) Tomme, P.; Kwan, E.; Gilkes, N. R.; Kilburn, D. G.; Antony, R.; Warren, J. Characterization of CenC, an Enzyme from Cellulomonas Fimi with Both Endo-and Exoglucanase Activities. *J Bacteriol* **1996**, *178* (14), 4216–4223.

(34) Koshland, D. E. Stereochemistry and the Mechanism of Enzymatic Reactions. *Biological Reviews* **1953**, *28* (4), 381–479.

(35) Zechel, D. L.; Withers, S. G. Glycosidase Mechanisms: Anatomy of a Finely Tuned Catalyst. *Acc Chem Res* **2000**, *33* (1), 11–18. https://doi.org/10.1021/ar970172.

(36) Jongkees, S. A. K.; Withers, S. G. Unusual Enzymatic Glycoside Cleavage Mechanisms. *Acc Chem Res* **2014**, *47* (1), 226–235. https://doi.org/10.1021/ar4001313.

(37) Mccarter, J. D.; Withers, S. G. Mechanisms of Enzymatic Glycoside Hydrolysis. *Curr Opin Struct Biol* **1994**, *4*, 885–892.

(38) White, A.; Tull, D.; Johns, K.; Withers, S. G.; Rose, D. R. Crystallographic Observation of a Covalent Catalytic Intermediate in a β-Glycosidase. *Nat Struct Mol Biol* **1996**, *3*, 149–154.

(39) Tanaka, I.; Nishinomiya, R.; Goto, R.; Shimazakic, S.; Chataked, T. Recent Structural Insights into the Mechanism of Lysozyme Hydrolysis. *Acta Crystallogr D Struct Biol* **2021**, *77*, 288–292. https://doi.org/10.1107/S2059798321000346.

(40) Vuong, T. v.; Wilson, D. B. Glycoside Hydrolases: Catalytic Base/Nucleophile Diversity. *Biotechnol Bioeng* **2010**, *107* (2), 195–205. https://doi.org/10.1002/bit.22838.

(41) Albesa-Jové, D.; Mendoza, F.; Rodrigo-Unzueta, A.; Gomollón-Bel, F.; Cifuente, J. O.; Urresti, S.; Comino, N.; Gómez, H.; Romero-García, J.; Lluch, J. M.; Sancho-Vaello, E.; Biarnés, X.; Planas, A.; Merino, P.; Masgrau, L.; Guerin, M. E. A Native Ternary Complex Trapped in a Crystal Reveals the Catalytic Mechanism of a Retaining Glycosyltransferase. *Angewandte Chemie - International Edition* **2015**, *127* (34), 10036–10040. https://doi.org/10.1002/ange.201504617.

(42) Hurtado-Guerrero, R. Recent Structural and Mechanistic Insights into Protein *O*-GalNAc Glycosylation. *Biochem Soc Trans* **2016**, *44*, 61–67. https://doi.org/10.1042/BST20150178.

(43) Moremen, K. W.; Haltiwanger, R. S. Emerging Structural Insights into Glycosyltransferase-Mediated Synthesis of Glycans. *Nat Chem Biol* **2019**, *15* (9), 853–864. https://doi.org/10.1038/s41589-019-0350-2.

(44) Yu, H.; Takeuchi, M.; LeBarron, J.; Kantharia, J.; London, E.; Bakker, H.; Haltiwanger, R. S.; Li, H.; Takeuchi, H. Notch-Modifying Xylosyltransferase Structures Support an SNi-like Retaining Mechanism. *Nat Chem Biol* **2015**, *11* (11), 847–854. https://doi.org/10.1038/nchembio.1927.

(45) Lira-Navarrete, E.; Iglesias-Fernández, J.; Zandberg, W. F.; Compañón, I.; Kong, Y.; Corzana, F.; Pinto, B. M.; Clausen, H.; Peregrina, J. M.; Vocadlo, D. J.; Rovira, C.; Hurtado-Guerrero, R. Substrate-Guided Front-Face Reaction Revealed by

Combined Structural Snapshots and Metadynamics for the Polypeptide *N*-Acetylgalactosaminyltransferase 2. *Angewandte Chemie - International Edition* **2014**, *126* (31), 8345–8349. https://doi.org/10.1002/ange.201402781.

(46) Ardèvol, A.; Iglesias-Fernández, J.; Rojas-Cervellera, V.; Rovira, C. The Reaction Mechanism of Retaining Glycosyltransferases. *Biochem Soc Trans* **2016**, *44*, 51–60. https://doi.org/10.1042/BST20150177.

(47) Smith, M. B.; March, J. The S_Ni Mechanism. In *March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*; John Wiley & Sons, Inc.: Hoboken, New Jersey, 2007; pp 468–469.

(48) Gómez, H.; Lluch, J. M.; Masgrau, L. Substrate-Assisted and Nucleophilically Assisted Catalysis in Bovine α1,3-Galactosyltransferase. Mechanistic Implications for Retaining Glycosyltransferases. *J Am Chem Soc* **2013**, *135* (18), 7053–7063. https://doi.org/10.1021/ja4024447.

(49) Bar-On, Y. M.; Phillips, R.; Milo, R. The Biomass Distribution on Earth. *Proc Natl Acad Sci U S A* **2018**, *115* (25), 6506–6511. https://doi.org/10.1073/pnas.1711842115.

(50) Ramage, M. H.; Burridge, H.; Busse-Wicher, M.; Fereday, G.; Reynolds, T.; Shah, D. U.; Wu, G.; Yu, L.; Fleming, P.; Densley-Tingley, D.; Allwood, J.; Dupree, P.; Linden, P. F.; Scherman, O. The Wood from the Trees: The Use of Timber in Construction. *Renewable and Sustainable Energy Reviews* **2017**, *68*, 333–359. https://doi.org/10.1016/j.rser.2016.09.107.

(51) John H. Seinfeld; Spyros N. Pandis. Global Cycles: Sulfur and Carbon. In *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*; John Wiley & Sons, Inc., 2016; pp 908–930.

(52) Pauly, M.; Keegstra, K. Plant Cell Wall Polymers as Precursors for Biofuels. *Curr Opin Plant Biol* **2010**, *13* (3), 304–311. https://doi.org/10.1016/j.pbi.2009.12.009.

(53) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc Natl Acad Sci U S A* **1977**, *74* (12), 5463–5467.

(54) Ghislain, B.; Nicolini, E. E.; Romain, I.; Ruelle, J.; Yoshinaga, A.; Alford, M. H.; Clair, B. Multilayered Structure of Tension Wood Cell Walls in Salicaceae Sensu Lato and

Its Taxonomic Significance. *Botanical Journal of the Linnean Society* **2016**, *182*, 744–756.

(55) Mackiewicz, P.; Gagat, P. Monophyly of Archaeplastida Supergroup and Relationships among Its Lineages in the Light of Phylogenetic and Phylogenomic Studies. Are We Close to a Consensus? *Acta Societatis Botanicorum Poloniae* **2014**, *83* (4), 263–280. https://doi.org/10.5586/asbp.2014.044.

(56) Crang, R.; Lyons-Sobaski, S.; Wise, R. The Nature of Plants. In *Plant Anatomy*; Springer International Publishing: Cham, 2018; pp 3–44. https://doi.org/10.1007/978-3-319-77315-5_1.

(57) Pryer, K. M.; Schuettpelz, E.; Wolf, P. G.; Schneider, H.; Smith, A. R.; Cranfill, R. Phylogeny and Evolution of Ferns (Monilophytes) with a Focus on the Early Leptosporangiate Divergences. *Am J Bot* **2004**, *91* (10), 1582–1598. https://doi.org/10.3732/ajb.91.10.1582.

(58) Schuettpelz, E.; Schneider, H.; Smith, A. R.; Hovenkamp, P.; Prado, J.; Rouhan, G.; Salino, A.; Sundue, M.; Almeida, T. E.; Parris, B.; Sessa, E. B.; Field, A. R.; de Gasper, A. L.; Rothfels, C. J.; Windham, M. D.; Lehnert, M.; Dauphin, B.; Ebihara, A.; Lehtonen, S.; Schwartsburd, P. B.; Metzgar, J.; Zhang, L. B.; Kuo, L. Y.; Brownsey, P. J.; Kato, M.; Arana, M. D.; Assis, F. C.; Barker, M. S.; Barrington, D. S.; Chang, H. M.; Chang, Y. H.; Chao, Y. S.; Chen, C. W.; Chen, D. K.; Chiou, W. L.; de Oliveira Dittrich, V. A.; Duan, Y. F.; Dubuisson, J. Y.; Farrar, D. R.; Fawcett, S.; Gabriel y Galán, J. M.; de Araújo Góes-Neto, L. A.; Grant, J. R.; Grusz, A. L.; Haufler, C.; Hauk, W.; He, H.; Hennequin, S.; Hirai, R. Y.; Huiet, L.; Kessler, M.; Korall, P.; Labiak, P. H.; Larsson, A.; León, B.; Li, C. X.; Li, F. W.; Link-Pérez, M.; Liu, H. M.; Lu, N. T.; Meza-Torres, E. I.; Miao, X. Y.; Moran, R.; Mynssen, C. M.; Nagalingum, N.; Øllgaard, B.; Paul, A. M.; Jovani, J. B.; Perrie, L. R.; Ponce, M.; Ranker, T. A.; Schulz, C.; Shinohara, W.; Shmakov, A.; Sigel, E. M.; de Souza, F. S.; Sylvestre, L. da S.; Testo, W.; Triana-Moreno, L. A.; Tsutsumi, C.; Tuomisto, H.; Valdespino, I. A.; Vasco, A.; Viveros, R. S.; Weakley, A.; Wei, R.; Weststrand, S.; Wolf, P. G.; Yatskievych, G.; Xu, X. G.; Yan, Y. H.; Zhang, L.; Zhang, X. C.; Zhou, X. M. A Community-

Derived Classification for Extant Lycophytes and Ferns. *J Syst Evol* **2016**, *54* (6), 563–603. https://doi.org/10.1111/jse.12229.

(59) Puttick, M. N.; Morris, J. L.; Williams, T. A.; Cox, C. J.; Edwards, D.; Kenrick, P.; Pressel, S.; Wellman, C. H.; Schneider, H.; Pisani, D.; Donoghue, P. C. J. The Interrelationships of Land Plants and the Nature of the Ancestral Embryophyte. *Current Biology* **2018**, *28*, 733–745. https://doi.org/10.1016/j.cub.2018.01.063.

(60) de La Torre, A. R.; Birol, I.; Bousquet, J.; Ingvarsson, P. K.; Jansson, S.; Jones, S. J. M.; Keeling, C. I.; MacKay, J.; Nilsson, O.; Ritland, K.; Street, N.; Yanchuk, A.; Zerbe, P.; Bohlmann, J. Insights into Conifer Giga-Genomes. *Plant Physiol* **2014**, *166*, 1724–1732. https://doi.org/10.1104/pp.114.248708.

(61) Lughadha, E. N.; Govaerts, R.; Belyaeva, I.; Black, N.; Lindon, H.; Allkin, R.; Magill, R. E.; Nicolson, N. Counting Counts: Revised Estimates of Numbers of Accepted Species of Flowering Plants, Seed Plants, Vascular Plants and Land Plants with a Review of Other Recent Estimates. *Phytotaxa* **2016**, *272* (1), 82–88. https://doi.org/10.11646/phytotaxa.272.1.5.

(62) Ronse De Craene, L. P. Eudicots. In *eLS*; John Wiley & Sons, Ltd: Chichester, 2012. https://doi.org/10.1002/9780470015902.a0003684.pub2.

(63) Byng, J. W.; Chase, M. W.; M Christenhusz, M. J.; Fay, M. F.; Judd, W. S.; Mabberley, D. J.; Sennikov, A. N.; Soltis, D. E.; Soltis, P. S.; Stevens, P. F.; Briggs, B.; Brockington, S.; Chautems, A.; Clark, J. C.; Conran, J.; Haston, E.; Moore, M.; Olmstead, R.; Perret, M.; Skog, L.; Smith, J.; Tank, D.; Vorontsova, M.; Weber, A. An Update of the Angiosperm Phylogeny Group Classification for the Orders and Families of Flowering Plants: APG IV. *Botanical Journal of the Linnean Society* **2016**, *181*, 1–20.

(64) Tuskan, G. A.; Difazio, S.; Jansson, S.; Bohlmann, J.; Grigoriev, I.; Hellsten, U.; Putnam, N.; Ralph, S.; Rombauts, S.; Salamov, A.; Schein, J.; Sterck, L.; Aerts, A.; Bhalerao, R. R.; Bhalerao, R. P.; Blaudez, D.; Boerjan, W.; Brun, A.; Brunner, A.; Busov, V.; Campbell, M.; Carlson, J.; Chalot, M.; Chapman, J.; Chen, G.-L.; Cooper, D.; Coutinho, P. M.; Couturier, J.; Covert, S.; Cronk, Q.; Cunningham, R.; Davis, J.; Degroeve, S.; Déjardin, A.; Depamphilis, C.; Detter, J.; Dirks, B.; Dubchak, I.;

Duplessis, S.; Ehlting, J.; Ellis, B.; Gendler, K.; Goodstein, D.; Gribskov, M.; Grimwood, J.; Groover, A.; Gunter, L.; Hamberger, B.; Heinze, B.; Helariutta, Y.; Henrissat, B.; Holligan, D.; Holt, R.; Huang, W.; Islam-Faridi, N.; Jones, S.; Jones-Rhoades, M.; Jorgensen, R.; Joshi, C.; Kangasjärvi, J.; Karlsson, J.; Kelleher, C.; Kirkpatrick, R.; Kirst, M.; Kohler, A.; Kalluri, U.; Larimer, F.; Leebens-Mack, J.; Leplé, J.-C.; Locascio, P.; Lou, Y.; Lucas, S.; Martin, F.; Montanini, B.; Napoli, C.; Nelson, D. R.; Nelson, C.; Nieminen, K.; Nilsson, O.; Pereda, V.; Peter, G.; Philippe, R.; Pilate, G.; Poliakov, A.; Razumovskaya, J.; Richardson, P.; Rinaldi, C.; Ritland, K.; Rouzé, P.; Ryaboy, D.; Schmutz, J.; Schrader, J.; Segerman, B.; Shin, H.; Siddiqui, A.; Sterky, F.; Terry, A.; Tsai, C.-J.; Uberbacher, E.; Unneberg, P.; Vahala, J.; Wall, K.; Wessler, S.; Yang, G.; Yin, T.; Douglas, C.; Marra, M.; Sandberg, G.; van de Peer, Y.; Rokhsar, D. The Genome of Black Cottonwood, *Populus Trichocarpa* (Torr. & Gray). *Science (1979)* **2006**, *313*, 1596–1604.

(65) Somerville, C.; Bauer, S.; Brininstool, G.; Facette, M.; Hamann, T.; Milne, J.; Osborne, E.; Paredez, A.; Persson, S.; Raab, T.; Vorwerk, S.; Youngs, H. Toward a Systems Approach to Understanding Plant Cell Walls. *Science (1979)* **2004**, *306* (5705), 2206–2211. https://doi.org/10.1126/science.1102765.

(66) Scheller, H. V.; Ulvskov, P. Hemicelluloses. *Annu Rev Plant Biol* **2010**, *61*, 263–289. https://doi.org/10.1146/annurev-arplant-042809-112315.

(67) Clair, B.; Déjardin, A.; Pilate, G.; Alméras, T. Is the G-Layer a Tertiary Cell Wall? *Front Plant Sci* **2018**, *9*, 623. https://doi.org/10.3389/fpls.2018.00623.

(68) Cosgrove, D. J.; Jarvis, M. C. Comparative Structure and Biomechanics of Plant Primary and Secondary Cell Walls. *Front Plant Sci* **2012**, *3*, 204. https://doi.org/10.3389/fpls.2012.00204.

(69) Meents, M. J.; Watanabe, Y.; Samuels, A. L. The Cell Biology of Secondary Cell Wall Biosynthesis. *Ann Bot* **2018**, *121*, 1107–1125. https://doi.org/10.1093/aob/mcy005.

(70) Røislien, J.; Frøslie, K. F. Jane and John Do Not Exist. *Tidsskrift for den Norske Laegeforening* **2019**, *139* (1). https://doi.org/10.4045/tidsskr.18.0824.

(71) Harholt, J.; Suttangkakul, A.; Scheller, H. V. Biosynthesis of Pectin. *Plant Physiol* **2010**, *153* (2), 384–395. https://doi.org/10.1104/pp.110.156588.

(72) Cosgrove, D. J. Growth of the Plant Cell Wall. *Nat Rev Mol Cell Biol* **2005**, *6* (11), 850–861. https://doi.org/10.1038/nrm1746.

(73) Vincken, J. P.; Schols, H. A.; Oomen, R. J. F. J.; McCann, M. C.; Ulvskov, P.; Voragen, A. G. J.; Visser, R. G. F. If Homogalacturonan Were a Side Chain of Rhamnogalacturonan I. Implications for Cell Wall Architecture. *Plant Physiol* **2003**, *132* (4), 1781–1789. https://doi.org/10.1104/pp.103.022350.

(74) Cosgrove, D. J. Re-Constructing Our Models of Cellulose and Primary Cell Wall Assembly. *Curr Opin Plant Biol* **2014**, *22*, 122–131. https://doi.org/10.1016/j.pbi.2014.11.001.

(75) Carpita, N. C.; Gibeaut, D. M. Structural Models of Primary Cell Walls in Flowering Plants: Consistency of Molecular Structure with the Physical Properties of the Walls during Growth. *me Plant Journal* **1993**, *3* (1), 1–30.

(76) Kubicki, J. D.; Yang, H.; Sawada, D.; O'Neill, H.; Oehme, D.; Cosgrove, D. The Shape of Native Plant Cellulose Microfibrils. *Sci Rep* **2018**, *8*, 13983. https://doi.org/10.1038/s41598-018-32211-w.

(77) Hill, J. L.; Hammudi, M. B.; Tien, M. The *Arabidopsis* Cellulose Synthase Complex: A Proposed Hexamer of CESA Trimers in an Equimolar Stoichiometry. *Plant Cell* **2014**, *26* (12), 4834–4842. https://doi.org/10.1105/tpc.114.131193.

(78) Burton, R. A.; Gidley, M. J.; Fincher, G. B. Heterogeneity in the Chemistry, Structure and Function of Plant Cell Walls. *Nat Chem Biol* **2010**, *6*, 724–732. https://doi.org/10.1038/nchembio.439.

(79) Vergara, C. E.; Carpita, N. C. β-D-Glycan Synthases and the CesA Gene Family: Lessons to Be Learned from the Mixed-Linkage (1→3),(1→4)β-D-Glucan Synthase. In *Plant Cell Walls*; Springer Netherlands: Dordrecht, 2001; Vol. 47, pp 145–160. https://doi.org/10.1007/978-94-010-0668-2_9.

(80) Vanzin, G. F.; Madson, M.; Carpita, N. C.; Raikhel, N. v.; Keegstra, K.; Reiter, W.-D. The *mur2* Mutant of *Arabidopsis Thaliana* Lacks Fucosylated Xyloglucan Because

of a Lesion in Fucosyltransferase *At*FUT1. *Proceedings of the National Academy of Sciences* **2002**, *99* (5), 3340–3345. https://doi.org/10.1073/pnas.052450699.

(81) Machmudah, S.; Wahyudiono; Kanda, H.; Goto, M. Hydrolysis of Biopolymers in Near-Critical and Subcritical Water. In *Water Extraction of Bioactive Compounds: From Plants to Drug Development*; Elsevier, 2017; pp 69–107. https://doi.org/10.1016/B978-0-12-809380-1.00003-6.

(82) Bulone, V.; Schwerdt, J. G.; Fincher, G. B. Co-Evolution of Enzymes Involved in Plant Cell Wall Metabolism in the Grasses. *Front Plant Sci* **2019**, *10*, 1009. https://doi.org/10.3389/fpls.2019.01009.

(83) Peña, M. J.; Kulkarni, A. R.; Backe, J.; Boyd, M.; O'Neill, M. A.; York, W. S. Structural Diversity of Xylans in the Cell Walls of Monocots. *Planta* **2016**, *244*, 589–606. https://doi.org/10.1007/s00425-016-2527-1.

(84) Little, A.; Lahnstein, J.; Jeffery, D. W.; Khor, S. F.; Schwerdt, J. G.; Shirley, N. J.; Hooi, M.; Xing, X.; Burton, R. A.; Bulone, V. A Novel (1,4)-β-Linked Glucoxylan Is Synthesized by Members of the *Cellulose Synthase-Like F* Gene Family in Land Plants. *ACS Cent Sci* **2019**, *5*, 73–84. https://doi.org/10.1021/acscentsci.8b00568.

(85) Ray, B.; Lahaye, M. Cell-Wall Polysaccharides from the Marine Green Alga Ulva "Rigida" (Ulvales, Chlorophyta). Extraction and Chemical Composition. *Carbohydr Res* **1995**, *274*, 251–261. https://doi.org/10.1016/0008-6215(95)00138-J.

(86) Timell, T. E. Recent Progress in the Chemistry of Wood Hemicelluloses. *Wood Sci Technol* **1967**, *1*, 45–70. https://doi.org/10.1007/BF00592255.

(87) Fischer, M. H.; Yu, N.; Gray, G. R.; Ralph, J.; Anderson, L.; Marlett, J. A. The Gel-Forming Polysaccharide of Psyllium Husk (Plantago Ovata Forsk). *Carbohydr Res* **2004**, *339* (11), 2009–2017. https://doi.org/10.1016/j.carres.2004.05.023.

(88) Edwards, S.; Chaplin, M. F.; Blackwood, A. D.; Dettmar, P. W. Primary Structure of Arabinoxylans of Ispaghula Husk and Wheat Bran. *Proceedings of the Nutrition Society* **2003**, *62* (1), 217–222. https://doi.org/10.1079/pns2003202.

(89) Ebringerová, A.; Heinze, T. Xylan and Xylan Derivatives-Biopolymers with Valuable Properties, 1. *Macromol Rapid Commun* **2000**, *21* (9), 542–556.

(90) Hsieh, Y. S. Y.; Harris, P. J. Xylans of Red and Green Algae: What Is Known about Their Structures and How They Are Synthesised? *Polymers (Basel)* **2019**, *11*, 354. https://doi.org/10.3390/POLYM11020354.

(91) Lunin, V. v.; Wang, H. T.; Bharadwaj, V. S.; Alahuhta, M.; Peña, M. J.; Yang, J. Y.; Archer-Hartmann, S. A.; Azadi, P.; Himmel, M. E.; Moremen, K. W.; York, W. S.; Bomble, Y. J.; Urbanowicz, B. R. Molecular Mechanism of Polysaccharide Acetylation by the Arabidopsis Xylan *O*-Acetyltransferase XOAT1. *Plant Cell* **2020**, *32* (7), 2367–2382. https://doi.org/10.1105/tpc.20.00028.

(92) Smith, P. J.; Wang, H. T.; York, W. S.; Peña, M. J.; Urbanowicz, B. R. Designer Biomass for Next-Generation Biorefineries: Leveraging Recent Insights into Xylan Structure and Biosynthesis. *Biotechnol Biofuels* **2017**, *10*, 286. https://doi.org/10.1186/s13068-017-0973-z.

(93) Bromley, J. R.; Busse-Wicher, M.; Tryfona, T.; Mortimer, J. C.; Zhang, Z.; Brown, D. M.; Dupree, P. GUX1 and GUX2 Glucuronyltransferases Decorate Distinct Domains of Glucuronoxylan with Different Substitution Patterns. *Plant Journal* **2013**, *74*, 423–434. https://doi.org/10.1111/tpj.12135.

(94) Busse-Wicher, M.; Gomes, T. C. F.; Tryfona, T.; Nikolovski, N.; Stott, K.; Grantham, N. J.; Bolam, D. N.; Skaf, M. S.; Dupree, P. The Pattern of Xylan Acetylation Suggests Xylan May Interact with Cellulose Microfibrils as a Twofold Helical Screw in the Secondary Plant Cell Wall of *Arabidopsis Thaliana*. *Plant Journal* **2014**, *79* (3), 492–506. https://doi.org/10.1111/tpj.12575.

(95) Busse-Wicher, M.; Li, A.; Silveira, R. L.; Pereira, C. S.; Tryfona, T.; Gomes, T. C. F.; Skaf, M. S.; Dupree, P. Evolution of Xylan Substitution Patterns in Gymnosperms and Angiosperms: Implications for Xylan Interaction with Cellulose. *Plant Physiol* **2016**, *171*, 2418–2431. https://doi.org/10.1104/pp.16.00539.

(96) Grantham, N. J.; Wurman-Rodrich, J.; Terrett, O. M.; Lyczakowski, J. J.; Stott, K.; Iuga, D.; Simmons, T. J.; Durand-Tardif, M.; Brown, S. P.; Dupree, R.; Busse-Wicher, M.; Dupree, P. An Even Pattern of Xylan Substitution Is Critical for Interaction with Cellulose in Plant Cell Walls. *Nat Plants* **2017**, *3* (11), 859–865. https://doi.org/10.1038/s41477-017-0030-8.

(97) Ye, Z.-H.; Zhong, R. Outstanding Questions on Xylan Biosynthesis. *Plant Science* **2022**, *325*, 111476. https://doi.org/10.1016/j.plantsci.2022.111476.

(98) Peña, M. J.; Zhong, R.; Zhou, G. K.; Richardson, E. A.; O'Neill, M. A.; Darvill, A. G.; York, W. S.; Yeb, Z. H. Arabidopsis Irregular Xylem8 and Irregular Xylem9: Implications for the Complexity of Glucuronoxylan Biosynthesis. *Plant Cell* **2007**, *19*, 549–563. https://doi.org/10.1105/tpc.106.049320.

(99) Simmons, T. J.; Mortimer, J. C.; Bernardinelli, O. D.; Pöppler, A. C.; Brown, S. P.; DeAzevedo, E. R.; Dupree, R.; Dupree, P. Folding of Xylan onto Cellulose Fibrils in Plant Cell Walls Revealed by Solid-State NMR. *Nat Commun* **2016**, *7*, 1–9. https://doi.org/10.1038/ncomms13902.

(100) Wohlert, M.; Benselfelt, T.; Wågberg, L.; Furó, I.; Berglund, L. A.; Wohlert, J. Cellulose and the Role of Hydrogen Bonds: Not in Charge of Everything. *Cellulose* **2022**, *29*, 1–23. https://doi.org/10.1007/s10570-021-04325-4.

(101) Martínez-Abad, A.; Berglund, J.; Toriz, G.; Gatenholm, P.; Henriksson, G.; Lindström, M.; Wohlert, J.; Vilaplana, F. Regular Motifs in Xylan Modulate Molecular Flexibility and Interactions with Cellulose Surfaces. *Plant Physiol* **2017**, *175* (4), 1579–1592. https://doi.org/10.1104/pp.17.01184.

(102) Terrett, O. M.; Lyczakowski, J. J.; Yu, L.; Iuga, D.; Franks, W. T.; Brown, S. P.; Dupree, R.; Dupree, P. Molecular Architecture of Softwood Revealed by Solid-State NMR. *Nat Commun* **2019**, *10*, 4978. https://doi.org/10.1038/s41467-019-12979-9.

(103) Gao, Y.; Lipton, A. S.; Wittmer, Y.; Murray, D. T.; Mortimer, J. C. A Grass-Specific Cellulose–Xylan Interaction Dominates in Sorghum Secondary Cell Walls. *Nat Commun* **2020**, *11*, 6081. https://doi.org/10.1038/s41467-020-19837-z.

(104) Nishiyama, Y. Molecular Interactions in Nanocellulose Assembly. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2018**, *376*, 20170047. https://doi.org/10.1098/rsta.2017.0047.

(105) Kang, X.; Kirui, A.; Dickwella Widanage, M. C.; Mentink-Vigier, F.; Cosgrove, D. J.; Wang, T. Lignin-Polysaccharide Interactions in Plant Secondary Cell Walls

Revealed by Solid-State NMR. *Nat Commun* **2019**, *10*, 347. https://doi.org/10.1038/s41467-018-08252-0.

(106) Anslyn, E. v.; Dougherty, D. A. The Origin of the Hydrophobic Effect. In *Modern Physical Organic Chemistry*; University Science Books, 2006; pp 192–194.

(107) *Arabidopsis Nomenclature*. https://www.arabidopsis.org/portals/nomenclature/guidelines.jsp (accessed 2022-11-11).

(108) Pearson, W. R. An Introduction to Sequence Similarity ("Homology") Searching. *Curr Protoc Bioinformatics* **2013**, *42*, 3.1.1-3.1.8. https://doi.org/10.1002/0471250953.bi0301s42.

(109) Jensen, R. A. Orthologs and Paralogs - We Need to Get It Right. *Genome Biol* **2001**, *2*, interactions1002.1. https://doi.org/doi.org/10.1186/gb-2001-2-8-interactions1002.

(110) *About Arabidopsis*. https://www.arabidopsis.org/portals/education/aboutarabidopsis.jsp (accessed 2022-11-10).

(111) *The Arabidopsis Information Resource (TAIR)*. https://www.arabidopsis.org (accessed 2022-11-10).

(112) Turner, S. R.; Somerville, C. R. Collapsed Xylem Phenotype of Arabidopsis Identifies Mutants Deficient in Cellulose Deposition in the Secondary Cell Wall. *Plant Cell* **1997**, *9* (5), 689–701. https://doi.org/10.1105/tpc.9.5.689.

(113) Brown, D. M.; Zeef, L. A. H.; Ellis, J.; Goodacre, R.; Turner, S. R. Identification of Novel Genes in Arabidopsis Involved in Secondary Cell Wall Formation Using Expression Profiling and Reverse Genetics. *Plant Cell* **2005**, *17* (8), 2281–2295. https://doi.org/10.1105/tpc.105.031542.

(114) Mitchell, R. A. C.; Dupree, P.; Shewry, P. R. A Novel Bioinformatics Approach Identifies Candidate Genes for the Synthesis and Feruloylation of Arabinoxylan. *Plant Physiol* **2007**, *144*, 43–53. https://doi.org/10.1104/pp.106.094995.

(115) Bischoff, V.; Nita, S.; Neumetzler, L.; Schindelasch, D.; Urbain, A.; Eshed, R.; Persson, S.; Delmer, D.; Scheible, W. R. TRICHOME BIREFRINGENCE and Its Homolog AT5G01360 Encode Plant-Specific DUF231 Proteins Required for Cellulose

Biosynthesis in Arabidopsis. *Plant Physiol* **2010**, *153* (2), 590–602. https://doi.org/10.1104/pp.110.153320.

(116) Brown, D. M.; Goubet, F.; Wong, V. W.; Goodacre, R.; Stephens, E.; Dupree, P.; Turner, S. R. Comparison of Five Xylan Synthesis Mutants Reveals New Insight into the Mechanisms of Xylan Synthesis. *Plant Journal* **2007**, *52* (6), 1154–1168. https://doi.org/10.1111/j.1365-313X.2007.03307.x.

(117) Brown, D. M.; Zhang, Z.; Stephens, E.; Dupree, P.; Turner, S. R. Characterization of IRX10 and IRX10-like Reveals an Essential Role in Glucuronoxylan Biosynthesis in Arabidopsis. *Plant Journal* **2009**, *57* (4), 732–746. https://doi.org/10.1111/j.1365-313X.2008.03729.x.

(118) Lee, C.; Teng, Q.; Huang, W.; Zhong, R.; Ye, Z.-H. The Arabidopsis Family GT43 Glycosyltransferases Form Two Functionally Nonredundant Groups Essential for the Elongation of Glucuronoxylan Backbone. *Plant Physiol* **2010**, *153* (2), 526–541. https://doi.org/10.1104/pp.110.155309.

(119) Wu, A. M.; Rihouey, C.; Seveno, M.; Hörnblad, E.; Singh, S. K.; Matsunaga, T.; Ishii, T.; Lerouge, P.; Marchant, A. The Arabidopsis IRX10 and IRX10-LIKE Glycosyltransferases Are Critical for Glucuronoxylan Biosynthesis during Secondary Cell Wall Formation. *Plant Journal* **2009**, *57*, 718–731. https://doi.org/10.1111/j.1365-313X.2008.03724.x.

(120) Wu, A. M.; Hörnblad, E.; Voxeur, A.; Gerber, L.; Rihouey, C.; Lerouge, P.; Marchant, A. Analysis of the Arabidopsis IRX9/IRX9-L and IRX14/IRX14-L Pairs of Glycosyltransferase Genes Reveals Critical Contributions to Biosynthesis of the Hemicellulose Glucuronoxylan. *Plant Physiol* **2010**, *153*, 542–554. https://doi.org/10.1104/pp.110.154971.

(121) Zeng, W.; Lampugnani, E. R.; Picard, K. L.; Song, L.; Wu, A. M.; Farion, I. M.; Zhao, J.; Ford, K.; Doblin, M. S.; Bacic, A. Asparagus IRX9, IRX10, and IRX14A Are Components of an Active Xylan Backbone Synthase Complex That Forms in the Golgi Apparatus. *Plant Physiol* **2016**, *171*, 93–109. https://doi.org/10.1104/pp.15.01919.

(122) Jiang, N.; Wiemels, R. E.; Soya, A.; Whitley, R.; Held, M.; Faik, A. Composition, Assembly, and Trafficking of a Wheat Xylan Synthase Complex. *Plant Physiol* **2016**, *170* (4), 1999–2023. https://doi.org/10.1104/pp.15.01777.

(123) Ren, Y.; Hansen, S. F.; Ebert, B.; Lau, J.; Scheller, H. V. Site-Directed Mutagenesis of IRX9, IRX9L and IRX14 Proteins Involved in Xylan Biosynthesis: Glycosyltransferase Activity Is Not Required for IRX9 Function in Arabidopsis. *PLoS One* **2014**, *9* (8), e105014. https://doi.org/10.1371/journal.pone.0105014.

(124) Jensen, J. K.; Johnson, N. R.; Wilkerson, C. G. *Arabidopsis Thaliana* IRX10 and Two Related Proteins from Psyllium and *Physcomitrella Patens* Are Xylan Xylosyltransferases. *Plant Journal* **2014**, *80*, 207–215. https://doi.org/10.1111/tpj.12641.

(125) Urbanowicz, B. R.; Peña, M. J.; Moniz, H. A.; Moremen, K. W.; York, W. S. Two Arabidopsis Proteins Synthesize Acetylated Xylan in Vitro. *Plant Journal* **2014**, *80*, 197–206. https://doi.org/10.1111/tpj.12643.

(126) Breton, C.; Šnajdrová, L.; Jeanneau, C.; Koča, J.; Imberty, A. Structures and Mechanisms of Glycosyltransferases. *Glycobiology* **2006**, *16* (2), 29R-37R. https://doi.org/10.1093/glycob/cwj016.

(127) Chiniquy, D.; Varanasi, P.; Oh, T.; Harholt, J.; Katnelson, J.; Singh, S.; Auer, M.; Simmons, B.; Adams, P. D.; Scheller, H. v.; Ronald, P. C. Three Novel Rice Genes Closely Related to the Arabidopsis IRX9, IRX9L, and IRX14 Genes and Their Roles in Xylan Biosynthesis. *Front Plant Sci* **2013**, *4*, 83. https://doi.org/10.3389/fpls.2013.00083.

(128) Rennie, E. A.; Scheller, H. V. Xylan Biosynthesis. *Current Opinion in Biotechnology*. April 2014, pp 100–107. https://doi.org/10.1016/j.copbio.2013.11.013.

(129) Mortimer, J. C.; Faria-Blanc, N.; Yu, X.; Tryfona, T.; Sorieul, M.; Ng, Y. Z.; Zhang, Z.; Stott, K.; Anders, N.; Dupree, P. An Unusual Xylan in Arabidopsis Primary Cell Walls Is Synthesised by GUX3, IRX9L, IRX10L and IRX14. *Plant Journal* **2015**, *83*, 413–426. https://doi.org/10.1111/tpj.12898.

(130) Urbanowicz, B. R.; Peña, M. J.; Ratnaparkhe, S.; Avci, U.; Backe, J.; Steet, H. F.; Foston, M.; Li, H.; O'Neill, M. A.; Ragauskas, A. J.; Darvill, A. G.; Wyman, C.; Gilbert, H. J.; York, W. S. 4-*O*-Methylation of Glucuronic Acid in *Arabidopsis*

Glucuronoxylan Is Catalyzed by a Domain of Unknown Function Family 579 Protein. *Proceedings of the National Academy of Sciences* **2012**, *109* (35), 14253–14258. https://doi.org/10.1073/pnas.1208097109.

(131) Lee, C.; Teng, Q.; Zhong, R.; Yuan, Y.; Haghighat, M.; Ye, Z.-H. Three Arabidopsis DUF579 Domain-Containing GXM Proteins Are Methyltransferases Catalyzing 4-*O*-Methylation of Glucuronic Acid on Xylan. *Plant Cell Physiol* **2012**, *53* (11), 1934–1949. https://doi.org/10.1093/pcp/pcs138.

(132) Jensen, J. K.; Kim, H.; Cocuron, J.-C.; Orler, R.; Ralph, J.; Wilkerson, C. G. The DUF579 Domain Containing Proteins IRX15 and IRX15-L Affect Xylan Synthesis in Arabidopsis. *The Plant Journal* **2011**, *66* (3), 387–400. https://doi.org/10.1111/j.1365-313X.2010.04475.x.

(133) Brown, D.; Wightman, R.; Zhang, Z.; Gomez, L. D.; Atanassov, I.; Bukowski, J.-P.; Tryfona, T.; McQueen-Mason, S. J.; Dupree, P.; Turner, S. Arabidopsis Genes IRREGULAR XYLEM (IRX15) and IRX15L Encode DUF579-Containing Proteins That Are Essential for Normal Xylan Deposition in the Secondary Cell Wall. *The Plant Journal* **2011**, *66* (3), 401–413. https://doi.org/10.1111/j.1365-313X.2011.04501.x.

(134) Engle, K. A.; Amos, R. A.; Yang, J.; Glushka, J.; Atmodjo, M. A.; Tan, L.; Huang, C.; Moremen, K. W.; Mohnen, D. Multiple Arabidopsis Galacturonosyltransferases Synthesize Polymeric Homogalacturonan by Oligosaccharide Acceptor-dependent or de Novo Synthesis. *The Plant Journal* **2022**, *109* (6), 1441–1456. https://doi.org/10.1111/tpj.15640.

(135) Lee, C.; Zhong, R.; Richardson, E. A.; Himmelsbach, D. S.; McPhail, B. T.; Ye, Z. H. The PARVUS Gene Is Expressed in Cells Undergoing Secondary Wall Thickening and Is Essential for Glucuronoxylan Biosynthesis. *Plant Cell Physiol* **2007**, *48* (12), 1659–1672. https://doi.org/10.1093/pcp/pcm155.

(136) Zhong, R.; Cui, D.; Ye, Z. H. Regiospecific Acetylation of Xylan Is Mediated by a Group of DUF231-Containing *O*-Acetyltransferases. *Plant Cell Physiol* **2017**, *58* (12), 2126–2138. https://doi.org/10.1093/pcp/pcx147.

(137) Yuan, Y.; Teng, Q.; Zhong, R.; Ye, Z. H. TBL3 and TBL31, Two Arabidopsis DUF231 Domain Proteins, Are Required for 3-*O*-Monoacetylation of Xylan. *Plant Cell Physiol* **2016**, *57* (1), 35–45. https://doi.org/10.1093/pcp/pcv172.

(138) Yuan, Y.; Teng, Q.; Zhong, R.; Haghighat, M.; Richardson, E. A.; Ye, Z. H. Mutations of Arabidopsis TBL32 and TBL33 Affect Xylan Acetylation and Secondary Wall Deposition. *PLoS One* **2016**, *11* (1). https://doi.org/10.1371/journal.pone.0146460.

(139) Yuan, Y.; Teng, Q.; Zhong, R.; Ye, Z. H. Roles of Arabidopsis TBL34 and TBL35 in Xylan Acetylation and Plant Growth. *Plant Science* **2016**, *243*, 120–130. https://doi.org/10.1016/j.plantsci.2015.12.007.

(140) Schultink, A.; Naylor, D.; Dama, M.; Pauly, M. The Role of the Plant-Specific ALTERED XYLOGLUCAN9 Protein in Arabidopsis Cell Wall Polysaccharide *O*-Acetylation. *Plant Physiol* **2015**, *167* (4), 1271–1283. https://doi.org/10.1104/pp.114.256479.

(141) Manabe, Y.; Nafisi, M.; Verhertbruggen, Y.; Orfila, C.; Gille, S.; Rautengarten, C.; Cherk, C.; Marcus, S. E.; Somerville, S.; Pauly, M.; Knox, J. P.; Sakuragi, Y.; Scheller, H. V. Loss-of-Function Mutation of *REDUCED WALL ACETYLATION2* in Arabidopsis Leads to Reduced Cell Wall Acetylation and Increased Resistance to *Botrytis Cinerea*. *Plant Physiol* **2011**, *155* (3), 1068–1078. https://doi.org/10.1104/pp.110.168989.

(142) Lee, C.; Teng, Q.; Zhong, R.; Ye, Z. H. The Four Arabidopsis REDUCED WALL ACETYLATION Genes Are Expressed in Secondary Wall-Containing Cells and Required for the Acetylation of Xylan. *Plant Cell Physiol* **2011**, *52*, 1289–1301. https://doi.org/10.1093/pcp/pcr075.

(143) Zhong, R.; Cui, D.; Ye, Z. Evolutionary Origin of *O*-Acetyltransferases Responsible for Glucomannan Acetylation in Land Plants. *New Phytologist* **2019**, *224*, 466–479. https://doi.org/10.1111/nph.15988.

(144) Rauwerdink, A.; Kazlauskas, R. J. How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: The Serine-Histidine-Aspartate Catalytic Triad of α/β-Hydrolase Fold Enzymes. *ACS Catal* **2015**, *5*, 6153–6176. https://doi.org/10.1021/acscatal.5b01539.

(145) Gille, S.; Pauly, M. *O*-Acetylation of Plant Cell Wall Polysaccharides. *Front Plant Sci* **2012**, *3*, 12. https://doi.org/10.3389/fpls.2012.00012.

(146) Pauly, M.; Ramírez, V. New Insights into Wall Polysaccharide *O*-Acetylation. *Front Plant Sci* **2018**, *9*, 1210. https://doi.org/10.3389/fpls.2018.01210.

(147) Anders, N.; Wilkinson, M. D.; Lovegrove, A.; Freeman, J.; Tryfona, T.; Pellny, T. K.; Weimar, T.; Mortimer, J. C.; Stott, K.; Baker, J. M.; Defoin-Platel, M.; Shewry, P. R.; Dupree, P.; Mitchell, R. A. C. Glycosyl Transferases in Family 61 Mediate Arabinofuranosyl Transfer onto Xylan in Grasses. *Proceedings of the National Academy of Sciences* **2012**, *109* (3), 989–993. https://doi.org/10.1073/pnas.1115858109.

(148) Zhong, R.; Cui, D.; Phillips, D. R.; Sims, N. T.; Ye, Z.-H. Functional Analysis of GT61 Glycosyltransferases from Grass Species in Xylan Substitutions. *Planta* **2021**, *254* (6), 131. https://doi.org/10.1007/s00425-021-03794-y.

(149) Plante, O. J.; Palmacci, E. R.; Seeberger, P. H. Automated Solid-Phase Synthesis of Oligosaccharides. *Science (1979)* **2001**, *291* (5508), 1523–1527. https://doi.org/10.1126/science.1057324.

(150) Wen, L.; Edmunds, G.; Gibbons, C.; Zhang, J.; Gadi, M. R.; Zhu, H.; Fang, J.; Liu, X.; Kong, Y.; Wang, P. G. Toward Automated Enzymatic Synthesis of Oligosaccharides. *Chem Rev* **2018**, *118*, 8151–8187. https://doi.org/10.1021/acs.chemrev.8b00066.

(151) Joseph, A. A.; Pardo-Vargas, A.; Seeberger, P. H. Total Synthesis of Polysaccharides by Automated Glycan Assembly. *J Am Chem Soc* **2020**, *142* (19), 8561–8564. https://doi.org/10.1021/jacs.0c00751.

(152) Schmidt, D.; Schuhmacher, F.; Geissner, A.; Seeberger, P. H.; Pfrengle, F. Automated Synthesis of Arabinoxylan-Oligosaccharides Enables Characterization of Antibodies That Recognize Plant Cell Wall Glycans. *Chemistry - A European Journal* **2015**, *21* (15), 5709–5713. https://doi.org/10.1002/chem.201500065.

(153) Senf, D.; Ruprecht, C.; de Kruijff, G. H. M.; Simonetti, S. O.; Schuhmacher, F.; Seeberger, P. H.; Pfrengle, F. Active Site Mapping of Xylan-Deconstructing Enzymes with Arabinoxylan Oligosaccharides Produced by Automated Glycan Assembly. *Chemistry - A European Journal* **2017**, *23* (13), 3197–3205. https://doi.org/10.1002/chem.201605902.

(154) Buchholz, K.; Seibel, J. Industrial Carbohydrate Biotransformations. *Carbohydr Res* **2008**, *343*, 1966–1979. https://doi.org/10.1016/j.carres.2008.02.007.

(155) Gagarinov, I. A.; Li, T.; Toraño, J. S.; Caval, T.; Srivastava, A. D.; Kruijtzer, J. A. W.; Heck, A. J. R.; Boons, G.-J. Chemoenzymatic Approach for the Preparation of Asymmetric Bi-, Tri-, and Tetra-Antennary *N*-Glycans from a Common Precursor. *J Am Chem Soc* **2017**, *139* (2), 1011–1018. https://doi.org/10.1021/jacs.6b12080.

(156) Tsai, T.-I.; Lee, H.-Y.; Chang, S.-H.; Wang, C.-H.; Tu, Y.-C.; Lin, Y.-C.; Hwang, D.-R.; Wu, C.-Y.; Wong, C.-H. Effective Sugar Nucleotide Regeneration for the Large-Scale Enzymatic Synthesis of Globo H and SSEA4. *J Am Chem Soc* **2013**, *135* (39), 14831–14839. https://doi.org/10.1021/ja4075584.

(157) Hayes, M. R.; Pietruszka, J. Synthesis of Glycosides by Glycosynthases. *Molecules* **2017**, *22* (9), 1434. https://doi.org/10.3390/molecules22091434.

(158) Bacon, J. S. D. The Oligosaccharides Produced by the Action of Yeast Invertase Preparations on Sucrose. *Biochemical Journal* **1954**, *57* (2), 320–328. https://doi.org/doi.org/10.1042/bj0570320.

(159) Danby, P. M.; Withers, S. G. Advances in Enzymatic Glycoside Synthesis. *ACS Chem Biol* **2016**, *11* (7), 1784–1794. https://doi.org/10.1021/acschembio.6b00340.

(160) Mackenzie, L. F.; Wang, Q.; Warren, R. A. J.; Withers, S. G. Glycosynthases: Mutant Glycosidases for Oligosaccharide Synthesis. *J Am Chem Soc* **1998**, *120*, 5583–5584.

(161) Malet, C.; Planas, A. From L-Glucanase to L-Glucansynthase: Glycosyl Transfer to K-Glycosyl £uorides Catalyzed by a Mutant Endoglucanase Lacking Its Catalytic Nucleophile. *FEBS Lett* **1998**, *440*, 208–212.

(162) Honda, Y.; Kitaoka, M. The First Glycosynthase Derived from an Inverting Glycoside Hydrolase. *Journal of Biological Chemistry* **2006**, *281* (3), 1426–1431. https://doi.org/10.1074/jbc.M511202200.

(163) Sugimura, M.; Nishimoto, M.; Kitaoka, M. Characterization of Glycosynthase Mutants Derived from Glycoside Hydrolase Family 10 Xylanases. *Biosci Biotechnol Biochem* **2006**, *70* (5), 1210–1217. https://doi.org/10.1271/bbb.70.1210.

Chapter 2
Chemoenzymatic Synthesis of Xylan Oligosaccharides

2.1. Introduction

The interest in the structure and biosynthesis of xylan polysaccharides has been growing in recent years due to evidence suggesting that these glycans play a central cohesive role in the secondary cell walls of angiosperms by interacting with both cellulose and lignin. Those interactions and therefore the role that xylan play in the cell wall seem to be highly

regulated by the plant through domains with specific substitution patterns. For that reason, a methodology to obtain defined xylan oligosaccharides is of potential interest.

2.1.1. Strategies to Obtain Xylan Oligo- and Polysaccharides

In the literature, several approaches to obtain xylan oligo- and polysaccharides have been described. Three strategies are reviewed here: (a) extraction, hydrolysis, and purification of plant raw materials to obtain xylan oligosaccharides (XOSs); (b) chemical synthesis (incl. automated solid-phase synthesis); (c) synthesis involving the use of glycosynthases.

Xylan oligosaccharides obtained from wood or other raw materials

There are several strategies to transform plant raw materials containing xylan into pure xylan oligosaccharides (XOSs). These strategies consist of one or more steps for the removal of xylan from the cell wall matrix and the hydrolysis of this glycan into shorter (unsubstituted) XOSs.

The process usually starts either with an acidic, basic, or enzymatic pretreatment step that prepares the raw material for the hydrolysis or with a step to fractionate the raw material into their cell wall components, usually by basic treatment.[1–3] These steps aim at removing any cell wall component or substituent that can hinder xylan hydrolysis. The hydrolysis of xylan (or xylan-enriched materials or liquors) is usually performed by acidic or enzymatic treatment. Acid hydrolysis has the advantage of being cheap and relatively fast.[4] Enzyme-mediated hydrolysis (usually done with xylanases), on the other hand, is a more selective process that minimizes the formation of side products of xylose degradation such as furfural.[5]

Hydrothermal treatment can also be used to simultaneously fractionate the cell wall components and hydrolyse xylan into XOSs.[6] This method uses water at temperatures between 150 and 230°C and at high pressure to both hydrolyse and extract XOSs or other cell wall components from plant biomass. It is usually the cheapest and most eco-friendly strategy.

Purification of the XOS-containing mixtures that result from any of these processes can be performed in several ways. Anion-exchange chromatography (AEC) allows for the separation of molecules by charge, such as those containing carboxylic acid substituents (e.g., MeGlcA) from neutral XOSs. Size-exclusion chromatography (SEC) allows for the separation of XOS by size, usually with a resin such as P2 gel from Bio-Rad Laboratories.[4] Alternatively, solid-phase extraction (SPE) with graphite carbon can also be used.

The main advantage of isolating xylans from natural resources is generally the reduce costs of the process. However, the diversity of available xylan structures is limited to the natural structures and by the possibility to isolate the desired target compounds. In addition, highly pure and monodisperse samples can sometimes be difficult to obtain as the physicochemical properties of fragments of comparable lengths tend to be very similar, complicating the purification. For all those reasons, this strategy to obtain xylan oligo- and polysaccharides is more common to be applied at industrial scale where these glycans can be used, for instance, as additives in the food and papermaking industries.

Chemical synthesis of xylan oligosaccharides (incl. AGA)

In the literature, there are several examples of the synthesis of defined homo- and heteroxylan oligosaccharides.[7] Via traditional chemical transformations, homoxylan oligosaccharides up to the decasaccharide have been prepared by stepwise synthesis using a disaccharide methyl thioxyloside as a glycosyl donor.[8] At the 4-OH glycosylation site, this donor was equipped with a chloroacetyl group that could be orthogonally removed. Glycosylation reactions were done using NIS/AgOTf in DCM and only the desired β-glycosylation product was obtained.

Homoxylan oligosaccharides up to the pentasaccharide have been also prepared by stepwise synthesis using 1,2,3-*O*-triacetyl-4-*O*-benzyl-D-xylosyl bromide as a glycosyl donor, which allowed for the selective removal of the 4-*O*-benzyl group at the glycosyl product by hydrogenolysis.[9] Instead of the Bn group, the chloroacetyl group was also considered, but purification of synthetic intermediates bearing this group was found to require chromatography, which was considered impractical due to the relatively large scale of the synthesis.

Despite the 2-*O*-acetyl group at the xylosyl bromide donor, glycosylation products were obtained as α/β mixtures of approx. 40:60.

Using automated glycan assembly (AGA) there are two examples for the synthesis of homo- and heteroxylan oligosaccharides,[10,11] which were reported from our laboratory. In both cases, dibutyl phosphate xylosides bearing an Fmoc group at the 4-OH were used as glycosyl donors for the automated glycan assembly (AGA). Homoxylan oligosaccharides up to the octasaccharide and heteroxylans up to a hexasaccharide backbone bearing arabinosyl and xylosyl substituents were prepared.

Xylan synthesis using glycosynthases

In the literature, there are several examples for the synthesis of xylan using glycosynthases. In 2004, AbgE358G-2F6, a glycosynthase obtained from the β-glucosidase of the *Agrobacterium* sp. (Abg; family GH1) was found to catalyse the coupling between α-xylosyl fluorides.[12,13] In 2005, a glycosynthase obtained from the inverting xylanase of *Bacillus halodurans* C-125 (Rex; family GH8) was not only the first glycosynthase obtained from a xylanase, but also the first one obtained from an inverting enzyme.[14] However, these glycosynthases can only produce xylotriose (a trisaccharide) and not larger xylan chains. The first glycosynthases capable of polymerising α-xylosyl fluorides were obtained from retaining xylanases belonging to family GH10 (described below).

Sugimura and colleagues obtained glycosynthases from *Thermotoga maritima* MSB8 XylB (TM-E259G), from *Clostridium stercorarium* F-9 XynB (CS-E293G), from *Bacillus halodurans* C-125 XynA (BH-E301G), and from *Cellulomonas fimi* Cex (CF-E233G) by replacing their glutamic acid catalytic residue by glycine. All these glycosynthases were able to produce polymeric 1,4-β-linked xylan precipitates and have no hydrolytic activity. However, while TM-E259G and CF-E233G were able to use β-xylobiose efficiently as an acceptor, CS-E293G and BH-E301G showed higher activity with β-xylotriose.[15]

Soon after, Kim and colleagues obtained glycosynthase CFXcd-E235G by replacing the catalytic glutamic acid also by glycine in the active site of the retaining *endo*-1,4-β-xylanase from *Cellulomonas fimi* (CFXcd). By using α-xylobiose fluoride and *para*-nitrophenyl β-

xylobiose as an acceptor, xylan tetra- to dodecasaccharides were obtained. *para*-Nitrophenyl β-xylobiose was found to be a better acceptor than α-xylobiose fluoride since it extends into the +3 subsite of the glycosynthase,[16] while α-xylobiose fluoride does not.[17]

Another example of a xylan glycosynthase is XynAE265G[18], also known as XT6E265G, which results from the replacement of the nucleophilic glutamic acid (Glu, E) residue 265 by glycine (Gly, G) in the active site of the extracellular *endo*-1,4-β-xylanase of *G. stearothermophilus*,[19] known as XynA, Xyn10A or XT6. The natural *endo*-xylanase (EC 3.2.1.8) belongs to family GH10 (clan GH-A) and has a relatively high stability at elevated temperatures and high pH. Enzymes in this family follow a retaining catalytic mechanism and both catalytic residues are glutamic acid; structurally, they form a $(\beta/\alpha)_8$ barrel fold (TIM barrel). The two catalytic residues of XynA, Glu159 (acid/base catalyst) and Glu265 (nucleophile), are at a 5.5 Å-distance, which is consistent with the retaining catalytic mechanism, and the TIM-barrel fold containing a deep active-site groove is consistent with its *endo*-activity.[19]

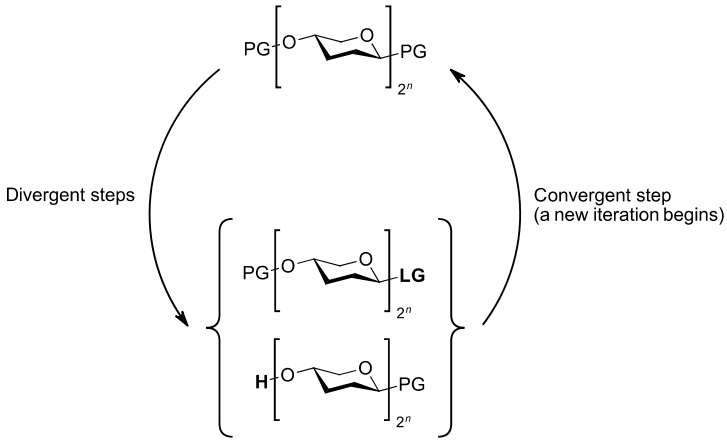
Glycosynthase XynAE265G has been shown to catalyse 1,4-β-glycosylation reactions with unprotected α-xylosyl fluoride donors of, at least, two sugar residues, e.g., α-xylobiosyl fluoride, with excellent regio- and stereoselectivity.[18,20] Previously in our laboratory, glycosynthase XynAE265G was used to obtain xylan polysaccharides with regular arabinosyl substitution patterns, emulating the structures of natural xylans.[20] Oligosaccharide α-xylosyl fluoride donors were allowed for self-condensation (by not blocking the reducing-end 4-OH group with a PG) producing xylans with a degree of polymerization (DP) that ranged from approx. 35 sugar residues for homoxylan to more than 200 residues for xylan bearing a 2-*O*-arabinosyl substituent at every third xylosyl residue.[20] The high disparity of these results and the fact that xylan chains with even or no substitution had an evidently lower degree of polymerization might be related to the conformations and interactions of the xylan molecules in solution and thus to the degree of access that the enzyme has to these molecules. Those results showed that this glycosynthase tolerates xylosyl substrates bearing substituents such as 2-*O*- and 3-*O*-arabinosyl groups at the terminal non-reducing xylosyl residue in both glycosyl donor and acceptor.

The advantages of using enzymes such as glycosynthases for the synthesis of xylan molecules are multiple. They can catalyse glycosylation reactions in aqueous buffer with unprotected substrates in high yield with high regio- and stereoselectivity, with the further advantage of a low toxicity. In addition, self-condensation of the glycosyl donor by the enzyme and therefore the formation of polymeric mixtures can be easily prevented by blocking the non-reducing end 4-OH with a protecting group.[21] In this way, only the desired glycosyl product between the glycosyl acceptor and donor is formed.

2.1.2. General Strategies for the Synthesis of Compounds with Repeating Units

The iterative divergent–convergent synthetic strategy

An iterative divergent–convergent (IDC) synthetic strategy[22] is a synthetic methodology where a molecule is partitioned into two portions and each portion is converted into a reactant for a subsequent coupling reaction (Scheme 2.1). This coupling product is again partitioned into two for the next coupling reaction. With each iteration, the length of the molecule grows exponentially. This strategy is ideal for the synthesis of linear molecules with a repeating structure (polymers) when monodisperse samples are wanted. With this method, *n* coupling reactions produce molecules of 2^n repeating units (exponential growth), as opposed to step-by-step synthesis where *n* couplings produce molecules of *n* repeating units (linear growth). This strategy was first reported in 1982.[23] In glycochemistry, examples of its implementation can be found in the work from the Kishi lab for the chemical synthesis of different 1,4-α-linked glycans;[24–26] or, more recently, from the Yu lab, in the chemical synthesis of a 128-mer glycan.[27]



Scheme 2.1. Iterative divergent-convergent synthetic scheme; *n* = number of iterations or cycles; 2^n = number of repeating units.

2.1.3. Challenges for the Synthesis of Long Unprotected Glycans

The limiting factor for the preparation of long glycan oligosaccharides from unprotected substrates in water is usually substrate solubility since the glycosylation reactions require the substrates to be dissolved. The solubility of glycans such as xylan tends to decrease with the increase in length of the glycan molecules, which can be explained in terms of entropic change since the aggregation of glycans seems to be an entropy-driven process.[28] Per mass, the entropic penalty of immobilizing xylan molecules should be smaller for a given mass of longer structures than for the same mass of shorter ones since the number of degrees of freedom lost upon aggregation of long structures should be smaller than upon aggregation of the same mass of shorter structures.

2.2. Aims

The aim of this chapter was to develop a methodology that produces pure and well-defined 1,4-β-linked xylan molecules of sufficient length to enable biosynthetic studies with xylan-modifying enzymes. Substrate length is a key factor for biosynthetic studies since the

activity of enzymes is often regulated by the length and the motifs present in the substrates and from longer substrates often information of higher quality about the enzymatic activity, such as the resulting substitution patterns, is obtained.

Two key ideas were considered to design the synthetic strategy towards obtaining these xylan molecules: (a) the implementation of a divergent-convergent iterative synthetic methodology where glycosylation products are converted into both glycosyl donors and glycosyl acceptors for the next glycosylation reaction; and (b) the used of glycosynthase XynAE265G for the coupling reactions between α-xylosyl fluoride donors and xylosyl acceptors. The xylosyl donors had to be equipped with a protecting group at the 4-OH of the non-reducing end xylosyl residue to prevent self-condensation during the enzyme-mediated glycosylation reactions. Xylan glycosynthase XynAE265G is known to catalyse glycosylation reaction with xylan substrates in aqueous medium with excellent stereo- and regioselectivity and was already employed with excellent results in our laboratory for the synthesis of a library of xylan polysaccharides with different arabinosyl substitution patterns.

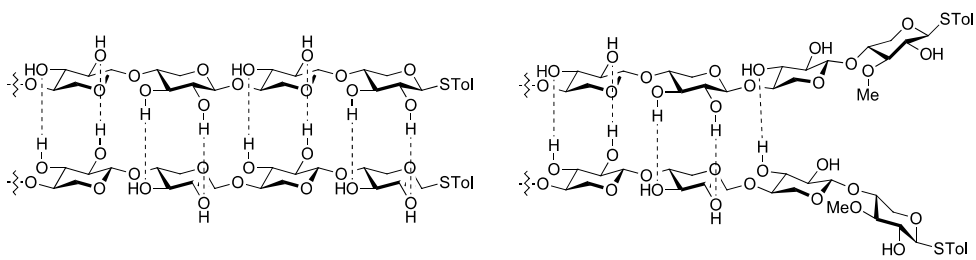
In this chapter, the synthesis of both a series of unsubstituted xylan oligosaccharides and a series of xylan oligosaccharides substituted with a methyl group at the 3-OH of the reducing end xylosyl residue was planned. Because the solubility of glycans decrease as they increase in chain length, the methyl substituent was included to potentially increase the water solubility of the respective long-chain xylan oligosaccharides. Because of its relatively small size, the methyl substituent was hypothesised to neither affect the activity of the xylan glycosynthase used for the glycosylation reactions or the activity of xylan-modifying enzymes, such as the glucuronosyltransferase *At*GUX3 from *Arabidopsis thaliana,* studied in later biochemical assays.

2.3. Results and Discussion

2.3.1. Synthetic Strategy and Structural Features of the Glycosyl Building Blocks

The synthetic approach to obtain xylan oligosaccharides applied in this thesis was designed around two central ideas: (a) an iterative diverge–convergent (IDC) synthetic strategy where glycosylation products are converted again into glycosyl donors and acceptors for the following glycosylation step, and (b) the use of an xylan glycosynthase for the glycosylation steps.

Due to the limited solubility of unprotected long xylan structures, two parallel xylan oligosaccharide syntheses were performed. One of the series of target oligosaccharides carry an unnatural methyl group substituent at the 3-OH of the reducing end xylosyl residue. This substituent improved the water solubility of the xylan oligosaccharides probably due to the disruption of H-bonding interactions between xylan chains, even if xylan aggregation seems to be an entropy-driven process[28] (Scheme 2.2).



Scheme 2.2. Possible conformations of and interactions between xylan oligosaccharide chains with and without the disruption of the intermolecular interactions of a methyl group substituent.

The synthesis of xylan oligosaccharides here reported consists of two parts: (a) a chemical synthesis of the disaccharide building blocks (BBs), α-xylosyl fluoride donor **22** and xylosyl acceptors **15** and **17** (Scheme 2.3), and (b) a chemoenzymatic iterative synthesis, with every iteration consisting of a divergent part, involving only chemical transformations, and a convergent part, consisting of an enzyme-mediated glycosylation.

The enzyme-mediated glycosylation step is performed in aqueous buffer, and, for that reason, it requires water-soluble unprotected substrates. However, to prevent polymerization of the glycosyl donors in the enzyme-mediated glycosylation step, i.e., self-condensation, the α-xylosyl fluoride donors (e.g., **22**) were equipped with a protecting group on the 4-OH of the non-reducing end xylosyl residue. The two xylosyl acceptor BBs (**15** and **17**) differ only in their substitution at the 3-OH of the reducing end xylosyl residue.

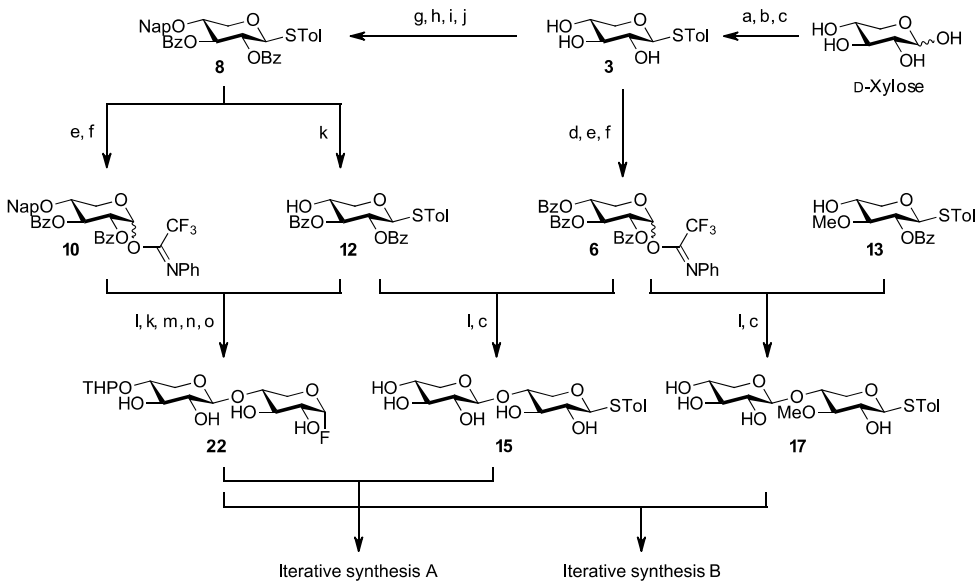
2.3.2. Synthetic Planning for Chemical BB Synthesis

In comparison with other natural products, the synthetic planning in glycochemistry, as it is also the case for peptides and nucleic acids, is relatively simple as repeating structures are very common and glycosidic linkages are arguably the only connection between glycosyl residues.

In order to minimize the waste of materials (i.e., SM, intermediates, etc.), it is important to calculate the partition ratios at the divergent nodes of the synthesis. A divergent node is a point in the synthesis where a reactant, e.g., an intermediate, must be partitioned between two synthetic pathways. Even though there are often changes in the synthetic plan and steps that require investigation and therefore additional material, calculating partition ratios facilitates decision making regarding how much material to employ in each reaction pathway in a divergent node. Three parameters must be taken into consideration for each pathway: (a) reported or expected yields, (b) number of steps, and (c) number of equivalents used per reaction, which is only in convergent steps between pathways not equal to a value of one, i.e., in coupling reactions between synthetic intermediates.[29]

In the synthetic planning of the BB synthesis in this work (Scheme 2.3), there are four divergent nodes that require batch partition: the nodes at intermediates **3**, **6**, **8** and **12**. There is an additional divergent node, but the partition of glycosyl donor **22** should be done either 1:1 since both iterative syntheses are equivalent, or in agreement with the obtained amounts of **15** and **17**.

Calculation of partition ratios during synthetic planning is done based on approximated yields and play only a guiding role. During calculation, the goal is to set up a system of linear equations where ultimately the number of moles of an intermediate that must be partitioned in two (or more) pathways is written as a function of the number of moles of the final products. When the system of equations is then simplified as a single equation showing the ratio between the moles of intermediate that should be employed in one or the other pathway ($n_{\text{pathway A}}/n_{\text{pathway B}}$), a numerical value should be obtained, i.e., the partition ratio. As an example, the calculations for the partition ratio (in moles) in the divergent node of intermediate **8** are explained below and shown in Scheme A.1 of the Appendix. The calculations for the divergent nodes of the iterative synthesis, e.g., **23,** are shown in Scheme A.2 of the Appendix.



Scheme 2.3. Synthetic tree for BB syntheses, showing only the structures prior to convergent or divergent nodes. Every letter corresponds to a chemical transformation; in Scheme 2.5 and Scheme 2.6A, reactions conditions and the structures of the remaining intermediates are given.

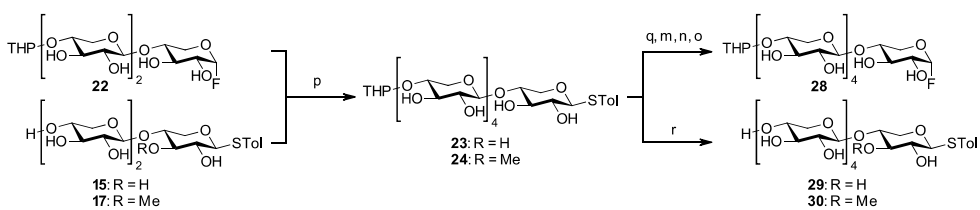
Here a detailed explanation for the construction of the system of linear equations shown in the Appendix is given. The number of moles of intermediate **8** (n_8) must be partitioned between the pathway towards **10** ($n_{8\text{to}10}$) and the pathway towards **12** ($n_{8\text{to}12}$) (Scheme 2.3).

n_{8to10} is equal to the moles of **10** divided by the product of the yields of the steps occurring between the two intermediates (**8** and **10**), steps *e* and *f* in this case (ϵ_e and ϵ_f). The calculation for the node in intermediate **12** (from which the partition ratio of **8** depends) is done in the same way, with the particularity that the equation for n_{12to22} involves a factor (1.2) because **12** is planned to be used in slight excess, 1.2 equiv, in the coupling reaction with **10**, and therefore an additional amount of **12** will be consumed in the process.

Precise partition ratios can be obtained if reproducible yields are known. For general guidance, the partitioning ratio for **8** (Scheme 2.3) was found to be 45:55 (n_{8to10}/n_{8to12}), and the ratio for **3**, between 15:85 and 20:80 (n_{3to6}/n_{3to8}) (numerical calculations not shown).

Donor design for glycosynthase-mediated reactions

The α-xylosyl fluorides employed in this thesis were chosen for being compatible with xylan glycosynthase XynAE265G. In order to avoid enzyme-mediated self-condensation, these donors were protected at the glycosylation site (4-OH group of the non-reducing end xylosyl residue). This protecting group had to be orthogonal to acyl groups and ideally withstand the fluorination conditions required for the formation of the α-xylosyl fluorides, so an additional reinstallation step can be avoided (Scheme 2.4). Previously, a THP group was used in glycosynthase-mediated glycosylation reactions[21] and because it has a pyranosyl ring structure, it is potentially compatible with the xylan glycosynthase. However, a THP group, even though it is orthogonal to acyl groups, cannot withstand the required fluorination conditions and must be reinstalled in an additional step. For that reason, other PGs were also investigated (PMB, allyl and Nap), which will be shown in the next section.



Scheme 2.4. A chemoenzymatic iteration consisting of the enzymatic convergent step (p) and the chemical divergent steps (q, m, n, o, r).

The partition ratio (in moles) of the product of each enzymatic glycosylation reaction, e.g, **23,** should be 60:40 between conversion to the donor and to the acceptor (system of linear equations shown in Scheme A.2 of the Appendix).

2.3.3. Synthesis

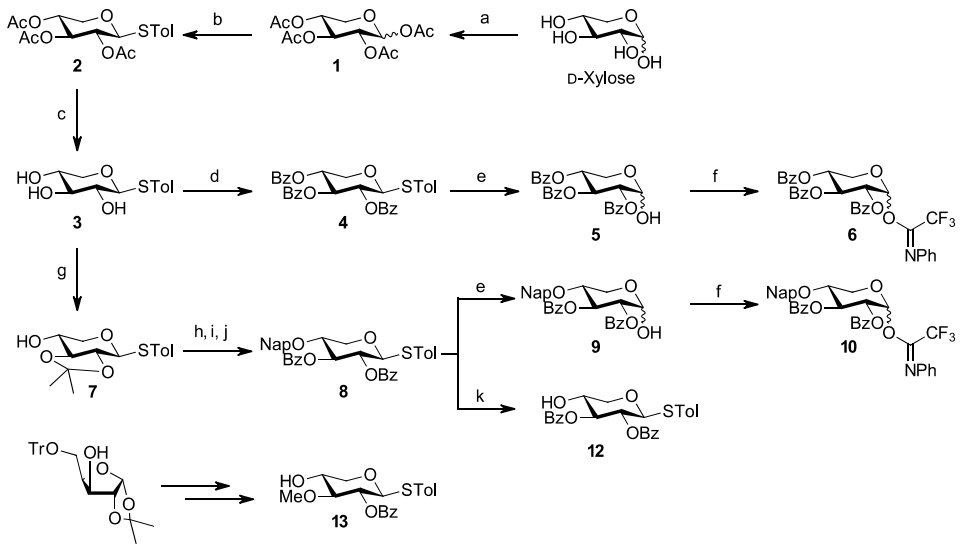
Building block chemical synthesis

Monosaccharide BBs

The chemical synthesis of the disaccharide building blocks started from D-xylose (Scheme 2.5). This monosaccharide was acetylated (**1**), then transformed into acetylated *p*-tolyl thio-oxyloside (**2**) and then deacetylated under Zemplén conditions[30,31] to afford **3,** which was purified by crystallization from warm EtOAc/hexane. Intermediate **3** was partitioned 1:4 between the pathways leading to **6** and to **8,** respectively. To obtain **6**, thioxyloside **3** was dissolved in Et_3N and benzoylated with BzCl to form **4**, and then hydrolysed with NIS/TFA in DCM/H_2O (10:1) in an 83% yield, following the method developed by the van der Marel group[32]; 2-OH benzoyl α-D-xyloside side-product was also isolated. At hemiacetal **5**, a 2,2,2-trifluoro-*N*-phenyl acetimidoyl group was installed to form **6**.[33] Alternatively, **6** can also be obtained from **1** or **2**. To obtain **8**, a 2,3-*O*-isopropylidene group was first installed by syringe-pump addition of 2-methoxypropene to CSA in DMF at 60°C in a 59% yield, with the 3,4-*O*-isoporpylidene isomer also being isolated in 18% yield (2,3-*O*-isomer/3,4-O-isomer, 3:1). Even though higher yields have been reported in the literature**,**[34] those could not be achieved under these conditions and increasing the temperature to 70°C gave slightly lower yields. 2,3-*O*-Protected intermediate **7** was alkylated using NapBr and NaH in DMF, known as the Williamson etherification, and then the 2,3-*O*-isopropylidene group was substituted by two benzoyl groups. This conversion was done by acidic treatment, with a few drops of 1M HCl(aq), followed by benzoylation using BzCl in Et_3N.[34]

Intermediate **8** was partitioned between the pathways towards **10** and towards **12**, approx. 55:45. To obtain **10**, thioglycoside **8** was hydrolysed using NIS/TFA in DCM[32] before the installation of the acetimidoyl group on intermediate **9**.[33] To obtain **12**, Nap-containing

thioglycoside **8** was selectively deprotected at the 4-OH group with DDQ in DCM/H_2O (10:1), over 120 min, to give the desired product in 82% yield; over prolonged reaction times, DDQ is known to slowly react with benzoyl groups as well, so long reaction times need to be avoided. Xylosyl acceptor **13** can be obtained from 1,2-*O*-isopropylidene-5-*O*-triphenylmethyl-α-D-xylofuranoside and was available in the working group.



Scheme 2.5. Chemical synthesis of monosaccharide BBs for chemical glycosylation reactions; (a) Ac_2O/py in Et_2O, 0°C to RT; (b) *p*-thiocresol, $BH_3.OEt_2$ in DCM, 0°C to RT; (c) MeONa in DCM/MeOH (1:3), 0°C to RT; (d) BzCl in Et_3N, 0°C to RT; (e) NIS/TFA in DCM/H_2O (10:1), 0°C; (f) 2,2,2-trifluoro-*N*-phenyl acetimidoyl chloride, Cs_2CO_3 in acetone, 0°C to RT; (g) 2-methoxypropene, CSA in DMF, 60°C; (h) NapBr, NaH in DMF, 0°C to RT; (i) gentle acidification with 1M HCl (aq); (j) BzCl in Et_3N, 0°C to RT; (k) DDQ, DCM/H_2O (10:1), 0°C.

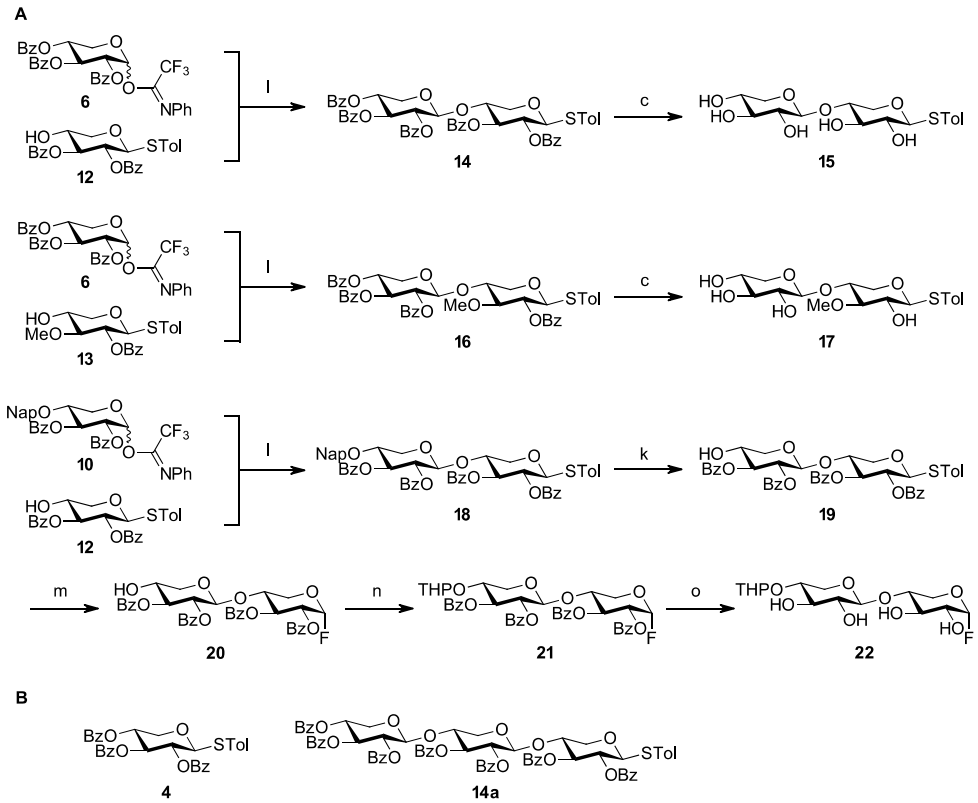
Disaccharide BBs

With xylosyl donors **6** and **10**, and xylosyl acceptors **12** and **13** in hand, glycosylation reactions between glycosyl donors and acceptors were performed using NIS/TfOH in DCM, at −50 to −10°C, to obtain the desired 1,4-β-linked disaccharides (Scheme 2.6A). The stereoselectivity of the reactions was controlled by the well-known neighbouring-group participation effect of the 2-*O*-acyl groups. These donors form 1,2-*trans* products since the

glycosylation reactions take place through the formation of a dioxalenium ion intermediate, which in some cases can be found to be in equilibrium with the covalent or contact ion-par glycosyl triflate counterpart.[35,36] In xyloses, this effect leads to the formation of β-xylosyl products.

Yields in the glycosylation reactions with donor **6** were only moderate since the reactivity of this xylosyl donor is compromised by the presence of benzoyl substituents in the molecule. Equatorially oriented, electron-withdrawing acyl groups are known to destabilize positively charged glycosylation intermediates more than other substituents such as alkyl groups would do. The formation of a poorly stabilized dioxalenium ion intermediate often leads to aglycone transfer from thioglycosides present in solution that can better stabilize the positive charge, e.g., glycosyl acceptor or product. This intermolecular group transfer leads to the formation of side-products and to an overall decrease in the yield of the desired product. In this case, in the reaction between **6** and **12**, the desired product **14** was obtained in a 50% yield and the monomer (**4**) and trimer (**14a**) that result from aglycon transfer were obtained in a 33% and 8% yield, respectively (Scheme 2.6B). The glycosylation reaction between donor **6** and acceptor **13** resulted in a similar yield (54%) and 19% of acceptor was recovered. In comparison, disaccharide **18** was obtained in 83% yield.

Acceptors **15** and **17** were obtained from the corresponding glycosylation products by diacylation of **14** and **16** using Zemplén conditions (Scheme 2.6A). Starting materials were first dissolved in one part of DCM before the addition of 3 parts of MeOH ensured a homogeneous medium. As a base catalyst, sodium methoxide was used, but other bases in MeOH (e.g., NaOH, NH_3) are known to function in the same way.[30]



Scheme 2.6. A. Chemical synthesis of disaccharide BBs for the xylan iterative synthesis; (c) cat. MeONa in DCM/MeOH (1:3); (k) DDQ in DCM/water (10:1); (l) TMSOTf in DCM; (m) NIS, HF/py (7:3; 100 equiv) in DCM; (n) DHP and CSA; (o) cat. MeONa, in DCM/MeOH (1:3). **B.** Side-products formed in the glycosylation reaction between **6** and **12**.

α-Xylosyl fluoride donor **22** was obtained from disaccharide thioxyloside **18** (Scheme 2.6A). Since the Nap group was found to be partially cleaved during the fluorination reaction, the sequence of transformations between **18** and **22** included the removal of Nap prior to the fluorination reaction, and the installation of the THP group after the fluorination reaction, which is known to be compatible with the xylan glycosynthase. The Nap group was removed from **18** using DDQ in DCM/H_2O (10:1) in an 84% yield, followed by fluorination of the resulting product **19** using NIS and HF/py (100 equiv) in DCM, at −50 to −3°C for 5.5h, to obtain **20** in an 89% yield with an α/β ratio of 93:7.[37] Shorter reaction times or lower temperatures (even if only slightly lower) led systematically to incomplete conversion into

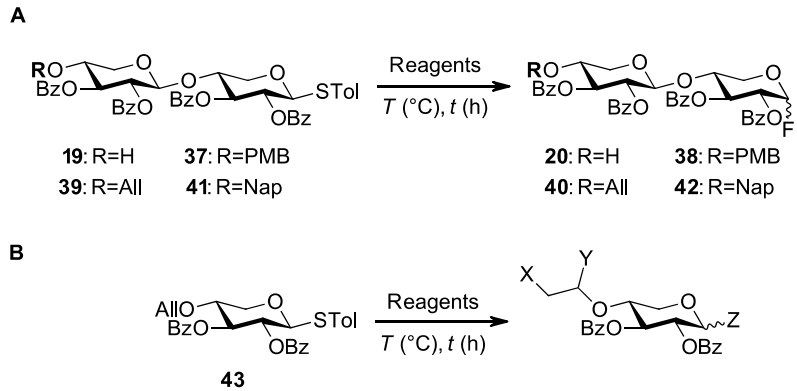
the desired α-anomer; details regarding the screening of reaction conditions are given at the following subsection and in Table 2.5. After this step, a THP group was installed at the unprotected 4-OH group and the glycan was carefully debenzoylated under Zemplén conditions at 0°C to form the desired disaccharide α-xylosyl fluoride donor **22**, which was purified during the workup by extraction and washing using Et_2O and water. Deprotected α-xylosyl fluorides such as **22** can be lyophilised and safely stored in a freezer at −20°C.

Protecting group investigation for the 4-OH glycosylation site

During the BB synthesis above described, the Nap group was installed to protect the 4-OH glycosylation site (on **7**), then removed prior to the fluorination step (from **18**) and then the THP group was installed (on **20**). During the design of the synthesis, finding a single orthogonal, permanent protecting group that could be installed in the beginning of the synthesis and removed with the final deprotection was intended to save one synthetic step during the synthetic sequence towards the formation of the glycosyl fluoride donors. The desired characteristics for this kind of PG would be: (a) being able to withstand fluorination conditions (NIS and HF/py in DCM), (b) being orthogonal to the other PG and stable under chemical glycosylation reactions, (c) being easily removable from xylan oligosaccharides in water, and (d) being accommodated by the xylan glycosynthase.

During previous studies, the THP group was found to be accepted by the enzyme, but its high lability to acidic conditions leads to its cleavage during the fluorination step. For that reason, PMB, allyl and Nap protecting groups were installed at the 4-OH glycosylation site and investigated.

Of all the chemical transformations during the building block synthesis, the fluorination step to obtain the thermodynamically favoured α-xylosyl fluoride (e.g., **20**) posed the greatest challenge since hydrogen fluoride (70% in pyridine) is acidic enough to cleave acid labile protecting groups.



Scheme 2.7. A. Disaccharide structures used to investigate different 4-OH protecting groups. **B.** Reactive points of the allyl-protected thioxyloside model **43** under hydrolysing conditions.

The first protecting group to be investigated was PMB (Table 2.1). In a first attempt, PMB was found to be completely cleaved after treatment of **37** with NIS and HF/py in DCM at temperatures from −65°C to −20°C for 3h (Scheme 2.7B). Alternatively, the hemiacetal obtained from **37** was treated with DAST in THP, but the β-xylosyl fluoride (**38β**) was exclusively formed and anomerization to the α-counterpart (**38α**) with 20 equiv of HF/py in pyridine (50 mM of substrate) overnight at 0°C to room temperature did not occur, leading to the full recovery of the β-xylosyl fluoride (**38β**). An attempt to anomerize the β-xylosyl fluoride using HF/py in DCM resulted in partial PMB cleavage at temperatures equal to or below −42°C.

Substrate	Reagent	Equiv	*T* (°C)	*t* (h)	Outcome
PMB-protected thioglycoside (**37**)	NIS, HF/py in DCM	>100 (HF/py)	−65°C to −20°C	3	**20**; PMB fully cleaved
PMB-protected hemiacetal	DAST in THF	1.2 (DAST)	−20°C to 5°C	5	**38β** (78%)
PMB-protected β-xylosyl fluoride (**38β**)	HF/py in py (50 mM)	20 (HF/py)	0°C to RT	o/n	SM recovered (**38β**)
	HF/py in DCM	A few drops (HF/py)	−50°C to −42°C	0.5	**20** and **38**; PMB almost fully cleaved

Table 2.1. Fluorination reactions on PMB-bearing substrates.

After the observation that the PMB group was easily cleaved under the fluorination conditions, the allyl group was investigated (Table 2.2 and Table 2.3). Substrates bearing this group were found to undergo electrophilic addition side-reactions on the allyl group under different conditions. Attempts to hydrolyse the thioxyloside **43** using NIS or trichloroisocyanuric acid (TCAA) as a promoter in acetone/water to obtain the corresponding hemiacetal gave a mix of products, including iodohydroxylation of the allyl group (Table 2.2; Scheme 2.7). Chloramine-T in acetone/H_2O (not in DCM or CH_3CN) did selectively hydrolyse the glycosidic bond on model thioxyloside **43**. However, when the same conditions (chloramine-T in acetone/H_2O) were applied to disaccharide **39**, precipitation occurred, and SM was recovered.

Attempts to fluorinate disaccharide **39** using HF/py in DCM, and NIS or chloramine-T as promoters also gave mixtures of products, again mostly due to electrophilic addition reactions; using NIS, at least small amounts of xylosyl fluoride **40** could be isolated (Table 2.3).

Substrate	Reagent	Equiv	*T* (°C)	*t* (h)	Outcome
Allyl-protected thioglycoside (**43**)	NIS in acetone/H_2O (9:1)	1.2 (NIS)	0°C	2	Mix, incl. iodohydroxylation of the allyl group
Allyl-protected thioglycoside (**43**)	TCCA in acetone/H_2O (9:1)	0.3 (TCCA)	rt	o/n	SM, iodohydroxylation of the allyl group and desired hydrolysis
Allyl-protected thioglycoside (**43**)	Chloramine-T in DCM or CH_3CN	1.5 (chloramine-T)	0°C	3	Very little to no desired product formation
Allyl-protected thioglycoside (**43**)	Chloramine-T in acetone/H_2O (9:1)	2 (chloramine-T)	0°C	3	Formation of desired hemiacetal
Allyl-protected thioglycoside (**39**)			0°C to RT	6	SM recovered (**39**) (78%)

Table 2.2. Hydrolysis reactions on allyl-bearing substrates.

Substrate	Reagent	HF/py (equiv)	*T* (°C)	*t* (h)	Outcome
Allyl-protected thioglycoside (**39**)	NIS, HF/py in DCM	>100	−50°C to RT	o/n	Fluoride product with iodofluorination of the allyl group
Allyl-protected thioglycoside (**43**)	NIS or chloramine-T, HF/py in DCM		−50°C to −10°C	4	Mix of products; with NIS, partial formation of the xylosyl fluoride

Table 2.3. Fluorination reactions with allyl-bearing substrates.

As expected, the Nap group was significantly less labile than PMB under acidic conditions (Table 2.4). The Nap group was stable in NIS and HF/py in DCM on benzoylated substrates

(**41** and **42**) up to −15°C (Scheme 2.7). However, higher reaction temperatures resulted in partial removal of the group. When benzoyl groups were substituted by acetyl groups, treatment with NIS, HF/py in DCM resulted in complete cleavage of the Nap group at temperatures equal to or below −20°C. The increase of the lability of Nap protecting group is potentially due to the reduction of the steric hindrance in the substrate.

Substrate	Reagent	HF/py (equiv)	*T* (°C)	*t* (h)	Outcome
Nap-protected thioglycoside (**41**)	NIS, HF/py in DCM	65	−50°C to −15°C	6	**42** α/β, 12:88
		65	−50°C to −5°C	5.5	**42** and **20**
		25	−50°C to 5°C	5	**20** (46%); 1-mer α-xylosyl fluoride (17%)
Nap-protected thioglycoside (acetylated)		65	−50°C to −20°C	2	Nap fully cleaved

Table 2.4. Fluorination reactions on Nap-bearing substrates.

Because all protecting groups investigated partially reacted under the fluorination conditions, the fluorination step was performed without a 4-OH protecting group, and a THP protecting group was installed in an additional next step. It has been found that temperatures up to −3°C for at least 5.5h are necessary for the complete equilibration of the initially formed β-xylosyl fluoride into the α-counterpart **20**, while temperatures above 0°C can result in cleavage of the glycosidic bond between xylosyl residues (Table 2.5). Under a controlled constant temperature of −5°C for prolonged reaction times (>16h), no glycosidic bond cleavage has been observed.

More precisely, when treating thioglycoside **19** with NIS, HF/py in DCM for 5.5h at a temperature oscillating between −3°C and −5°C, the desired xylosyl fluoride **20** (α/β, 93:7) was obtained in 89% yield. However, treatment of the same material (**19**) overnight at −5°C using a cooling liquid circulator only gave xylosyl fluoride **20** in a α/β ratio of 85:15. Therefore, workup of the reaction was done and, after resubmission at the same temperature

for one additional night, the desired product **20** was obtained (α/β, 93:7) in 90% yield in a 1.5-gram scale. Important to note that with the reaction having reached temperatures between 0°C and 5°C in another batch using Nap-containing substrate **41**, 2,3-*O*-dibenzoyl-α-D-xylosyl fluoride was isolated in 17% yield (Table 2.4). Therefore, the optimal temperature for the conversion of thioglycoside **19** into the desired α-xylosyl fluoride (**20**) seem to be around −3°C for approx. 6h. Alternatively, treatment at −5°C for approx. 30h can give the same results (Table 2.5).

Substrate	Reagent	Equiv of HF/py	*T* (°C)	*t* (h)	Outcome
4-OH thioglycoside (**19**) (1.5-gram scale)	NIS, HF/py in DCM	50	−50°C to −5°C	o/n	**20;** α/β, 85:15
				o/n + o/n	**20** (90%); α/β, 93:7
4-OH thioglycoside (**19**)		100	−50°C to −3°C	5.5 h	**20** (89%); α/β, 93:7

Table 2.5. Fluorination reactions on substrates without a 4-OH protecting group.

Chemoenzymatic iterative synthesis

With α-xylosyl fluoride **22** and xylosyl acceptors **15** and **17** in hand, the chemoenzymatic iterative syntheses of xylan oligosaccharides was started. Two parallel iterative syntheses were done, one with **15** and one with **17** as xylosyl acceptors (Scheme 2.8). While the use of **15** provides access to unsubstituted xylan oligosaccharides, acceptor **17** affords xylan oligomers bearing a methyl group at the 3-OH of the reducing end xylosyl residue. In both iterative syntheses, only α-xylosyl fluoride donors without methylation were used.

Glycosynthase-mediated glycosylation reactions

Enzyme-mediated glycosylation reactions were catalysed by xylan glycosynthase XynAE265G[18] and occurred regio- and stereoselectively without formation of sugar side-products. In order to avoid self-condensation of the α-xylosyl fluoride donors and thus polymerization, these donors were equipped with a THP group at the 4-OH glycosylation

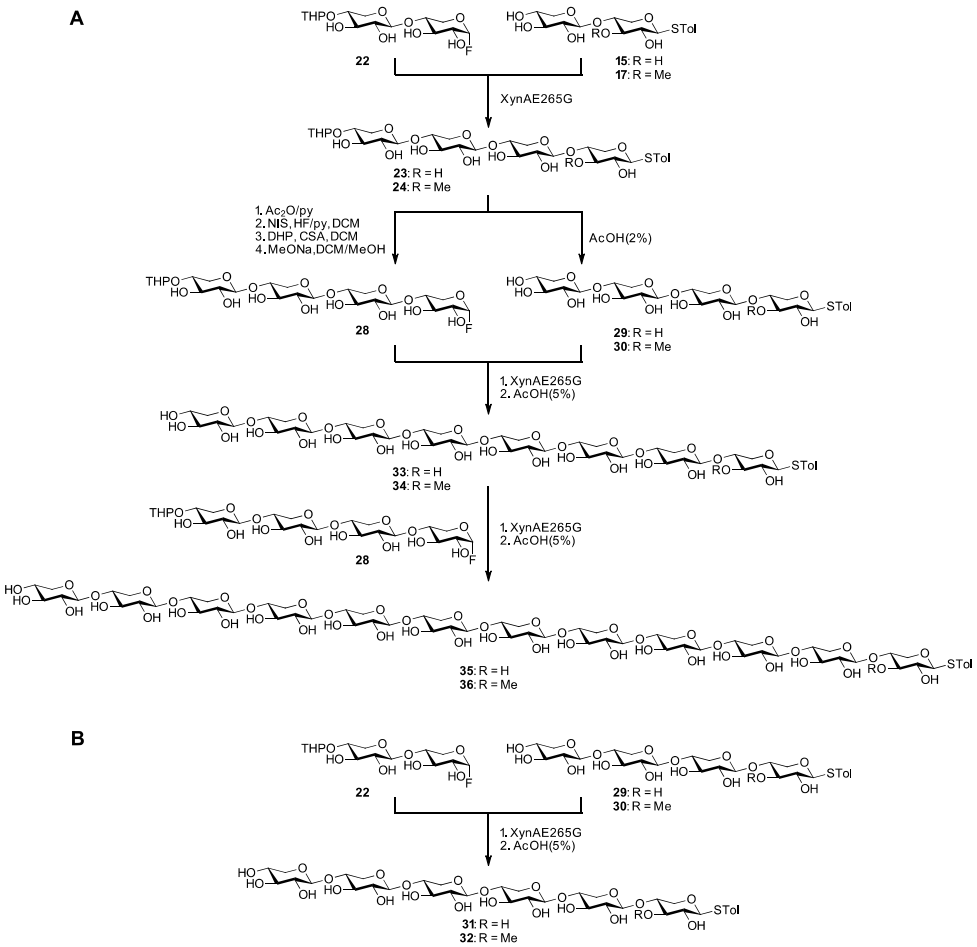
site, which was found to be accepted by the glycosynthase and was easily removed in water by mild acidic treatment.

For the enzyme-mediated glycosylation reaction, 1.2 equiv of α-xylosyl fluoride donor and 1 equiv of xylosyl acceptor were dissolved in phosphate buffer (100 mM; pH 7.4), and XynAE265 in phosphate buffer was added so that a final concentration of 500 to 1000 μg/mL was reached. The donor was usually added in excess because, once the reaction is complete, separating glycosyl fluorides or hemiacetals from thioglycosides using size exclusion chromatography (SEC) is easier than separating thioglycosides of different chain lengths, e.g., glycosyl acceptor and product; with the donor in excess, virtually all glycosyl acceptor is converted into the desired glycosyl product.

The iterative synthesis began with dissolving deprotected α-xylosyl fluoride **22** and acceptors **15** or **17** in phosphate buffer (100 mM; pH 7.4) followed by incubation with XynAE265G at 30°C. After 16h, tetrasaccharide products **23** and **24** were obtained in 96% and 81%, respectively. Tetrasaccharide **24** was then divided into two batches, with approx. 60% yield for the synthesis of tetramer xylosyl fluoride donor **28** and 40% for the formation of acceptor **29**. The following fluorination reaction towards the synthesis of tetrasaccharide **28** was complete at lower temperatures and shorter reaction times than in the case towards disaccharide **22**, which is probably due to the type of acyl groups present in each molecule. Acetyl groups are less bulky than benzoyl groups and they can enable the anomeric carbon atom to be more electrophilic, thus facilitating the equilibration to the α-anomer.

Removal of the THP group from **23** or **24** was performed with acetic acid (2% in water), providing **29** and **30** in 91% and 88% yield, respectively. There are examples in the literature of 1M HCl(aq) being used for this deprotection on 1,4-β-linked glucosyl oligosaccharides,[21] but those glycosidic bonds seem to be significantly less reactive than the ones in 1,4-β-linked xylosyl oligosaccharides. In the literature, controlled cleavage of glycosidic bonds in xylan by acidic treatment is reported[4] and there are even procedures for the decomposition of xylose at higher temperatures into products such as furfural.[5] After trying, 50 mM HCl(aq) for THP removal, 2 to 5% AcOH in water was found to provide the best results.

Samples lyophilised from DMSO seemed to dissolve better in water, which can help for the deprotection of long xylan structures.



Scheme 2.8. Chemoenzymatic iterative synthesis of xylan oligosaccharides. **A.** Synthesis of the dodecasaccharides **35** and **36**. **B.** Synthesis of hexasaccharides **31** and **32**.

Tetrasaccharide acceptors **29** and **30** were then coupled to tetrasaccharide α-xylosyl fluoride donor **28** following the previously described enzymatic procedure using XynAE265G. Subsequent acidic treatment with 5% acetic acid in water yielded octasaccharides **33** and **34** in 53% and 43% yield, respectively, over two steps. Octasaccharide xylosyl acceptors **33** and **34** were then coupled with α-xylosyl fluoride donor **28** to furnish the corresponding

dodecasaccharide products. Final treatment with 5% acetic acid in water yielded dodecasaccharides **35** and **36** in 38% and 33% yield, respectively, over two steps.

Additionally, xylosyl acceptors **29** and **30** were coupled with disaccharide α-xylosyl fluoride donor **22** to form the corresponding hexasaccharides, which after THP removal yielded oligosaccharides **31** and **32** in 49% and 68% yield, respectively, over two steps.

Reaction monitorization and xylan oligosaccharide purification

Monitorization of the reactions and of the SEC purifications was performed by thin layer chromatography (TLC) and reversed-phase liquid chromatography (also known as hydrophobic chromatography) coupled to a mass spectrometer (RFLC-MS).

For TLC, eluents mixtures of high polarity where used. For polar molecules, common eluent mixtures are, e.g., EtOAc/AcOH/H_2O, $CHCl_3$/MeOH/H_2O (e.g., 10:10:1, 10:10:3), EtOAc/MeOH/H_2O (e.g., 7:2:1), *i*-PrOH/EtOAc/H_2O (e.g., 3:1:1), EtOAc/MeOH, DCM/MeOH, etc. Here it was found that mixtures of EtOAc/AcOH/H_2O between 7:2:1 and 2:2:1 gave resolved TLCs that enable the clear differentiation between oligomers of different lengths. For RFLC-MS, a C4 column using a water/acetonitrile gradients from 95:5 to 80:20 was found to give the best results. While fluoride donors or their hemiacetal counterparts run almost with the solvent front, glycosyl acceptors and products have different retention times, potentially due to the hydrophobic interactions between the 4-methylphenyl substituent and the stationary phase.

Purification of the glycosyl products was done in the first instance based on the differences in solubility between the glycosyl product and the glycosyl donor or acceptor. Due to the longer chain length of the glycosylation products, they were less soluble than donors and acceptors. Interestingly, the THP group has also a significant effect on solubility; THP-containing glycosyl products were less soluble than the deprotected acceptors of the same length. This observation is in agreement with the model used in Chapter 1 to explain glycan aggregation in water; THP-containing xylan oligosaccharides are less soluble because aggregation of the non-polar surfaces of the THP group increases the entropy of the water

molecules that were in contact with the non-polar surfaces of the THP groups and thus the entropy of the system also increases.

Once the insoluble fraction of the glycosyl product was separated by carrying out one or two cycles of centrifugation followed by decantation of the supernatant solution, the aqueous solutions were combined. After lyophilization followed by dissolution in a minimum amount of water, the remaining glycosylation product was purified by size exclusion chromatography (SEC) using Bio-Gel P-2 gel from Bio-Rad laboratories. This SEC resin is a polyacrylamide gel, a copolymer of *N*,*N*'-methylenebisacrylamide and acrylamide, that can be used to separate molecules dissolved in water with masses between 100 and 1800 u.[38] Sephadex G-10 resin was also tried for product purification by SEC but did not provide superior results in purification of oligomers of this size.

2.3.4. Studies with Xylan-Modifying Enzymes on Xylan Substrates

With dodecasaccharide **36** in hand, studies with xylan-modifying enzymes were performed by Dr Colin Ruprecht in our laboratory. The catalytic domains of four xylan-modifying enzymes were heterologously expressed in HEK293 cells and tested against the newly synthesised acceptor substrate **36**, which was selected due to its higher solubility in comparison with the non-methylated substrate **35**. The four enzymes investigated were: (a) the well-characterized acetyltransferase *At*XOAT1 (from *Arabidopsis thaliana*), which served as a proof of concept for the newly synthesised substrate; the arabinosyltransferases (b) *Os*XAT3 (from *Oryza sativa*; rice) and (c) *Pv*XAT3 (from *Panicum virgatum*; switchgrass); and (d) the glucuronosyltransferase *At*GUX3 (from *A. thaliana*), which had not been heterologously expressed so far.

The acetyltransferase *At*XOAT1 is known to exclusively acetylate at the 2-OH group of xylan molecules.[39] Incubation of **36** with acetylsalicylic acid and *At*XOAT1 resulted in the formation of xylan oligosaccharides with five to ten acetyl groups attached to the xylan backbone, as determined by MALDI mass spectrometry. *Os*XAT3 and *Pv*XAT3 belong to family GT61, in which several other enzymes were found to be involved in the 3-*O*-Ara*f*

substitution of grass xylans. *Os*XAT3 transferred up to two Ara*f* to synthetic dodecasaccharide **36** and *Pv*XAT3 transferred up to five Ara*f* substituents. These two examples of grass XATs show how these enzymes can generate highly substituted arabinoxylan polysaccharides. The catalytic domain of *At*GUX3 (family GT8) has been here for the first time successfully expressed. In family GT8, *At*GUX1 and *At*GUX2 are reported to add glucuronic acid to xylan in the SCW of Arabidopsis, providing substitution patterns.[40] More recently, *At*GUX3 was identified as the only glucuronosyltransferase required for the installation of GlcA at the PCW xylan of Arabidopsis and, from studies on Arabidopsis *gux3* mutants, this enzyme was found to produce a unique substitution pattern in which every sixth xylose residue is substituted with a glucuronic acid.[41] However, biochemical characterization of *At*GUX3 had not yet been done. Here, we incubated the newly expressed catalytic domain of *At*GUX3 with dodecasaccharide **36** and observed that only one equivalent of GlcA from UDP-GlcA was transferred. Subsequently, the position of the GlcA-containing xylan oligosaccharide substrate was investigated.

MALDI studies on the xylan substrate modified by glucuronosyltransferase *At*-GUX3

To investigate the position of the GlcA substituent on the xylan backbone of **36** after the incubation with *At*GUX3, MALDI-TOF MS/MS (matrix-assisted laser desorption/ionization, time-of-flight, tandem mass spectroscopy) was used. The MALDI-TOF MS/MS analysis showed that *At*GUX3 specifically installs one GlcA substituent at one of the two middle xylose residues of the dodecasaccharide, i.e., the 6^{th} or 7^{th} xylosyl residue from either end.

The structural information obtained from the MS/MS analysis was extracted from the relative intensity of the B- and Y fragments formed during the analysis (Figure 2.1).[42] These two types of fragments are the result of a molecule getting split at a glycosidic bond. The newly formed fragment towards the non-reducing end is the B-fragment and the remaining fragment towards the reducing end is the Y-fragment; the subindex assigned to the letters B or Y refers to the number of sugar residues in the given fragment. Since in this case the Y-fragments include the anomeric STol motif at the reducing end, these fragments are

readily identified and cannot be mistaken by internal fragments formed by the cleavage of multiple glycosidic bonds in one molecule. Given that the GlcA residue can be cleaved off during fragmentation, as observed from the unsubstituted Y_{12}-fragment (the most intense peak in the spectrum), the GlcA-containing Y-fragments (blue in Figure 2.1) represent the most reliable source of information. Of these, the smallest is the GlcA-substituted Y_6-fragment, suggesting that no GlcA-substitutions happened at the last five xylosyl residues towards the reducing end. In agreement with this, Y-fragments with seven (Y_7) or more xylose units are predominantly substituted (blue) and little is found of the unsubstituted kind (red); unsubstituted Y-fragments (red) peaked in intensity at Y_5 and Y_6 and then they decreased drastically. These observations are supported by the B-fragments found. B-fragments with seven (B_7) or more xylosyl residues are predominantly substituted (blue). Unsubstituted B-fragments (yellow) peak at B_5 and B_6 and then decrease rapidly in intensity. The relatively intensities of the B_6-fragments in the substituted (blue) and unsubstituted (yellow) forms suggest that B-fragments with six sugar residues might or might not contain a GlcA, whereas B_5-fragments are more often than not unsubstituted, and B_7-fragments are more often than not substituted. The presence of small amounts of small, substituted B-fragments (blue) may be explained by the formation of internal fragments. However, we cannot exclude the presence of small amounts of oligosaccharides with GlcA-substitution in the xylosyl residues close to the non-reducing end.

These results suggest that *At*GUX3 requires five GlcA-free xylosyl residues at both sides of a given xylosyl residue in order to install a GlcA substituent. This *in vitro* substitution pattern of *At*GUX3 is consistent with the one proposed in the previously reported *in vivo* studies.[41]

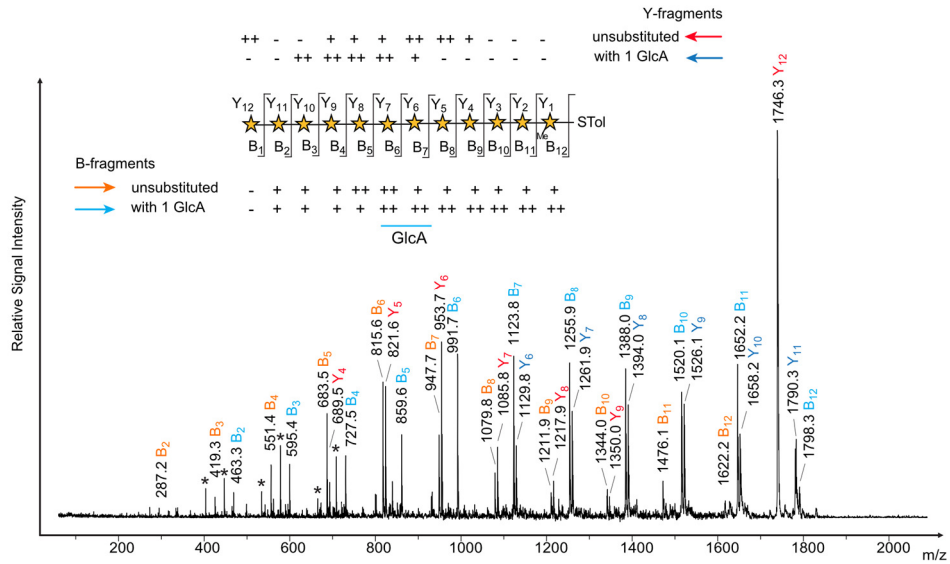


Figure 2.1.[1] MALDI-TOF MS/MS analysis of the products of the reaction between substrate **36** and glycosyltransferase *At*GUX3.

2.4. Conclusions

In this chapter, the syntheses of two 1,4-β-linked xylan dodecasaccharides (**35** and **36**) are reported. The divergent-convergent iterative design of the syntheses was found to be very convenient for the preparation of glycans having a repeating structure since it can provide an exponential growth of the xylan molecules with each iteration/cycle. The convergent steps of the syntheses consisted of enzyme-mediated glycosylation reactions using the glycosynthase XynAE265G, and unprotected α-xylosyl fluoride donors and xylosyl acceptors in aqueous buffer, which proceeded with excellent regio- and stereoselectivity. For the synthesis of **36**, xylosyl acceptors containing a 3-*O*-methyl substituent on the reducing-end

[1] This figure was prepared by Dr Colin Ruprecht.

xylosyl residue were used, which provided these structures with a higher solubility in comparison with their unsubstituted counterparts.

During the building block synthesis, a Nap protecting group was found to be orthogonal to the other protecting groups and stable under glycosylation conditions. Due to the partial cleavage of this group during the fluorination reactions to obtain the α-xylosyl donors, the fluorination step was performed with a free 4-OH group on the substrate, which did not cause any side-reaction, and, in the next step, a THP group was installed which was found to be compatible with the xylan glycosynthase. While the conditions to obtain α-xylosyl fluorides in benzoylated substrates in excellent yields were found, the fluorination of acetylated substrates providing the same yields was possible at lower temperatures.

Subsequent enzymatic studies performed in our laboratory with xylan-modifying enzymes (*At*XOAt1, *Os*XAT3, *Pv*XAT3 and *At*GUX3) using the xylan dodecasaccharide **36** showed that the enzymes were active on this xylan oligosaccharide. During this work, *At*GUX3 was heterologously expressed for the first time, and the MS analysis showed the installation of a single GlcA substituent on the xylan backbone of the xylan dodecasaccharide **36**. The data obtained by MALDI-TOF MS/MS analysis was found to be compatible with the installation of this substituent at the 6^{th} or 7^{th} xylosyl residue of the dodecasaccharide substrate, which is consistent with previously reported results from *in vivo* studies.

2.5. Experimental Section

General procedure for enzyme-mediated glycosylation reaction

α-Xylosyl fluoride donor (1.2 equiv) and xylosyl acceptor (1 equiv) were dissolved in phosphate buffer (100 mM; pH 7.4). XynAE265G in phosphate buffer (100 mM; pH 7.4) was added in order to achieve a reaction concentration of 1 mg of enzyme/mL of solution (unless otherwise stated). The mixture was allowed to react in a rotating shaker at 30°C for 16h. Often, the reaction got cloudy due to the partial precipitation of the desired product. For that reason, once the reaction was complete, the reaction mixture was centrifuged, the

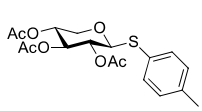
liquid phase was decanted, and the insoluble material was twice dispersed in H_2O and centrifuged. The solid usually did not require further purification; the aqueous phases, however, were combined and lyophilised. The soluble solid remaining after lyophilization was dissolved in a minimum amount of H_2O and purified by size-exclusion chromatography (SEC) using Bio-Gel® P-2 polyacrylamide gel with pure H_2O as an eluent. The product obtained by centrifugation and the one obtained by SEC were combined. The overall procedure yielded the desired product as a white fluffy solid.

General procedure for THP removal

Xylosyl oligosaccharide was dissolved in water; if necessary, prior lyophilization from DMSO or the addition of a small amount of DMSO before the addition of water helped making a homogeneous solution. AcOH was added and the reaction mixture was stirred for 16h or until the reaction was complete; monitorization was done by TLC (EtOAc/AcOH/H_2O; 7:2:1−2:2:1) or LCMS (C-4 column; H_2O/CH_3CN, 95:5 to 80:20). The reaction mixture was then lyophilised, and the residue was purified, if necessary, by size-exclusion chromatography (SEC) using Bio-Gel® P-2 polyacrylamide gel with H_2O as an eluent. The procedure yielded the desired product as a white fluffy solid.

Phosphate buffer (100 mM; 1 L; pH 7.4): Na_2HPO_4 (10.704 g) + $NaH_2PO_4.H_2O$ (3.395 g)

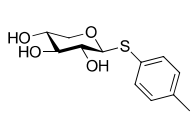
4-Methylphenyl 2,3,4-tri-*O*-acetyl-1-thio-β-D-xylopyranoside (2). D-Xylose (15.0 g; 100 mmol) was suspended in Et_2O (100 mL) and acetic anhydride (57.0 mL; 600 mmol; 6 equiv), pyridine (48.0 mL; 600 mmol; 6 equiv) and DMAP (100 mg; 800 µmol; 0.8 equiv) were added. The mixture was stirred at RT for 20 h. The homogeneous reaction mixture was washed with 2M HCl(aq) (3 × 150 mL) and this aqueous phase was then extracted with Et_2O (3 × 80 mL). The organic layers were combined and dried over $Na_2SO_4.H_2O$. After filtration, evaporation of the solvent and co-evaporation with toluene, the crude acetyl 2,3,4-tri-*O*-acetyl-1-D-xylopyranoside (**1**) (31.4 g) was obtained as a colourless oil.



The crude acetyl 2,3,4-tri-*O*-acetyl-1-D-xylopyranoside (**1**) (31.4 g) was dissolved in DCM (200 mL). *p*-Thiocresol (13.8 g; 109 mmol; 1.1 equiv) was added in one portion and the

solution was cooled down in an ice-water bath. $BF_3.OEt_2$ (18.6 mL; 148 mmol; 1.5 equiv) was added dropwise over 60-90 seconds. The reaction mixture was allowed to rise to RT and stirred for 20 h. The vivid-red reaction mixture was quenched with $NaHCO_3$(aq) (200 mL), what made the reaction mixture loose its colour to a yellowish colour. After separation, the organic layer was concentrated over reduced pressure. For purification, a recrystallization was done as follows. The crude was co-evaporated with ethyl acetate and then heated up to 40-50°C till the crude with the remaining EtOAc become a liquid. At this point, 1-2 volumes of warm hexane were added slowly. After allowing to cool down to r.t. the first crystals were form and were allowed to grow overnight at 4°C. Ethanol could also be used for recrystallization, but neither DCM/MeOH nor EtOH/water seemed to work. Pale brown crystals were collected by filtration. Desired β-thioxyloside **2** was obtained as brown crystals (17.9 g; 46.8 mmol; 47% over two steps). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.36 (d, *J* = 8.1 Hz, 2H; *H*C_{Ar}), 7.12 (d, *J* = 8.1 Hz, 2H; *H*C_{Ar}), 5.16 (dd, *J* = 8.5, 8.2 Hz, 1H; H-3), 4.90 (ddd, *J* = 9.0, 8.5, 5.0 Hz, 1H; H-4), 4.90 (ddd, *J* = 8.5, 8.2 Hz, 1H; H-2), 4.70 (d, *J* = 8.5 Hz, 1H; H-1), 4.24 (dd, *J* = 11.7, 5.0 Hz, 1H; H-5a), 3.38 (dd, *J* = 11.7, 9.0 Hz, 1H; H-5b), 2.33 (s, 3H; C*H*$_3$$C_{Ar}$), 2.09 (s, 3H; C*H*$_3$CO), 2.03 (s, 3H; C*H*$_3$CO), 2.03 (s, 3H; C*H*$_3$CO). ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.1 (CH_3*C*O), 169.9 (CH_3*C*O), 169.4 (CH_3*C*O), 138.7 ($C_{Ar,q}$), 133.6 (2×C_{Ar}H), 129.9 (2×C_{Ar}H), 128.1 ($C_{Ar,q}$), 86.4 (*C*-1), 72.3 (*C*-3), 69.9, 68.5 [2 peaks: C-2, C-4], 65.5 (*C*-5), 21.3 (*C*$H_3$$C_{Ar}$), 20.9 (*C*$H_3$CO), 20.8 (*C*$H_3$CO), 20.8 (*C*$H_3$CO).

4-Methyphenyl 1-thio-β-D-xylopyranoside (3). Thioglycoside **2** (17.9 g; 46.8 mmol) was dissolved in DCM/MeOH (1:3; 100 mL). Sodium hydroxide (560 mg; 14 mmol; 0.3 equiv) was added in one portion and the solution was stirred for 2 h at room temperature. The reaction was quenched with an acidic ion-exchange resin in H form (Amberlite® IRC-120(H)). The resin was filter off and the solution was concentrated under reduced pressure. For purification, a reprecipitation was done. The crude that formed while concentrating under reduced pressure was gently heated up so that a solution with the remaining solvent is formed again. To it, 2 or 3 volumes of a warm mixture hexane/Et_2O (1:1) was added, what results in the precipitation of the desired product. The flask was placed at 4°C for 30 min approx. and then the solid was



collected by filtration. The desired product was obtained as white crystals (11.3 g; 43.9 mmol; 94%). ^{1}H NMR (400 MHz, CD_3OD_SPE) δ 7.35 (d, *J* = 8.1 Hz, 2H; *H*C_{Ar}), 7.07 (bd, *J* = 8.1 Hz, 2H; *H*C_{Ar}), 4.59 (s, 2H; *H*O) [3 OH; proton exchange may reduce the integral], 4.40 (d, *J* = 9.3 Hz, 1H; H-1), 3.86 (dd, *J* = 11.3, 5.2 Hz, 1H; H-5a), 3.39 (ddd, *J* = 10.1, 8.7, 5.2 Hz, 1H; H-4), 3.30 – 3.23 (m, 1H; H-3), 3.17 – 3.08 (m, 2H; H-5b, H-2), 2.26 (s, 3H; C*H*$_3$$C_{Ar}$). ^{13}C NMR (101 MHz, CD_3OD_SPE) δ 139.1 (Me-*C*$_{Ar,q}$), 133.9 (2×*C*$_{Ar}$H), 130.7 (*C*$_{Ar,q}$), 130.6 (2×*C*$_{Ar}$H), 90.3 (C-1), 79.2 (C-3), 73.6 (C-2), 70.8 (C-4), 70.5 (C-5), 21.1 (*C*H$_3$$C_{Ar}$).

4-Methyphenyl 2,3,4-tri-*O*-benzoyl-1-thio-D-xylopyranoside (4). 4-Methyphenyl 1-thio-D-xylopyranoside (**3**) (4.49 g; 17.5 mmol) was dissolved in Et_3N (73 mL) and cooled down to 0°C in an ice-water bath. Under vigorous stirring, benzoyl chloride (11.1 g; 78.8 mmol; 9.2 mL; 4.5 equiv) was added. After 120 min, the TLC shows that the reaction was complete. Then $NaHCO_3$(aq) (200 mL) was added, and the mixture was extracted with DCM (3 x 300 mL). The organic layers were combined, dried over $MgSO_4$, concentrated under reduced pressure, and purified by flash column chromatography (toluene). The procedure yielded the desired product **4** (9.39 g; 16.5 mmol; 94%; α/β, 15:85). R_f(Hex/EtOAc, 7:1) = 0.43. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.19 – 7.93 (m, 8H; *H*C_{Ar}), 7.65 – 7.31 (m, 15H; *H*C_{Ar}), 7.21 – 7.07 (m, 3H; *H*C_{Ar}), 6.04 (dd, *J* = 9.3, 8.8 Hz, 0.2H; H-3-α), 5.93 (d, *J* = 5.2 Hz, 0.2H; H-1-α), 5.77 (dd, *J* = 6.9, 6.7 Hz, 1H; H-3-β), 5.51 (dd, *J* = 9.3, 5.6 Hz, 0.2H; H-2-α), 5.44 (dd, *J* = 6.7, 6.4 Hz, 1H; H-2-β), 5.41 (ddd, *J* = 9.6, 8.8, 5.4 Hz, 1H; H-4-α), 5.28 (ddd, *J* = 6.9, 6.8, 4.1 Hz, 1H; H-4-β), 5.19 (d, *J* = 6.4 Hz, 1H; H-1-β), 4.69 (dd, *J* = 12.2, 4.1 Hz, 1H; H-5a-β), 4.41 (dd, *J* = 11.5, 9.6 Hz, 0.2H; H-5a-α), 4.15 (dd, *J* = 11.5, 5.4 Hz, 0.2H; H-5b-α), 3.79 (dd, *J* = 12.2, 6.8 Hz, 1H; H-5b-β), 2.35 (s, 3H; C*H*$_3$-β), 2.31 (s, 0.5H; C*H*$_3$-α). ^{13}C NMR (101 MHz, $CDCl_3$) δ 171.0, 165.6, 165.6, 165.5, 165.5, 165.3, 165.1 [carbonyl signals], 138.5, 138.1, 133.8, 133.5, 133.4, 133.4, 133.4, 133.3, 132.6, 130.2, 130.1, 130.0, 130.0, 129.9, 129.9, 129.9, 129.8, 129.2, 129.2, 129.1, 129.0, 128.9, 128.5, 128.5, 128.5, 128.5, 128.4, 128.4, 128.3 [aromatic signals], 86.6 (C-1-β), 86.4 (C-1-α), 71.5 (C-2-α), 70.7 (C-3-β), 70.0 (C-2-β), 69.9, 69.7, 68.8 (C-4-β), 63.8 (C-5-β), 21.2 (*C*H$_3$-β), 21.2 (*C*H$_3$-α). HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{33}H_{28}O_7S$: 591.1453; found: 591.1460.

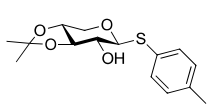
2,3,4-Tri-*O*-benzoyl-D-xylopyranose (5). 4-Methyphenyl 2,3,4-tri-*O*-benzoyl-1-thio-D-xylopyranoside (**4**) (9.95 g; 17.5 mmol) was dissolved in DCM/H_2O (10:1; 250 mL) and was cooled down to 0°C in an ice-water bath. Under vigorous stirring, NIS (3.94 g; 17.5 mmol) and TFA (2.00 g; 17.5 mmol) were added. The reaction was monitor by TLC (toluene/EtOAc, 19:1), and, after 120 min, quenched with a few drops of Et_3N. The reaction mixture was diluted in EtOAc (650 mL) and washed with $NaHCO_3$(aq) (sat.; 300 mL), $Na_2S_2O_3$(aq) (10%; 300 mL) and brine (150 mL). The organic phase was dried with $MgSO_4$, concentrated under reduced pressure and purified by flash column chromatography (toluene/EtOAc, 19:1). The procedure yielded the desired product **5** (6.69 g; 14.5 mmol; 83%; α/β, 68:32). R_f(Tol/EtOAc, 9:1) = 0.33. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.03 – 7.91 (m, 9H; *H*C_{Ar}), 7.57 – 7.44 (m, 5H; *H*C_{Ar}), 7.43 – 7.30 (m, 9H; *H*C_{Ar}), 6.20 (t, *J* = 9.5 Hz, 1H; H-3-α), 5.92 (t, *J* = 9.0 Hz, 0.5H; H-3-β), 5.68 (d, *J* = 3.4 Hz, 1H; H-1-α), 5.44 – 5.36 (m, 1.5H; H-4-α, H-4-β), 5.32 – 5.26 (m, 1.5H; H-2-α, H-2-β), 4.99 (bs, 0.5H; H-1-β), 4.46 (dd, *J* = 11.8, 5.3 Hz, 0.5H; H-5a-β), 4.16 – 4.10 (m, 1.5H; H-5a-α, H-5b-α), 3.79 (d, *J* = 5.8 Hz, 0.5H; OH-β), 3.65 (dd, *J* = 11.8, 9.4 Hz, 0.5H; H-5b-β), 3.04 (bs, 1H; OH-α). ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.0, 165.9, 165.8, 165.8, 165.7 [carbonyl signals], 133.8, 133.7, 133.6, 133.6, 133.4, 130.3, 130.1, 130.1, 130.0, 130.0, 129.9, 129.9, 129.4, 129.2, 129.1, 129.0, 129.0, 128.6, 128.6, 128.6, 128.5 [aromatic signals], 96.3 (C-1-β), 90.8 (C-1-α), 73.7 (C-2-β), 72.1 (C-2-α), 71.4 (C-3-β), 70.1 (C-4-α), 69.8 (C-4-β), 69.6 (C-3-α), 62.9 (C-5-β), 59.2 (C-5-α).

2,2,2-Trifluoro-*N*-phenylacetimidoyl 2,3,4-tri-*O*-benzoyl-D-xylopyranoside (**6**) (anomer′/anomer″, 4/1). 2,3,4-Tri-*O*-benzoyl-D-xylopyranose (**5**) (3.46 g; 7.47 mmol) was dissolved in acetone (75 mL) and was cooled down to 0°C in an ice-water bath. Under vigorous stirring, 2,2,2-trifluoro-*N*-phenylacetimidoyl chloride (4.65 g; 22.4 mmol; 3 equiv) and caesium carbonate (4.87 g; 14.9 mmol; 2 equiv) were added. The reaction mixture was allowed to stir for 2h while allowing to warm up to room temperature. Once complete, the reaction mixture was filtered through a pad of celite and concentrated under reduced pressure. The crude was then purified by flash column chromatography (hexane/EtOAc, 9:1). The procedure yielded the desired product **6** (4.11 g; 6.48 mmol; 87%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.14 – 8.08 (m, 0.6H; *H*C_{Ar}), 8.07 –

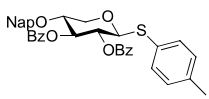
7.97 (m, 5H; *H*C$_{Ar}$), 7.97 – 7.91 (m, 2H; *H*C$_{Ar}$), 7.61 – 7.52 (m, 4H; *H*C$_{Ar}$), 7.51 – 7.22 (m, 12H; *H*C$_{Ar}$), 7.19 – 7.08 (m, 2H; *H*C$_{Ar}$), 7.06 – 7.00 (m, 1H; *H*C$_{Ar}$), 6.88 – 6.82 (m, 0.7H; *H*C$_{Ar}$), 6.82 – 6.70 (m, 0.6H; H-1), 6.57 – 6.35 (m, 2H; *H*C$_{Ar}$), 6.22 (dd, *J* = 9.8, 10.0 Hz, 1H; H-3'), 5.73 (dd, *J* = 3.7, 4.3 Hz, 1H; H-3''), 5.62 – 5.46 (m, 2.25H; H-4'; H-2'; H-2''), 5.29 (dd, *J* = 3.2, 6.6 Hz, 1H; H-4''), 4.57 (dd, *J* = 2.2, 12.9 Hz, 1H; H-5a''), 4.31 (dd, *J* = 5.7, 10.9 Hz, 1H; H-5'a), 4.04 (dd, *J* = 9.9, 10.9 Hz, 1.25H; H-5'b; H-5''b). ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.8 (C=O'), 165.7 (C=O'), 165.6 (C=O''), 165.5 (C=O'), 165.1 (C=O''), 164.9 (C=O''), 143.3, 143.0, 135.2, 133.9, 133.7, 133.6, 133.5, 130.2, 130.1, 130.1, 130.0, 129.9, 129.5, 129.3, 129.1, 128.9, 128.9, 128.8, 128.7, 128.6, 126.5, 124.5, 120.6, 119.4, 119.2 [arom. signals], 92.5 (br; C-1), 70.5 (C-2'), 69.7 (C-3'), 69.4 (C-4'), 67.4 (C-2''), 67.1 (C-4''), 67.0 (C-3''), 61.4 (C-5'), 60.8 (C-5'').

4-Methyphenyl 2,3-*O*-isopropylidene-1-thio-β-D-xylopyranoside (7). 4-Methyphenyl 1-thio-β-D-xylopyranoside (**3**) (117 mg; 456 µmol) and CSA (15.9 mg; 68 µmol; 0.15 equiv) were dissolved in DMF (9 mL). The solution was heated up to 60°C and 2-methoxypropene (131 µL; 1.368 mmol; 3 equiv) was added dropwise over 30 s. The solution was stirred for 35 min, allowed to cooled down to r.t. for 15 min and quenched with Et_3N. The solution was concentrated under reduced pressure, co-evaporated with toluene and purified by flash column chromatography (toluene/acetone, 9:1). The product (**7**) was obtained as a colourless solid (79.4 mg; 270 µmol; 59%). 3,4-*O*-Isopropylidene side-product (**7a**) was also isolated (23.9 mg; 81 µmol; 18%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.45 (d, *J* = 8.1 Hz, 2H; *H*C$_{Ar}$), 7.12 (bd, *J* = 8.1 Hz, 2H; *H*C$_{Ar}$), 4.71 (d, *J* = 9.5 Hz, 1H; H-1), 4.09 (dd, *J* = 11.7, 5.2 Hz, 1H; H-5a), 3.94 (dddd, *J* = 9.4, 9.1, 5.2, 4.0 Hz, 1H; H-4), 3.50 (dd, *J* = 9.1, 9.1 Hz, 1H; H-3), 3.23 – 3.16 (m, 2H; H-5, H-2), 2.88 (d, *J* = 4.0 Hz, 1H; *H*O), 2.34 (s, 3H; C*H*$_3$C$_{Ar}$), 1.48 (s, 3H; C*H*$_3$C), 1.43 (s, 3H; C*H*$_3$C). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.6 (*C*$_{Ar,q}$), 133.7 (2×*C*$_{Ar}$H), 129.7 (2×*C*$_{Ar}$H), 127.7 (*C*$_{Ar,q}$), 111.3 (*C*(CH$_3$)$_2$), 85.6 (C-1), 82.9, 75.1, 70.0, 68.9, 26.7 (*C*H$_3$C), 26.6 (*C*H$_3$C), 21.3 (*C*H$_3$C$_{Ar}$).

4-Methyphenyl 3,4-*O*-isopropylidene-1-thio-β-D-xylopyranoside (7a). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.46 – 7.41 (m, 2H; *H*C_{Ar}), 7.16 – 7.11 (m, 2H; *H*C_{Ar}), 4.36 (d, *J* = 8.8 Hz, 1H; H-1), 4.25 (dd, *J* = 9.2, 3.4 Hz, 1H; H-5a), 3.63 (ddd, *J* = 9.1, 8.8, 2.0 Hz, 1H), 3.59 – 3.42 (m, 3H; H-5b, H-3, H-4), 2.68 (d, *J* = 2.1 Hz, 1H; O*H*), 2.35 (s, 3H; C*H*$_3$C$_{Ar}$), 1.45 (s, 3H; C*H*$_3$C), 1.44 (s, 3H; C*H*$_3$C). ^{13}C NMR (101 MHz, $CDCl_3$) δ 138.9 (*C*$_{Ar,q}$), 133.5 (2×*C*$_{Ar}$H), 130.0 (2×*C*$_{Ar}$H), 127.9 (*C*$_{Ar,q}$), 112.0 (*C*(CH$_3$)$_2$), 90.1 (*C*-1), 82.4 (*C*-3), 73.4 (*C*-2), 71.5 (*C*-4), 68.7 (*C*-5), 26.8 (*C*H$_3$C), 26.7 (*C*H$_3$C), 21.3 (*C*H$_3$C$_{Ar}$).



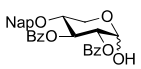
4-Methyphenyl 2,3-di-*O*-benzoyl-4-*O*-(naphthalen-2-ylmethyl)-1-thio-β-D-xylopyranoside (8). 4-Methyphenyl 2,3-*O*-isopropylidene-1-thio-β-D-xylopyranoside (**7**) (3.56 g; 12.0 mmol) and 2-(bromomethyl)naphthalene (3.50 g; 15.6 mmol; 1.25 equiv) were dissolved in DMF (100 mL). The solution was cooled down to 0°C in an ice-water bath. Under vigorous stirring, NaH 60% in mineral oil (1.2 g; 29 mmol; 2.4 equiv) was added in small portions over 10 min. After approx. 120 min, the reaction shows complete consumption of SM on TLC (Tol/acetone, 9:1). Then, a few drops of 1M HCl(aq) were added until pH of the solution became acidic. After 16h, the reaction showed complete consumption of the 4-ONap synthetic intermediate on TLC. To quench, $NaHCO_3$(aq) was added. The mixture was then extracted with 3 x 200 mL of EtOAc. The organic layers were then combined and washed with 100 mL of brine. The organic solution was then concentrated under reduced pressure and co-evaporated with toluene.



The crude mixture was dissolved in Et_3N (100 mL) and cooled down to 0°C in an ice-water bath. Under vigorous stirring, benzoyl chloride (6.7 g; 48 mmol; 4 equiv) was added by syringe. After 16h, the TLC shows that the reaction was complete. Then $NaHCO_3$(aq) was added and the mixture was extracted with DCM (3 x 200 mL). The organic layers were combined, concentrated under reduced pressure, co-evaporated with toluene and purified by flash column chromatography (Toluene/EtOAc, 49:1). The procedure yielded the desired product **8** (4.64 g; 7.68 mmol; 64%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.00 – 7.94 (m, 4H), 7.79 – 7.06 (m, 20H) [21 *H*C_{Ar} expected], 5.66 (dd, *J* = 8.1, 8.0 Hz, 1H; H-3), 5.32 (dd, *J* = 8.1, 8.1

Hz, 1H; H-2), 4.98 (d, *J* = 8.1 Hz, 1H; H-1), 4.79 (d, *J* = 12.2 Hz, 1H; C*H*H, Nap), 4.74 (d, *J* = 12.2 Hz, 1H; CH*H*, Nap), 4.34 (dd, *J* = 11.9, 4.6 Hz, 1H; H-5a), 3.84 (ddd, *J* = 8.6, 8.0, 4.6 Hz, 1H; H-4), 3.61 (dd, *J* = 11.9, 8.6 Hz, 1H; H-5b), 2.33 (s, 3H; CH_3). ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.7 (*C*=O), 165.4 (*C*=O), 138.4 ($C_{Ar,q}$Me), 135.0 ($C_{Ar,q}CH_2$), 133.4 (C_{Ar}H), 133.3 (C_{Ar}H), 133.2, 133.2, 133.1, 130.0, 129.9, 129.5, 129.4, 129.2, 129.1, 128.5, 128.5, 128.4, 128.4, 128.0, 127.8, 126.9, 126.3, 126.1, 125.8 [22 C_{Ar} signals expected], 87.2 (C-1), 74.3 (C-4), 73.7 (C-3), 72.9 (CH_2, Nap), 70.6 (C-2), 66.4 (C-5), 21.3 (CH_3). HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{37}H_{32}O_6S$: 627.1817; found: 627.1818.

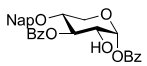
2,3-Di-*O*-benzoyl-4-*O*-(naphthalen-2-ylmethyl)-α-D-xylopyranose (9) (). 4-Methyphenyl



2,3-di-*O*-benzoyl-4-*O*-(naphthalen-2-ylmethyl)-1-thio-β-D-xylopyranoside (**8**) (3.28 g; 5.42 mmol) was dissolved in DCM/H_2O (10:1; 110 mL) and cooled down to 0°C in an ice-water bath. Under vigorous stirring, NIS (1.22 g; 5.42 mmol) and TFA (620 mg; 5.42 mmol) were added. The reaction was monitor by TLC (toluene/acetone, 9:1), and, after 120 min, quenched with a few drops of Et_3N. The reaction mixture was diluted in EtOAc (400 mL) and washed with $NaHCO_3$(aq) (sat.; 200 mL), $Na_2S_2O_3$(aq) (10%; 200 mL) and brine (100 mL). The organic phase was concentrated under reduced pressure and purified by flash column chromatography (Toluene/acetone, 9:1). The procedure yielded the desired product **9** (2.13 g; 4.28 mmol; 79%; α/β, 65:35) as well as the benzoyl-migration side-product **9a** (110.2 mg; 220 µmol; 4%). ^{1}H NMR (400 MHz, CDCl3) δ 8.00 – 7.90 (m, 6H; *H*C_{Ar}), 7.78 – 7.72 (m, 2H; *H*C_{Ar}), 7.68 – 7.59 (m, 5H; *H*C_{Ar}), 7.55 – 7.40 (m, 7H; *H*C_{Ar}), 7.39 – 7.24 (m, 9H; *H*C_{Ar}), 7.21 – 7.14 (m, 1H; *H*C_{Ar}), 6.01 (dd, *J* = 9.9, 8.6 Hz, 1H; H-3-α), 5.72 (dd, *J* = 9.5, 9.0 Hz, 0.5H; H-3-β), 5.56 (d, *J* = 3.4 Hz, 1H; H-1-α), 5.13 – 5.07 (m, 1.5H; H-2-α, H-2-β), 4.86 (d, *J* = 7.6 Hz, 0.5H; H-1-β), 4.80 (d, *J* = 12.3 Hz, 1H; C*H*HC_{Ar}-α), 4.79 (d, *J* = 12.3 Hz, 0.5H; C*H*HC_{Ar}-β), 4.71 (d, *J* = 12.3 Hz, 1H; CH*H*C_{Ar}-α), 4.70 (d, *J* = 12.3 Hz, 0.5H; CH*H*C_{Ar}-β), 4.19 (dd, *J* = 11.9, 5.2 Hz, 0.5H; H-5a-β), 4.10 – 4.02 (m, 1H; H-5a-α), 3.94 – 3.83 (m, 3H; H-4-β; H-4-α; H-5b-α), 3.54 (dd, *J* = 11.9, 10.1 Hz, 0.5H; H-5b-β), 3.05 (s, 0.8H; O*H*). ^{13}C NMR (101 MHz, $CDCl_3$) δ 167.2 (*C*=O), 166.1 (*C*=O), 165.9 (*C*=O), 165.8 (*C*=O), 135.1, 134.9, 133.7, 133.5, 133.4, 133.2, 133.2, 133.1, 133.1, 130.1, 130.0, 129.9, 129.9, 129.8, 129.4, 129.2, 129.1, 128.9, 128.6, 128.5, 128.5, 128.5, 128.4, 128.4, 128.0, 127.8, 126.9, 126.8, 126.3,

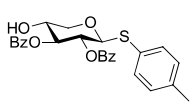
126.2, 126.2, 126.1, 125.8, 125.8, 125.4 [aromatic signals], 96.5 (C-1-β), 90.7 (C-1-α), 75.1 (C-4-α), 75.0 (C-4-β), 74.3 (C-2-β), 73.4 (C-3-β), 73.1 (CH_2C_{Ar}-β), 73.0 (CH_2C_{Ar}-α), 72.4 (C-2-α), 71.4 (C-3-α), 64.2 (C-5-β), 60.1 (C-5-α).

Benzoyl 3-*O*-benzoyl-4-*O*-(naphthalen-2-ylmethyl)-α-D-xylopyranoside (9a). ^{1}H NMR (400



MHz, $CDCl_3$) δ 8.18 – 8.09 (m, 2H), 8.08 – 8.00 (m, 2H), 7.81 – 7.40 (m, 13H), 7.37 – 7.31 (m, 1H) [17 HC_{Ar} expected], 6.41 (d, *J* = 3.5 Hz, 1H; H-1), 5.70 – 5.64 (m, 1H; H-3), 4.84 (d, *J* = 12.1 Hz, 1H; C*H*H, Nap), 4.78 (d, *J* = 12.1 Hz, 1H; CH*H*, Nap), 4.04 – 3.81 (m, 4H; H-2, H-4, H-5a, H-5b), 2.60 (s, 1H; O*H*). ^{13}C NMR (101 MHz, $CDCl_3$) δ 167.4 (*C*=O), 165.2 (*C*=O), 134.8, 134.0, 133.7, 133.2, 130.1, 130.0, 129.5, 129.2, 128.8, 128.6, 128.6, 128.0, 127.8, 127.0, 126.4, 126.3, 125.8 [aromatic signals], 92.6 (C-1), 75.4 (C-3), 74.1 (C-4), 73.4 (CH_2C_{Ar}), 70.8 (C-2), 62.5 (C-5).

4-Methyphenyl 2,3-di-*O*-benzoyl-1-thio-β-D-xylopyranoside (12). 4-Methyphenyl 2,3-di-



O-benzoyl-4-*O*-(naphthalen-2-ylmethyl)-1-thio-β-D-xylopyranoside (**8**) (6.5 g; 10.7 mmol) was dissolved in DCM/H_2O (10:1; 110 mL) and cooled down to 0°C in an ice-water bath. Under vigorous stirring, DDQ (4.86 g; 21.4 mmol; 2 equiv) was added. The reaction was monitored by TLC (toluene/acetone, 20:1). After 120 min, the reaction mixture was diluted with DCM and washed with $NaHCO_3$(aq) (2 × 200 mL) until the washing layers are colourless. The organic layers were then combined, concentrated under reduced pressure and purified by flash column chromatography (Hexane/EtOAc, 4:1 to 3:1). The procedure yielded the desired product **12** (4.07 g; 8.77 mmol; 82%). R_f(Tol/EtOAc, 9:1) = 0.29. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.06 – 7.98 (m, 4H; Ar), 7.58 – 7.51 (m, 2H; Ar), 7.44 – 7.37 (m, 6H; Ar), 7.15 – 7.10 (m, 2H; Ar), 5.41 (dd, *J* = 7.5, 7.5 Hz, 1H; H-2), 5.33 (dd, *J* = 7.5, 7.5 Hz, 1H; H-3), 5.01 (d, *J* = 7.5 Hz, 1H; H-1), 4.42 (dd, *J* = 11.9, 4.5 Hz, 1H; H-5a), 3.98 (ddd, *J* = 8.1, 7.5, 4.5 Hz, 1H; H-4), 3.58 (dd, *J* = 11.9, 8.1 Hz, 1H; H-5b), 3.11 (s, 1H; O*H*), 2.34 (s, 3H; CH_3). ^{13}C NMR (101 MHz, $CDCl_3$) δ 167.1 (*C*=O), 165.2 (*C*=O), 138.6 ($C_{Ar,q}$Me), 133.8 (C_{Ar}H), 133.6 (C_{Ar}H), 133.5 (2×C_{Ar}H), 130.2 (2×C_{Ar}H), 129.9 (2×C_{Ar}H), 129.9 (2×C_{Ar}H), 129.3 ($C_{Ar,q}$), 128.9 ($C_{Ar,q}$), 128.9 ($C_{Ar,q}$), 128.6 (2×C_{Ar}H), 128.6 (2×C_{Ar}H), 87.0 (C-1), 76.2 (C-3), 70.2 (C-2), 68.5 (C-4), 67.7 (C-5), 21.3 (CH_3). HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{26}H_{24}O_6S$: 487.1191; found: 487.1192.

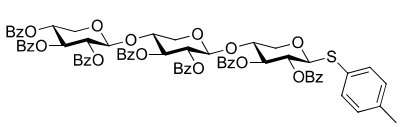
4-Methyphenyl 2-*O*-benzoyl-4-*O*-(9H-fluoren-9-yl)methoxycarbonyl-3-*O*-methyl-1-thio-β-D-xylopyranoside. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.21 – 8.13 (m, 2H; *H*C$_{Ar}$), 7.79 – 7.74 (m, 2H; *H*C$_{Ar}$), 7.63 – 7.59 (m, 1H; *H*C$_{Ar}$), 7.57 – 7.50 (m, 2H; *H*C$_{Ar}$), 7.44 – 7.37 (m, 7H; *H*C$_{Ar}$), 7.32 – 7.21 (m, 2H; *H*C$_{Ar}$), 7.13 – 7.09 (m, 2H; *H*C$_{Ar}$), 5.35 (dd, *J* = 5.2, 4.6 Hz, 1H; H-2), 5.20 (d, *J* = 4.6 Hz, 1H; H-1), 4.79 (ddd, *J* = 5.3, 4.9, 3.2 Hz, 1H; H-4), 4.56 (dd, *J* = 3.2, 12.6 Hz, 1H; H-5a), 4.41 – 4.37 (m, 2H; Fmoc), 4.20 (t, *J* = 7.4 Hz, 1H; Fmoc), 3.73 (dd, *J* = 4.9, 12.6 Hz, 1H; H-5b), 3.73 – 3.70 (m, 1H; H-3), 3.56 (s, 3H; C*H*$_3$O), 2.32 (s, 3H; C*H*$_3$C$_{Ar}$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.4 (*C*=O), 154.6 (*C*=O), 143.4, 143.2, 141.4, 141.4, 138.0, 133.6, 132.37, 130.8, 130.1, 129.9, 129.7, 128.6, 128.1, 127.4, 127.3, 125.3, 125.2, 120.2 [aromatic signals; 24 C atoms], 86.5 (C-1), 77.4, 72.2, 70.3, 69.8 [C-2, C-3, C-4, Fmoc], 61.7 (C-5), 59.4 (*C*H$_3$O), 46.8 (Fmoc), 21.3 (*C*H$_3$C$_{Ar}$).

4-Methylphenyl 2-*O*-benzyl-3-*O*-methyl-1-thio-β-D-xylopyranoside (13). ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.08 – 8.05 (m, 2H; *H*C$_{Ar}$), 7.62 – 7.58 (m, 1H; *H*C$_{Ar}$), 7.49 – 7.45 (m, 2H; *H*C$_{Ar}$), 7.39 – 7.35 (m, 2H; *H*C$_{Ar}$), 7.11 – 7.08 (m, 2H; *H*C$_{Ar}$), 5.26 (dd, *J* = 6.4, 6.2 Hz, 1H; H-2), 5.01 (d, *J* = 6.2 Hz, 1H; H-1), 4.39 (dd, *J* = 11.8, 3.8 Hz, 1H; H-5a), 3.78 (ddd, *J* = 7.0, 6.3, 3.8 Hz, 1H; H-4), 3.53 (s, 3H; C*H*$_3$O), 3.50 (dd, *J* = 11.8, 7.0 Hz, 1H; H-5b), 3.50 (dd, *J* = 6.4, 6.3 Hz, 1H; H-3), 2.32 (s, 3H; C*H*$_3$C$_{Ar}$). ^{13}C NMR (151 MHz, $CDCl_3$) δ 138.1 (*C*=O), 133.6, 132.8, 130.5, 129.9, 129.9, 129.7, 128.7 [aromatic signals], 87.4 (C-1), 82.0 (C-3), 71.4 (C-2), 68.3 (C-4), 66.1 (C-5), 59.7 (*C*H$_3$O), 21.2 (*C*H$_3$C$_{Ar}$). HRMS (ESI/Q-TOF) m/z: [M+HCOO]$^-$ Calcd for $C_{20}H_{22}O_5S$: 419.1169; found: 419.1173.

4-Methyphenyl 2',3',4'-tri-*O*-benzoyl-β-D-xylopyranosyl-(1'→4)-2,3-di-*O*-benzoyl-1-thio-β-D-xylopyranoside (14). 2,2,2-Trifluoro-*N*-phenylacetimidoyl 2,3,4-tri-*O*-benzoyl-D-xylopyranoside (**6**) (3.46 g; 5.46 mmol; 1.25 equiv) and 4-methyphenyl 2,3-di-*O*-benzoyl-1-thio-β-D-xylopyranoside (**12**) (2.03 g; 4.37 mmol) were dissolved in DCM (105 mL), to which freshly activated 4 Å molecular sieves were added. The mixture was cooled down to −78°C in an acetone–dry-ice bath. TMSOTf (97.1 mg; 440 µmol; 79 µL) was added by

syringe. The reaction mixture was monitored by TLC and allowed to warm up slowly over 120 min to −20°C. Once the reaction is complete, it was quenched with a few drops of Et_3N. The mixture was concentrated under reduced pressure and purified by flash column chromatography (hexane/EtOAc, 5:1 to 2:1). The procedure yielded the desired product (**14**) (1.96 g; 2.16 mmol; 50%), as well as the monosaccharide, aglycon-transfer product (**4**) (831 mg; 1.46 mmol; 33%) and the trisaccharide, aglycon-transfer product (**14a**) (423.5 mg; 340 µmol; 8%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.06 – 7.85 (m, 10H; *H*C$_{Ar}$), 7.59 – 7.47 (m, 5H; *H*C$_{Ar}$), 7.46 – 7.29 (m, 12H; *H*C$_{Ar}$), 7.12 – 7.06 (m, 2H; *H*C$_{Ar}$), 5.66 (dd, *J* = 8.2, 8.0 Hz, 1H; H-3), 5.63 (dd, *J* = 6.7, 6.6 Hz, 1H; H-3'), 5.31 (dd, *J* = 8.2, 8.2 Hz, 1H; H-2), 5.24 (dd, *J* = 6.7, 4.8 Hz, 1H; H-2'), 5.03 (ddd, *J* = 6.6, 6.3, 3.9 Hz, 1H; H-4'), 4.94 (d, *J* = 4.8 Hz, 1H; H-1'), 4.92 (d, *J* = 8.2 Hz, 1H; H-1), 4.24 (dd, *J* = 12.1, 4.8 Hz, 1H; H-5a), 4.08 (ddd, *J* = 8.8, 8.0, 4.8 Hz, 1H; H-4), 4.01 (dd, *J* = 12.4, 3.9 Hz, 1H; H-5a'), 3.52 (dd, *J* = 12.1, 8.8 Hz, 1H; H-5b), 3.41 (dd, *J* = 12.4, 6.3 Hz, 1H; H-5b'), 2.32 (s, 3H; C*H*$_3$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.5 (*C*=O), 165.4 (*C*=O), 165.3 (*C*=O), 165.3 (*C*=O), 165.1 (*C*=O), 138.5 (*C*$_{Ar,q}$), 133.5, 133.4, 133.3, 133.3, 130.0, 129.9, 129.9, 129.8, 129.8, 129.8, 129.4 (*C*$_{Ar}$H), 129.3 (*C*$_{Ar}$H), 129.2 (*C*$_{Ar}$H), 129.0 (*C*$_{Ar}$H), 128.9 (*C*$_{Ar}$H), 128.6 (*C*$_{Ar}$H), 128.5, 128.5, 128.5, 128.4, 128.4, 99.7 (C-1'), 87.0 (C-1), 75.2 (C-4), 73.2 (C-3), 70.5 (C-2), 70.1 (C-2'), 69.5 (C-3'), 68.4 (C-4'), 65.7 (C-5), 60.8 (C-5'), 21.2 (C*H*$_3$).

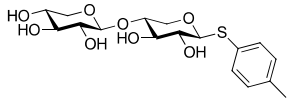
4-Methyphenyl 2'',3'',4''-tri-*O*-benzoyl-β-D-xylopyranosyl-(1''→4')-2',3'-di-*O*-benzoyl-(1'→4)-2,3-di-*O*-benzoyl-1-thio-β-D-xylopyranoside (**14a**). ^{1}H NMR (400 MHz, $CDCl_3$) δ



8.02 – 7.89 (m, 19H; *H*C$_{Ar}$), 7.59 – 7.48 (m, 10H; *H*C$_{Ar}$), 7.44 – 7.28 (m, 23H; *H*C$_{Ar}$), 5.60 (dd, *J* = 8.3, 8.0 Hz, 1H; H-3), 5.60 (dd, *J* = 6.8, 6.7 Hz, 1H; H-3''), 5.54 (dd, *J* = 8.4, 8.0 Hz, 1H; H-3'), 5.28 (dd, *J* = 8.3, 8.3 Hz, 1H; H-2), 5.18 (dd, *J* = 8.4, 6.4 Hz, 1H; H-2'), 5.16 (dd, *J* = 6.8, 4.9 Hz, 1H; H-2''), 5.02 (ddd, *J* = 6.7, 6.3, 4.0 Hz, 1H; H-4''), 4.88 (d, *J* = 8.3 Hz, 1H; H-1), 4.73 (d, *J* = 6.4 Hz, 1H; H-1'), 4.72 (d, *J* = 4.9 Hz, 1H; H-1''), 4.12 (dd, *J* = 12.0, 4.9 Hz, 1H; H-5a), 3.98 (ddd, *J* = 8.9, 8.0, 4.9 Hz, 1H; H-4), 3.97 (dd, *J* = 12.5, 4.0 Hz, 1H; H-5a''), 3.84 (ddd, *J* = 8.6, 8.0, 4.8 Hz, 1H; H-4'), 3.54 (dd, *J* = 12.3, 4.8 Hz, 1H; H-5a'), 3.44 (dd, *J* = 12.0, 8.9 Hz, 1H; H-5b), 3.36 (dd, *J* = 12.5, 6.3 Hz, 1H; H-5b''), 3.15 (dd,

J = 12.3, 8.6 Hz, 1H; H-5b'), 2.31 (s, 3H; *C*H_3). ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.5 (*C*=O), 165.4 (*C*=O), 165.4 (*C*=O), 165.3 (*C*=O), 165.2 (*C*=O), 165.1 (*C*=O), 164.9 (*C*=O), 138.43, 133.5, 133.4, 133.4, 133.3, 133.1, 130.0, 129.9, 129.9, 129.8, 129.8, 129.7, 129.6, 129.4, 129.2, 129.1, 128.9, 128.6, 128.5, 128.5, 128.4, 128.4, 128.3 [aromatic signals], 101.0 (C-1'), 99.4 (C-1''), 86.9 (C-1), 75.8 (C-4), 74.8 (C-4'), 73.3 (C-3), 72.0 (C-3'), 71.4 (C-2'), 70.4 (C-2), 70.0 (C-2''), 69.6 (C-3''), 68.5 (C-4''), 66.0 (C-5), 62.12 (C-5'), 60.8 (C-5''), 21.2 (*C*H_3C_{Ar}). HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{71}H_{60}O_{19}S$: 1271.3347; found: 1271.3335.

4-Methyphenyl β-D-xylopyranosyl-(1'→4)-1-thio-β-D-xylopyranoside (15). 4-



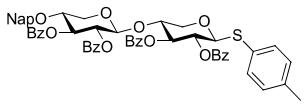
Methyphenyl 2',3',4'-tri-*O*-benzoyl-β-D-xylopyranosyl-(1'→4)-2,3-di-*O*-benzoyl-1-thio-β-D-xylopyranoside (**14**) (1027 mg; 1.13 mmol) was dissolved in DCM (6 mL). To that, MeOH (18 mL) was added. The reaction mixture was cooled down to 0°C in an ice-water bath and 500 mM sodium methoxide in MeOH (2.26 mL; 1.13 mmol; 1.0 equiv) were added. The reaction was allowed to stir for 16h while allowing to warm up to room temperature. The reaction was monitored by LCMS and when intermediates no longer present, it was quenched with an ion-exchange resin in H form (Amberlite® IRC-120(H)). The mixture was then concentrated under reduce pressure. The crude was taken up in Et_2O/H_2O (1:1; 45 mL) and the aqueous layer was then washed with Et_2O (45 mL). The organic layers were combined and extracted with H_2O (2 × 12 mL). Aqueous layers were then combined and lyophilised. Without further purification, the procedure yielded the desired product **15** (428 mg; 1.10 mmol; 97%). ^{1}H NMR (400 MHz, D_2O) δ 7.41 – 7.37 (m, 2H; Ar), 7.23 – 7.14 (m, 2H; Ar), 4.61 (d, *J* = 9.7 Hz, 1H; H-1), 4.36 (d, *J* = 7.8 Hz, 1H; H-1'), 4.03 (dd, *J* = 11.6, 5.2 Hz, 1H; H-5a), 3.88 (dd, *J* = 11.6, 5.5 Hz, 1H; H-5a'), 3.67 (ddd, *J* = 10.4, 9.1, 5.2 Hz, 1H; H-4), 3.54 (ddd, *J* = 10.5, 9.2, 5.5 Hz, 1H; H-4'), 3.50 (dd, *J* = 9.1, 9.0 Hz, 1H; H-3), 3.34 (dd, *J* = 9.4, 9.2 Hz, 1H; H-3'), 3.17 (dd, *J* = 9.4, 7.8 Hz, 1H; H-2'), 3.33 – 3.18 (m, 3H; H-5b, H-2, H-5b'), 2.27 (s, 3H; C*H*$_3$). ^{13}C NMR (101 MHz, D_2O) δ 139.2 ($C_{Ar,q}$), 132.7 (2×C_{Ar}H), 129.9 (2×C_{Ar}H), 127.3 ($C_{Ar,q}$), 101.6 (C-1'), 88.1 (C-1), 75.9 (C-4), 75.4 (C-3'), 75.1 (C-3), 72.6 (C-2'), 71.5 (C-2), 69.0 (C-4'), 66.4 (C-5), 65.1 (C-5'), 20.1 (*C*H_3). HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{17}H_{24}O_8S$: 411.1090; found: 411.1088.

4-Methyphenyl 2',3',4'-tri-*O*-benzoyl-β-D-xylopyranosyl-(1'→4)-2-*O*-benzoyl-3-*O*-methyl-1-thio-β-D-xylopyranoside (16). 2,2,2-Trifluoro-*N*-phenylacetimidoyl 2,3,4-tri-*O*-benzoyl-D-xylopyranoside (**6**) (441 mg; 696 µmol; 1.2 equiv.) and xylosyl acceptor **13** (217 mg; 580 µmol) were dissolved in DCM (15 mL), and freshly activated 4 Å molecular sieves were added. The mixture was cooled down to −65°C in an acetone–dry-ice bath. TMSOTf (12.9 mg; 58 µmol; 11 µL) was added by syringe. The reaction mixture was allowed to warm up slowly to −20°C over 120 min. Once the reaction was complete, it was quenched with a few drops of Et_3N. The mixture was concentrated under reduced pressure and purified by flash column chromatography (toluene/EtOAc, 49:1 to 9:1). The procedure yielded the desired product **16** (257 mg; 314 µmol; 54%) as a colourless oil. Additionally, acceptor (**13**) was recovered (42 mg; 111 µmol; 19%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.20 – 7.80 (m, 9H; *H*C$_{Ar}$), 7.64 – 7.45 (m, 7H; *H*C$_{Ar}$), 7.41 – 7.32 (m, 7H; *H*C$_{Ar}$), 7.32 – 7.28 (m, 2H; *H*C$_{Ar}$), 7.08 – 7.03 (m, 2H; *H*C$_{Ar}$), 5.72 (dd, *J* = 7.5, 7.5 Hz, 1H; H-3'), 5.34 (dd, *J* = 7.5, 5.6 Hz, 1H; H-2'), 5.25 (ddd, *J* = 7.5, 7.3, 4.4 Hz, 1H; H-4'), 5.17 (dd, *J* = 8.0, 7.6 Hz, 1H; H-2), 4.93 (d, *J* = 5.6 Hz, 1H; H-1'), 4.79 (d, *J* = 8.0 Hz, 1H; H-1), 4.44 (dd, *J* = 12.2, 4.4 Hz, 1H; H-5a'), 4.17 (dd, *J* = 12.0, 4.6 Hz, 1H; H-5a), 3.90 (ddd, *J* = 8.5, 7.6, 4.7 Hz, 1H; H-4), 3.67 (dd, *J* = 12.2, 7.3 Hz, 1H; H-5b'), 3.57 (dd, *J* = 7.6, 7.6 Hz, 1H; H-3), 3.52 (s, 3H; C*H*$_3$O), 3.33 (dd, *J* = 12.0, 8.5 Hz, 1H; H-5b), 2.30 (s, 3H; C*H*$_3$C$_{Ar}$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.6 (*C*=O), 165.5 (*C*=O), 165.3 (*C*=O), 165.2 (*C*=O), 138.2, 133.6, 133.6, 133.5, 133.4, 133.0, 130.0, 129.9, 129.9, 129.8, 129.4, 129.2, 129.2, 129.2, 129.1, 128.6, 128.6, 128.6, 128.5, 128.4 [C$_{Ar}$], 99.6 (C-1'), 87.0 (C-1), 82.0 (C-3), 75.3 (C-4), 71.3 (C-2), 70.7, 70.7 [C-2'; C-3'], 69.2 (C-4'), 65.3 (C-5), 61.7 (C-5'), 60.3 (C*H*$_3$O), 21.3 (*C*H$_3$C$_{Ar}$). HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{46}H_{42}O_{12}S$: 841.2295; found: 841.2290.

4-Methylphenyl β-D-xylopyranosyl-(1'→4)-3-*O*-methyl-1-thio-β-D-xylopyranoside (17). Disaccharide **16** (249 mg; 304 µmol) was dissolved in DCM (2.5 mL) and MeOH (7.5 mL) was added. The reaction mixture was cooled down to 0°C in an ice-water bath and 500 mM sodium methoxide in MeOH (970 µL; 486 µmol; 1.6 equiv.) was

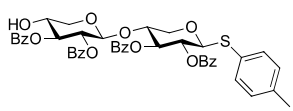
added. After 30 min, the ice-water bath was removed, and the reaction mixture was allowed to stir for additional 6.5 h until intermediates were no longer detected in LC-MS. The reaction was then quenched with an ion-exchange resin in H form (Amberlite® IRC-120(H)). The mixture was concentrated under reduced pressure. The crude was taken up in Et_2O/H_2O (1:1; 15 mL) and the aqueous layer was washed with Et_2O (15 mL). The organic layers were combined and extracted twice with H_2O (7 and 3 mL). The aqueous layers were then combined and lyophilised. Without further purification, the procedure yielded the desired product **17** (108 mg; 885 µmol; 88%) as a white fluffy solid. ^{1}H NMR (400 MHz, D_2O) δ 7.39 (d, *J* = 8.1 Hz, 2H; *H*C$_{Ar}$), 7.19 (d, *J* = 8.1 Hz, 2H; *H*C$_{Ar}$), 4.62 (d, *J* = 9.0 Hz, 1H; H-1), 4.35 (d, *J* = 7.8 Hz, 1H; H-1'), 4.04 (dd, *J* = 11.6, 5.2 Hz, 1H; H-5a), 3.88 (dd, *J* = 11.6, 5.5 Hz, 1H; H-5a'), 3.78 – 3.70 (m, 1H; H-4), 3.54 (ddd, *J* = 10.5, 9.1, 5.5 Hz, 1H; H-4'), 3.51 (s, 3H; C*H*$_3$O), 3.38 – 3.26 (m, 4H; H-2, H-3', H-3, H-5b), 3.21 (dd, *J* = 11.6, 10.5 Hz, 1H; H-5b'), 3.17 (dd, *J* = 9.3, 7.8 Hz, 1H; H-2'), 2.27 (s, 3H; C*H*$_3$C$_{Ar}$). ^{13}C NMR (101 MHz, D_2O) δ 139.2 (*C*$_{Ar,q}$), 132.9 (2x*C*$_{Ar}$H), 129.9 (2x*C*$_{Ar}$H), 127.1 (*C*$_{Ar,q}$), 101.7 (C-1'), 88.0 (C-1), 84.2 (C-3), 75.5 [C-3' or C-2], 74.8 (C-4), 72.7 (C-2'), 70.6 [C-2 or C-3'], 69.1 (C-4'), 66.1 (C-5), 65.1 (C-5'), 59.5 (*C*H$_3$O), 20.1 (*C*H$_3$C$_{Ar}$). HRMS (ESI/Q-TOF) m/z: [M+Na]$^+$ Calcd for $C_{18}H_{26}O_8S$: 425.1246; found: 425.1246.

4-Methyphenyl 2',3'-di-*O*-benzoyl-4'-*O*-(naphthalen-2-ylmethyl)-β-D-xylopyranosyl-(1'→4)-2,3-di-*O*-benzoyl-1-thio-β-D-xylopyranoside (18). 2,2,2-Trifluoro-*N*-phenylacetimidoyl 2,3-di-*O*-benzoyl-4'-*O*-(naphthalen-2-ylmethyl)-D-xylopyranoside (**10**) (5.2 g; 7.7 mmol) and 4-methyphenyl 2,3-di-*O*-benzoyl-1-thio-β-D-xylopyranoside (**12**) (3.9 g; 8.4 mmol) were dissolved in DCM (150 mL), to which freshly activated 4 Å molecular sieves were added. The mixture was cooled down to −78°C in an acetone–dry-ice bath. TMSOTf (170 mg; 780 µmol) was added by syringe. The reaction mixture was monitored by TLC (toluene/acetone, 9:1) and allowed to warm up slowly over 120 min to −20°C. Once the reaction is complete, it was quenched with a few drops of Et_3N. The mixture was concentrated under reduced pressure and purified by flash column chromatography (toluene/acetone, 49:1). The procedure yielded the desired product **18** (6.04 g; 6.39 mmol; 83%). ^{1}H



NMR (400 MHz, $CDCl_3$) δ 8.02 – 7.85 (m, 8H), 7.78 – 7.71 (m, 1H), 7.68 – 7.02 (m, 32H) [31 *H*C_{Ar} expected], 5.60 (dd, *J* = 8.1, 8.0 Hz, 1H; H-3), 5.51 (dd, *J* = 8.3, 7.5 Hz, 1H; H-3'), 5.27 (dd, *J* = 8.3, 8.1 Hz, 1H; H-2), 5.14 (dd, *J* = 8.3, 6.3 Hz, 1H; H-2'), 4.88 (d, *J* = 8.3 Hz, 1H; H-1), 4.73 (d, *J* = 6.3 Hz, 1H; H-1'), 4.61 (d, *J* = 12.2 Hz, 1H; C*H*H, Nap), 4.56 (d, *J* = 12.2 Hz, 1H; CH*H*, Nap), 4.12 (dd, *J* = 12.0, 4.9 Hz, 1H; H-5a), 3.99 (ddd, *J* = 8.8, 8.0, 4.9 Hz, 1H; H-4), 3.64 – 3.50 (m, 2H; H-5a'; H-4'), 3.44 (dd, *J* = 12.0, 8.8 Hz, 1H; H-5b), 3.20 (dd, *J* = 11.3, 7.7 Hz, 1H; H-5b'), 2.31 (s, 3H; C*H*$_3$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.7 (*C*=O), 165.5 (*C*=O), 165.3 (*C*=O), 165.2 (*C*=O), 138.5, 134.9, 133.4, 133.4, 133.3, 133.2, 133.2, 133.1, 130.1, 129.9, 129.9, 129.8, 129.7, 129.5, 129.5, 129.4, 129.3, 129.2, 128.8, 128.6, 128.5, 128.5, 128.4, 128.4, 128.0, 127.8, 126.8, 126.5, 126.3, 126.1, 125.7, 120.6 [30 C_{Ar} signals expected], 101.3 (C-1'), 87.0 (C-1), 76.1 (C-4), 74.2 (C-4)', 73.4 (C-3), 72.8 (*C*H_2, Nap), 72.4 (C-3'), 71.3 (C-2'), 70.5 (C-2), 66.1 (C-5), 62.8 (C-5'), 21.3 (*C*H_3).

4-Methyphenyl 2',3'-di-*O*-benzoyl-β-D-xylopyranosyl-(1'→4)-2,3-di-*O*-benzoyl-1-thio-β-D-xylopyranoside (19). 4-Methyphenyl 2',3'-di-*O*-benzoyl-4'-*O*-(naphthalen-2-ylmethyl)-β-



D-xylopyranosyl-(1'→4)-2,3-di-*O*-benzoyl-1-thio-β-D-xylopyranoside (**18**) (6.04 g; 6.39 mmol) was dissolved in DCM/H_2O (10:1; 250 mL) and cooled down to 0°C in an ice-water bath. Under vigorous stirring, DDQ (2.90 g; 12.8 mmol; 2 equiv) was added. The reaction was monitored by TLC (hexane/EtOAc, 2:1). After 180 min, the reaction mixture was diluted with DCM and washed with $NaHCO_3$(aq) (2 × 300 mL) until the washing layers are colourless. The organic layers were then combined, concentrated under reduced pressure, and purified by flash column chromatography (Hexane/EtOAc, 3:1 to 2:1). The procedure yielded the desired product **19** (4.34 g; 5.39 mmol; 84%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.10 – 7.85 (m, 8H), 7.59 – 7.47 (m, 4H), 7.45 – 7.34 (m, 8H), 7.36 – 7.28 (m, 2H), 7.11 – 7.04 (m, 2H), 5.63 (dd, *J* = 8.3, 8.0 Hz, 1H; H-3), 5.30 (dd, *J* = 8.4, 8.3 Hz, 1H; H-2), 5.25 (dd, *J* = 8.2, 6.0 Hz, 1H; H-2'), 5.17 (dd, *J* = 8.2, 6.9 Hz, 1H; H-3'), 4.89 (d, *J* = 8.4 Hz, 1H; H-1), 4.78 (d, *J* = 6.0 Hz, 1H; H-1'), 4.14 (dd, *J* = 12.0, 4.9 Hz, 1H; H-5a), 4.03 (ddd, *J* = 8.9, 8.0, 4.9 Hz, 1H; H-4), 3.78 – 3.61 (m, 2H; H-4'; H-5a'), 3.46 (dd, *J* = 12.0, 8.9 Hz, 1H; H-5b), 3.17 (dd, *J* = 11.7, 7.4 Hz, 1H; H-5b'), 2.98 (s, 1H; *H*O), 2.32 (s, 3H; C*H*$_3$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 167.2

(*C*=O), 165.6 (*C*=O), 165.3 (*C*=O), 165.1 (*C*=O), 138.5 ($C_{Ar,q}CH_3$), 133.8 ($C_{Ar}H$), 133.6 ($C_{Ar}H$), 133.4, 133.3 ($C_{Ar}H$), 130.1 (2×$C_{Ar}H$), 130.1 (2×$C_{Ar}H$), 129.9, 129.8 (2×$C_{Ar}H$), 129.7 ($C_{Ar}H$), 129.4 ($C_{Ar}H$), 129.1 ($C_{Ar}H$), 128.8 ($C_{Ar}H$), 128.7 (2×$C_{Ar}H$), 128.6 (2×$C_{Ar}H$), 128.5, 101.1 (C-1'), 87.0 (C-1), 76.1 (C-4), 75.5 (C-3'), 73.4 (C-3), 70.7 (C-2), 70.5 (C-2'), 68.4 (C-4'), 66.2 (C-5'), 64.5 (C-5), 21.3 (CH_3).

2',3'-Di-*O*-benzoyl-β-D-xylopyranosyl-(1'→4)-2,3-di-*O*-benzoyl-α-D-xylopyranosyl fluoride (20). To a PP or PFA conical tube, disaccharide **19** (500 mg; 621 µmol) and DCM (12 mL) were added. Once the sugar was dissolved, NIS (158 mg; 681 µmol) was added and the mixture was immediately cooled down to −50°C in an acetone–dry-ice bath. Under vigorous stirring, HF/py (7:3; 3.0 mL; 100 equiv.) was added by syringe (with a Luer-Lock fitting). The reaction mixture was allowed to warm up to −5°C over 90 min. Then, it was kept stirring for 5.5 h at a temperature between −5°C and −3°C, as temperatures above 0°C can lead to glycosidic bond cleavage. Monitoring of the reaction was done by ^{19}F NMR, by quenching an aliquot in NH_3(aq) (25%) and extracting it with EtOAc. Once the reaction was complete, the reaction mixture was cooled down to −60°C. Under vigorous stirring, the reaction mixture was very slowly added to a PP conical tube containing NH_3 (aq) (25%, 20 mL) at −60°C using a LDPE Pasteur pipette. The mixture was then transferred to a separatory funnel and washed with NH_3 (aq) (25%; 2 × 50 mL), $Na_2S_2O_3$ (aq) (10%; 50 mL) and brine (20 mL). The organic layer was dried over $MgSO_4$ and then concentrated under reduced pressure. The crude was purified by flash column chromatography (hexane/EtOAc, 2:1). The procedure yielded the desired product **20** (385 mg; 549 µmol; 88%; α/β, 93:7) as a colourless solid. ^1H NMR (400 MHz, $CDCl_3$) of α-product: δ 8.09 – 7.85 (m, 8H), 7.62 – 7.48 (m, 4H), 7.47 – 7.34 (m, 9H), 5.94 (dd, *J* = 10.2, 9.4 Hz, 1H; H-3), 5.83 (dd, *J* = 53.1, 2.7 Hz, 1H; H-1), 5.25 (dd, *J* = 7.9, 5.7 Hz, 1H; H-2'), 5.20 – 5.16 (m, 1H; H-3'), 5.16 (ddd, *J* = 24.0, 10.2, 2.7 Hz, 1H; H-2), 4.80 (d, *J* = 5.7 Hz, 1H; H-1'), 4.16 (ddd, *J* = 9.6, 9.4, 7.2 Hz, 1H; H-4), 3.87 – 3.82 (m, 2H; H-5a; H-5b), 3.78 (ddd, *J* = 7.2, 6.7, 4.4 Hz, 1H; H-4'), 3.72 (dd, *J* = 12.0, 4.4 Hz, 1H; H-5a'), 3.17 (dd, *J* = 12.0, 7.2 Hz, 1H; H-5b'). ^{13}C NMR (101 MHz, $CDCl_3$) δ 167.2 (*C*=O), 166.0 (*C*=O), 165.5 (*C*=O), 165.1 (*C*=O), 133.9, 133.8, 133.8, 133.3, 130.1, 129.7, 129.2, 129.1, 128.8, 128.8, 128.7, 128.7, 128.6, 128.5 (16 C_{Ar}),

104.2 (d, $^{1}J_{C1\text{-}F}$ = 229.6 Hz; C-1), 101.0 (C-1'), 76.0 (C-4), 75.2 (C-3'), 71.7 (d, $^{2}J_{C2\text{-}F}$ = 25.1 Hz; C-2), 70.6 (C-2'), 70.3 (C-3), 68.3 (C-4'), 64.3 (C-5'), 61.5 (d, $^{3}J_{C5\text{-}F}$ = 4.8 Hz; C-5). HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{38}H_{33}FO_{12}$: 723.1853; found: 723.1855.

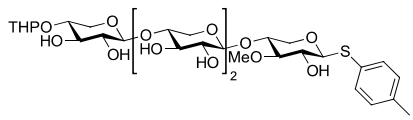
2',3'-Di-*O*-benzoyl-4'-*O*-tetrahydropyranyl-β-D-xylopyranosyl-(1'→4)-2,3-di-*O*-benzoyl-α-D-xylopyranosyl fluoride (21). α-Xylosyl fluoride **20** (946 mg; 1.35 mmol), DHP (585 mg; 6.75 mmol; 640 µL) and CSA (63 mg; 270 µmol) were dissolved in DCM (27 mL). The reaction was allowed to stir at room temperature for 2 h while being monitored by TLC (Hex/EtOAc, 3:1). Once complete, the reaction mixture was diluted in DCM (50 mL) and washed with $NaHCO_3$(aq) (2 × 50 mL) and brine (25 mL). The organic layer was dried over $MgSO_4$, concentrated under reduced pressure, and purified by flash column chromatography (hexane/EtOAc, 3:1). The procedure yielded the desired product **21** (1.01 g; 1.29 mmol; 96%; diastereomeric ratio at the THP group: 44/56) as a colourless oil. R_f (Hex/EtOAc, 3:1) = 0.26. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.04 – 7.90 (m, 8H; *H*C_{Ar}), 7.56 – 7.48 (m, 4H; *H*C_{Ar}), 7.45 – 7.34 (m, 8H; *H*C_{Ar}), 5.93 (m, 1H; H-3), 5.82 (dd, *J* = 53.2, 1.3 Hz, 0.6H; H-1), 5.82 (dd, *J* = 53.2, 1.3 Hz, 0.4H; H-1), 5.49 (dd, *J* = 9.4, 7.3 Hz, 0.4H; H-3'), 5.42 (dd, *J* = 8.7, 7.6 Hz, 0.5H; H-3'), 5.22 – 5.12 (m, 2H; H-2, H-2'), 4.78 (d, *J* = 6.0 Hz, 0.4H; H-1), 4.74 (d, *J* = 6.4 Hz, 0.6H; H-1), 4.58 – 4.53 (m, 1H, H-1_{THP}), 4.19 – 4.10 (m, 1H; H-4), 3.91 – 3.58 (m, 5H; C*H*, C*H*H), 3.57 – 3.48 (m, 0.5H, C*H*H), 3.38 – 3.30 (m, 0.5H, C*H*H), 3.27 – 3.14 (m, 1.4H, C*H*H), 1.74 – 1.30 (m, 6H; C*H*H_{THP}). ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.9, 165.7, 165.6, 165.5, 165.5, 165.2, 165.2 [*C*=O signals], 133.7, 133.5, 133.4, 133.3, 133.2, 133.1, 130.1, 129.9, 129.8, 129.5, 129.3, 129.3, 129.3, 128.7, 128.6, 128.6, 128.6, 128.5, 128.5, 128.4 [C_{Ar} signals], 104.1 (d, $J_{C1\text{-}F}$ = 229.7 Hz; C-1), 101.4 (C-1'), 101.2 (C-1'), 100.4 (C-1_{THP}), 95.9 (C-1_{THP}), 76.1, 75.9, 73.9, 72.8, 71.9, 71.8, 71.5, 71.5, 71.3, 71.1, 70.3, 70.3, 70.2 [*C*H signals], 64.1, 62.7, 62.0, 61.5 (d, $^{3}J_{C5\text{-}F}$ = 4.6 Hz; C-5), 61.5, 60.5 [*C*H_2 signals], 30.6, 30.3, 25.2, 25.1, 19.4, 18.4. [*C*$H_{2(THP)}$ signals]. HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{43}H_{41}FO_{13}$: 807.2429; found: 807.2441.

4'-*O*-Tetrahydropyranyl-β-D-xylopyranosyl-(1'→4)-α-D-xylopyranosyl fluoride (22). α-Xylosyl fluoride **21** (150 mg; 191 µmol) was dissolved in DCM (1.9 mL). Then, MeOH (5.7 mL) was added. The reaction mixture was cooled down to 0°C in an ice-water bath and 500 mM sodium methoxide in MeOH (570 µL; 287 µmol; 1.5 equiv.) was added. The reaction was allowed to stir for 9 h while allowing to warm up to room temperature. The reaction was monitored by LC-MS and when intermediates were no longer detected, the reaction mixture was quenched with an ion-exchange resin in H form (Amberlite® IRC-120(H)). The mixture was then concentrated under reduce pressure at 25°C. The crude was taken up in Et_2O/H_2O (1:1; 40 mL) and the aqueous layer was washed with Et_2O (20 mL). The organic layers were combined and extracted with H_2O (2 × 10 mL). The aqueous layers were combined and lyophilised. Without further purification, the procedure yielded α-xylosyl fluoride donor **22** (68 mg; 184 µmol; 96%) as a white fluffy solid. ^{1}H NMR (400 MHz, D_2O) δ 5.57 (dd, *J* = 53.3, 2.8 Hz, 1H; H-1), 4.43 – 4.34 (m, 1H; H-1'), 4.10 – 3.88 (m, 2H), 3.88 – 3.48 (m, 5H), 3.45 (ddd, *J* = 9.3, 9.2, 2.7 Hz, 1H), 3.33 – 3.15 (m, 2H), 1.92 – 1.60 (m, 2H; C*H*H_{THP}), 1.60 – 1.25 (m, 4H; C*H*H_{THP}). ^{13}C NMR (101 MHz, D_2O) δ 107.1 (d, $^1J_{C1\text{-}F}$ = 223.8 Hz; C-1), 101.9, 101.6, 101.5 [C-1'; C-1'; C-1$_{THP}$], 97.7 (*C*-1$_{THP}$), 77.1 (*C*H), 75.4 (*C*H), 75.4 (*C*H), 74.6 (*C*H), 74.0 (*C*H), 73.7 (*C*H), 72.7 (*C*H), 72.6 (*C*H), 71.0 (d, $^2J_{C2\text{-}F}$ = 25.0 Hz; C-2), 70.7 (*C*H), 64.4 (*C*H_2), 63.9 (*C*H_2), 63.3 (*C*H_2), 62.8 (*C*H_2), 60.9 (d, $^3J_{C5\text{-}F}$ = 4.1 Hz; C-5), 30.1 (*C*H_2), 30.0 (*C*H_2), 24.4 (*C*H_2), 24.4 (*C*H_2), 19.5 (*C*H_2), 18.8 (*C*H_2). HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{15}H_{25}FO_9$: 391.1380; found: 391.1381

4-Methylphenyl 4'''-*O*-tetrahydropyranyl-β-D-xylopyranosyl-(1'''→4'')-bis[β-D-xylopyranosyl-(1→4)]-1-thio-β-D-xylopyranoside (23). According to general procedure A, α-xylosyl fluoride donor **22** (100 mg; 271 µmol; 1.2 equiv.) and disaccharide acceptor **15** (87.9 mg; 226 µmol; 1.0 equiv.) were dissolved in phosphate buffer (7.8 mL), and XynAE265G (1.3 mL; from 1.80 mg/mL) was added [c_{enzyme} = 250 µg/mL]. The procedure yielded the desired product **23** (160 mg; 217 µmol; 96%) as a white fluffy solid. TLC eluent for reaction monitoring: EtOAc/AcOH/H_2O; 7:2:1. ^{1}H NMR (400 MHz, D_2O) δ 7.44 – 7.35 (m, 2H; Ar), 7.22 – 7.14 (m, 2H; Ar), 4.77 – 4.75 [under H_2O peak; H-1$_{THP}$; H-1$_{THP}$],

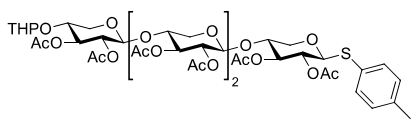
4.61 (d, *J* = 9.7 Hz, 1H; H-1), 4.42 – 4.35 (m, 3H; H-1'; H-1''; H-1'''), 4.09 – 3.99 (m, 5H; C*H*H), 3.98 – 3.91 (m, 0.2H; C*H*H), 3.86 – 3.79 (m, 1H; C*H*H), 3.76 – 3.62 (m, 4H; C*H*), 3.62 – 3.56 (m, 0.4H; C*H*), 3.56 – 3.40 (m, 4H), 3.35 – 3.16 (m, 9H), 2.27 (s, 3H; CH_3), 1.82 – 1.64 (m, 3H; C*H*H_{THP},H-$2_{eq,\ THP}$; H-$3_{eq,\ THP}$), 1.56 – 1.38 (m, 5H; C*H*H_{THP}; H-$2_{ax,\ THP}$; H-$3_{ax,\ THP}$; H-$4_{ax,\ THP}$; H-$4_{eq,\ THP}$). ^{13}C NMR (101 MHz, D_2O) δ 139.1 ($C_{Ar,q}$), 132.8 (2×C_{Ar}H), 129.9 (2×C_{Ar}H), 127.3 ($C_{Ar,q}$), 101.9 (C-1_{THP}), 101.6 (*C*H), 101.5 (*C*H), 101.5 (*C*H), 101.5 (*C*H), 97.7 (C-1_{THP}), 88.1 (C-1, S), 77.2 (*C*H), 76.2 (*C*H), 76.2 (*C*H), 75.9 (*C*H), 75.1 (*C*H), 74.6 (*C*H), 74.0 (*C*H), 73.7 (*C*H), 73.5 (*C*H), 72.7 (*C*H), 72.6 (*C*H), 72.6 (*C*H), 72.5 (*C*H), 71.5 (*C*H), 66.4 (*C*H_2), 64.4 (*C*H_2), 63.9 (*C*H_2), 63.3 (*C*H_2), 62.8 (*C*H_2), 30.1 (*C*H_2), 24.4 (*C*H_2), 20.1 (*C*H_3), 19.5 (*C*H_2), 18.9 (*C*H_2). HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{32}H_{48}O_{17}S$: 759.2510; found: 759.2501.

4-Methylphenyl 4'''-*O*-tetrahydropyranyl-β-D-xylopyranosyl-(1'''→4'')-[β-D-xylopyranosyl-(1→4)]-3-*O*-methyl-1-thio-β-D-xylopyranoside (24). According to general procedure



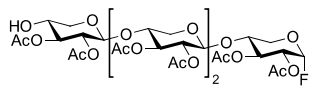
A, glycosyl fluoride donor **22** (57.0 mg; 155 µmol; 1.2 equiv.) and disaccharide acceptor **17** (51.9 mg; 129 µmol; 1 equiv.) were dissolved in phosphate buffer (3.74 mL), and XynAE265G (1.42 mL; from 1.80 mg/mL) was added [c_{enzyme} = 500 µg/mL]. The procedure yielded the desired product **7** (78.9 mg; 105 µmol; 81%) as a white fluffy solid. R_f (EtOAc/AcOH/H_2O; 7:2:1) = 0.50. ^{1}H NMR (400 MHz, D_2O) δ 7.40 (d, *J* = 8.1 Hz, 2H; *H*C_{Ar}), 7.20 (d, *J* = 8.1 Hz, 2H; *H*C_{Ar}), 4.79 – 4.76 [under H_2O peak; H-1_{THP}; H-1_{THP}], 4.64 (d, *J* = 9.0 Hz, 1H; H-1), 4.44 – 4.37 (m, 3H; H-1', H-1'', H-1'''), 4.10 – 3.80 (m, 5H; C*H*H), 3.80 – 3.58 (m, 4H; C*H*), 3.58 – 3.43 (m, 7H; CH_3O, C*H*H), 3.39 – 3.18 (m, 9H; C*H*H, C*H*), 2.28 (s, 3H; CH_3C_{Ar}), 1.83 – 1.66 (m, 2H; C*H*H_{THP}), 1.57 – 1.39 (m, 4H; C*H*H_{THP}). ^{13}C NMR (101 MHz, D_2O) δ 139.3 ($C_{Ar,q}$), 133.0 (2xC_{Ar}H), 129.9 (2xC_{Ar}H), 127.2 ($C_{Ar,q}$), 102.0 (C-1_{THP}), 101.6, 101.6, 101.5 [C-1, C-1'', C-1'''], 97.7 (C-1_{THP}), 88.0 (C-1), 84.2 (C-3), 77.2, 76.2, 76.2, 74.8, 74.6, 74.1, 73.8, 73.5, 72.7, 72.6, 72.6, 72.6, 72.6, 70.6 [*C*H signals], 66.1, 64.4, 63.9, 63.3, 62.8 [*C*H_2 signals], 59.5 (*C*H_3O), 30.1 (*C*$H_{2(THP)}$), 24.4 (*C*$H_{2(THP)}$), 20.1 (*C*H_3C_{Ar}), 19.5 (*C*$H_{2(THP)}$), 18.9 (*C*$H_{2(THP)}$)**.** HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{33}H_{50}O_{17}S$: 795.2745; found: 795.2762.

4-Methylphenyl 2‴,3‴-di-*O*-acetyl-4‴-*O*-tetrahydropyranyl-β-D-xylopyranosyl-(1‴→4″)-bis[2,3-di-*O*-acetyl-β-D-xylopyranosyl-(1→4)]-2,3-di-*O*-acetyl-1-thio-β-D-xylopyranose (25). Tetrasaccharide **23** (150 mg; 204 µmol) was dissolved in pyridine (1.7 mL).



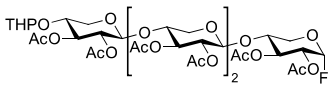
The reaction mixture was cooled down to 0°C in an ice-water bath and acetic anhydride (190 µL; 1.96 mmol; 9.6 equiv.) was added. The reaction mixture was stirred for 16 h while allowing to warm up to room temperature. The reaction was monitored by TLC (toluene/EtOAc, 1:1). When the reaction was complete, the reaction mixture was diluted in toluene, concentrated under reduced pressure, and purified by flash column chromatography (toluene/EtOAc, 2:1). The procedure yielded the desired product **23** (205 mg; 191 µmol; 94%) as a colourless oil. ^{1}H NMR (400 MHz, acetone-d6) δ 7.37 (d, *J* = 8.2 Hz, 2H; *H*C$_{Ar}$), 7.16 (d, *J* = 7.9 Hz, 2H; *H*C$_{Ar}$), 5.10 (dd, *J* = 8.8, 8.8 Hz, 1H; C*H*), 5.06 – 4.97 (m, 2H; C*H*), 4.87 (d, *J* = 9.5 Hz, 1H; H-1), 4.80 – 4.73 (m, 1.3H; H-2; H-1$_{THP}$), 4.72 – 4.61 (m, 6H; H-1’; H-1’’; H-1‴; 3×C*H*), 4.13 – 3.96 (m, 4H; 4×C*H*H), 3.90 – 3.67 (m, 4H), 3.52 – 3.30 (m, 5H; 5×C*H*H), 2.31 (s, 3H; *H*$_3$CC$_{Ar}$), 2.00 – 1.95 (m, 31H; *H*$_3$CCO) [expected 24H], 1.77 – 1.64 (m, 1H; C*H*H$_{THP}$), 1.65 – 1.53 (m, 1H; C*H*H$_{THP}$), 1.56 – 1.36 (m, 4H; 4×C*H*H$_{THP}$). ^{13}C NMR (101 MHz, acetone-d6) δ 170.2 (*C*=O), 170.1 (*C*=O), 170.0 (*C*=O), 170.0 (*C*=O), 169.8 (*C*=O), 169.7 (*C*=O), 169.7 (*C*=O), 138.9 (*C*$_{Ar,q}$), 133.7 (2×*C*$_{Ar}$H), 130.5 (2×*C*$_{Ar}$H), 129.6 (*C*$_{Ar,q}$), 101.8, 101.5, 101.4, 101.3, 101.3 [C-1 signals], 96.7 (*C*-1$_{THP}$), 86.6 (*C*-1, S), 76.4 (*C*H), 76.4 (*C*H), 76.3 (*C*H), 76.2 (*C*H), 75.6 (*C*H), 74.4 (*C*H), 74.3 (*C*H), 73.6 (*C*H), 73.1 (*C*H), 73.1 (*C*H), 72.3 (*C*H), 72.2 (*C*H), 72.0 (*C*H), 71.9 (*C*H), 71.0 (*C*H), 67.2 (*C*H$_2$), 65.5 (*C*H$_2$), 63.9 (*C*H$_2$), 63.7 (*C*H$_2$), 63.6 (*C*H$_2$), 63.1 (*C*H$_2$), 61.9 (*C*H$_2$), 31.2 (*C*H$_{2(THP)}$), 26.0 (*C*H$_{2(THP)}$), 26.0 (*C*H$_{2(THP)}$), 21.1, 20.9, 20.9, 20.8, 20.8, 20.7, 20.7, 20.7 [*C*H$_3$ signals], 20.3 (*C*H$_{2(THP)}$), 19.4 (*C*H$_{2,THP}$). HRMS (ESI/Q-TOF) m/z: [M+Na]$^+$ Calcd for $C_{48}H_{64}O_{25}S$: 1095.3355; found: 1095.3342.

2‴,3‴-Di-*O*-acetyl-β-D-xylopyranosyl-(1‴→4″)-bis[2,3-di-*O*-acetyl-β-D-xylopyranosyl-(1→4)]-2,3-di-*O*-acetyl-α-D-xylopyranosyl fluoride (26). To a PP or PFA conical tube, **25**



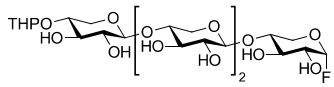
(85.8 mg; 80.0 µmol) and DCM (1.6 mL) were added. Once the compound was dissolved, NIS (19.8 mg; 88.0 µmol; 1.1 equiv.) was added and the mixture was immediately cooled down to −50°C in an acetone–dry-ice bath. Under vigorous stirring, HF/py (7:3; 200 µL; 50 equiv.) was added using a syringe (with a Luer-Lock fitting). The reaction mixture was allowed to warm up to –5°C over 90 min and stirred for another 4.5 h at a constant temperature of −5°C. The reaction mixture was cooled down to −60°C and, under vigorous stirring, very slowly added to a 100-mL round-bottomed flask containing NH_3(aq) (33%, 20 mL) at −60°C using a LDPE Pasteur pipette. The mixture was transferred to a separatory funnel and washed with NH_3(aq) (33%; 2 × 25 mL) and $Na_2S_2O_3$(aq) (10%; 25 mL). The organic layer was concentrated under reduced pressure. The crude product was then purified by flash column chromatography (hexane/EtOAc, 40:60 to 33:67) to yield the desired product **26** (52 mg; 58 µmol; 73%; α/β, 93:7) as a colourless oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.63 (dd, *J* = 53.1, 2.8 Hz, 1H; H-1), 5.39 (dd, *J* = 10.2, 8.6 Hz, 1H; C*H*), 5.12 – 4.98 (m, 2H), 4.89 – 4.68 (m, 5H; H-2; H-2′; H-2″; H-2‴; C*H*), 4.53 – 4.45 (m, 3H; H-1′; H-1″; H-1‴), 4.03 (dd, *J* = 11.9, 4.4 Hz, 1H; C*H*H), 3.98 – 3.89 (m, 2H; 2×C*H*H), 3.89 – 3.66 (m, 5H), 3.44 – 3.23 (m, 3H; 3×C*H*H), 2.10 – 2.01 (m, 27H; 8×CH_3) [expected 24H]. ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.4 (*C*=O), 170.1 (*C*=O), 170.0 (*C*=O), 169.7 (*C*=O), 169.5 (*C*=O), 169.5 (*C*=O), 169.2 (*C*=O), 104.1 (d, $^1J_{C1\text{-}F}$ = 229.0 Hz; C-1), 100.4, 99.7 [C-1 signals], 75.0, 74.6, 74.3, 74.2, 72.1, 71.7, 71.0, 70.8, 70.5, 70.1, 69.6, 67.9 [CH_2 signals], 64.3, 62.6, 62.5, 61.1 (d, $^3J_{C5\text{-}F}$ = 4.8 Hz; C-5) [*C*H signals], 21.0, 20.9, 20.9, 20.9, 20.8, 20.8 [CH_3 signals]. HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{36}H_{49}FO_{24}$: 907.2496; found: 907.2503.

2‴,3‴-Di-*O*-acetyl-4‴-*O*-tetrahydropyranyl-β-D-xylopyranosyl-(1‴→4″)-bis[2,3-di-*O*-acetyl-β-D-xylopyranosyl-(1→4)]-2,3-di-*O*-acetyl-α-D-xylopyranosyl fluoride (27). **26** (131



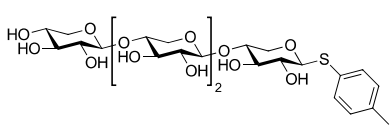
mg; 148 µmol), DHP (64.2 mg; 740 µmol; 70 µL; 5 equiv.) and CSA (6.9 mg; 30 µmol; 0.2 equiv.) were dissolved in DCM (3.0 mL). The reaction mixture was allowed to stir at room temperature for 3 h. Upon completion, the reaction mixture was diluted in EtOAc (20 mL) and washed with $NaHCO_3$(aq) (2 × 20 mL). The aqueous layers were combined and extracted with EtOAc (20 mL). The organic layer was washed with brine (10 mL), dried over $MgSO_4$, concentrated under reduced pressure and purified by flash column chromatography (toluene/EtOAc, 2:1). The procedure yielded the desired product **27** (135 mg; 139 µmol; 94%). R_f (Tol/EtOAc, 1:1) = 0.35. ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.63 (dd, *J* = 53.1, 2.8 Hz, 1H; H-1), 5.39 (dd, *J* = 10.2, 8.6 Hz, 1H; H-3), 5.12 – 4.94 (m, 3H), 4.86 – 4.81 (m, 0.7H), 4.80 – 4.67 (m, 4H), 4.64 – 4.58 (m, 0.7H; H-1_{THP}), 4.54 – 4.40 (m, 3H; H-1’; H-1’’; H-1’’’), 4.10 (dd, *J* = 11.7, 5.3 Hz, 1H; C*H*H), 4.00 (dd, *J* = 11.7, 4.9 Hz, 0.4H; C*H*H), 3.95 – 3.87 (m, 2H), 3.90 – 3.66 (m, 6H), 3.54 – 3.43 (m, 1H; C*H*H), 3.39 – 3.23 (m, 3H; 3×C*H*H), 2.12 – 1.94 (m, 29H; 8×CH_3), 1.83 – 1.70 (m, 1H; C*H*H_{THP}), 1.70 – 1.59 (m, 2H; C*H*H_{THP}), 1.58 – 1.42 (m, 5H; C*H*H_{THP}). ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.4, 170.3, 170.2, 170.1, 170.0, 169.7, 169.5, 169.5, 169.5, 104.0 (d, $^1J_{C1\text{-}F}$ = 229.2 Hz; C-1), 100.8, 100.7, 100.7, 100.4, 100.3, 100.2 [C-1 signals], 96.4 (C-1_{THP}), 75.0, 75.0, 74.8, 74.5, 74.5, 74.1, 73.2, 72.5, 72.0, 71.9, 71.7, 71.4, 71.3, 70.9, 70.8, 70.8, 70.7, 70.5, 69.6 (C-3) [*C*H signals], 64.7, 63.1, 63.0, 62.7, 62.6, 62.5, 61.8, 61.1 (d, $^3J_{C5\text{-}F}$ = 4.5 Hz; C-5) [*C*H_2 signals], 30.8 (C*H*H_{THP}), 30.5 (C*H*H_{THP}), 29.8 (C*H*H_{THP}), 25.3 (C*H*H_{THP}), 25.3 (C*H*H_{THP}), 21.0, 20.9, 20.9, 20.9, 20.9, 20.8, 20.8, 20.7 [*C*H_3 signals], 19.7 (C*H*H_{THP}). HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{41}H_{57}FO_{25}$: 991.3070; found: 991.3067.

4'''-*O*-Tetrahydropyranyl-β-D-xylopyranosyl-(1'''→4'')-bis[β-D-xylopyranosyl-(1→4)]-α-D-xylopyranosyl fluoride (28). Tetrasaccharide **27** (76 mg; 79 µmol) was dissolved in DCM



(600 µL) and MeOH (2.0 mL) was added. The reaction mixture was cooled down to 0°C in an ice-water bath and 500 mM sodium methoxide in MeOH (251 µL; 126 µmol; 1.6 equiv.) was added. The reaction was allowed to stir for 7 h while allowing to warm up to room temperature. The reaction was monitored by LC/MS and when no intermediates were present anymore, the reaction was quenched with an ion-exchange resin in H form (Amberlite® IRC-120(H)). Upon dilution with MeOH, part of the desired product precipitated and was collected by filtration. The rest of the liquid was concentrated under reduced pressure at 25°C. Without further purification, the procedure yielded the desired product **28** (44 mg; 69 µmol; 88%) as a white fluffy solid. ^{1}H NMR (400 MHz, D_2O) δ 5.60 (dd, *J* = 53.2, 2.7 Hz, 1H), 4.82 – 4.77 [under H_2O peak; H-1_{THP}; H-1_{THP}], 4.46 – 4.38 (m, 3H; H-1', H-1'', H-1'''), 4.06 (m, *J* = 2.6 Hz, 3H; C*H*H), 3.98 – 3.89 (m, *J* = 5.5 Hz, 2H; C*H*H), 3.87 – 3.44 (m, 13H), 3.41 – 3.17 (m, 9H), 1.86 – 1.64 (m, 2H; C*H*H), 1.61 – 1.33 (m, 4H; C*H*H). ^{13}C NMR (101 MHz, D_2O) δ 107.1 (d, $^{1}J_{C1\text{-}F}$ = 223.8 Hz; C-1), 101.7, 101.6, 101.6 [C-1', C-1'', C-1'''], 94.5 (C-1_{THP}), 77.2, 76.2, 76.2, 75.5, 75.4, 73.5, 73.5, 72.7, 72.6, 72.6, 71.0 (d, $^{1}J_{C2\text{-}F}$ = 25.0 Hz; C-2), 70.7, 69.1 [*C*H signals], 65.1, 64.8, 62.9, 61.0 (d, $^{3}J_{C5\text{-}F}$ = 3.7 Hz, C-5), 57.1 [CH_2 signals], 31.6, 30.1, 24.4, 20.2, 18.9 [$CH_{2(THP)}$]. HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{25}H_{41}FO_{17}$: 677.2304; found: 677.2354.

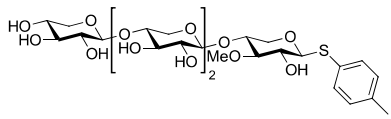
4-Methyphenyl β-D-xylopyranosyl-(1'''→4'')-bis[β-D-xylopyranosyl-(1→4)]-1-thio-β-D-xylopyranoside (29). According to general procedure B for THP removal, tetrasaccharide **23**



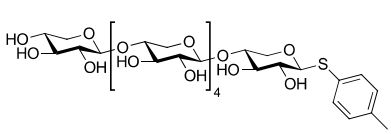
(10.8 mg; 14.7 µmol) was dissolved in H_2O (4.2 mL) and acetic acid was added (91 mg; 1.51 mmol; 86 µL). The procedure yielded the desired product **29** (8.7 mg; 13.3 µmol; 91%) as a white fluffy solid. R_f (EtOAc/AcOH/H_2O, 5:2:1) = 0.26. ^{1}H NMR (400 MHz, D_2O) δ 7.40 (d, *J* = 8.1 Hz, 2H; *H*C_{Ar}), 7.20 (d, *J* = 8.0 Hz, 2H; *H*C_{Ar}), 4.62 (d, *J* = 9.7 Hz, 1H; H-1), 4.43 – 4.37 (m, 3H; H-1', H-1'', H-1'''), 4.09 – 3.98 (m, 3H), 3.91 (dd, *J* = 11.6, 5.4 Hz, 1H), 3.83 – 3.43 (m, 9H), 3.40 – 3.15 (m, 9H), 2.28 (s, 3H; CH_3). ^{13}C NMR (101 MHz,

D_2O) δ 139.2 ($C_{Ar,q}$), 132.7 (2x$C_{Ar}H$), 129.9 (2x$C_{Ar}H$), 127.3 ($C_{Ar,q}$), 101.7, 101.5, 101.5 [C-1', C-1'', C-1'''], 88.1 (C-1), 76.2, 76.2, 75.9, 75.5, 75.1, 73.5, 73.5, 72.6, 72.6, 72.5, 71.5, 70.6, 69.1 [*C*-H signals], 66.4, 65.1, 62.8 [C-5 signals], 20.1 (CH_3). HRMS (ESI/Q-TOF) m/z: $[M+Na]^+$ Calcd for $C_{27}H_{40}O_{16}S$: 675.1935; found: 675.1934.

4-Methylphenyl β-D-xylopyranosyl-(1^{IV}→4^{III})-bis[β-D-xylopyranosyl-(1→4)]-3^{I}-*O*-methyl-1^{I}-thio-β-D-xylopyranoside (30). According to general procedure B for THP removal, tetrasaccharide **24** (8.3 mg; 11.1 µmol) was dissolved in H_2O (3.3 mL), and acetic acid was added (70 mg; 1.16 mmol; 66 µL). The procedure yielded the desired product **30** (6.5 mg; 9.8 µmol; 88%) as a white fluffy solid. R_f (EtOAc/AcOH/H_2O, 5:2:1) = 0.38. ^{1}H NMR (600 MHz, D_2O) δ 7.46 (d, *J* = 8.1 Hz, 2H; *H*C_{Ar}), 7.26 (d, *J* = 8.0 Hz, 2H; *H*C_{Ar}), 4.69 (d, *J* = 8.9 Hz, 1H; H-1^{I}), 4.49 – 4.43 (m, 3H; H-1^{II}; H-1^{III}; H-1^{IV}), 4.14 – 4.05 (m, 3H; 3 × H-5), 3.97 (dd, *J* = 5.5, 11.6 Hz, 1H; H-5), 3.85 – 3.74 (m, 3H), 3.62 (ddd, *J* = 5.5, 9.2, 10.4 Hz, 1H), 3.57 (s, 3H; CH_3O), 3.56 (d, *J* = 9.1 Hz, 1H), 3.53 (d, *J* = 8.8 Hz, 1H), 3.44 – 3.33 (m, 6H), 3.30 (dd, *J* = 10.3, 11.2 Hz, 1H), 3.29 (dd, *J* = 7.0, 7.6 Hz, 1H), 3.27 (dd, *J* = 6.7, 7.6 Hz, 1H), 3.25 (dd, *J* = 7.9, 9.3 Hz, 1H), 2.34 (s, 3H; CH_3C_{Ar}). ^{13}C NMR (151 MHz, D_2O) δ 139.4 ($C_{Ar,q}$), 133.1 (2x$C_{Ar}H$), 130.1 (2x$C_{Ar}H$), 127.5 ($C_{Ar,q}$), 102.0, 101.8, 101.7 [C-1^{II}, C-1^{III}, C-1^{IV}], 88.2 (C-1^{I}), 84.3 (C-3^{I}), 76.5, 76.5, 75.7, 75.0, 73.8, 72.9, 72.8, 70.8, 69.3 [*C*H signals], 66.3, 65.3, 63.1, 63.0 [C-5 signals], 59.6 (*C*H_3O), 20.3 (*C*H_3C_{Ar}). HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{28}H_{42}O_{16}S$: 711.2170; found: 711.2173.



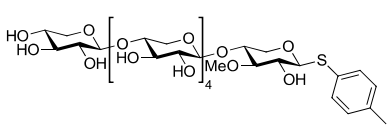
4-Methylphenyl β-D-xylopyranosyl-(1^{VI}→4^{V})-tetrakis[β-D-xylopyranosyl-(1→4)]-1^{I}-thio-β-D-xylopyranoside (31). According to general procedure A for enzymatic reactions, α-xylosyl fluoride donor **22** (6.9 mg; 18.7 µmol; 1.1 equiv.) and tetrasaccharide acceptor **29** (11.2 mg; 17.0 µmol; 1.0 equiv.) were dissolved in phosphate buffer (1.37 mL), and XynAE265G (530 µL; from 1.80 mg/mL) was added [c_{enzyme} = 500 µg/mL]. TLC eluent: EtOAc/AcOH/H_2O; 5:2:1. The procedure yielded the desired product (11.0 mg; 11.0 µmol; 65%) as a white fluffy solid. ^{1}H NMR (600 MHz, DMSO-d6) δ 7.33 (d, *J* = 8.1 Hz, 1H; *H*C_{Ar}), 7.14 (d, *J* = 8.1 Hz, 1H; *H*C_{Ar}), 5.42 – 5.31 (m, 1H), 5.20 – 5.04 (m, 7H),



5.04 – 4.94 (m, 5H), 4.78 – 4.74 (m, 1H; H-1_{THP}), 4.74 – 4.69 (m, 1H; H-1_{THP}), 4.53 (d, *J* = 9.2 Hz, 1H; H-1^{I}), 4.30 – 4.19 (m, 5H; H-1 signals), 4.06 – 3.97 (m, 1H; C*H*H), 3.96 – 3.89 (m, 2H; C*H*H), 3.89 – 3.78 (m, 5H; C*H*H), 3.78 – 3.69 (m, 1H; C*H*H), 3.53 – 3.43 (m, 7H), 3.27 – 3.20 (m, 10H), 3.20 – 3.12 (m, 7H), 3.07 – 2.98 (m, 8H), 2.55 – 2.53 (m, 2H), 2.28 (s, 3H; CH_3C_{Ar}), 1.79 – 1.71 (m, 1H), 1.71 – 1.54 (m, 3H), 1.51 – 1.36 (m, 6H). ^{13}C NMR (151 MHz, DMSO-d6) δ 136.6 ($C_{Ar,q}$), 131.6 (2xC_{Ar}H), 129.8 ($C_{Ar,q}$), 129.5 (2xC_{Ar}H), 101.9, 101.7, 101.6, 99.9, 95.4 [C-1 signals], 87.8 (C-1^{I}), 76.2, 75.4, 75.4, 75.2, 75.1, 74.1, 74.0, 72.7, 72.6, 72.1 [*C*H signals], 66.4, 64.6, 63.2, 62.8, 61.9 [C-5 signals], 40.4, 30.3 ($CH_{2,THP}$), 25.0 ($CH_{2,THP}$), 20.6 (CH_3C_{Ar}), 19.4 ($CH_{2,THP}$), 18.7 ($CH_{2,THP}$). HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{42}H_{64}O_{25}S$: 1045.3434; found: 1045.3466.

According to general procedure B for THP removal, the hexasaccharide product of the previous reaction (6.0 mg; 6.0 µmol) was dissolved in H_2O (1.9 mL), and acetic acid was added (100 µL). The procedure yielded the desired product **31** (4.1 mg; 4.5 µmol; 75%) as a white fluffy solid. R_f (EtOAc/ AcOH/H_2O, 3:2:1) = 0.33. ^{1}H NMR (600 MHz, D_2O) δ 7.47 (d, *J* = 8.0 Hz, 2H; *H*C_{Ar}), 7.26 (d, *J* = 7.9 Hz, 2H; *H*C_{Ar}), 4.68 (d, *J* = 9.7 Hz, 1H; H-1^{I}), 4.52 – 4.43 (m, 5H; H-1^{II}, H-1^{III}, H-1^{IV}, H-1^{V}, H-1^{VI}), 4.15 – 4.06 (m, 5H; 5×H-5), 3.97 (dd, *J* = 5.3, 11.6 Hz, 1H; H-5), 3.89 – 3.71 (m, 6H), 3.70 – 3.51 (m, 7H), 3.47 – 3.23 (m, 14H), 2.34 (s, 3H; CH_3C_{Ar}). ^{13}C NMR (151 MHz, D_2O) δ 139.4 ($C_{Ar,q}$), 132.9 (2xC_{Ar}H), 130.1 (2xC_{Ar}H), 127.5 ($C_{Ar,q}$), 101.9, 101.7, 101.7 [C-1 signals], 88.2 [C-1^{I}], 76.5, 76.4, 76.2, 75.7, 75.3, 73.8, 72.9, 72.8, 72.8, 71.7, 69.4, 69.3 [*C*H signals], 66.6, 65.3, 63.3, 63.1 [C-5 signals], 20.3 (CH_3C_{Ar}).HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{37}H_{56}O_{24}S$: 961.2859; found: 961.2867.

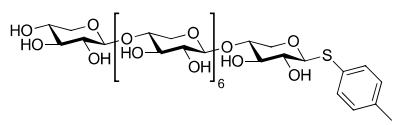
4-Methylphenyl β-D-xylopyranosyl-(1^{VI}→4^{V})-tetrakis[β-D-xylopyranosyl-(1→4)]-3^{I}-*O*-methyl-1^{I}-thio-β-D-xylopyranoside (32). According to general procedure A for enzymatic reaction, α-xylosyl fluoride donor **22** (3.0 mg; 8.1 µmol; 1.1 equiv.) and tetrasaccharide acceptor **30** (4.9 mg; 7.4 µmol; 1.0 equiv.) were dissolved in phosphate buffer (862 µL), and XynAE265G (115 µL; from 8.5 mg/mL) was added [c_{enzyme} = 1.0 mg/mL]. The procedure yielded the desired product (6.4 mg; 6.3 µmol; 85%) as a white fluffy solid. R_f (EtOAc/AcOH/H_2O; 5:2:1) = 0.35. ^{1}H NMR (600 MHz, D_2O) δ 7.46 (d, *J* = 8.0



Hz, 2H; *H*C$_{Ar}$), 7.27 (d, *J* = 8.0 Hz, 2H; *H*C$_{Ar}$), 4.70 (d, *J* = 8.9 Hz, 1H; H-1^{I}), 4.51 – 4.43 (m, 5H), 4.15 – 4.06 (m, 6H), 4.05 – 4.00 (m, 0.7H), 3.93 – 3.87 (m, 0.9H), 3.86 – 3.73 (m, 6H), 3.67 (ddd, *J* = 5.5, 9.7, 9.7 Hz, 1H), 3.63 – 3.59 (m, 1.5H), 3.58 (s, 3H, C*H*$_3$O), 3.57 – 3.50 (m, 5.5H), 3.44 – 3.33 (m, 8H), 3.32 – 3.26 (m, 5H), 2.34 (s, 3H; C*H*$_3$C$_{Ar}$), 1.86 – 1.75 (m, 2H; C*H*H$_{THP}$), 1.61 – 1.49 (m, 4H; C*H*H$_{THP}$). ^{13}C NMR (151 MHz, D$_2$O) δ 139.5 (*C*$_{Ar,q}$), 133.1 (2x*C*$_{Ar}$H), 130.2 (2x*C*$_{Ar}$H), 127.6 (*C*$_{Ar,q}$), 102.0, 101.9, 101.8, 101.8, 98.2 [C-1 signals], 88.3 (C-1^{I}), 84.4 (C-3^{I}), 77.3, 76.6, 76.6, 76.5, 75.0, 74.9, 74.6, 74.1, 73.9, 73.0, 72.9, 72.9, 70.9 [*C*H signals], 66.3, 64.7, 64.1, 63.8, 63.2 [C-5 signals], 59.7 (*C*H$_3$O), 30.4 (*C*H$_{2,THP}$), 30.3 (*C*H$_{2,THP}$), 24.7 (*C*H$_{2,THP}$), 20.3 (*C*H$_3$C$_{Ar}$), 19.6 (*C*H$_{2,THP}$), 19.2 (*C*H$_{2,THP}$). HRMS (ESI/Q-TOF) m/z: [M+HCOO]$^-$ Calcd for $C_{43}H_{66}O_{25}S$: 1059.3591; found: 1059.3611.

According to general procedure B for THP removal, the hexasaccharide product of the previous reaction (3.4 mg; 3.4 µmol) was dissolved in H$_2$O (1.33 mL), and acetic acid was added (27 µL). The procedure yielded the desired product **32** (2.5 mg; 2.7 µmol; 80%) as a white fluffy solid. *R*$_f$(EtOAc/ AcOH/H$_2$O, 3:2:1) = 0.41. ^{1}H NMR (600 MHz, D$_2$O) δ 7.47 (d, *J* = 8.1 Hz, 2H; *H*C$_{Ar}$), 7.27 (d, *J* = 8.1 Hz, 2H; *H*C$_{Ar}$), 4.70 (d, *J* = 8.4 Hz, 1H; H-1^{I}), 4.54 – 4.44 (m, 5H; H-1^{II}, H-1^{III}, H-1^{IV}, H-1^{V}, H-1^{VI}), 4.15 – 4.07 (m, 5H; 5 × H-5), 3.97 (dd, *J* = 5.5, 11.7 Hz, 1H; H-5), 3.86 – 3.75 (m, 6H), 3.66 – 3.60 (m, 1H), 3.58 (s, 3H; C*H*$_3$O), 3.60 – 3.52 (m, 5H), 3.46 – 3.34 (m, 9H), 3.34 – 3.24 (m, 7H), 2.35 (s, 3H; C*H*$_3$C$_{Ar}$). ^{13}C NMR (151 MHz, D$_2$O) δ 139.5 (*C*$_{Ar,q}$), 133.1 (2x*C*$_{Ar}$H), 130.1 (2x*C*$_{Ar}$H), 127.4 (*C*$_{Ar,q}$), 101.9, 101.8, 101.7 [C-1 signals], 88.2 (C-1^{I}), 84.3 (C-3^{I}), 76.5, 76.5, 76.4, 75.7, 75.0, 73.8, 72.9, 72.8, 70.8, 69.3 [*C*H signals], 66.3, 65.3, 63.1, 63.0 [C-5 signals], 59.6 (*C*H$_3$O), 20.3 (*C*H$_3$C$_{Ar}$). HRMS (ESI/Q-TOF) m/z: [M+HCOO]$^-$ Calcd for $C_{38}H_{58}O_{24}S$: 975.3015; found: 975.3027.

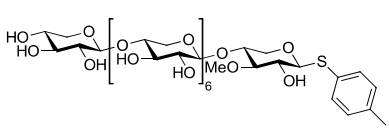
4-Methylphenyl β-D-xylopyranosyl-(1^{VIII}→4^{VII})-hexakis[β-D-xylopyranosyl-(1→4)]-1^{I}-thio-β-D-xylopyranoside (33). According to general procedure A for enzymatic reactions, α-xylosyl fluoride donor **28** (20.8 mg; 33 µmol; 1.2 equiv.) and tetrasaccharide acceptor **29** (17.9 mg; 27 µmol; 1 equiv.) were dissolved in phosphate buffer (782 µL), and XynAE265G (302 µL; from 1.80 mg/mL) was added [c_{enzyme} = 500 µg/mL]. The procedure yielded the desired product (25 mg; 20 µmol; 72%) as a white fluffy



solid. R_f(EtOAc/AcOH/H_2O, 3:2:1) = 0.39. HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{52}H_{80}O_{33}S$: 1309.4279; found: 1309.4282.

According to general procedure B for THP removal, the octasaccharide product of the previous reaction (5.0 mg; 4.0 µmol) was dissolved in H_2O (1.7 mL), and acetic acid was added (88 µL). The procedure yielded the desired product **33** (2.8 mg; 2.4 µmol; 60%) as a white fluffy solid. R_f(EtOAc/AcOH/H_2O, 3:2:1) = 0.19. ^{1}H NMR (600 MHz, D_2O) δ 7.47 (d, *J* = 7.9 Hz, 2H; *H*C_{Ar}), 7.27 (d, *J* = 7.9 Hz, 2H; *H*C_{Ar}), 4.69 (d, *J* = 9.7 Hz, 1H; H-1^{I}), 4.50 – 4.44 (m, 7H; H-1^{II}, H-1^{III}, H-1^{IV}, H-1^{V}, H-1^{VI}, H-1^{VII}, H-1^{VIII}), 4.14 – 4.08 (m, 8H; 7×H-5), 3.98 (dd, *J* = 5.4, 11.6 Hz, 1H; H-5), 3.83 – 3.73 (m, 8H), 3.63 (ddd, *J* = 5.5, 9.7, 9.7 Hz, 1H), 3.60 – 3.52 (m, 8H), 3.43 (dd, *J* = 8.9, 9.6 Hz, 2H), 3.41 – 3.24 (m, 19H), 2.35 (s, 3H; C*H*$_3$C$_{Ar}$). ^{13}C NMR (151 MHz, D_2O) δ 139.4 (*C*$_{Ar,q}$), 132.9 (2x*C*$_{Ar}$H), 130.1 (2x*C*$_{Ar}$H), 127.5 (*C*$_{Ar,q}$), 101.9, 101.8, 101.7 [C-1 signals], 88.2 (C-1^{I}), 76.5, 76.4, 76.2, 75.7, 75.3, 73.8, 72.9, 72.8, 71.7, 69.3 [*C*H signals], 66.6, 65.3, 63.1 [C-5 signals], 20.3 (*C*H$_3$C$_{Ar}$). HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{47}H_{72}O_{32}S$: 1225.3704; found: 1225.3685.

4-Methylphenyl β-D-xylopyranosyl-(1^{VIII}→4^{VII})-hexakis[β-D-xylopyranosyl-(1→4)]-3^{I}-*O*-methyl-1^{I}-thio-β-D-xylopyranoside (34). According to general procedure A for enzymatic

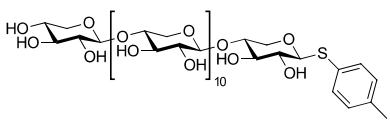


reactions, α-xylosyl fluoride donor **28** (5.0 mg; 7.9 µmol; 1.0 equiv.) and tetrasaccharide acceptor **30** (5.8 mg; 8.7 µmol; 1.1 equiv.) were dissolved in phosphate buffer (1279 µL), and XynAE265G (171 µL; from 8.5 mg/mL) was added [c_{enzyme} = 1.0 mg/mL]. The procedure yielded the desired product (7.4 mg; 5.8 µmol; 73%) as a white fluffy solid. R_f(EtOAc/ AcOH/H_2O, 3:2:1) = 0.44. HRMS (ESI/Q-TOF) m/z: $[M+Cl]^-$ Calcd for $C_{53}H_{82}O_{33}S$: 1313.4148; found: 1313.4139.

According to general procedure B for THP removal, the octasaccharide product of the previous reaction (9.0 mg; 7.0 µmol) was dissolved in H_2O (5.9 mL), and acetic acid was added (300 µL). The procedure yielded the desired product **34** (6.1 mg; 5.1 µmol; 73%) as a white fluffy solid. R_f(EtOAc/AcOH/H_2O, 3:2:1) = 0.29. ^{1}H NMR (600 MHz, DMSO-d6) δ 7.33 (d, *J* = 8.1 Hz, 2H; *H*C_{Ar}), 7.15 (d, *J* = 8.1 Hz, 2H; *H*C_{Ar}), 5.60 – 5.43 (br, 2H; *H*O), 5.28 – 5.11 (br, 7H; *H*O), 5.11 – 4.97 (br, 10H; *H*O), 4.55 (d, *J* = 8.8 Hz, 1H; H-1^{I}), 4.28 – 4.20 (m, 7H; H-1 signals),

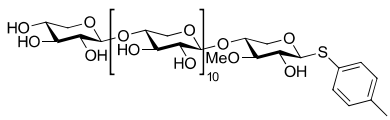
3.97 (dd, *J* = 4.8, 11.4 Hz, 1H; H-5), 3.90 – 3.82 (m, 6H; H-5 signals), 3.69 (dd, *J* = 5.3, 11.3 Hz, 1H; H-5), 3.58 – 3.53 (m, 3H), 3.53 – 3.47 (m, 8H), 3.46 (s, 3H; CH_3O), 3.30 – 3.19 (m, 10H), 3.19 – 3.12 (m, 7H), 3.12 – 3.06 (m, 4H), 3.06 – 2.96 (m, 8H), 2.96 – 2.91 (m, 1H), 2.28 (s, 3H; CH_3C_{Ar}). ^{13}C NMR (151 MHz, DMSO-d6) δ 136.7 ($C_{Ar,q}$), 131.6 (2xC_{Ar}H), 129.6 ($C_{Ar,q}$), 129.4 (2xC_{Ar}H), 102.0, 101.8, 101.7, 101.6 [C-1 signals], 87.9 (C-1^{I}), 85.0 (C-3^{I}), 76.3, 75.4, 75.3, 75.0, 74.1, 74.0, 73.2, 72.6, 72.5, 71.2, 69.5 [*C*H signals], 66.0, 65.8, 63.3, 63.2, 63.1 [C-5 signals], 59.7 (*C*H_3O), 20.5 (*C*H_3C_{Ar}).HRMS (ESI/Q-TOF) m/z: [M+HCOO]$^-$ Calcd for $C_{48}H_{74}O_{32}S$: 1239.3861; found: 1239.3863.

4-Methylphenyl β-D-xylopyranosyl-(1^{XII}→4^{XI})-decakis[β-D-xylopyranosyl-(1→4)]-1^{I}-thio-β-D-xylopyranoside (35). According to general procedure A for enzymatic reactions, α-xylosyl fluoride donor **28** (540 µg; 850 nmol; 1.2 equiv.) and octasaccharide acceptor **33** (840 µg; 710 nmol; 1.0 equiv.) were dissolved in phosphate buffer (466 µL), and XynAE265G (180 µL; from 1.80 mg/mL) was added [c_{enzyme} = 500 µg/mL]. The obtained product was directly used in the next reaction without further characterization.



According to general procedure B for THP removal, the dodecasaccharide product of the previous reaction was dissolved in DMSO, lyophilized and then dissolved in H_2O (950 µL), before acetic acid was added (50 µL). The procedure yielded the desired product **35** (460 µg; 270 nmol; 38% over two steps) as a white fluffy solid. ^{1}H NMR (600 MHz, DMSO-d6) δ 7.33 (d, *J* = 8.0 Hz, 2H; *H*C$_{Ar}$), 7.14 (d, *J* = 8.6 Hz, 2H; *H*C$_{Ar}$), 5.42 – 5.33 (m, 1H), 5.20 – 5.06 (m, 11H), 5.06 – 4.90 (m, 13H), 4.53 (d, *J* = 9.2 Hz, 1H; H-1^{I}), 4.31 – 4.19 (m, 11H; H-1 signals), 4.11 – 4.04 (m, 50H), 3.93 (dd, *J* = 4.7, 11.9 Hz, 1H; C*H*H), 3.90 – 3.83 (m, 11H; C*H*H), 3.70 (dd, *J* = 5.3, 11.1 Hz, 2H; C*H*H), 3.28 – 3.20 (m, 30H; C*H*), 3.20 – 3.12 (m, 19H; C*H*H), 3.07 – 2.97 (m, 19H; C*H*), 2.28 (s, 3H; CH_3C_{Ar}). [^{13}C NMR (DMSO-d6) δ 101.5 (H-1), 87.5 (H-1), 76.0 (C*H*), 75.3 (C*H*), 75.1 (C*H*), 73.8 (C*H*), 72.5 (C*H*), 65.7 (CH*H*), 63.1 (CH*H*), 63.1 (CH*H*), 40.2 (C*H*), 20.5 (*C*H_3C_{Ar}). This data was extracted from an HSQC-spectrum. No 13C-spectrum could be obtained due to the small amounts of material available.] HRMS (ESI/Q-TOF) m/z: [M+HCOO]$^-$ Calcd for $C_{67}H_{104}O_{48}S$: 1753.5394; found: 1753.5463.

4-Methylphenyl β-D-xylopyranosyl-(1^{XII}→4^{XI})-decakis[β-D-xylopyranosyl-(1→4)]-3^{I}-*O*-methyl-1^{I}-thio-β-D-xylopyranoside (36). According to general procedure A for enzymatic reactions, glycosyl fluoride donor **28** (3.4 mg; 5.4 µmol) and octasaccharide acceptor **34** (6.3 mg; 5.3 µmol) were dissolved in phosphate buffer (1.1 mL), and XynAE265G (148 µL; from 8.5 mg/mL) was added [c_{enzyme} = 1.0 mg/mL]. The procedure yielded the desired product (4.7 mg; 2.6 µmol; 48%) as a white fluffy solid. TLC eluent: EtOAc/AcOH/H_2O = 3:2:1. ^{1}H NMR (600 MHz, D_2O) δ 7.33 (d, *J* = 8.1 Hz, 2H; *H*C$_{Ar}$), 7.14 (d, *J* = 8.2 Hz, 2H; *H*C$_{Ar}$), 5.48 (br, 1H), 5.10 (br, 11H), 5.00 (br, 11H), 4.76 (br, 0.7H; H-1$_{THP}$), 4.71 (br, 0.6H; H-1$_{THP}$), 4.55 (d, *J* = 8.6 Hz, 1H; H-1^{I}), 4.29 – 4.22 (m, 11H; H-1 signals), 4.04 – 4.00 (m, 0.5H), 3.97 (dd, *J* = 5.3, 11.8 Hz, 1H; H-5), 3.94 – 3.79 (m, 12H; H-5 signals), 3.77 – 3.70 (m, 1H), 3.53 – 3.47 (m, 16H), 3.46 (s, 3H; C*H*$_3$O), 3.27 – 3.21 (m, 17H), 3.20 – 3.13 (m, 14H), 3.12 – 3.07 (m, 5H), 3.07 – 3.00 (m, 12H), 2.97 – 2.92 (m, 2H), 2.28 (s, 3H; C*H*$_3$C$_{Ar}$), 1.80 – 1.71 (m, 1H), 1.71 – 1.53 (m, 2H), 1.54 – 1.36 (m, 5H). ^{13}C NMR (151 MHz, D_2O) δ 136.7 (C$_{Ar,q}$), 131.7 (2x*C*$_{Ar}$H), 129.6 (C$_{Ar,q}$), 129.5 (2x*C*$_{Ar}$H), 101.9, 101.8, 101.7, 101.6 [C-1 signals], 99.9 (C-1$_{THP}$), 95.4 (C-1$_{THP}$), 87.9 (C-1^{I}), 85.0 (C-3^{I}), 75.4, 75.3, 75.0, 74.1, 74.0, 73.2, 72.8, 72.6, 71.2 [*C*H signals], 66.0, 63.2 [*C*H$_2$ signals], 59.7 (*C*H$_3$O), 30.3 (*C*H$_{2,THP}$), 24.9 (*C*H$_{2,THP}$), 20.6 (C*H*$_3$C$_{Ar}$).



According to general procedure B for THP removal, the dodecasaccharide product of the previous reaction (4.1 mg; 2.3 µmol) was dissolved in H_2O (950 µL), and acetic acid was added (50 µL). The desired product **36** was obtained (2.69 mg; 1.56 µmol; 69 %) as a white fluffy solid. R_f(EtOAc/ AcOH/H_2O, 2:2:1) = 0.25. ^{1}H NMR (600 MHz, DMSO-d6) δ 7.29 (d, *J* = 8.0 Hz, 2H; *H*C$_{Ar}$), 7.12 (d, *J* = 8.0 Hz, 2H; *H*C$_{Ar}$), 4.52 (d, *J* = 8.4 Hz, 1H; H-1^{I}), 3.91 – 3.81 [under H_2O peak; CH*H*], 3.96 – 3.83 (m, 10H; H-5 signals), 3.72 (dd, *J* = 5.3, 11.3 Hz, 1H; H-5), 3.56 – 3.46 (m, 11H), 3.41 (s, 3H; C*H*$_3$O), 3.35 – 3.25 (m, 11H), 3.22 – 3.11 (m, 15H), 3.11 – 3.06 (m, 2H), 3.06 – 3.01 (m, 9H), 3.01 – 2.96 (m, 3H), 2.22 (s, 3H; C*H*$_3$C$_{Ar}$). ^{13}C NMR (151 MHz, DMSO) δ 137.6 (*C*$_{Ar,q}$), 132.3 (2x*C*$_{Ar}$H), 130.1 (2x*C*$_{Ar}$H), 129.8 (*C*$_{Ar,q}$), 102.4, 102.1 [C-1 signals], 88.4 (C-1^{I}), 85.1 (C-3^{I}), 76.4, 76.2, 75.5, 74.2, 73.5, 73.0, 71.6, 69.8 [*C*H signals],

66.3, 66.1, 63.5, 63.4 [C-5 signals], 60.2 (CH_3O), 21.0 (CH_3C_{Ar}). HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{68}H_{106}O_{48}S$: 1767.5551; found: 1767.5543.

2.6. Appendix

$$n_8 = n_{8_{12}} + n_{8_{10}};\ \begin{cases} n_{8_{12}} = n_{12}\dfrac{1}{\epsilon_k} \\ n_{8_{10}} = n_{10}\dfrac{1}{\epsilon_e\epsilon_f} \end{cases};\ \begin{cases} n_{12} = n_{12_{22}} + n_{12_{15}};\ \begin{cases} n_{12_{22}} = 1.2n_{22}\dfrac{1}{\epsilon_l\epsilon_k\epsilon_m\epsilon_n\epsilon_o} \\ n_{12_{15}} = n_{15}\dfrac{1}{\epsilon_l\epsilon_c} = \dfrac{n_{22}}{2.4}\dfrac{1}{\epsilon_l\epsilon_c} \end{cases} \\ n_{10} = n_{22}\dfrac{1}{\epsilon_l\epsilon_k\epsilon_m\epsilon_n\epsilon_o} \end{cases}$$

$$\frac{n_{8_{12}}}{n_{8_{10}}} = \frac{\left(1.2n_{22}\dfrac{1}{\epsilon_l\epsilon_k\epsilon_m\epsilon_n\epsilon_o} + \dfrac{n_{22}}{2.4}\dfrac{1}{\epsilon_l\epsilon_c}\right)\dfrac{1}{\epsilon_k}}{n_{22}\dfrac{1}{\epsilon_l\epsilon_k\epsilon_m\epsilon_n\epsilon_o}\dfrac{1}{\epsilon_e\epsilon_f}} = 1.2\frac{\epsilon_e\epsilon_f}{\epsilon_k} + \frac{\epsilon_m\epsilon_n\epsilon_o\epsilon_e\epsilon_f}{2.4\epsilon_c}$$

Scheme A.1. The system of linear equations for the partition ratio in the divergent node of intermediate **8**. ϵ is the expected or reported yield of each reaction and n_{8sub10} and n_{8sub12} are the moles of intermediate **8** that are earmarked for the pathways to **10** and to **12,** respectively (Scheme 2.3).

$$n_{23} = n_{23_{27}} + n_{23_{28}};\ \begin{cases} n_{23_{27}} = n_{27}\dfrac{1}{\epsilon_q\epsilon_m\epsilon_n\epsilon_o} \\ n_{23_{28}} = n_{28}\dfrac{1}{\epsilon_r} \\ n_{27} = 1.2n_{28} \end{cases};\ \frac{n_{23_{27}}}{n_{23_{28}}} = \frac{1.2\epsilon_r}{\epsilon_q\epsilon_m\epsilon_n\epsilon_o}$$

Scheme A.2. The system of linear equations for the partition ratio of the glycosynthase-mediated glycosylation product **23**. This system of equations can be applied to any glycosynthase-mediated glycosylation product of the iterative synthesis. ϵ is the expected or reported yield of each reaction and $n_{23sub27}$ and $n_{23sub28}$ are the moles of intermediate **23** that are earmarked for the pathways to **27** and to **28,** respectively (Scheme 2.4).

2.7. References

(1) Chen, H.; Liu, J.; Chang, X.; Chen, D.; Xue, Y.; Liu, P.; Lin, H.; Han, S. A Review on the Pretreatment of Lignocellulose for High-Value Chemicals. *Fuel Processing Technology* **2017**, *160*, 196–206. https://doi.org/10.1016/j.fuproc.2016.12.007.

(2) Kim, J. S.; Lee, Y. Y.; Kim, T. H. A Review on Alkaline Pretreatment Technology for Bioconversion of Lignocellulosic Biomass. *Bioresour Technol* **2016**, *199*, 42–48. https://doi.org/10.1016/j.biortech.2015.08.085.

(3) Hakala, T. K.; Liitiä, T.; Suurnäkki, A. Enzyme-Aided Alkaline Extraction of Oligosaccharides and Polymeric Xylan from Hardwood Kraft Pulp. *Carbohydr Polym* **2013**, *93*, 102–108. https://doi.org/10.1016/j.carbpol.2012.05.013.

(4) Yang, Z.-Y.; Wu, D.-T.; Chen, C.-W.; Cheong, K.-L.; Deng, Y.; Chen, L.-X.; Han, B.-X.; Chen, N.-F.; Zhao, J.; Li, S.-P. Preparation of Xylooligosaccharides from Xylan by Controlled Acid Hydrolysis and Fast Protein Liquid Chromatography Coupled with Refractive Index Detection. *Sep Purif Technol* **2016**, *171*, 151–156. https://doi.org/10.1016/j.seppur.2016.06.051.

(5) Danon, B.; Marcotullio, G.; de Jong, W. Mechanistic and Kinetic Aspects of Pentose Dehydration towards Furfural in Aqueous Media Employing Homogeneous Catalysis. *Green Chem.* **2014**, *16*, 39–54. https://doi.org/10.1039/C3GC41351A.

(6) Zhao, S.; Dien, B. S.; Lindemann, S. R.; Chen, M.-H. Controlling Autohydrolysis Conditions to Produce Xylan-Derived Fibers That Modulate Gut Microbiota Responses and Metabolic Outputs. *Carbohydr Polym* **2021**, *271*, 118418. https://doi.org/10.1016/j.carbpol.2021.118418.

(7) Kinnaert, C.; Daugaard, M.; Nami, F.; Clausen, M. H. Chemical Synthesis of Oligosaccharides Related to the Cell Walls of Plants and Algae. *Chem Rev* **2017**, *117* (17), 11337–11405. https://doi.org/10.1021/acs.chemrev.7b00162.

(8) Takeo, K.; Ohguchi, Y.; Hasegawa, R.; Kitamura, S. Synthesis of (1→4)-β-D-Xylo-Oligosaccharides of Dp 4–10 by a Blockwise Approach. *Carbohydr Res* **1995**, *278* (2), 301–313. https://doi.org/10.1016/0008-6215(95)00259-6.

(9) Hirsch, J.; Kováč, P.; Petráková, E. An Approach to the Systematic Synthesis of (1→4)-β-D-Xylo-Oligosaccharides. *Carbohydr Res* **1982**, *106* (2), 203–216. https://doi.org/10.1016/S0008-6215(00)81074-5.

(10) Schmidt, D.; Schuhmacher, F.; Geissner, A.; Seeberger, P. H.; Pfrengle, F. Automated Synthesis of Arabinoxylan-Oligosaccharides Enables Characterization of Antibodies That Recognize Plant Cell Wall Glycans. *Chemistry - A European Journal* **2015**, *21* (15), 5709–5713. https://doi.org/10.1002/chem.201500065.

(11) Senf, D.; Ruprecht, C.; de Kruijff, G. H. M.; Simonetti, S. O.; Schuhmacher, F.; Seeberger, P. H.; Pfrengle, F. Active Site Mapping of Xylan-Deconstructing Enzymes with Arabinoxylan Oligosaccharides Produced by Automated Glycan Assembly. *Chemistry - A European Journal* **2017**, *23* (13), 3197–3205. https://doi.org/10.1002/chem.201605902.

(12) Kim, Y.-W.; Lee, S. S.; Warren, R. A. J.; Withers, S. G. Directed Evolution of a Glycosynthase from Agrobacterium Sp. Increases Its Catalytic Activity Dramatically and Expands Its Substrate Repertoire. *Journal of Biological Chemistry* **2004**, *279* (41), 42787–42793. https://doi.org/10.1074/jbc.M406890200.

(13) Kim, Y.-W.; Chen, H.; Withers, S. G. Enzymatic Transglycosylation of Xylose Using a Glycosynthase. *Carbohydr Res* **2005**, *340* (18), 2735–2741. https://doi.org/10.1016/j.carres.2005.09.010.

(14) Honda, Y.; Kitaoka, M. The First Glycosynthase Derived from an Inverting Glycoside Hydrolase. *Journal of Biological Chemistry* **2006**, *281* (3), 1426–1431. https://doi.org/10.1074/jbc.M511202200.

(15) Sugimura, M.; Nishimoto, M.; Kitaoka, M. Characterization of Glycosynthase Mutants Derived from Glycoside Hydrolase Family 10 Xylanases. *Biosci Biotechnol Biochem* **2006**, *70* (5), 1210–1217. https://doi.org/10.1271/bbb.70.1210.

(16) Davies, G. J.; Wilson, K. S.; Henrissat, B. Nomenclature for Sugar-Binding Subsites in Glycosyl Hydrolases. *Biochemical Journal* **1995**, *321*, 557–559.

(17) Kim, Y.-W.; Fox, D. T.; Hekmat, O.; Kantner, T.; McIntosh, L. P.; Warren, R. A. J.; Withers, S. G. Glycosynthase-Based Synthesis of Xylo-Oligosaccharides Using an

Engineered Retaining Xylanase from Cellulomonas Fimi. *Org Biomol Chem* **2006**, *4* (10), 2025–2032. https://doi.org/10.1039/b601667g.

(18) Ben-David, A.; Bravman, T.; Balazs, Y. S.; Czjzek, M.; Schomburg, D.; Shoham, G.; Shoham, Y. Glycosynthase Activity of *Geobacillus Stearothermophilus* GH52 β-Xylosidase: Efficient Synthesis of Xylooligosaccharides from α-D-Xylopyranosyl Fluoride through a Conjugated Reaction. *ChemBioChem* **2007**, *8* (17), 2145–2151. https://doi.org/10.1002/cbic.200700414.

(19) Teplitsky, A.; Mechaly, A.; Stojanoff, V.; Sainz, G.; Golan, G.; Feinberg, H.; Gilboa, R.; Reiland, V.; Zolotnitsky, G.; Shallom, D.; Thompson, A.; Shoham, Y.; Shoham, G. Structure Determination of the Extracellular Xylanase from *Geobacillus Stearothermophilus* by Selenomethionyl MAD Phasing. *Acta Crystallogr D Biol Crystallogr* **2004**, *60* (5), 836–848. https://doi.org/10.1107/S0907444904004123.

(20) Senf, D.; Ruprecht, C.; Kishani, S.; Matic, A.; Toriz, G.; Gatenholm, P.; Wågberg, L.; Pfrengle, F. Tailormade Polysaccharides with Defined Branching Patterns: Enzymatic Polymerization of Arabinoxylan Oligosaccharides. *Angewandte Chemie* **2018**, *130* (37), 12163–12168. https://doi.org/10.1002/ange.201806871.

(21) Fort, S.; Christiansen, L.; Schülein, M.; Cottaz, S.; Drigueza, H. Stepwise Synthesis of Cellodextrins Assisted by a Mutant Cellulase. *Isr J Chem* **2000**, *40*, 217–221. https://doi.org/10.1560/5NCY-L9K2-JND4-MB6Q.

(22) Binauld, S.; Damiron, D.; Connal, L. A.; Hawker, C. J.; Drockenmuller, E. Precise Synthesis of Molecularly Defined Oligomers and Polymers by Orthogonal Iterative Divergent/Convergent Approaches. *Macromol Rapid Commun* **2011**, *32* (2), 147–168. https://doi.org/10.1002/marc.201000548.

(23) Paynter, O. I.; Simmonds, D. J.; Whiting, M. C. The Synthesis of Long-Chain Unbranched Aliphatic Compounds by Molecular Doubling. *J Chem Soc Chem Commun* **1982**, No. 20, 1165. https://doi.org/10.1039/c39820001165.

(24) Meppen, M.; Wang, Y.; Cheon, H.-S.; Kishi, Y. Synthetic 6-*O*-Methylglucose-Containing Polysaccharides (*s*MGPs): Design and Synthesis. *J Org Chem* **2007**, *72* (6), 1941–1950. https://doi.org/10.1021/jo061990v.

(25) Hsu, M. C.; Lee, J.; Kishi, Y. Synthetic 3-*O*-Methylmannose-Containing Polysaccharides (*s*MMPs): Design and Synthesis. *J Org Chem* **2007**, *72* (6), 1931–1940. https://doi.org/10.1021/jo061991n.

(26) Cheon, H.-S.; Lian, Y.; Kishi, Y. Highly Stereoselective and Iterative Synthesis of α-(1→4)-Linked Polysaccharides Composed of 3-*O*-Methyl-D-Mannose. *Org Lett* **2007**, *9* (17), 3323–3326. https://doi.org/10.1021/ol0713335.

(27) Zhu, Q.; Shen, Z.; Chiodo, F.; Nicolardi, S.; Molinaro, A.; Silipo, A.; Yu, B. Chemical Synthesis of Glycans up to a 128-Mer Relevant to the *O*-Antigen of *Bacteroides Vulgatus*. *Nat Commun* **2020**, *11*, 4142. https://doi.org/10.1038/s41467-020-17992-x.

(28) Wohlert, M.; Benselfelt, T.; Wågberg, L.; Furó, I.; Berglund, L. A.; Wohlert, J. Cellulose and the Role of Hydrogen Bonds: Not in Charge of Everything. *Cellulose* **2022**, *29*, 1–23. https://doi.org/10.1007/s10570-021-04325-4.

(29) Andraos, J. On Using Tree Analysis to Quantify the Material, Input Energy, and Cost Throughput Efficiencies of Simple and Complex Synthesis Plans and Networks: Towards a Blueprint for Quantitative Total Synthesis and Green Chemistry. *Org Process Res Dev* **2006**, *10* (2), 212–240. https://doi.org/10.1021/op0501904.

(30) Ren, B.; Wang, M.; Liu, J.; Ge, J.; Zhang, X.; Dong, H. Zemplén Transesterification: A Name Reaction That Has Misled Us for 90 Years. *Green Chemistry* **2015**, *17* (3), 1390–1394. https://doi.org/10.1039/C4GC02006E.

(31) Pezzetta, C.; Tshepelevitsh, S.; Leito, I.; Poli, G. Comment on "Zemplén Transesterification: A Name Reaction That Has Misled Us for 90 Years" by B. Ren, M. Wang, J. Liu, J. Ge, X. Zhang and H. Dong, *Green Chemistry*, 2015, **17**, 1390–1394. *Green Chemistry* **2018**, *20* (10), 2392–2394. https://doi.org/10.1039/C7GC03795C.

(32) Dinkelaar, J.; Witte, M. D.; van den Bos, L. J.; Overkleeft, H. S.; van der Marel, G. A. NIS/TFA: A General Method for Hydrolyzing Thioglycosides. *Carbohydr Res* **2006**, *341* (10), 1723–1729. https://doi.org/10.1016/j.carres.2006.03.009.

(33) Yu, B.; Tao, H. Glycosyl Trifluoroacetimidates. 2. Synthesis of Dioscin and Xiebai Saponin I. *J Org Chem* **2002**, *67* (25), 9099–9102. https://doi.org/10.1021/jo026103c.

(34) Yang, B.; Yoshida, K.; Yin, Z.; Dai, H.; Kavunja, H.; El-Dakdouki, M. H.; Sungsuwan, S.; Dulaney, S. B.; Huang, X. Chemical Synthesis of a Heparan Sulfate Glycopeptide: Syndecan-1. *Angewandte Chemie* **2012**, *124* (40), 10332–10336. https://doi.org/10.1002/ange.201205601.

(35) Zeng, Y.; Wang, Z.; Whitfield, D.; Huang, X. Installation of Electron-Donating Protective Groups, a Strategy for Glycosylating Unreactive Thioglycosyl Acceptors Using the Preactivation-Based Glycosylation Method. *J Org Chem* **2008**, *73* (20), 7952–7962. https://doi.org/10.1021/jo801462r.

(36) Crich, D. En Route to the Transformation of Glycoscience: A Chemist's Perspective on Internal and External Crossroads in Glycochemistry. *J Am Chem Soc* **2021**, *143*, 17–34. https://doi.org/10.1021/jacs.0c11106.

(37) Hayashi, M.; Hashimoto, S.; Noyori, R. Simple Synthesis of Glycosyl Fluorides. *Chem Lett* **1984**, *13* (10), 1747–1750. https://doi.org/10.1246/cl.1984.1747.

(38) *Bio-Gel® P Polyacrylamide Gel, Instruction Manual (LIT174 Rev B)*. https://www.bio-rad.com/sites/default/files/webroot/web/pdf/lsr/literature/LIT-174B.pdf (accessed 2022-11-28).

(39) Lunin, V. v.; Wang, H. T.; Bharadwaj, V. S.; Alahuhta, M.; Peña, M. J.; Yang, J. Y.; Archer-Hartmann, S. A.; Azadi, P.; Himmel, M. E.; Moremen, K. W.; York, W. S.; Bomble, Y. J.; Urbanowicz, B. R. Molecular Mechanism of Polysaccharide Acetylation by the Arabidopsis Xylan *O*-Acetyltransferase XOAT1. *Plant Cell* **2020**, *32* (7), 2367–2382. https://doi.org/10.1105/tpc.20.00028.

(40) Bromley, J. R.; Busse-Wicher, M.; Tryfona, T.; Mortimer, J. C.; Zhang, Z.; Brown, D. M.; Dupree, P. GUX1 and GUX2 Glucuronyltransferases Decorate Distinct Domains of Glucuronoxylan with Different Substitution Patterns. *Plant Journal* **2013**, *74*, 423–434. https://doi.org/10.1111/tpj.12135.

(41) Mortimer, J. C.; Faria-Blanc, N.; Yu, X.; Tryfona, T.; Sorieul, M.; Ng, Y. Z.; Zhang, Z.; Stott, K.; Anders, N.; Dupree, P. An Unusual Xylan in Arabidopsis Primary Cell Walls Is Synthesised by GUX3, IRX9L, IRX10L and IRX14. *Plant Journal* **2015**, *83*, 413–426. https://doi.org/10.1111/tpj.12898.

(42) Domon, B.; Costello, C. E. A Systematic Nomenclature for Carbohydrate Fragmentations in FAB-MS/MS Spectra of Glycoconjugates. *Glycoconj J* **1988**, *5* (4), 397–409. https://doi.org/10.1007/BF01049915.

Chapter 3
Chemical Synthesis of Glucoxylan Oligosaccharides

3.1. Introduction

3.1.1. Glucoxylan, Structure and Biosynthesis

Glucoxylan polysaccharides are a type of glycan composed of an unsubstituted and unramified linear 1,4-β-linked backbone of glucosyl and xylosyl residues that was first found in the cell wall of the species *Ulva rigida*,[1] which is a type of green algae in the Chlorophyta division, belonging to the genus of sea lettuces. Recently, it has also been found in land

plants, precisely in barley (*Hordeum vulgare*), where it is synthesised by the enzymes encoded in the *HvCslF3* and *HvCslF10* genes.[2]

These two genes are part of the *Cellulose synthase-like* (*Csl*) subfamily F, which, in turn, is part of a bigger family of genes that is often known as the (Archaeplastida) cellulose synthase gene superfamily. This superfamily is divided into the *CesA* subfamily and ten *Csl* subfamilies, i.e., from *CslA* to *CslM*, excl., *CslI*, *CslK* and *CslL*.[3] The genes in this superfamily are involved in the synthesis of glycans in plants[4] and all are reported to encode membrane-bound processive enzymes that belong to family GT2. These enzymes, therefore, contain a GT-A fold and are inverting glycosyltransferases. The first genes of this superfamily were identified in rice and cotton because of their similarities with bacterial cellulose synthases.[5] Now, identification of genes that belong to this superfamily can be done by sampling plant genomes for genes with assignments to the PF03552 (*Cellulose_synt*), PF00535 (*Glycos_transf_2*) and PF13632 (*Glyco_trans_2_3*) Pfam hidden Markov models (HMMs).[3]

Of the eleven subfamilies of the cellulose synthase gene superfamily, *CslB* and *CslM* are eudicot-specific subfamilies and *CslF*, *CslH*, *CslJ* are monocot-specific subfamilies.[3] The genes of the *CesA* subfamily have been linked to the synthesis of cellulose and the genes of the *Csl* subfamilies mostly to the synthesis of non-cellulosic glycans, e.g., *CslA* genes are linked to the synthesis of mannan and glucomannan, *CslC* genes to the 1,4-β-glucan backbone of xyloglucan[6] and *CslD* genes to cellulose (Figure 3.1).[7] The genes of the *CslF*, *CslH* and *CslJ* subfamilies have been linked to the synthesis of 1,3;1,4-β-glucan[3,8,9] and, as mentioned above, the *CslF3* and *CslF10* genes of barley to glucoxylan.[2] It was expected for some proteins encoded within the *Csl* subfamilies to be involved in the synthesis of xylan; however, no example has been found so far.

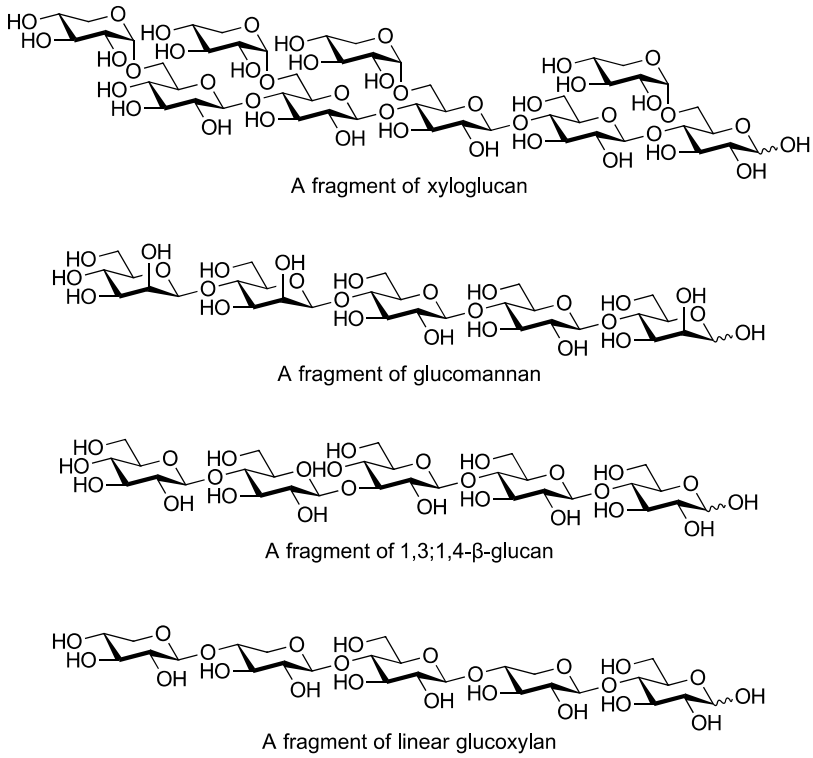


Figure 3.1. Examples of fragments of glycans synthesised by proteins encoded in the *Csl* genes.

Even though all the enzymes encoded in the cellulose synthase gene superfamily are classified in the CAZy database as part of the GT2 family[10], plants can encode proteins that only belong to the GT2 family, but not to the cellulose synthase gene superfamily. For instance, *Arabidopsis thaliana* encode 40 proteins belonging to this gene superfamily[11] and at least two additional proteins that belong only to the GT2 family, which are two non-Leloir glycosyltransferases: a dolichylphosphate β-D-mannosyltransferase (DPMS1; At1g20575) and a dolichylphosphate β-glucosyltransferase (At2g39630).[12]

In the genome of barley (*Hordeum vulgare*; a grass species), there are 41 genes belonging to the cellulose synthase gene superfamily, divided into 8 clades or subfamilies: *CesA* (8 genes; *CesA1-6*, *8*, *11*), *CslA* (8 genes; *CslA1-8*), *CslC* (5 genes; *CslC1-5*), *CslD* (5 genes; *CslD1-5*), *CslE* (3 genes; *CslE1-2*, *6*), *CslF* (10 genes; *CslF3-4*, *6-13*), *CslH* (1 gene; *CslH1*), and *CslJ* (1 gene; *CslJ1*).[13] Additionally, *HvCesA7* and *HvCesA9* genes have been mentioned as

homologs of *HvCesA5* and *HvCesA6* genes, respectively.[14,15] This species has genes of all subfamilies of the gene superfamily except in the eudicot-specific subfamilies (*CslB*, *CslM*) and in the *CslG* subfamily (which was believed to be a eudicot-specific clade but has also been found in monocots).[3]

Structural domains and motifs in the proteins of the cellulose synthase gene superfamily

Enzymes in the *CesA* subfamily are localized both to the plasma membrane and to the Golgi apparatus but synthesise cellulose only at the former, while enzymes encoded in the *Csl* subfamilies, including *HvCslF3* and *HvCslF10*, are localized to the Golgi apparatus and synthesise glycans there.

The proteins of the *CesA* subfamily are known to form a six-subunit rosette-like structure at the plasma membrane, which is known as the cellulose synthase complex (CSC). At least in *Arabidopsis thaliana*, each of these subunits is arranged as a heterogeneous trimer of CESA proteins in a 1:1:1 stoichiometry,[16] which seem to synthesise a protofibril with three cellulose chains that will end up forming part of a larger 18-chain cellulose microfibril.[17] However, in Aspen (*Populus tremula*), a 3:2:1 stoichiometry of CESA proteins has been found, which would be in agreement with, for instance, each subunit being formed by hexamers of CESA proteins.[18] The proteins of the *Csl* subfamilies, which are mostly involved in the synthesis of non-cellulosic glycans, have not been found to form oligomeric structures.[19] Proteins encoded by genes like those in the *CesA* or the CslF subfamilies use UDP-glucose and require a cation such as Mg^{2+} or Mn^{2+} for their synthetic activity.[19]

The proteins of the *CesA*, *CslD* and *CslF* subfamilies show a relative high sequence similarity among them and share at least the same membrane topology.[20] All the proteins of the *CesA* gene superfamily have several transmembrane-domains, and they generally conserve three aspartic acid residues (including aspartic acid at the TED motif) and a QxxRW motif,[4,21] however, there are exceptions[15,22]. Bacterial cellulose synthases are often a good model to study the structural motifs and interactions of cellulose synthases.[23,24]

The proteins belonging to the *CesA* subfamily, which form part of the cellulose synthase complex (CSC), are the most studied enzymes within the cellulose synthase gene superfamily and have a relatively high sequence similarity with the genes of some *Csl* subfamilies such as the *CslF* subfamily. The sequences of these proteins generally include seven transmembrane (TM) domains and three interface (IF) helixes, including IF3, which was previously believed to be a TM domain.[17] As an example, the structural sequence of *Ptt*CESA8 (from *Populus tremulata x tremuloides*, a eudicot tree*)* is described here[17]. Starting from the terminal NH_2, there is a RING (really interesting new gene) finger domain, which chelates Zn atoms, and has been found to be critical for the protein-protein interactions that stabilize the CSC.[25] After that, there is a variable region (VR1) followed by two TM domains (TM1 and TM2). The cytoplasmatic region that follows (between TM2 and TM3) includes a conserved region among bacteria and plants, and within it, a plant-conserved region (PCR), conserved within Archaeplastida (Plantae *sensu lato*). This part of the cytoplasmatic region also contains IF1 and two coordinating motifs (DDG and DCD). After that, there is a second variable region (VR2), which is specific for each class or isomorph of CESA protein and therefore is also known as class-specific region (CSR). Before the next TM domains, there is another conserved region among bacteria and plants that includes the TED motif (which includes the aspartic acid (D) base catalyst), and IF2, which contains the QxxRW motif with the W residue that has been found to coordinate the acceptor via CH-π stacking interactions.[17] After that, there are two TM domains (TM3 and TM4), IF3, and then the last three TM domains before the terminal COOH.[17,26,27]

The monocot-specific subfamily *CslF*

The *CslF* subfamily, together with the *CslH* and *CslJ* subfamilies, are the only monocot-specific subfamilies of the *CesA* gene superfamily; the *CslF* subfamily is more precisely specific to the graminid and restiid clades of plants in the Poales order in monocots.[3] Proteins in these subfamilies have been found to be involved in the synthesis of 1,3;1,4-β-glucan, which is a linear, unsubstituted and unramified hemicellulose composed of 1,3-β-linked and 1,4-β-linked glucosyl residues; generally, a 1,3-β-linkage is found every three or four

linkages. A model to explain the mechanism of formation of this glycan by the CSLF6 protein was recently published.[19]

Instead of synthesising 1,3;1,4-β-glucan, the *HvCslF3* and *HvCslF10* genes of barley, which form a clade themselves within the *CslF* subfamily, encode proteins that synthesise 1,4-β-linked glucoxylan and no detectable amounts of 1,3;1,4-β-glucan.[2] This discovery was made by Little and colleagues in 2019.[2] Their study began by expressing each of the genes of the monocot-specific *CslF* subfamily from barely (*Hordeum vulgare*) in the leaves of *Nicotiana benthamiana*, a eudicot plant. They found that leaves expressing *HvCslF6*, *HvCslF7*, *HvCslF8* or *HvCslF9* contained 1,3;1,4-β-glucan (with the levels in the leaves of the plant expressing *HvCslF6* being much higher), while the leaves expressing *HvCslF3* or *HvCslF10* contained the newly discovered linear glucoxylan polysaccharide. The analysis of the glycan content of the leaves was done by treating the leaves with different glycosidase preparations (Driselase, E-CELAN and E-CELTR), and then separating and analysing the released molecules from each treatment. The separation and analysis of the samples was done by HPAEC-PAD (high-performance anion-exchange chromatography with pulsed amperometric detection), carbon SPE (solid-phase extraction), MS (LC-ESI-qTOF-MS), LC-MS^2, ^{13}C NMR, etc. These efforts led first to the identification of diagnostic disaccharides Xyl*p*-(1,4)-β-Glc*p* and Glc*p*-(1,4)-β-Xyl*p* and then to the identification of longer glucoxylan molecules. Finally, several parts of the barley plant were treated with E-CELTR (which hydrolyse most of the 1,4-β-linkages of glucan and xylan fragments) and the glycan fragments released, which were mostly disaccharides, were separated and analysed by LC-ESI-qTOF-MS as previously done. This procedure showed the presence of glucoxylan oligosaccharides, predominantly in the coleoptile (which is a part of the plant particular of monocotyledons), but also, in leaves and roots.

3.1.2. The Synthesis of β-1,4-Linked Oligosaccharides

There are several examples in the literature for the synthesis of other linear 1,4-linked oligosaccharides. To protect the hydroxyl glycosylation site in these types of syntheses, silyl groups are a common choice since they are generally orthogonal to other common

hydroxyl protecting groups such as benzyl and benzoyl groups and they are usually easy to install and easy to remove. In order of increasing bulkiness and hence the robustness as protecting groups, some silyl groups can be found listed as: trimethylsilyl (TMS), triethylsilyl (TES), *tert*-butyldimethylsilyl (TBS), triisopropylsilyl (TIPS) and *tert*-butyldiphenylsilyl (TBDPS). Among them, TBS and TIPS are often the most popular choices since they offer a good balance between robustness and ease of installation and removal. For the 4-OH glycosylation site, TBS (or TIPS) are normally the choice when, e.g., a 4,6-*O*-benzylidene ring cannot be installed because the 4-OH is in axial (e.g., galacto configuration), the 6-OH is absent (e.g., xylo configuration), or the 6-OH needs to also be orthogonally protected to the rest of hydroxyl groups (e.g., for a later oxidation into a carboxylic acid).

Examples of the synthesis of 1,4-linked oligosaccharides using TBS can be found for heparin[28], heparan[29,30], 1,4-β-mannuronates[31], 1,4-α-linked aminogalacturonates[32] as well as for a recently synthesised 1,3;1,4-linked 128-mer glycan[33]. TIPS is more robust against acidic and basic cleavage and, therefore, it may be the preferred choice in combination with, e.g., glycosyl phosphate donors and TBSOTf as a promoter, as otherwise desilylation may occur.[34]

To the best of our knowledge, no syntheses of glucoxylan oligosaccharides have been reported in the literature. An example of a somewhat related plant cell wall glycan with a linear 1,4-β-linked backbone containing alternating residues of different configurations is glucomannan. The synthesis of this glycan is particularly challenging due to the difficulties of forming a β-mannosyl linkage and, thus far, only the synthesis of a glucomannan trisaccharide has been reported in the literature.[35,36]

3.2. Aims

This work has been designed with the aim of obtaining a small library of the glucoxylan oligosaccharides for the study, for instance, of the substrate specificity of glycosyltransferases involved in glucoxylan biosynthesis. The target oligosaccharides were designed to be

equipped with an aminoalkyl linker for later immobilization on glycan arrays or attachment of tags.

The synthesis of these glycan structures was designed taking to minimize the number of necessary chemical transformation. To enable a convergent synthesis, the anomeric leaving groups of the building blocks had to be chosen carefully to allow for a sequential activation according to the needs of the synthesis. As glucoxylan polysaccharides are 1,4-β-linked glycans, the glycosylation reactions had to be highly stereoselective towards the β-glycosylation products and an orthogonal protecting group had to be chosen for the 4-OH glycosylation site, which could be installed and removed independently to the rest of hydroxyl protecting groups.

3.3. Results and Discussion

The glucoxylan target oligosaccharides synthesised were selected among the fragments identified by Little and colleagues[2] in the leaves of *Nicotiana benthamiana* expressing the barley genes *HvCslF3* or *HvCslF10*.

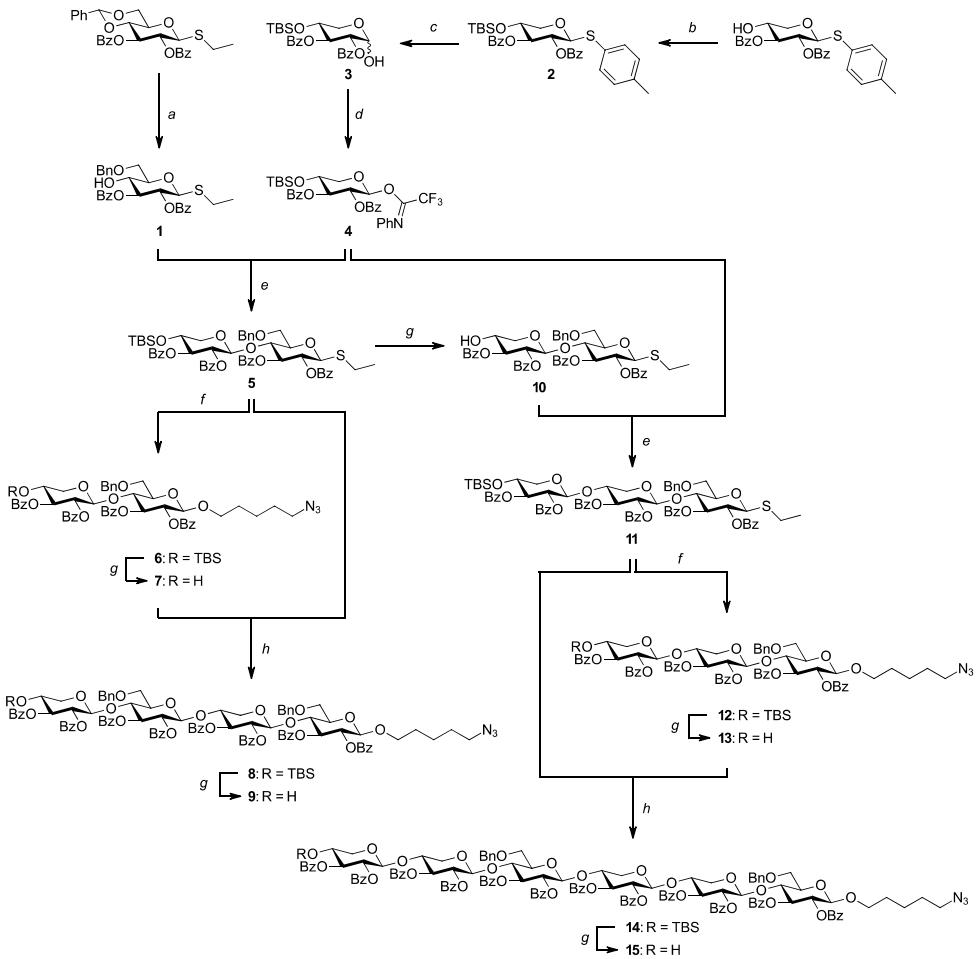
Similar to the xylan syntheses in Chapter 2, the synthesis of the glucoxylan oligosaccharides includes the synthetic motifs where an intermediate is transformed into both glycosyl donor and glycosyl acceptor for the following glycosylation reaction. For the 4-OH glycosylation site, a *tert*-butyldimethylsilyl (TBS) group has been chosen as a temporary protecting group, since it offers the desired balance between inertness and ease of installation and removal.

Ethyl thioglucoside **1** and TBS-protected 2,2,2-trifluoro-*N*-phenylacetimidoyl xyloside **4** were used as monosaccharide building blocks (Scheme 3.1). Xylosyl donor **4** can be activated in the presence of thioglucoside **1,** and the resulting coupling product can in turn be readily used as a glycosyl donor or be transformed into an acceptor for further elongation.

Ethyl thioglucoside **1** was obtained from the commercially available ethyl 4,6-*O*-benzylidine-1-thio-β-D-glucopyranoside by benzoylation of the two free hydroxy groups with BzCl

in Et_3N followed by selective ring opening of the benzylidene group using Et_3SiH and TfOH in DCM at −78°C to obtain the 4-OH glucoside **1** in 81% yield (Scheme 3.1). Xylosyl donor **4** was obtained from 4-methylphenyl 2,3-di-*O*-benzoyl-1-thio-β-D-xylopyranoside, for which the synthesis is described in Chapter 2. This thioxyloside was silylated at the 4-OH group with TBSCl and imidazole in DMF[37,38] to form **2** in 94% yield. Then hydrolysis of the C-S bond was performed using NIS/TFA in DCM/H_2O (10:1), providing **3** in 96% yield. Hemiacetal **3** was transformed into the desired xylosyl donor **4** in 93% yield by treatment with 2,2,2-trifluoro-*N*-phenylacetimidoyl chloride and Cs_2CO_3 in acetone[39] (Scheme 3.1).

Monosaccharides BBs **1** and **4** were then readily coupled to form the key intermediate disaccharide **5** using TMSOTf in DCM in 71% yield, without detection of any product resulting from the cleavage of the TBS group by TMSOTf (Scheme 3.1). This molecule (**5**) was then partitioned into three unequal portions, two for the route towards tetrasaccharide **9** (XGXG) and one for the route towards hexasaccharide **15** (XXGXXG). On the route towards **9**, one portion of **5** is kept for its use as glycosyl donor and the other is transformed into the glycosyl acceptor (**7**) for the next glycosylation reaction. Thus, the portion used towards the synthesis of **7** was glycosylated with 5-azidopentanol using NIS/TfOH in DCM to obtain **6** in 81% yield. 5-Azidopentanol was obtained from 5-aminopentanol by reaction with 1*H*-imidazole-1-sulfonyl azide hydrochloride.[40,41] Disaccharide **6** was then transformed into acceptor **7** by TBS removal with TBAF/AcOH in THF[42] in 91% yield. Acetic acid was used as a buffer to prevent cleavage of the benzoyl groups. During the workup procedure of this reaction, TBAF was easily removed by treatment with $CaCO_3$ (2 equiv) and Dowex® 50W X8 in H form (3 times the mass of $CaCO_3$ used) in THF (1.5 times the volume of initial solvent) for 1h.[43] The original workup procedure calls for dilution in MeOH, but in the presence of a basic ion such as fluoride (even if buffered with AcOH), the presence of methanol might lead to the cleavage of the benzoyl groups. With donor **5** and acceptor **7** in hand, they were coupled using the NIS/TfOH system in DCM to form the desired tetrasaccharide **8** in 60% yield, which was subsequently desilylated as described above to obtain **9** in 80% yield. The moderate yield of the glycosylation reaction is presumably due to the poor donor reactivity of **5** caused by the electron-withdrawing benzoyl groups.



Scheme 3.1. Synthetic scheme for the synthesis of glucoxylan oligosaccharides. (a) Et_3SiH, TfOH in DCM at −78°C; (b) TBSCl, imidazol in DMF at 0°C to room temperature (RT); (c) NIS/TFA in DCM/H_2O (10:1) at 0°C; (d) ClC(NPh)CF_3, Cs_2CO_3 in acetone at 0°C to RT; (e) TMSOTf in DCM at −60°C to −20°C; (f) 5-azidopentanol with NIS/TfOH in DCM at −60°C to −20°C; (g) TBAF/AcOH in THF at 0°C to RT; (h) NIS/TfOH in DCM at −60°C to −20°C.

On the route towards hexasaccharide **15**, disaccharide **5** was desilylated to form glycosyl acceptor **10** in 81% yield, which was glycosylated with xylosyl acetimidate **4** using TMSOTf in DCM in 62% yield. The resulting trisaccharide **11** was then partitioned into two portions. One was kept as a glycosyl donor, and the other was transformed into acceptor **13** by glycosylation with 5-azidopentanol in 84% yield followed by desilylation in 92% yield. Both

compounds (**11** and **13**) were coupled using NIS/TfOH in DCM to form hexasaccharide **14** in 46% yield, which was desilylated to form **15** in a 46% yield. There might be two reasons for the relatively poor yield in the glycosylation reaction: (a) longer reaction times were required for this transformation because of the electron-withdrawing benzoyl groups and the relatively bulkiness of the glycosyl donor and acceptor, or (b) the leaving group and activation system used, i.e., thioglycosides using NIS/TfOH in DCM, does not form intermediate species that are reactive enough for this glycosylation reaction and, for that reason, other leaving groups and activation systems[33] are sometimes preferred.

With the four compounds in hand (**7**, **9**, **13** and **15**), removal of the protecting groups would allow for the use of these molecules as substrates for biosynthetic studies.

3.4. Conclusions

A small library of glucoxylan oligosaccharides have been synthesised. These molecules, which range from di- to hexasaccharides (compounds **7**, **9**, **13** and **15**), are among the glucoxylan polysaccharide fragments identified by Little and colleagues[2] during the study of the *HvCslF3* and *HvCslF10* genes of barley.

The design of the synthesis was done taking into consideration the structural similarity between the glucoxylan oligosaccharide fragments, and this allowed for the implementation of common synthetic pathways to different target molecules, which is characteristic of divergent syntheses. In addition, similar to the syntheses in chapter 2, the synthetic motifs where a single intermediate is transformed into both the glycosyl donor and the acceptor for the following glycosylation reaction were also implemented here.

To obtain the desired β-linkage between glycosyl residues, the glycosyl donors were equipped with a 2-*O*-benzoyl group, which, due to their neighbouring-group participating effect, led exclusively to the formation of the desired β-glycosylation products.

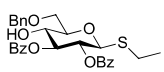
Different protecting groups and anomeric leaving groups were used for the synthesis of the glucoxylan oligosaccharides. The PG for the 4-OH glycosylation site, *tert*-butyldimethylsilyl

(TBS) group, was found to be orthogonal to the rest of hydroxyl protecting groups, easy to install, easy to remove and stable during all the chemical transformations performed, including glycosylation reactions with equimolar amounts of TMSOTf. The 5-azidopentyl linker installed at the exocyclic anomeric oxygen atom and all the rest of protecting groups used during the syntheses, i.e., benzoyl and benzyl groups, and the azido group in the linker, have not been affected during the synthesis. Overall, this synthesis represents a robust and efficient method for preparing glucoxylan oligosaccharides.

After deprotection of the hydroxyl and amino groups, these synthesised glucoxylan oligosaccharides can be printed as a microarray and to study, for instance, the substrate specificity of plant cell wall biosynthetic enzymes.

3.5. Experimental Section

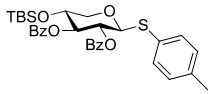
Ethyl 2,3-di-*O*-benzoyl-6-*O*-benzyl-1-thio-β-D-glucopyranoside (1). Ethyl 2,3-di-*O*-benzoyl-



4,6-*O*-benzylidene-1-thio-β-D-glucopyranoside (3.33 g; 6.40 mmol) was dissolved in DCM (64 mL). The mixture was cooled down to −78°C in an acetone–dry-ice bath. Under vigorous stirring, Et_3SiH (3.07 mL; 2.23 g; 19.2 mmol) and TfOH (2.35 mL; 2.88 g; 19.2 mmol) were consecutively added by syringe. The reaction mixture was stirred for 90 min at −78°C, while monitored by TLC (toluene/EtOAc, 9:1). Once complete, Et_3N (8 mL) and MeOH (8 mL) were added. The mixture was then concentrated under reduce pressure and purified by flash column chromatography (toluene/EtOAc, 19:1). The procedure yielded the desired product **1** (2.70 g; 6.40 mmol; 81%). R_f(toluene/EtOAc, 9:1) = 0.42. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.98 – 7.93 (m, 4H; *H*C_{Ar}), 7.53 – 7.48 (m, 2H; *H*C_{Ar}), 7.39 – 7.33 (m, 8H; *H*C_{Ar}), 7.33 – 7.28 (m, 1H; *H*C_{Ar}), 5.51 – 5.43 (m, 2H; H-2, H-3), 4.70 (d, *J* = 9.3 Hz, 1H; H-1), 4.64 (d, *J* = 11.9 Hz, 1H; C*H*HPh), 4.60 (d, *J* = 11.9 Hz, 1H; CH*H*Ph), 3.98 (dd, *J* = 9.1, 9.1 Hz, 1H; H-4), 3.89 – 3.83 (m, 2H; H-6), 3.72 (ddd, *J* = 4.4, 4.9, 9.5 Hz, 1H; H-5), 3.23 (br, 1H; *H*O), 2.81 – 2.69 (m, 2H; CH_2S), 1.26 (t, *J* = 7.5 Hz, 3H; CH_3CH$_2$). ^{13}C NMR (151 MHz, $CDCl_3$) δ 167.2 (*C*=O), 165.5 (*C*=O), 137.8, 133.5, 133.4, 130.1, 130.0, 129.5, 129.2, 129.2, 128.6, 128.5, 128.5, 128.0, 127.9 [aromatic signals], 83.7 (C-1), 78.8

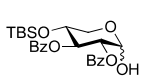
(C-5), 77.8 (C-3), 74.0 (*C*H_2Ph), 71.3 (C-4), 70.4 [C-2 or C-6], 70.3 [C-2 or C-6], 24.4 (*C*H_2S), 15.1 (*C*H_3CH$_2$). HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{29}H_{30}O_7S$: 567.1689; found: 567.1701.

4-Methylphenyl 2,3-di-*O*-benzoyl-4-*O*-(*tert*-butyl)(dimethyl)silyl-1-thio-β-D-xylopyranoside (2). 4-methylphenyl 2,3-di-*O*-benzoyl-1-thio-β-D-xylopyranoside (2.00 g; 4.31 mmol)



was dissolved in DMF (43 mL). The mixture was cooled down to 0°C in a water-ice bath. TBSCl (973 mg; 6.46 mmol) and imidazole (733 mg; 10.8 mmol) were added. The reaction mixture was stirred for 20 h while allowing to warm up to room temperature (RT). Once complete according to TLC (hexane/EtOAc, 9:1), sat. NH_4Cl(aq) (50 mL) was added, and the mixture was extracted with DCM (3 × 100 mL). The organic layers were combined, washed with brine (50 mL), dried over $MgSO_4$, concentrated under reduced pressure, and purified by flash column chromatography (hexane/EtOAc, 19:1). The desired product **2** was obtained (1.41 g; 2.43 mmol; 94%). R_f(hexane/EtOAc, 9:1) = 0.34. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.97 – 7.92 (m, 4H; *H*C$_{Ar}$), 7.53 – 7.47 (m, 2H; *H*C$_{Ar}$), 7.40 – 7.34 (m, 6H; *H*C$_{Ar}$), 7.12 – 7.09 (m, 2H; *H*C$_{Ar}$), 5.51 (dd, *J* = 9.3, 8.3 Hz, 1H; H-3), 5.31 (dd, *J* = 9.3, 9.3 Hz, 1H; H-2), 4.86 (d, *J* = 9.3 Hz, 1H; H-1), 4.12 (dd, *J* = 11.6, 5.3 Hz, 1H; H-5a), 3.99 (ddd, *J* = 9.9, 8.3, 5.3 Hz, 1H; H-4), 3.46 (dd, *J* = 11.6, 9.9 Hz, 1H; H-5b), 2.33 (s, 3H; C*H*$_3$C$_{Ar}$), 0.75 (s, 9H; (C*H*$_3$)$_3$C), 0.01 (s, 3H; C*H*$_3$Si), −0.14 (s, 3H; C*H*$_3$Si). ^{13}C NMR (151 MHz, $CDCl_3$) δ 165.7 (*C*=O), 165.5 (*C*=O), 138.5, 133.4, 133.3, 133.2, 130.0, 129.9, 128.5, 128.4 [aromatic signals], 87.6 (C-1), 76.6 (C-3), 70.8 (C-2), 70.1 (C-5), 69.2 (C-4), 25.6 ((*C*H$_3$)$_3$C), 21.3 (*C*H$_3$C$_{Ar}$), 17.9 (*C*(CH$_3$)$_3$), −4.6 (*C*H$_3$Si), −4.9 (*C*H$_3$Si). HRMS (ESI/Q-TOF) m/z: $[M-H]^-$ Calcd for $C_{32}H_{38}O_6SSi$: 577.2080; found: 577.2127.

2,3-Di-*O*-benzoyl-4-*O*-(*tert*-butyl)(dimethyl)silyl-D-xylopyranose (3). TBS-containing thi-



oxyloside **2** (2.24 g; 3.88 mmol) was dissolved in DCM (35.2 mL) and H_2O (3.5 mL) was added. The reaction mixture was cooled down to 0°C in an ice-water bath. NIS (872 mg; 3.88 mmol) was added. Under vigorous stirring, TFA (297 µL; 442 mg; 3.88 mmol) was added. The reaction was allowed to stir for 4 h at 0°C, monitored by TLC and, when complete, quenched by the addition of Et_3N (0.3 mL). Then, it was diluted with EtOAc (100 mL), washed with $Na_2S_2O_3$ (aq) (3×75 mL) and brine (50 mL), dried over

$MgSO_4$ and concentrated under reduce pressure. The mixture was then purified by flash column chromatography (toluene/EtOAc, 20:1). The procedure yielded the desired hemi-acetal **3** (1.76 g; 3.73 mmol; 96%; α/β, 56:44). R_f(Hex/EtOAc, 2:1) = 0.16. ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.01 – 7.92, (m, 7H; *H*C$_{Ar}$), 7.51 – 7.45, (m, 4H; *H*C$_{Ar}$), 7.39 – 7.32, (m, 7H; *H*C$_{Ar}$), 5.88 (dd, *J* = 8.7, 10.0 Hz, 1H; H-3–α), 5.62 – 5.57, (m, 2H; H-3–β; H-1–α), 5.16 – 5.10, (m, 2H; H-2–β; H-2–α), 4.85 (d, *J* = 7.7 Hz, 0.8H; H-1–β), 4.09 – 3.96, (m, 5H; H-4–α; H-4–β; H-5a–α; H-5a–β; O*H*), 3.74 – 3.69, (m, 1H; H-5b–α), 3.47 (dd, *J* = 10.3, 11.4 Hz, 0.8H; H-5b–β), 3.23 (s, 1H; O*H*), 0.77 (s, 9H; (C*H*$_3$)$_3$C–α), 0.76 (s, 6H; (C*H*$_3$)$_3$C–β), 0.05 (s, 3H; C*H*$_3$Si–α), 0.04 (s, 2H; C*H*$_3$Si–β), −0.10 (s, 3H; C*H*$_3$Si–α), −0.11 (s, 2H; C*H*$_3$Si–β). ^{13}C NMR (151 MHz, $CDCl_3$) δ 167.2, 166.2, 165.8, 165.8 [carbonyl signals], 133.6, 133.4, 133.3, 133.1, 130.2, 130.1, 130.0, 130.0, 129.8, 129.8, 129.7, 129.3, 129.0, 128.8, 128.5, 128.5, 128.4 [aromatic signals], 96.7 (C-1-β), 90.8 (C-1-α), 75.1 (C-3-β), 74.5 (C-2-β), 72.9 (C-3-α), 72.5 (C-2-α), 69.7 (C-4-α), 69.6 (C-4-β), 66.8 (C-5-β), 62.4 (C-5-α), 25.6 ((*C*H$_3$)$_3$C), 25.5 ((*C*H$_3$)$_3$C), 17.9 (*C*$_q$(CH$_3$)$_3$), 17.9 (*C*$_q$(CH$_3$)$_3$), −4.6 (*C*H$_3$Si), −4.6 (*C*H$_3$Si), −4.8 (*C*H$_3$Si), −4.9 (*C*H$_3$Si). HRMS (ESI/Q-TOF) m/z: [M+HCOO]$^-$ Calcd for $C_{25}H_{32}O_7Si$: 517.1894; found: 517.1922.

2,2,2-Trifluoro-*N*-phenylacetimidoyl 2,3-di-*O*-benzoyl-4-*O*-(*tert*-butyl)(dimethyl)silyl-β-D-xylopyranoside (4). Protected xylose derivative **3** (1.74 g; 3.68 mmol) was dissolved in acetone (37 mL) and the solution was cooled down to 0°C in an ice-water bath. Cs_2CO_3 (2.40 g; 7.35 mmol) was added and then, under vigorous stirring, 2,2,2-trifluoro-*N*-phenylethanimidoyl chloride (1.18 mL; 1.53 g; 7.35 mmol) was added. The reaction mixture was stirred for 3 h while allowing to warm up to RT. It was monitored by TLC (toluene/EtOAc, 9:1). Once complete, it was filtered over a pad of Celite® and concentrated under reduced pressure. The mixture was then purified by flash column chromatography (hexane/EtOAc, 19:1; for packing the column, 0.1% Et_3N in the eluent was used). The procedure yielded the desired product (2.21 g; 3.43 mmol; 93%). R_f(toluene/EtOAc, 9:1) = 0.61. ^{1}H NMR (600 MHz, CD_2Cl_2) δ 8.02 – 7.98 (m, 2H; *H*C$_{Ar}$), 7.96 – 7.92 (m, 2H; *H*C$_{Ar}$), 7.59 – 7.51 (m, 2H; *H*C$_{Ar}$), 7.44 – 7.38 (m, 4H; *H*C$_{Ar}$), 7.19 – 7.14 (m, 2H; *H*C$_{Ar}$), 7.06 – 7.01 (m, 1H; *H*C$_{Ar}$), 6.66 (br, 1H; H-1), 6.48 (br, 2H; *H*C$_{Ar}$), 5.85 (dd, *J* = 9.4, 10.1 Hz, 1H; H-3), 5.39 (d, *J* = 8.4 Hz, 1H; H-2), 4.21 – 4.13 (m, 1H; H-4), 3.96 – 3.85 (m, 1H; H-5),

0.79 (s, 9H; (CH_3)$_3$C), 0.09 (s, 3H; CH_3Si), −0.07 (s, 3H; CH_3Si). ^{13}C NMR (151 MHz, CD_2Cl_2) δ 165.9 (*C*=O), 165.8 (*C*=O), 143.7 (*C*=N), 134.0, 133.5, 130.2, 130.1, 129.8, 129.3, 129.0, 129.0, 128.8, 126.7, 124.6, 121.0, 119.5 [aromatic signals], 93.4 (br; C-1), 73.1 (C-3), 71.1 (C-2), 69.3 (C-4), 65.2 (C-5), 25.6 ((*C*H$_3$)$_3$C), 18.1 (*C*$_q$(CH$_3$)$_3$), −4.6 (*C*H$_3$Si), −4.8 (*C*H$_3$Si).

Ethyl 2',3'-di-*O*-benzoyl-4'-*O*-(*tert*-butyl)(dimethyl)silyl-β-D-xylopyranosyl-(1'→4)-2,3-di-*O*-benzoyl-6-*O*-benzyl-1-thio-β-D-glucopyranoside (5). 2,2,2-Trifluoro-*N*-phenylacetimidoyl xyloside **4** (2.25 mg; 3.49 mmol) and ethyl thioglucoside **1** (2.19 g; 4.19 mmol) were dissolved in DCM (70 mL). Freshly activated 4 Å molecular sieves were added. The mixture was cooled down to −78°C in an acetone–dry-ice bath. Under vigorous stirring, TMSOTf (63 µL; 77.6 mg; 349 µmol) was added. The reaction was allowed to warm up to −20°C over 90 min. Once complete according to TLC (hexane/EtOAc, 4:1), the reaction was quenched with Et$_3$N (0.5 mL). The mixture was then concentrated under reduce pressure and purified by flash column chromatography (hexane/EtOAc, 5:1). The procedure yielded the desired product **5** (2.44 g; 2.50 mmol; 71%). *R*$_f$(Hex/EtOAc, 4:1) = 0.36. ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.03 – 7.99 (m, 2H; *H*C$_{Ar}$), 7.97 – 7.93 (m, 2H; *H*C$_{Ar}$), 7.90 – 7.85 (m, 4H; *H*C$_{Ar}$), 7.53 – 7.44 (m, 5H; *H*C$_{Ar}$), 7.44 – 7.31 (m, 13H; *H*C$_{Ar}$), 7.30 – 7.26 (m, 2H; *H*C$_{Ar}$), 5.66 (dd, *J* = 9.5, 9.1 Hz, 1H; H-3), 5.39 (dd, *J* = 10.0, 9.5 Hz, 1H; H-2), 5.30 (dd, *J* = 9.9, 8.8 Hz, 1H; H-3'), 5.17 (dd, *J* = 9.9, 7.7 Hz, 1H; H-2'), 4.60 (dd, *J* = 10.0 Hz, 1H; H-1), 4.57 (dd, *J* = 7.7 Hz, 1H; H-1'), 4.41 (d, *J* = 12.0 Hz, 1H; C*H*HPh), 4.28 (d, *J* = 12.0 Hz, 1H; CH*H*Ph), 4.11 (dd, *J* = 9.1, 8.5 Hz, 1H; H-4), 3.60 (ddd, *J* = 10.1, 8.8, 5.4 Hz, 1H; H-4'), 3.52 (m, 3H; H-5, H-6a, H-6b), 3.00 (dd, *J* = 11.8, 5.4 Hz, 1H; H-5'a), 2.87 (dd, *J* = 11.8, 10.1 Hz, 1H; H-5'b), 2.69 (m, 2H; C*H*$_2$S), 1.21 (t, *J* = 7.4 Hz, 3H; C*H*$_3$CH$_2$S), 0.65 (s, 9H; (C*H*$_3$)$_3$C), −0.23 (s, 3H; C*H*$_3$Si), −0.29 (s, 3H; C*H*$_3$Si). ^{13}C NMR (151 MHz, $CDCl_3$) δ 165.7, 165.7, 165.5, 165.1 [carbonyl signals], 138.2, 133.3, 133.3, 133.1, 132.8, 130.6, 130.1, 129.9, 129.9, 129.8, 129.6, 129.5, 128.6, 128.6, 128.5, 128.4, 128.2, 128.0, 127.9 [aromatic signals], 101.8 (C-1'), 83.7 (C-1), 78.8 (C-5), 76.8 (C-4), 75.6 (C-3'), 75.0 (C-3), 73.4 (*C*H$_2$Ph), 72.1 (C-2'), 70.8 (C-2), 69.2 (C-4'), 67.9 (C-6), 65.8 (C-5'), 25.5 ((*C*H$_3$)$_3$C), 24.3 (*C*H$_2$S), 17.8 (*C*$_q$(CH$_3$)$_3$), 15.0 (*C*H$_3$CH$_2$S), −4.9 (*C*H$_3$Si), −5.1 (*C*H$_3$Si). HRMS (ESI/Q-TOF) m/z: [M+HCOO]$^-$ Calcd for $C_{54}H_{60}O_{13}SSi$: 1021.3501; found: 1021.3494.

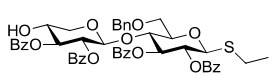
5-Azidopentyl 2',3'-di-*O*-benzoyl-4'-*O*-(*tert*-butyl)(dimethyl)silyl-β-D-xylopyranosyl-(1'→4)-2,3-di-*O*-benzoyl-6-*O*-benzyl-β-D-glucopyranoside (6). Disaccharide **5** (23.8 mg; 22.9 µmol) and 5-azidopentanol (4.4 mg; 34.4 µmol) were dissolved in DCM (500 µL). Freshly activated 4 Å molecular sieves were added and the reaction was stirred for 30 min. NIS (5.7 mg; 25.2 µmol) was added and the mixture was cooled down to −60°C in an acetone–dry-ice bath. Under vigorous stirring, TfOH (0.3 µL; 0.3 mg; 2.3 µmol) was added. The reaction was allowed to warm up to −10°C over 90 min. Once complete according to TLC (hexane/EtOAc, 3:1), the reaction was quenched with Et_3N (50 µL). Then, it was diluted with EtOAc (7 mL), washed with $Na_2S_2O_3$ (aq) (3×7 mL) and brine (4 mL), dried over $MgSO_4$ and concentrated under reduce pressure. The mixture was purified by flash column chromatography (hexane/EtOAc, 9:1 to 5:1). The procedure yielded the desired product **6** (19.3 mg; 18.5 µmol; 81%). R_f(Hex/EtOAc, 5:1) = 0.23. ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.05 – 8.01 (m, 2H; *H*C$_{Ar}$), 7.98 – 7.94 (m, 2H; *H*C$_{Ar}$), 7.91 – 7.85 (m, 4H; *H*C$_{Ar}$), 7.54 – 7.43 (m, 4H; *H*C$_{Ar}$), 7.43 – 7.37 (m, 6H; *H*C$_{Ar}$), 7.37 – 7.30 (m, 5H; *H*C$_{Ar}$), 7.30 – 7.25 (m, 2H; *H*C$_{Ar}$), 5.65 (dd, *J* = 9.1, 9.8 Hz, 1H; H-3), 5.35 (dd, *J* = 8.1, 9.8 Hz, 1H; H-2), 5.29 (dd, *J* = 8.9, 9.8 Hz, 1H; H-3'), 5.18 (dd, *J* = 7.7, 9.8 Hz, 1H; H-2'), 4.56 (d, *J* = 8.0 Hz, 1H; H-1), 4.55 (d, *J* = 7.7 Hz, 1H; H-1'), 4.43 (d, *J* = 12.1 Hz, 1H; C*H*HPh), 4.28 (d, *J* = 12.1 Hz, 1H; CH*H*Ph), 4.09 (dd, *J* = 9.1, 9.4 Hz, 1H; H-4), 3.85 (ddd, *J* = 5.8, 6.0, 10.0 Hz, 1H; C*H*$_2$O), 3.61 (ddd, *J* = 10.2, 8.9, 5.4 Hz, 1H; H-4'), 3.57 – 3.48 (m, 3H; H-5, H-6'a, H-6'b), 3.41 (ddd, *J* = 5.8, 7.1, 9.8 Hz, 1H; C*H*$_2$O), 3.01 (dd, *J* = 11.7, 5.4 Hz, 1H; H-5'a), 2.98 – 2.88 (m, 2H, CH$_{2,agly.}$N$_3$), 2.87 (dd, *J* = 11.7, 10.2 Hz, 1H; H-5'b), 1.57 – 1.43 (m, 2H; C*H*$_{2,agly.}$), 1.43 – 1.32 (m, 2H; C*H*$_{2,agly.}$), 1.32 – 1.16 (m, 2H; C*H*$_{2,agly.}$), 0.65 (s, 9H; (C*H*$_3$)$_3$C), −0.22 (s, 3H; C*H*$_3$Si), −0.28 (s, 3H; C*H*$_3$Si). ^{13}C NMR (151 MHz, $CDCl_3$) δ 165.7, 165.7, 165.3, 165.1 [carbonyl signals], 138.1, 133.3, 133.2, 133.1, 132.8, 130.6, 129.9, 129.8, 129.5, 128.6, 128.5, 128.5, 128.4, 128.2, 128.0, 127.9 [aromatic signals], 101.8, 101.2 [C-1, C-1'], 77.2 (C-4), 75.6 (C-3'), 74.6 (C-5), 73.8 (C-3), 73.4 (*C*H$_2$Ph), 72.1 (C-2'), 72.0 (C-2), 69.6 (*C*H$_{2,agly.}$O), 69.2 (C-4'), 67.8 (C-6), 65.8 (C-5'), 51.2 (*C*H$_{2,agly.}$N$_3$), 29.0 (*C*H$_{2,agly.}$), 28.5 (*C*H$_{2,agly.}$), 25.5 ((*C*H$_3$)$_3$C), 23.2 (*C*H$_{2,agly.}$), 17.8 (*C*$_q$(CH$_3$)$_3$), −4.9 (*C*H$_3$Si), −5.1 (*C*H$_3$Si). HRMS (ESI/Q-TOF) m/z: [M+HCOO]$^-$ Calcd for $C_{57}H_{65}N_3O_{14}Si$: 1088.4212; found: 1088.4208.

5-Azidopentyl 2^{II},3^{II}-di-*O*-benzoyl-β-D-xylopyranosyl-(1^{II}→4^{I})-2^{I},3^{I}-di-*O*-benzoyl-6^{I}-*O*-benzyl-β-D-glucopyranoside (7). TBS-containing disaccharide **6** (9.0 mg; 8.6 μmol) was dissolved in THF (172 μL). The mixture was cooled down to 0°C in an ice/water bath. Under vigorous stirring, acetic acid (1.0 μL; 1.0 mg; 17.2 μmol) and TBAF (1 M in DCM; 25.9 μL; 25.9 μmol) were added. The reaction mixture was stirred for 16h while allowing to warm up to RT. Once complete according to TLC (toluene/EtOAc, 9:1), the reaction was diluted in THF (0.30 mL). $CaCO_3$ (10.4 mg; 103 μmol) and Dowex® 50W X8 (31 mg) were added, and the mixture was stirred for 1 h. After filtering all solids out, the mixture was concentrated under reduce pressure and purified by flash column chromatography (toluene/EtOAc, 9:1). The procedure yielded the desired disaccharide **7** (8.2 mg; 7.86 μmol; 91%). R_f(hexane/EtOAc, 2:1) = 0.16. ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.02 – 7.97 (m, 2H; *H*C$_{Ar}$), 7.97 – 7.89 (m, 6H; *H*C$_{Ar}$), 7.56 – 7.47 (m, 4H; *H*C$_{Ar}$), 7.42 – 7.35 (m, 10H; *H*C$_{Ar}$), 7.34 – 7.28 (m, 3H; *H*C$_{Ar}$), 5.66 (dd, *J* = 9.1, 9.7 Hz, 1H; H-3'), 5.34 (dd, *J* = 8.0, 9.7 Hz, 1H; H-2'), 5.27 (dd, *J* = 6.8, 8.8 Hz, 1H; H-2''), 5.06 (dd, *J* = 7.9, 8.8 Hz, 1H; H-3''), 4.69 (d, *J* = 6.8 Hz, 1H; H-1''), 4.57 (d, *J* = 8.0 Hz, 1H; H-1'), 4.52 (d, *J* = 12.1 Hz, 1H; *C*H$_2$Ph), 4.35 (d, *J* = 12.1 Hz, 1H; *C*H$_2$Ph), 4.17 (dd, *J* = 9.1, 9.7 Hz, 1H; H-4'), 3.85 (ddd, *J* = 6.0, 6.0, 9.9 Hz, 1H; C*H*HO), 3.69 (ddd, *J* = 4.9, 7.9, 8.4 Hz, 1H; H-4''), 3.66 (dd, *J* = 3.5, 10.8 Hz, 1H; H-6'a), 3.56 (dd, *J* = 1.4, 10.8 Hz, 1H; H-6'b), 3.52 (ddd, *J* = 1.4, 3.5, 9.7 Hz, 1H; H-5'), 3.48 (dd, *J* = 4.9, 12.0 Hz, 1H; H-5''a), 3.42 (ddd, *J* = 5.7, 7.2, 9.8 Hz, 1H; CH*H*O), 3.01 (dd, *J* = 8.4, 12.0 Hz, 1H; H-5''b), 2.98 – 2.88 (m, 2H; C*H*$_2$N$_3$), 1.54 – 1.43 (m, 2H; C*H*$_{2,agly.}$), 1.43 – 1.35 (m, 2H; C*H*$_{2,agly.}$), 1.28 – 1.17 (m, 2H; C*H*$_{2,agly.}$). ^{13}C NMR (151 MHz, $CDCl_3$) δ 167.5, 165.8, 165.30, 165.0 [carbonyl signals], 138.2, 133.8, 133.6, 133.3, 133.1, 130.2, 130.1, 129.9, 129.8, 129.7, 129.3, 128.9, 128.7, 128.7, 128.6, 128.5, 128.4, 128.0, 127.9 [aromatic signals], 101.2, 101.1 [C-1; C-1'], 76.6, 76.6, 74.9, 73.7, 73.5 (*C*H$_2$Ph), 72.2, 71.0, 69.6 (*C*H$_{2,agly.}$O), 69.0, 67.8 (C-6'), 64.7, 51.3 (*C*H$_2$N$_3$), 29.0 (*C*H$_{2,agly.}$), 28.5 (*C*H$_{2,agly.}$), 23.2 (*C*H$_{2,agly.}$). HRMS (ESI/Q-TOF) m/z: [M+HCOO]$^-$ Calcd for $C_{51}H_{51}N_3O_{14}$: 974.3348; found: 974.3355.

5-Azidopentyl 2^{IV},3^{IV}-di-*O*-benzoyl-β-D-xylopyranosyl-(1^{IV}→4^{III})-2^{III},3^{III}-di-*O*-benzoyl-6^{III}-*O*-benzyl-β-D-glucopyranosyl-(1^{III}→4^{II})-2^{II},3^{II}-di-*O*-benzoyl-β-D-xylopyranosyl-(1^{II}→4^{I})-2^{I},3^{I}-di-*O*-benzoyl-6^{I}-*O*-benzyl-β-D-glucopyranoside (9). Disaccharide glycosyl donor **5** (15.2 mg; 14.6 µmol) and disaccharide glycosyl acceptor **7** (10.9 mg; 11.7 µmol) were dissolved in DCM (730 µL). Freshly activated 4 Å molecular sieves were added. NIS (3.6 mg; 16.1 µmol) was added and the mixture was cooled down to −60°C in an acetone–dry-ice bath. Under vigorous stirring, TfOH (0.2 µL; 0.2 mg; 1.46 µmol) was added. The reaction mixture was allowed to warm up to −5°C over 60 min. Once complete according to TLC (toluene/EtOAc, 9:1), the reaction was quenched with Et_3N (50 µL). Then, it was diluted with EtOAc (5 mL), washed with $Na_2S_2O_3$ (aq) (3 × 5 mL) and brine (3 mL), dried over $MgSO_4$ and concentrated under reduce pressure. The mixture was purified by flash column chromatography (toluene/EtOAc, 19:1). The procedure yielded the desired tetrasaccharide **8** (13.0 mg; 7.05 µmol; 60%). R_f(Tol/EtOAc, 9:1) = 0.39. Tetrasaccharide **8** (13.0 mg; 7.0 µmmol) was dissolved in THF (705 µL). The mixture was cooled down to 0°C in an ice/water bath. Under vigorous stirring, acetic acid (1.2 µL; 1.3 mg; 21.1 µmol) and TBAF (1M in DCM; 28.2 µL; 28.2 µmol) were added. The reaction mixture was stirred for 16h while allowing to warm up to RT. Once complete according to TLC (toluene/EtOAc, 5:1), the reaction mixture was diluted in THF (1.1 mL). $CaCO_3$ (5.6 mg; 56 µmol) and Dowex® 50W X8 (16.9 mg) were added, and the mixture was stirred for 1 h. After filtering all solids off, the mixture was concentrated under reduced pressure and purified by flash column chromatography (toluene/EtOAc, 9:1). The procedure yielded the desired tetrasaccharide **9** (9.7 mg; 5.6 µmol; 80%). R_f(Tol/EtOAc, 5:1) = 0.22. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.95 – 7.80 (m, 15H; *H*C$_{Ar}$), 7.60 – 7.46 (m, 6H; *H*C$_{Ar}$), 7.45 – 7.41 (m, 2H; *H*C$_{Ar}$), 7.39 – 7.33 (m, 12H; *H*C$_{Ar}$), 7.33 – 7.28 (m, 5H; *H*C$_{Ar}$), 7.25 – 7.21 (m, 2H; *H*C$_{Ar}$), 7.21 – 7.13 (m, 6H; *H*C$_{Ar}$), 5.57 (dd, *J* = 9.1, 9.7 Hz, 1H; H-3^{I}), 5.47 (dd, *J* = 9.1, 9.4 Hz, 1H; H-3^{III}), 5.33 (dd, *J* = 8.6, 9.0 Hz, 1H; H-3^{II}), 5.27 (dd, *J* = 8.0, 9.7 Hz, 1H; H-2^{I}), 5.16 (dd, *J* = 7.0, 8.9 Hz, 1H; H-2^{IV}), 5.13 (dd, *J* = 7.9, 9.4 Hz, 1H; H-2^{III}), 5.08 (dd, *J* = 7.3, 9.0 Hz, 1H; H-2^{II}), 4.96 (dd, *J* = 8.0, 8.9 Hz, 1H; H-3^{IV}), 4.51 (d, *J* = 8.0 Hz, 1H; H-1^{I}), 4.49 (d, *J* = 7.0 Hz, 1H; H-1^{IV}), 4.44 (d, *J* = 12.1 Hz, 1H; C*H*HPh), 4.43 (d, *J* = 7.3 Hz, 1H; H-1^{II}), 4.37 (d, *J* = 7.9 Hz, 1H; H-1^{III}), 4.26 (d, *J* = 12.1 Hz, 1H; CH*H*Ph), 4.19 (d, *J* = 11.9 Hz,

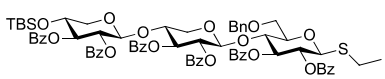
1H; C*H*HPh), 4.02 (dd, *J* = 9.1, 9.8 Hz, 1H; H-4^{I}), 3.99 (dd, *J* = 9.1, 9.7 Hz, 1H; H-4^{III}), 3.91 (d, *J* = 11.9 Hz, 1H; CH*H*Ph), 3.82 (ddd, *J* = 5.9, 5.9, 10.0 Hz, 1H; C*H*H$_{agly.}$O), 3.67 (ddd, *J* = 5.3, 8.4, 9.6 Hz, 1H; H-4^{II}), 3.62 (dd, *J* = 8.0, 14.2 Hz, 1H; H-4^{IV}), 3.53 (dd, *J* = 3.5, 10.6 Hz, 1H; H-6^{I}a), 3.47 (dd, *J* = 1.6, 10.6 Hz, 1H; H-6^{I}b), 3.43 (ddd, *J* = 1.6, 3.5, 9.8 Hz, 1H; H-5^{I}), 3.41 – 3.34 (m, 2H; CH*H*O; H-5^{IV}a), 3.31 (dd, *J* = 3.3, 11.2 Hz, 1H; H-6^{III}a), 3.16 (ddd, *J* = 1.6, 3.3, 9.7 Hz, 1H; H-5^{III}), 3.12 (dd, *J* = 5.3, 12.1 Hz, 1H; H-5^{II}a), 3.01 (dd, *J* = 1.6, 11.2 Hz, 1H; H-6^{III}b), 2.97 – 2.86 (m, 3H; H-5^{IV}b; C*H*HN$_3$; CH*H*N$_3$), 2.82 (br, 1H; *H*O), 2.79 (dd, *J* = 9.6, 12.1 Hz, 1H; H-5^{II}b), 1.53 – 1.40 (m, 2H; C*H*$_{2,agly.}$), 1.40 – 1.32 (m, 2H; C*H*$_{2,agly.}$), 1.24 – 1.14 (m, 2H; C*H*$_{2,agly.}$). ^{13}C NMR (151 MHz, $CDCl_3$) δ 167.5, 165.7, 165.3, 165.1, 165.0, 164.9, 164.9 [carbonyl signals], 138.2, 138.0, 133.8, 133.5, 133.4, 133.3, 133.2, 133.0, 132.9, 130.3, 130.1, 130.0, 129.9, 129.8, 129.8, 129.8, 129.7, 129.7, 129.5, 129.4, 129.2, 128.9, 128.7, 128.6, 128.6, 128.5, 128.5, 128.3, 128.2, 128.2, 128.1, 128.0, 127.9 [aromatic signals], 101.2, 101.1, 100.9, 100.6 [anomeric signals], 76.8 (C-3^{IV}), 76.5 (C-4^{I}), 75.8 (C-4^{III}), 75.6 (C-4^{II}), 74.9 (C-5^{III}), 74.7 (C-5^{I}), 73.6 (C-3^{I}), 73.5, 73.5, 73.4 [C-3^{III}; *C*H$_2$Ph; *C*H$_2$Ph], 72.9 (C-3^{II}), 72.1, 72.1 [C-2^{I}; C-2^{III}], 71.9 (C-2^{II}), 70.8 (C-2^{IV}), 69.6 (*C*H$_{2,agly.}$O), 69.1 (C-4^{IV}), 67.6, 67.5 [C-6^{I}, C-6^{IV}], 64.7 (C-5^{IV}), 62.6 (C-5^{II}), 51.3 (*C*H$_2$N$_3$), 29.0 (*C*H$_{2,agly.}$), 28.5 (*C*H$_{2,agly.}$), 23.2 (*C*H$_{2,agly.}$). HRMS (ESI/Q-TOF) m/z: [M+HCOO]$^-$ Calcd for $C_{97}H_{91}N_3O_{27}$: 1774.5817; found: 1774.5818.

Ethyl 2^{II},3^{II}-di-*O*-benzoyl-β-D-xylopyranosyl-(1^{II}→4^{I})-2^{I},3^{I}-di-*O*-benzoyl-6^{I}-*O*-benzyl-1^{I}-thio-β-D-glucopyranoside (10). TBS-containing disaccharide **5** (1.27 g; 1.30 mmol) was dissolved in THF (13 mL). The mixture was cooled down to 0°C in an ice/water bath. Under vigorous stirring, acetic acid (220 µL; 235 mg; 3.91 mmol) and TBAF (1 M in DCM; 5.21 mL; 5.21 mmol) were added. The reaction mixture was stirred for 16h while allowing to warm up to RT. Once complete according to TLC (toluene/EtOAc, 9:1), the reaction was diluted in THF (27 mL). $CaCO_3$ (1.04 g; 10.4 mmol) and Dowex® 50W X8 (3.13 g) were added, and the mixture was stirred for 1 h. After filtering all solids out, the mixture was concentrated under reduced pressure and purified by flash column chromatography (toluene/EtOAc, 20:1 to 5:1). The procedure yielded the desired disaccharide **10** (982 mg; 1.14 mmol; 87%). R_f(toluene/EtOAc, 9:1) = 0.23. ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.02 – 7.96 (m, 2H; *H*C$_{Ar}$), 7.96 – 7.89 (m, 6H; *H*C$_{Ar}$), 7.56 – 7.46 (m, 4H;



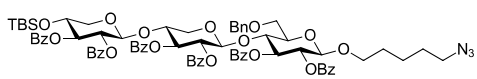
*H*C$_{Ar}$), 7.42 – 7.34 (m, 10H; *H*C$_{Ar}$), 7.34 – 7.28 (m, 3H; *H*C$_{Ar}$), 7.28 – 7.24 (m, 1H; *H*C$_{Ar}$), 5.68 (dd, *J* = 9.2, 9.5 Hz, 1H; H-3'), 5.39 (dd, *J* = 9.5, 10.0 Hz, 1H; H-2'), 5.28 (dd, *J* = 6.7, 8.7 Hz, 1H; H-2''), 5.07 (dd, *J* = 7.9, 8.7 Hz, 1H; H-3''), 4.72 (d, *J* = 6.7 Hz, 1H; H-1''), 4.61 (d, *J* = 10.0 Hz, 1H; H-1'), 4.51 (d, *J* = 12.1 Hz, 1H; C*H*HPh), 4.36 (d, *J* = 12.1 Hz, 1H; CH*H*Ph), 4.19 (dd, *J* = 9.3, 9.6 Hz, 1H; H-4'), 3.71 – 2.66 (m, 1H; H-4''), 3.66 (dd, *J* = 3.5, 11.1 Hz, 1H; H-6'a), 3.59 (dd, *J* = 1.4, 11.1 Hz, 1H; H-6'b), 3.54 (ddd, *J* = 1.4, 3.5, 9.6 Hz, 1H; H-5'), 3.49 (dd, *J* = 4.8, 11.9 Hz, 1H; H-5''a), 3.01 (dd, *J* = 8.3, 12.1 Hz, 1H; H-5''b), 2.89 – 2.87 (m, 1H; OH), 2.76 – 2.64 (m, 2H; C*H*$_2$S), 1.22 (dd, *J* = 7.2, 7.7 Hz, 1H; *C*H$_3$CH$_2$S). ^{13}C NMR (151 MHz, $CDCl_3$) δ 167.5, 165.8, 165.5, 165.0 [carbonyl signals], 138.2, 133.8, 133.6, 133.3, 133.0, 130.2, 130.1, 130.0, 129.9, 129.8, 129.5, 129.3, 129.2, 128.9, 128.7, 128.7, 128.6, 128.5, 128.4, 128.0, 127.9 [aromatic signals], 101.1 (C-1''), 83.6 (C-1'), 79.1 (C-5'), 76.6 (C-3''), 76.5 (C-4'), 74.9 (C-3'), 73.5 (*C*H$_2$Ph), 71.0, 70.9 [C-2'; C-2''], 69.0 (C-4''), 67.9 (C-6'), 64.7 (C-5''), 24.2 (*C*H$_2$S), 15.0 (*C*H$_3$CH$_2$S). HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{48}H_{46}O_{13}S$: 907.2636; found: 907.2648.

Ethyl 2'',3''-di-*O*-benzoyl-4''-*O*-(*tert*-butyl)(dimethyl)silyl-β-D-xylopyranosyl-(1''→4')-2',3'-di-*O*-benzoyl-β-D-xylopyranosyl-(1'→4)-2,3-di-*O*-benzoyl-6-*O*-benzyl-1-thio-β-D-glucopyranoside (11). Monosaccharide glycosyl donor **4** (800 mg; 1.24 mmol) and disaccharide glycosyl acceptor **10** (976 mg; 1.13 mmol) were dissolved in DCM (25 mL). Freshly activated 4 Å molecular sieves were added. The mixture was cooled down to −78°C in an acetone–dry-ice bath. Under vigorous stirring, TMSOTf (23 µL; 27.6 mg; 124 µmol) was added. The reaction mixture was allowed to warm up to −20°C over 90 min. Once complete according to TLC (toluene/EtOAc, 9:1), the reaction mixture was quenched with Et_3N (0.2 mL). The mixture was then concentrated under reduced pressure and purified by flash column chromatography (toluene/EtOAc, 19:1). The procedure yielded the desired trisaccharide **11** (1.02 g; 773 µmol; 62%). R_f(toluene/EtOAc, 19:1) = 0.23. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.96 – 7.89 (m, 6H; *H*C$_{Ar}$), 7.89 – 7.82 (m, 6H; *H*C$_{Ar}$), 7.59 – 7.55 (m, 1H; *H*C$_{Ar}$), 7.54 – 7.45 (m, 4H; *H*C$_{Ar}$), 7.43 – 7.29 (m, 13H; *H*C$_{Ar}$), 7.29 – 7.24 (m, 6H; *H*C$_{Ar}$), 7.17 – 7.12 (m, 2H; *H*C$_{Ar}$), 5.61 (dd, *J* = 9.4, 9.2 Hz, 1H; H-3), 5.38 (dd, *J* = 9.1, 8.6 Hz, 1H; H-3'), 5.32 (dd, *J* = 10.0, 9.4 Hz, 1H; H-2),



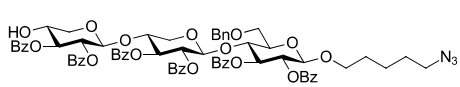
5.28 (dd, *J* = 9.6, 8.7 Hz, 1H; H-3''), 5.14 (dd, *J* = 9.1, 7.3 Hz, 1H; H-2'), 5.07 (dd, *J* = 9.6, 7.5 Hz, 1H; H-2''), 4.57 (dd, *J* = 10.0 Hz, 1H; H-1), 4.53 (d, *J* = 7.3 Hz, 1H; H-1'), 4.43 (d, *J* = 12.0 Hz, 1H; C*H*HPh), 4.35 (d, *J* = 7.5 Hz, 1H; H-1''), 4.28 (d, *J* = 12.0 Hz, 1H; CH*H*Ph), 4.07 (dd, *J* = 9.8, 9.2 Hz, 1H; H-4), 3.70 – 3.62 (m, 2H; H-4', H-4''), 3.56 – 3.46 (m, 3H; H-5, H-6a, H-6b), 3.17 (dd, *J* = 11.9, 5.3 Hz, 1H; H-5''a), 3.07 (dd, *J* = 12.1, 5.4 Hz, 1H; H-5'a), 2.90 (dd, *J* = 11.9, 9.9 Hz, 1H; H-5''b), 2.84 (dd, *J* = 12.1, 9.7 Hz, 1H; H-5'b), 2.67 (m, 2H; CH_2S), 1.20 (t, *J* = 7.4 Hz, 3H; CH_3CH$_2$S), 0.65 (s, 9H; (CH_3)$_3$C), −0.18 (s, 3H; CH_3Si), −0.26 (s, 3H; CH_3Si). ^{13}C NMR (151 MHz, $CDCl_3$) δ 165.6, 165.5, 165.4, 165.3, 164.9, 164.9 [carbonyl signals], 137.9, 133.2, 133.1, 133.0, 132.8, 132.7, 130.1, 130.0, 129.9, 129.9, 129.7, 129.7, 129.6, 129.6, 129.4, 129.4, 129.3, 128.5, 128.4, 128.3, 128.3, 128.2, 128.0, 128.0, 127.9 [aromatic signals], 102.0 (C-1''), 101.1 (C-1'), 83.5 (C-1), 78.7 (C-5), 76.8 (C-4'), 76.3 (C-4), 75.1 (C-3''), 74.7 (C-3), 73.3 (*C*H$_2$Ph), 73.0 (C-3'), 71.7 (C-2''), 71.6 (C-2'), 70.7 (C-2), 68.9 (C-4''), 67.7 (C-6), 65.8 (C-5''), 62.7 (C-5'), 25.4 ((*C*H$_3$)$_3$C), 24.1 (*C*H$_2$S), 17.7 (*C*$_q$(CH$_3$)$_3$), 14.9 (*C*H$_3$CH$_2$S), −5.0 (*C*H$_3$Si), −5.2 (*C*H$_3$Si). HRMS (ESI/Q-TOF) m/z: [M+HCOO]$^-$ Calcd for $C_{73}H_{76}O_{19}SSi$: 1361.4447; found: 1361.4455.

5-Azidopentyl 2'',3''-di-*O*-benzoyl-4''-*O*-(*tert*-butyl)(dimethyl)silyl-β-D-xylopyranosyl-(1''→4')-2',3'- di-*O*-benzoyl-β-D-xylopyranosyl-(1'→4)-2,3-di-*O*-benzoyl-6-*O*-benzyl-β-D-glucopyranoside (12). Trisaccharide **11** (100 mg; 75.9 µmol) and 5-azidopentanol (14.7 mg; 75.9 µmol) were dissolved in DCM (1.5 mL). Freshly activated 4 Å molecular sieves were added. NIS (18.8 mg; 83.5 µmol) was added and the mixture was immediately cooled down to −50°C in an acetone–dry-ice bath. Under vigorous stirring, TfOH (0.9 µL; 1.1 mg; 7.6 µmol) was added. The reaction mixture was allowed to warm up to −10°C over 90 min. Once complete according to TLC (toluene/EtOAc, 9:1), the reaction was quenched with Et_3N (100 µL). Then, it was diluted with EtOAc (6 mL), washed with $Na_2S_2O_3$ (aq) (3×6 mL) and brine (3 mL), dried over $MgSO_4$ and concentrated under reduce pressure. The mixture was purified by flash column chromatography (toluene/EtOAc, 19:1). The procedure yielded the desired trisaccharide **12** (88.2 mg; 63.7 µmol; 84%). R_f(toluene/EtOAc, 9:1) = 0.44. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.97 – 7.89 (m, 6H; *H*C$_{Ar}$), δ 7.90 – 7.83 (m, 6H; *H*C$_{Ar}$), 7.60



– 7.55 (m, 1H; HC_{Ar}), 7.54 – 7.44 (m, 4H; HC_{Ar}), 7.44 – 7.30 (m, 14H; HC_{Ar}), 7.30 –7.22 (m, 12H; HC_{Ar}), 7.18 – 7.13 (m, 2H; HC_{Ar}), 5.59 (dd, *J* = 9.8, 9.1 Hz, 1H; H-3), 5.37 (dd, *J* = 9.2, 8.6 Hz, 1H; H-3'), 5.28 (dd, *J* = 9.6, 8.8 Hz, 1H; H-3''), 5.27 (dd, *J* = 9.8, 8.0 Hz, 1H; H-2), 5.14 (dd, *J* = 9.2, 7.3 Hz, 1H; H-2'), 5.07 (dd, *J* = 9.6, 7.5 Hz, 1H; H-2''), 4.52 (d, *J* = 8.0 Hz, 1H; H-1), 4.49 (d, *J* = 7.3 Hz, 1H; H-1'), 4.45 (d, *J* = 12.1 Hz, 1H; C*H*HPh), 4.35 (d, *J* = 7.5 Hz, 1H; H-1''), 4.26 (d, *J* = 12.1 Hz, 1H; CH*H*Ph), 4.04 (dd, *J* = 9.5, 9.1 Hz, 1H; H-4), 3.82 (ddd, *J* = 9.8, 6.2, 5.7 Hz, 1H; C*H*$H_{agly.}$O), 3.67 (ddd, *J* = 9.6, 8.8, 5.3 Hz, 1H; H-4''), 3.65 (ddd, *J* = 9.8, 8.6, 5.4 Hz, 1H; H-4'), 3.54 (dd, *J* = 10.8, 3.6 Hz, 1H; H-6a), 3.49 (dd, *J* = 10.8, 1.6 Hz, 1H; H-6b), 3.46 (ddd, *J* = 9.5, 3.6, 1.6 Hz, 1H; H-5), 3.38 (ddd, *J* = 9.8, 7.2, 5.8 Hz, 1H; CH*H*$_{agly.}$O), 3.17 (dd, *J* = 11.9, 5.3 Hz, 1H; H-5''a), 3.06 (dd, *J* = 5.4, 12.1 Hz, 1H; H-5'a), 2.92 – 2.87 (m, 2H; C*H*$_{2,agly.}$$N_3$), 2.90 (dd, *J* = 11.9, 9.6 Hz, 1H; H-5'a), 2.84 (dd, *J* = 12.1, 9.8 Hz, 1H; H-5'b), 1.53 – 1.41 (m, 2H; C*H*$_{2,agly.}$), 1.41 – 1.32 (m, 2H; C*H*$_{2,agly.}$), 1.27 – 1.15 (m, 2H; C*H*$_{2,agly.}$), 0.65 (s, 3H; (C*H*$_3$)$_3$C), −0.18 (s, 3H; C*H*$_3$Si), −0.26 (s, 3H; C*H*$_3$Si). ^{13}C NMR (151 MHz, $CDCl_3$) δ 165.7, 165.7, 165.6, 165.3, 165.1, 165.0 [carbonyl signals], 138.0, 133.4, 133.2, 133.1, 133.0, 132.8, 130.3, 130.1, 130.0, 129.9, 129.8, 129.8, 129.8, 129.7, 129.5, 129.5, 128.7, 128.6, 128.5, 128.4, 128.3, 128.2, 128.1, 128.1 [aromatic signals}, 102.1 (C-1''), 101.3, 101.2 [C-1; C-1'], 77.0 (C-4'), 76.6 (C-4), 75.2 (C-3''), 74.7 (C-5), 73.7 (C-3), 73.5 (*C*H_2Ph), 73.2 (C-3'), 72.1 (C-2), 71.8, 71.8 [C-2'; C-2''], 69.6 (*C*H$_{2,agly.}$O), 69.0 (C-4''), 67.6 (C-6), 65.9 (C-5''), 62.9 (C-5'), 51.2 (*C*H$_{2,agly.}$$N_3$), 29.0 (*C*H$_{2,agly.}$), 28.5 (*C*H$_{2,agly.}$), 25.5 ((*C*H$_3$)$_3$C), 23.2 (*C*H$_{2,agly.}$), 17.8 (*C*$_q$(CH$_3$)$_3$), −4.8 (*C*H$_3$Si), −5.0 (*C*H$_3$Si). HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{76}H_{81}N_3O_{20}Si$: 1428.5159; found: 1428.5149.

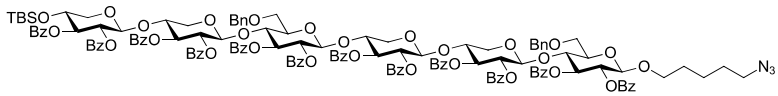
5-Azidopentyl 2'',3''-di-*O*-benzoyl-β-D-xylopyranosyl-(1''→4')-2',3'-di-*O*-benzoyl-β-D-xylopyra nosyl-(1'→4)-2,3-di-*O*-benzoyl-6-*O*-benzyl-β-D-glucopyranoside (13). Trisaccharide



12 (332 mg; 239 µmol) was dissolved in THF (2.4 mL). The mixture was cooled down to 0°C in an ice/water bath. Under vigorous stirring, acetic acid (41 µL; 43.1 mg; 718 µmol) and TBAF (1 M in DCM; 958 µL; 958 µmol) were added. The reaction mixture was stirred for 16h while allowing to warm up to RT. Once complete according to TLC (toluene/EtOAc, 7:1), the reaction mixture was diluted in THF (5.0 mL). $CaCO_3$ (192 mg; 1.92 mmol) and Dowex® 50W

X8 (575 mg) were added, and the mixture was stirred for 1 h. After filtering all solids off, the mixture was concentrated under reduced pressure and purified by flash column chromatography (toluene/EtOAc, 5:1). The procedure yielded the desired trisaccharide **13** (280 mg; 221 µmol; 92%). R_f(Tol/EtOAc, 5:1) = 0.27. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.95 – 7.88 (m, 10H; *H*C$_{Ar}$), 7.63 – 7.58 (m, 1H; *H*C$_{Ar}$), 7.53 – 7.46 (m, 5H; *H*C$_{Ar}$), 7.43 – 7.28 (m, 11H; *H*C$_{Ar}$), 7.28 – 7.23 (m, 12H; *H*C$_{Ar}$), 7.23 – 7.18 (m, 2H; *H*C$_{Ar}$), 5.60 (dd, *J* = 9.5, 9.1 Hz, 1H; H-3), 5.39 (dd, *J* = 9.1, 8.7 Hz, 1H; H-3'), 5.29 (dd, *J* = 9.9, 8.1 Hz, 1H; H-2), 5.15 (dd, *J* = 9.3, 7.3 Hz, 1H; H-2'), 5.14 (dd, *J* = 8.2, 6.0 Hz, 1H; H-2''), 5.09 (dd, *J* = 8.4, 7.1 Hz, 1H; H-3''), 4.53 (d, *J* = 8.0 Hz, 1H; H-1), 4.52 (d, *J* = 7.4 Hz, 1H; H-1'), 4.45 (d, *J* = 12.1 Hz, 1H; C*H*HPh), 4.45 (d, *J* = 6.4 Hz, 1H; H-1''), 4.27 (d, *J* = 12.1 Hz, 1H; CH*H*Ph), 4.06 (dd, *J* = 9.5, 9.1 Hz, 1H; H-4), 3.83 (ddd, *J* = 9.8, 6.1, 5.7 Hz, 1H; C*H*H$_{agly.}$O), 3.68 – 3.74 (m, 2H; H-4', H-4''), 3.60 (dd, *J* = 12.2, 4.6 Hz, 1H; H-5''a), 3.55 (dd, *J* = 10.7, 3.7 Hz, 1H; H-6a), 3.50 (dd, *J* = 10.7, 1.4 Hz, 1H; H-6b), 3.47 (ddd, *J* = 9.9, 3.7, 1.4 Hz, 1H; H-5), 3.39 (ddd, *J* = 9.8, 7.3, 5.7 Hz, 1H; CH*H* $_{agly.}$O), 3.14 (dd, *J* = 12.1, 5.3 Hz, 1H; H-5'a), 3.07 (dd, *J* = 12.2, 7.8 Hz, 1H; H-5''b), 2.97 – 2.84 (m, 3H; C*H*HN$_3$, CH*H*N$_3$, H-5'b), 1.55 – 1.41 (m, 2H; C*H*$_{2,agly.}$), 1.41 – 1.32 (m, 2H; C*H*$_{2,agly.}$), 1.27 – 1.14 (m, 2H; C*H*$_{2,agly.}$). ^{13}C NMR (151 MHz, $CDCl_3$) δ 167.3, 165.7, 165.6, 165.3, 165.1, 164.9 [carbonyl signals], 138.0, 133.8, 133.7, 133.4, 133.2, 133.2, 132.8, 130.3, 130.1, 130.0, 129.9, 129.8, 129.7, 129.4, 129.3, 128.9, 128.7, 128.7, 128.6, 128.5, 128.5, 128.2, 128.1, 128.0 [aromatic signals], 101.3, 101.2 [H-1; H-1'], 100.7 (H-1''), 76.6 (C-4), 76.1 (C-4'), 75.7 (C-3''), 74.7 (C-5), 73.7 (C-3), 73.5 (*C*H$_2$Ph), 73.0 (C-3'), 72.1 (C-2), 71.9 (C-2'), 70.6 (C-2''), 69.7 (*C*H$_{2,agly.}$O), 68.6 (C-4''), 67.7 (C-6), 64.4 (C-5''), 62.8 (C-5'), 51.3 (*C*H$_{2,agly.}$N$_3$), 29.0 (*C*H$_{2,agly.}$), 28.5 (*C*H$_{2,agly.}$), 23.2 (*C*H$_{2,agly.}$). HRMS (ESI/Q-TOF) m/z: [M+HCOO]$^-$ Calcd for $C_{70}H_{67}N_3O_{20}$: 1314.4295; found: 1314.4291.

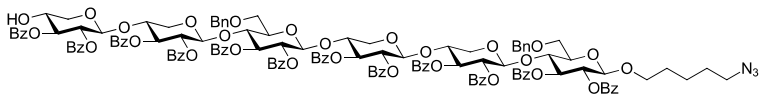
5-Azidopentyl 2^{VI},3^{VI}-di-*O*-benzoyl-4^{VI}-*O*-(*tert*-butyl)(dimethyl)silyl-β-D-xylopyranosyl-(1^{VI}→4^{V})-2^{V},3^{V}-di-*O*-benzoyl-β-D-xylopyranosyl-(1^{V}→4^{IV})-2^{IV},3^{IV}-di-*O*-benzoyl-6^{IV}-*O*-benzyl-β-D-glucopyranosyl-(1^{IV}→4^{III})-2^{III},3^{III}-di-*O*-benzoyl-β-D-xylopyranosyl-(1^{III}→4^{II})-2^{II},3^{II}-di-*O*-benzoyl-β-D-xylopyranosyl-(1^{II}→4^{I})-2^{I},3^{I}-di-*O*-benzoyl-6^{I}-*O*-benzyl-β-D-glucopyranoside (14). Trisaccharide glycosyl donor **11** (30 mg; 22.8 µmol) and trisaccharide glycosyl



acceptor **13** (26 mg; 20.5 µmol) were dissolved in DCM (500 µL). Freshly activated 4 Å molecular sieves were added. NIS (6.1 mg; 27.3 µmol) was added and the mixture was cooled down to −70°C in an acetone–dry-ice bath. Under vigorous stirring, TfOH (0.3 µL; 0.3 mg; 2.3 µmol) was added. The reaction mixture was allowed to warm up to −17°C over 120 min. Once the donor was consumed according to TLC (toluene/EtOAc, 7:1), the reaction was quenched with Et_3N (50 µL). The reaction mixture was diluted with EtOAc (5 mL), washed with $Na_2S_2O_3$ (aq) (3 × 5 mL) and brine (3 mL), dried over $MgSO_4$ and concentrated under reduced pressure. The mixture was purified by flash column chromatography (toluene/EtOAc, 19:1 to 9:1). The procedure yielded the desired hexasaccharide **14** (23.8 mg; 9.42 µmol; 46%). Additionally, acceptor **13** was partially recovered (11.7 mg; 9.2 µmol; 45%). R_f(Tol/EtOAc, 7:1) = 0.35. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.95 – 7.77 (m, 24H; *H*C$_{Ar}$), 7.59 – 7.44 (m, 9H; *H*C$_{Ar}$), 7.44 – 7.26 (m, 27H; *H*C$_{Ar}$), 7.25 – 7.08 (m, 13H; *H*C$_{Ar}$), 5.57 (dd, *J* = 9.2, 9.9 Hz, 1H; H-3^{I}), 5.39 (dd, *J* = 8.2, 8.9 Hz, 1H; H-3), 5.36 (dd, *J* = 7.8, 8.2 Hz, 1H; H-3), 5.31 – 5.24 (m, 4H; H-3; H-3; H-3; H-2^{I}), 5.09 – 5.01 (m, 4H; H-2; H-2; H-2; H-2), 4.95 (dd, *J* = 6.3, 8.6 Hz, 1H; H-2), 4.51 (d, *J* = 8.0 Hz, 1H; H-1^{I}), 4.46 – 4.42 (m, 2H; H-1; C*H*HPh), 4.40 (d, *J* = 7.7 Hz, 1H; H-1), 4.30 (d, *J* = 7.4 Hz, 1H; H-1), 4.28 (d, *J* = 7.5 Hz, 1H; H-1), 4.27 – 4.23 (m, 2H; CH*H*Ph; H-1), 4.14 (d, *J* = 11.8 Hz, 1H; C*H*HPh), 4.01 (dd, *J* = 8.5, 10.3 Hz, 1H; H-4^{I}), 3.88 – 3.79 (m, 3H; CH*H*Ph; H-4; C*H*H$_{agly}$O), 3.70 (ddd, *J* = 5.3, 8.0, 8.0 Hz, 1H; H-4), 3.65 (ddd, *J* = 5.5, 9.2, 9.2 Hz, 1H; H-4^{VI}), 3.60 – 3.53 (m, 2H; H-4*; H-4**), 3.52 (dd, *J* = 3.4, 10.7 Hz, 1H; H-6^{I}a), 3.49 – 3.42 (m, 2H; H-6^{I}b; H-5^{I}), 3.38 (ddd, *J* = 6.5, 6.7, 9.7 Hz, 1H; CH*H*$_{agly}$O), 3.30 (dd, *J* = 4.8, 12.1 Hz, 1H; H-5*a), 3.19 – 3.10 (m, 3H; H-5^{VI}a; H-5^{IV}; H-6^{IV}a), 3.02 (dd, *J* = 5.2, 12.1 Hz, 1H; H-5**a), 2.98 (dd, *J* = 5.4, 12.4 Hz, 1H; H-5***a), 2.95 – 2.84 (m, 5H; H-5*b; H-5^{VI}b; H-6^{IV}b; C*H*H$_{agly.}$N$_3$; CH*H*$_{agly.}$N$_3$), 2.78 (dd, *J* = 10.3, 11.5 Hz, 1H; H-5**b), 2.72 (dd, *J* = 10.3, 11.7 Hz, 1H; H-5***b), 1.54 – 1.40 (m,

2H; C$H_{2,aglγ.}$), 1.40 – 1.32 (m, 2H; C$H_{2,aglγ.}$), 1.32 – 1.14 (m, 2H; C$H_{2,aglγ.}$), 0.64 (s, 9H; (CH_3)$_3$C), −0.20 (s, 3H; CH_3Si), −0.27 (s, 3H; CH_3Si). ^{13}C NMR (151 MHz, $CDCl_3$) δ 165.7, 165.7, 165.6, 165.5, 165.5, 165.3, 165.0, 165.0, 165.0, 164.9 [carbonyl signals], 138.0, 137.9, 133.4, 133.4, 133.2, 133.1, 133.1, 132.9, 132.8, 132.8, 130.3, 130.1, 129.9, 129.9, 129.9, 129.8, 129.8, 129.7, 129.7, 129.5, 129.5, 129.4, 129.4, 128.7, 128.6, 128.5, 128.4, 128.3, 128.1, 128.0, 128.0 [aromatic signals], 102.1, 101.2 (C-1^{I}), 101.1, 101.0, 100.7, 100.4 [anomeric signals], 76.9 [C-4], 76.5 [C-4], 75.9 [C-4], 75.7 [C-4; C-2 or C-3], 75.2 (C-5^{I}; C-5^{IV}), 74.7 [C-2 or C-3], 73.6 [C-2 or C-3], 73.5 [C-2 or C-3], 73.4 (*C*H$_2$Ph), 73.4 (*C*H$_2$Ph), 73.1 [C-2 or C-3], 72.9 [C-2 or C-3], 72.2 [C-2 or C-3], 72.1 [C-2 or C-3], 72.1 [C-2 or C-3], 71.9 [C-2 or C-3], 71.8 [C-2 or C-3], 71.6 [C-2 or C-3], 71.4 [C-2 or C-3], 69.6 (*C*H$_{2,aglγ.}$O), 69.0 (C-4^{VI}), 67.6 [C-6], 67.3 [C-6], 65.9 [C-5], 62.8 [C-5], 62.6 [C-5], 62.4 [C-5], 51.2 (*C*H$_2$N$_3$), 29.0 (*C*H$_{2,aglγ.}$), 28.5 (*C*H$_{2,aglγ.}$), 25.5 ((C*H*$_3$)$_3$C), 23.2 (*C*H$_{2,aglγ.}$), 17.8 (*C*$_q$(CH$_3$)$_3$)), −4.9 (C*H*$_3$Si), −5.0 (C*H*$_3$Si). HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{141}H_{137}N_3O_{39}Si$: 2568.8575; found: 2568.8515.

5-Azidopentyl 2^{VI},3^{VI}-di-*O*-benzoyl-β-D-xylopyranosyl-(1^{VI}→4^{V})-2^{V},3^{V}-di-*O*-benzoyl-β-D-xylopyranosyl-(1^{V}→4^{IV})-2^{IV},3^{IV}-di-*O*-benzoyl-6^{IV}-*O*-benzyl-β-D-glucopyranosyl-(1^{IV}→4^{III})-2^{III},3^{III}-di-*O*-benzoyl-β-D-xylopyranosyl-(1^{III}→4^{II})-2^{II},3^{II}-di-*O*-benzoyl-β-D-xylopyranosyl-(1^{II}→4^{I})-2^{I},3^{I}-di-*O*-benzoyl-6^{I}-*O*-benzyl-β-D-glucopyranoside (15). Hexasaccharide **14** (54



mg; 21.4 µmol) was dissolved in THF (428 µL). The mixture was cooled down to 0°C in an ice/water bath. Under vigorous stirring, acetic acid (3.7 µL; 3.9 mg; 64.1 µmol) and TBAF (1 M in DCM; 85.5 µL; 85.5 µmol) were added. The reaction mixture was stirred for 16h while allowing to warm up to RT. Once complete according to TLC (toluene/EtOAc, 5:1), the reaction was diluted in THF (770 µL). $CaCO_3$ (17.1 mg; 171 µmol) and Dowex® 50W X8 (51.4 mg) were added, and the mixture was stirred for 1 h. After filtering all solids off, the mixture was concentrated under reduce pressure and purified by flash column chromatography (hexane/EtOAc, 60:40 to 55:45). The procedure yielded the desired hexasaccharide **15** (23.9 mg; 9.91 µmol; 46%). R_f(Tol/EtOAc, 5:1) = 0.36. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.95 – 7.77 (m, 25H; *H*C$_{Ar}$), 7.61 – 7.47 (m, 8H; *H*C$_{Ar}$), 7.47 – 7.40 (m, 6H; *H*C$_{Ar}$), 7.40 – 7.27 (m, 22H; *H*C$_{Ar}$), 7.25 – 7.14 (m, 12H;

HC_{Ar}), 7.13 – 7.09 (m, 2H; *H*C_{Ar}), 5.58 (dd, *J* = 9.3, 9.6 Hz, 1H; H-3^{Glu}), 5.41 (dd, *J* = 8.8, 9.5 Hz, 1H; H-2 or H-3), 5.37 (dd, *J* = 7.5, 8.5 Hz, 1H; H-2 or H-3), 5.32 – 5.25 (m, 3H; H-2 or H-3), 5.13 – 5.03 (m, 5H; H-2 or H-3), 4.96 (dd, *J* = 6.3, 8.1 Hz, 1H; H-2 or H-3), 4.52 (d, *J* = 8.0 Hz, 1H; H-1), 4.48 – 4.40 (m, 4H; C*H*HPh; 3×H-1), 4.33 (d, *J* = 7.2 Hz, 1H; H-1), 4.27 (d, *J* = 6.2 Hz, 1H; H-1), 4.26 (d, *J* = 12.1 Hz, 1H; CH*H*Ph), 4.15 (d, *J* = 11.9 Hz, 1H; C*H*HPh), 4.02 (dd, *J* = 9.7, 8.9 Hz, 1H; H-4^{Glu}), 3.89 (d, *J* = 11.9 Hz, 1H; CH*H*Ph), 3.87 (dd, *J* = 9.6, 8.9 Hz, 1H; H-4^{Glu}), 3.82 (ddd, *J* = 5.7, 6.2, 10.0 Hz, 1H; C*H*$H_{agly.}$O), 3.73 – 3.63 (m, 3H; H-4^{VI}; 2×H-4^{Xyl}), 3.60 – 3.51 (m, 3H; H-4^{Xyl}; H-5^{VI}a; H-6), 3.50 – 3.46 (m, 1H; H-6), 3.44 (ddd, *J* = 9.3, 3.0, 1.9 Hz, 1H; H-5^{Glu}), 3.38 (ddd, *J* = 9.9, 7.3, 5.7 Hz, 1H; CH*H*$_{agly.}$O), 3.33 (dd, *J* = 12.2, 4.9 Hz, 1H; H-5^{Xyl}), 3.20 (dd, *J* = 10.9, 3.5 Hz, 1H; H-6), 3.15 (ddd, *J* = 9.8, 3.3, 1.5 Hz, 1H; H-5^{Glu}), 3.10 – 3.02 (m, 3H; H-5^{VI}b; 2×H-5^{Xyl}), 2.98 – 2.88 (m, 4H; 2×C*H*HN_3; H-5^{Xyl}; H-6), 2.78 (ddd, *J* = 16.1; 12.1; 9.7 Hz, 2H; 2×H-5^{Xyl}), 1.55 – 1.41 (m, 2H; C*H*$_{2,agly.}$), 1.41 – 1.33 (m, 2H; C*H*$_{2,agly.}$), 1.24 – 1.14 (m, 2H; C*H*$_{2,agly.}$). ^{13}C NMR (151 MHz, $CDCl_3$) δ 167.2, 165.7, 165.6, 165.5, 165.5, 165.2, 165.0, 165.0, 165.0, 164.9, 164.9, 164.9 [*C*=O signals], 138.0, 138.0, 133.8, 133.6, 133.3, 133.2, 133.1, 133.1, 132.8, 132.8, 130.3, 130.2, 130.1, 129.9, 129.9, 129.9, 129.8, 129.7, 129.7, 129.5, 129.4, 129.4, 129.3, 129.3, 128.9, 128.7, 128.6, 128.6, 128.6, 128.5, 128.5, 128.4, 128.4, 128.3, 128.1, 128.1, 128.0, 127.9 [Ar. signals], 101.2, 101.1, 101.0, 100.7, 100.6, 100.4 [anomeric signals], 76.5 (C-4^{Glu}), 76.1 (C-4^{Xyl}), 75.9 (C-4^{Glu}), 75.7 (C-4^{Xyl}), 75.6 (C-2 or C-3), 75.2 (C-4^{Xyl}), 74.7 (C-5^{Glu}), 74.7 (C-5^{Glu}), 73.6 (C-2 or C-3), 73.5 (C-2 or C-3), 73.4 (*C*H_2Ph), 73.4 (*C*H_2Ph), 73.0, 72.9, 72.2, 72.1, 72.1, 71.9, 71.7, 71.4, 70.6 [C-2; C-3], 69.6 (*C*$H_{2,agly.}$O), 68.5 (C-4^{VI}), 67.6 (C-6), 67.4 (C-6), 64.4 (C-5^{VI}), 62.7 (C-5^{Xyl}), 62.6 (C-5 ^{Xyl}), 62.4 (C-5 ^{Xyl}), 51.2 (*C*H_2N_3), 29.0 (*C*$H_{2,agly.}$), 28.5 (*C*$H_{2,agly.}$), 23.2 (*C*$H_{2,agly.}$). HRMS (ESI/Q-TOF) m/z: $[M+HCOO]^-$ Calcd for $C_{135}H_{123}N_3O_{39}$: 2454.7710; found: 2454.7655.

3.6. References

(1) Ray, B.; Lahaye, M. Cell-Wall Polysaccharides from the Marine Green Alga Ulva "Rigida" (Ulvales, Chlorophyta). Extraction and Chemical Composition. *Carbohydr Res* **1995**, *274*, 251–261. https://doi.org/10.1016/0008-6215(95)00138-J.

(2) Little, A.; Lahnstein, J.; Jeffery, D. W.; Khor, S. F.; Schwerdt, J. G.; Shirley, N. J.; Hooi, M.; Xing, X.; Burton, R. A.; Bulone, V. A Novel (1,4)-β-Linked Glucoxylan Is Synthesized by Members of the *Cellulose Synthase-Like F* Gene Family in Land Plants. *ACS Cent Sci* **2019**, *5*, 73–84. https://doi.org/10.1021/acscentsci.8b00568.

(3) Little, A.; Schwerdt, J. G.; Shirley, N. J.; Khor, S. F.; Neumann, K.; O'Donovan, L. A.; Lahnstein, J.; Collins, H. M.; Henderson, M.; Fincher, G. B.; Burton, R. A. Revised Phylogeny of the *Cellulose Synthase* Gene Superfamily: Insights into Cell Wall Evolution. *Plant Physiol* **2018**, *177* (3), 1124–1141. https://doi.org/10.1104/pp.17.01718.

(4) Richmond, T. A.; Somerville, C. R. The Cellulose Synthase Superfamily. *Plant Physiol* **2000**, *124* (2), 495–498. https://doi.org/10.1104/pp.124.2.495.

(5) Pear, J. R.; Kawagoet, Y.; Schreckengost, W. E.; Delmert, D. P.; Stalker, D. M. *Higher Plants Contain Homologs of the Bacterial CelA Genes Encoding the Catalytic Subunit of Cellulose Synthase*; 1996; Vol. 93. https://www.pnas.org.

(6) Scheller, H. V.; Ulvskov, P. Hemicelluloses. *Annu Rev Plant Biol* **2010**, *61*, 263–289. https://doi.org/10.1146/annurev-arplant-042809-112315.

(7) Yang, J.; Bak, G.; Burgin, T.; Barnes, W. J.; Mayes, H. B.; Peña, M. J.; Urbanowicz, B. R.; Nielsen, E. Biochemical and Genetic Analysis Identify CSLD3 as a Beta-1,4-Glucan Synthase That Functions during Plant Cell Wall Synthesis. *Plant Cell* **2020**, *32* (5), 1749–1767. https://doi.org/10.1105/tpc.19.00637.

(8) Burton, R. A.; Wilson, S. M.; Hrmova, M.; Harvey, A. J.; Shirley, N. J.; Medhurst, A.; Stone, B. A.; Newbigin, E. J.; Bacic, A.; Fincher, G. B. Cellulose Synthase-Like *CslF* Genes Mediate the Synthesis of Cell Wall (1,3;1,4)-ß-D-Glucans. *Science (1979)* **2006**, *311*, 1940–1942. https://doi.org/10.1126/science.1122975.

(9) Doblin, M. S.; Pettolino, F. A.; Wilson, S. M.; Campbell, R.; Burton, R. A.; Fincher, G. B.; Newbigin, E.; Bacic, A. A Barley *Cellulose Synthase-like CSLH* Gene Mediates (1,3;1,4)-β-D-Glucan Synthesis in Transgenic *Arabidopsis*. *Proceedings of the National Academy of Sciences* **2009**, *106* (14), 5996–6001. https://doi.org/10.1073/pnas.0902019106.

(10) Drula, E.; Garron, M. L.; Dogan, S.; Lombard, V.; Henrissat, B.; Terrapon, N. The Carbohydrate-Active Enzyme Database: Functions and Literature. *Nucleic Acids Res* **2022**, *50*, D571–D577. https://doi.org/10.1093/nar/gkab1045.

(11) Hamann, T.; Osborne, E.; Youngs, H. L.; Misson, J.; Nussaume, L.; Somerville, C. Global Expression Analysis of CESA and CSL Genes in Arabidopsis. *Cellulose* **2004**, *11*, 279–286. https://doi.org/10.1023/B:CELL.0000046340.99925.57.

(12) Yin, Y.; Johns, M. A.; Cao, H.; Rupani, M. A Survey of Plant and Algal Genomes and Transcriptomes Reveals New Insights into the Evolution and Function of the Cellulose Synthase Superfamily. *BMC Genomics* **2014**, *15* (1). https://doi.org/10.1186/1471-2164-15-260.

(13) Schreiber, M.; Wright, F.; MacKenzie, K.; Hedley, P. E.; Schwerdt, J. G.; Little, A.; Burton, R. A.; Fincher, G. B.; Marshall, D.; Waugh, R.; Halpin, C. The Barley Genome Sequence Assembly Reveals Three Additional Members of the CslF (1,3;1,4)-β-Glucan Synthase Gene Family. *PLoS One* **2014**, *9* (3), e90888. https://doi.org/10.1371/journal.pone.0090888.

(14) Burton, R. A.; Shirley, N. J.; King, B. J.; Harvey, A. J.; Fincher, G. B. The *CesA* Gene Family of Barley. Quantitative Analysis of Transcripts Reveals Two Groups of Co-Expressed Genes. *Plant Physiol* **2004**, *134*, 224–236. https://doi.org/10.1104/pp.103.032904.

(15) Schwerdt, J. G.; MacKenzie, K.; Wright, F.; Oehme, D.; Wagner, J. M.; Harvey, A. J.; Shirley, N. J.; Burton, R. A.; Schreiber, M.; Halpin, C.; Zimmer, J.; Marshall, D. F.; Waugh, R.; Fincher, G. B. Evolutionary Dynamics of the Cellulose Synthase Gene Superfamily in Grasses. *Plant Physiol* **2015**, *168* (3), 968–983. https://doi.org/10.1104/pp.15.00140.

(16) Hill, J. L.; Hammudi, M. B.; Tien, M. The *Arabidopsis* Cellulose Synthase Complex: A Proposed Hexamer of CESA Trimers in an Equimolar Stoichiometry. *Plant Cell* **2014**, *26* (12), 4834–4842. https://doi.org/10.1105/tpc.114.131193.

(17) Purushotham, P.; Ho, R.; Zimmer, J. Architecture of a Catalytically Active Homotrimeric Plant Cellulose Synthase Complex. *Science (1979)* **2020**, *369* (6507), 1089–1094. https://doi.org/10.1126/science.abb2978.

(18) Zhang, X.; Dominguez, P. G.; Kumar, M.; Bygdell, J.; Miroshnichenko, S.; Sundberg, B.; Wingsle, G.; Niittylä, T. Cellulose Synthase Stoichiometry in Aspen Differs from Arabidopsis and Norway Spruce. *Plant Physiol* **2018**, *177* (3), 1096–1107. https://doi.org/10.1104/pp.18.00394.

(19) Purushotham, P.; Ho, R.; Yu, L.; Fincher, G. B.; Bulone, V.; Zimmer, J. Mechanism of Mixed-Linkage Glucan Biosynthesis by Barley Cellulose Synthase–like CslF6 (1,3;1,4)-β-Glucan Synthase. *Sci Adv* **2022**, *8* (45), 1596. https://doi.org/10.1126/sciadv.add1596.

(20) Yin, Y.; Huang, J.; Xu, Y. The Cellulose Synthase Superfamily in Fully Sequenced Plants and Algae. *BMC Plant Biol* **2009**, *9* (1), 99. https://doi.org/10.1186/1471-2229-9-99.

(21) Dimitroff, G.; Little, A.; Lahnstein, J.; Schwerdt, J. G.; Srivastava, V.; Bulone, V.; Burton, R. A.; Fincher, G. B. (1,3;1,4)-β-Glucan Biosynthesis by the CSLF6 Enzyme: Position and Flexibility of Catalytic Residues Influence Product Fine Structure. *Biochemistry* **2016**, *55* (13), 2054–2061. https://doi.org/10.1021/acs.biochem.5b01384.

(22) Handakumbura, P. P.; Matos, D. A.; Osmont, K. S.; Harrington, M. J.; Heo, K.; Kafle, K.; Kim, S. H.; Baskin, T. I.; Hazen, S. P. Perturbation OfBrachypodium Distachyon CELLULOSE SYNTHASE A4or7results in Abnormal Cell Walls. *BMC Plant Biol* **2013**, *13*, 131. https://doi.org/10.1186/1471-2229-13-131.

(23) Morgan, J. L. W.; McNamara, J. T.; Fischer, M.; Rich, J.; Chen, H.-M.; Withers, S. G.; Zimmer, J. Observing Cellulose Biosynthesis and Membrane Translocation in Crystallo. *Nature* **2016**, *531* (7594), 329–334. https://doi.org/10.1038/nature16966.

(24) Morgan, J. L. W.; Strumillo, J.; Zimmer, J. Crystallographic Snapshot of Cellulose Synthesis and Membrane Translocation. *Nature* **2013**, *493*, 181–186. https://doi.org/10.1038/nature11744.

(25) Park, S.; Ding, S. The N-terminal Zinc Finger of CELLULOSE SYNTHASE6 Is Critical in Defining Its Functional Properties by Determining the Level of Homodimerization in Arabidopsis. *The Plant Journal* **2020**, *103* (5), 1826–1838. https://doi.org/10.1111/tpj.14870.

(26) Ramírez-Rodríguez, E. A.; McFarlane, H. E. Insights from the Structure of a Plant Cellulose Synthase Trimer. *Trends Plant Sci* **2021**, *26* (1), 4–7. https://doi.org/10.1016/j.tplants.2020.09.010.

(27) Kumar, M.; Turner, S. Plant Cellulose Synthesis: CESA Proteins Crossing Kingdoms. *Phytochemistry* **2015**, *112*, 91–99. https://doi.org/10.1016/j.phytochem.2014.07.009.

(28) Wang, Z.; Xu, Y.; Yang, B.; Tiruchinapally, G.; Sun, B.; Liu, R.; Dulaney, S.; Liu, J.; Huang, X. Preactivation-Based, One-Pot Combinatorial Synthesis of Heparin-like Hexasaccharides for the Analysis of Heparin-Protein Interactions. *Chemistry - A European Journal* **2010**, *16* (28), 8365–8375. https://doi.org/10.1002/chem.201000987.

(29) Xu, Y.; Wang, Z.; Liu, R.; Bridges, A. S.; Huang, X.; Liu, J. Directing the Biological Activities of Heparan Sulfate Oligosaccharides Using a Chemoenzymatic Approach. *Glycobiology* **2012**, *22* (1), 96–106. https://doi.org/10.1093/glycob/cwr109.

(30) Dulaney, S. B.; Xu, Y.; Wang, P.; Tiruchinapally, G.; Wang, Z.; Kathawa, J.; El-Dakdouki, M. H.; Yang, B.; Liu, J.; Huang, X. Divergent Synthesis of Heparan Sulfate Oligosaccharides. *J Org Chem* **2015**, *80* (24), 12265–12279. https://doi.org/10.1021/acs.joc.5b02172.

(31) Tang, S.-L.; Pohl, N. L. B. Automated Solution-Phase Synthesis of β-1,4-Mannuronate and β-1,4-Mannan. *Org Lett* **2015**, *17* (11), 2642–2645. https://doi.org/10.1021/acs.orglett.5b01013.

(32) Zhang, G.-L.; Wei, M.-M.; Song, C.; Ma, Y.-F.; Zheng, X.-J.; Xiong, D.-C.; Ye, X.-S. Chemical Synthesis and Biological Evaluation of Penta- to Octa- Saccharide Fragments of Vi Polysaccharide from *Salmonella Typhi*. *Organic Chemistry Frontiers* **2018**, *5* (14), 2179–2188. https://doi.org/10.1039/C8QO00471D.

(33) Zhu, Q.; Shen, Z.; Chiodo, F.; Nicolardi, S.; Molinaro, A.; Silipo, A.; Yu, B. Chemical Synthesis of Glycans up to a 128-Mer Relevant to the *O*-Antigen of *Bacteroides Vulgatus*. *Nat Commun* **2020**, *11*, 4142. https://doi.org/10.1038/s41467-020-17992-x.

(34) Cheon, H.-S.; Lian, Y.; Kishi, Y. Highly Stereoselective and Iterative Synthesis of α-(1→4)-Linked Polysaccharides Composed of 3-*O*-Methyl-D-Mannose. *Org Lett* **2007**, *9* (17), 3323–3326. https://doi.org/10.1021/ol0713335.

(35) Ekholm, F. S.; Ardá, A.; Eklund, P.; André, S.; Gabius, H.-J.; Jiménez-Barbero, J.; Leino, R. Studies Related to Norway Spruce Galactoglucomannans: Chemical Synthesis, Conformation Analysis, NMR Spectroscopic Characterization, and Molecular Recognition of Model Compounds. *Chemistry - A European Journal* **2012**, *18* (45), 14392–14405. https://doi.org/10.1002/chem.201200510.

(36) Kinnaert, C.; Daugaard, M.; Nami, F.; Clausen, M. H. Chemical Synthesis of Oligosaccharides Related to the Cell Walls of Plants and Algae. *Chem Rev* **2017**, *117* (17), 11337–11405. https://doi.org/10.1021/acs.chemrev.7b00162.

(37) Corey, E. J.; Venkateswarlu, A. Protection of Hydroxyl Groups as Tert-Butyldimethylsilyl Derivatives. *J Am Chem Soc* **1972**, *94* (17), 6190–6191. https://doi.org/10.1021/ja00772a043.

(38) Patschinski, P.; Zhang, C.; Zipse, H. The Lewis Base-Catalyzed Silylation of Alcohols—A Mechanistic Analysis. *J Org Chem* **2014**, *79* (17), 8348–8357. https://doi.org/10.1021/jo5016568.

(39) Yu, B.; Tao, H. Glycosyl Trifluoroacetimidates. 2. Synthesis of Dioscin and Xiebai Saponin I. *J Org Chem* **2002**, *67* (25), 9099–9102. https://doi.org/10.1021/jo026103c.

(40) Goddard-Borger, E. D.; Stick, R. v. An Efficient, Inexpensive, and Shelf-Stable Diazotransfer Reagent: Imidazole-1-Sulfonyl Azide Hydrochloride. *Org Lett* **2007**, *9* (19), 3797–3800. https://doi.org/10.1021/ol701581g.

(41) Fischer, N.; Goddard-Borger, E. D.; Greiner, R.; Klapötke, T. M.; Skelton, B. W.; Stierstorfer, J. Sensitivities of Some Imidazole-1-Sulfonyl Azide Salts. *J Org Chem* **2012**, *77* (4), 1760–1764. https://doi.org/10.1021/jo202264r.

(42) Nicolaou, K. C.; Chakraborty, T. K.; Piscopio, A. D.; Minowa, N.; Bertinato, P. Total Synthesis of Rapamycin. *J Am Chem Soc* **1993**, *115* (10), 4419–4420. https://doi.org/10.1021/ja00063a093.

(43) Kaburagi, Y.; Kishi, Y. Operationally Simple and Efficient Workup Procedure for TBAF-Mediated Desilylation: Application to Halichondrin Synthesis. *Org Lett* **2007**, *9* (4), 723–726. https://doi.org/10.1021/ol063113h.