Structure and function of Xyloglucan Xylosyltransferases

by

Alan Thomas Culbertson

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Biochemistry

Program of Study Committee:
Olga Zabotina, Major Professor
Adam Barb
Thomas Bobik
Reuben Peters
Vincenzo Venditti

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2018

Table of Contents

ABSTRACT

The plant cell wall is a complex network composed mainly of polysaccharides, the most abundant biopolymers on earth and a rich source of biorenewable materials. Biosynthesis of these plant polysaccharides is poorly understood, largely due to difficulties in the structural characterization of glycosyltransferases and lack of suitable substrates for *in vitro* analysis. Xyloglucan Xylosyltransferases (XXTs) initiate side-chain extensions from a linear glucan polymer by transferring the xylosyl group from UDP-xylose during xyloglucan biosynthesis. Here, we optimized protein expression and enzymatic activity conditions of XXTs through numerous N- and C-terminal truncations, various *E. coli* strains, solubility tags, and storage conditions. This procedure was used for protein expression of three XXTs (XXT1, XXT2, and XXT5) and we show that XXT5 is catalytically active *in vitro*, though at a significantly slower rate compared to XXT1 or XXT2. As no structural information was available for any of the XXTs, we built a homology model of XXT2. This model was used to predict amino acids involved in UDP-xylose binding that were verified through mutagenesis. We subsequently solved the crystal structure of XXT1 without ligands and in complexes with UDP and cellohexaose. XXT1, a homodimer and member of the GT-A fold family of glycosyltransferases, binds UDP analogously to other GT-A fold enzymes. The structures detailed here combined with the properties of mutant XXT1s are consistent with a SN_i-like catalytic mechanism. Distinct from other systems is the recognition of cellohexaose by way of an extended cleft. The crystal structure of XXT1 demonstrates that XXT1 alone cannot produce

xylosylation patterns observed for native xyloglucans because of steric constraints imposed within the acceptor binding cleft. Homology modeling of XXT2 and XXT5, using the crystal structure of XXT1 as template, reveals a structurally altered cleft in XXT5 that could accommodate a partially xylosylated glucan chain produced by XXT1 and/or XXT2. This suggests that XXT1 and XXT2 xylosylate a growing glucan chain to produce the GXXG repeat, which is then utilized by XXT5 to produce the biologically observed XXXG repeat of native xyloglucan present in most of the plants. These results allowed us to propose a model of sequential xylosylation of glucan chain synthesized by glucan synthase and support the synthesis of xyloglucan via multiprotein complex localized in plant Golgi as proposed previously.

The work described in this dissertation was supported by NSF-MCB grant #1121163 and partially by the seed-grant from Roy J. Carver Trust.

CHAPTER 1. INTRODUCTION

Plant Cell Wall

The plant cell wall is a complex and dynamic assembly of polysaccharides, lignin, and glycoproteins (Cosgrove, 2005; Burton et al., 2010; Keegstra, 2010). The cell wall strengthens the plant body but also plays key roles in plant growth, cell differentiation, intercellular communication, water movement, and defense against biotic or abiotic attack (Keegstra, 2010). The plant cell wall is an attractive resource due to the fact that two of the three most abundant biopolymers on earth, cellulose and lignin, are major components of the cell wall. This demonstrates its enormous potential for the use of the plant cell wall for industrial applications including biofuels and biomaterials (Boerjan et al., 2003; Carroll and Somerville, 2009; Pauly and Keegstra, 2010; McFarlane et al., 2014). However, degradation of the plant cell wall is energy intensive and often very expensive due to the recalcitrant nature of cell walls. It is generally considered that the production of cellulases is the major bottleneck of complete degradation of the plant cell wall, and a large effort to increase cellulase production efficiency has yielded little success (Wilson, 2009). An alternative approach is to engineer the plant cell wall that can be more easily degraded with no detrimental effects on plant fitness (McCann and Carpita, 2008; Carpita, 2012; Ciesielski et al., 2014). Engineering a plant cell wall for human applications requires a complete understanding of the biosynthesis of the plant cell wall polysaccharides.

The composition of the plant cell wall is dependent on tissue, species, cell type, and environment. Depending on these factors, there is varying abundance of cellulose, hemicellulose, pectin, glycoproteins, and lignin. Generally, there are two types of plant cell wall: primary cell walls that surround growing cells and contain mostly cellulose, hemicellulose, and pectins, and secondary cell walls that are abundant in cellulose and lignin and play more of structural and defensive roles. The composition of the cell wall is likely more of a spectrum of these types, with the primary and secondary walls being the two extremes of the spectrum (Keegstra, 2010).

Cellulose is the most abundant biopolymer on earth. It consists of a β-(1,4) linked glucan backbone with numerous inter- and intra-hydrogen bonds to form an insoluble crystalline microfibril. Cellulose serves as the load-bearing structure in the cell wall. Cellulose biosynthesis is performed by cellulose synthases (CesA)(Pear et al., 1996). Arabidopsis has ten CesA genes consisting of two groups; CesA1, -2, -3, -5, -6, and -9, which are involved in primary cell wall biosynthesis, and CesA4, -7, and -8, which are involved in secondary cell wall biosynthesis (McFarlane et al., 2014). CesAs belong to GT family 2 and contain eight transmembrane domains that channel the cellulose chain through the plasma membrane (Morgan et al., 2013). CesAs form a rosette multi-protein structure with six-fold symmetry and are the functional unit of cellulose synthesis (Mueller and Brown, 1980). These complexes typically align with microtubules, suggesting that microtubules guide cellulose synthesis (McFarlane et al., 2014).

Recently, a crystal structure of cellulose synthase from bacteria (BcsA-BcsB) was revealed (Morgan et al., 2013). This structure demonstrates that cellulose synthase contains eight transmembrane domains, six of which directly interact with the glucan chain (Morgan et al., 2013). The catalytic domain is a GT-A fold and contains the D, D, D, Q(Q/R)XRW motif that is conserved among GT family 2 members (Morgan et al., 2013). Further, a homology model of CesA from *Gossypium hirsutum* was reported and revealed the plant-specific subdomains (Sethaphong et al., 2013). These plant-specific subdomains are predicted to be important for CesA oligomerization to form the rosette complex (Sethaphong et al., 2013).

Hemicellulose polysaccharides are loosely defined as those that require a strong alkali for solubilization and directly interact with cellulose. These polysaccharides can cross-link or act as spacers between cellulose microfibrils and influence cellulose crystallinity. The main types of hemicellulose are xyloglucan, xylan, mannan, and mixed linked glucans. The first three polysaccharides contain a β-(1,4) linked glycan backbone chain of either glucan, mannan, or xylan, whereas the fourth contains a β-(1,3:1,4) linked glucan backbone chain. These polysaccharides form the “gel-like” layer of the plant cell wall and consist of roughly a third of the dry mass of the plant primary cell wall (Pauly et al., 2013).

Mannans are mainly found in green algae (*Charophyceae*) or in the seeds of plants. The backbone is either a linear mannan chain or alternating mannose and glucose backbone. The mannan backbone is biosynthesized by CSLCA enzymes. Mannans are an essential part of seed development in Arabidopsis seeds (Goubet 2003). Mannans are grouped into four classes based on backbone and side-chain

composition: mannan, glucomannan, galactomannan, and galactoglucomannan (Scheller and Ulvskov, 2010).

Xylans are most abundant hemicellulose in secondary cell walls of monocots. They consist of a β-(1,4)-linked xylose backbone which is substituted with a wide variety of other sugars depending on species and tissue (Pauly et al., 2013). The main types of xylan are glucoronoxylan or arabinoxylan, which are substituted with α-(1,2) glucoronic acid or arabinose, respectively (Scheller and Ulvskov, 2010; Pauly et al., 2013). Mixed linked glucans are found only in grasses and not in dicot species. They consist of a β-(1,3) and β-(1,4) linkages and are a non-branched homopolymer.

Pectins are a major component in the primary cell wall and consists of three major polysaccharides: homogalacturonan (HG), rhamnogalactoronan I (RG-I), and rhamnogalacturonan II (RG-II), with HG being the most abundant (Atmodjo et al., 2013). The diversity and complexity of these polysaccharides make them difficult to study. Additionally, isolation from cell wall requires chemical or enzymatic treatments to release the pectic polysaccharides that can alter their structure (Atmodjo et al., 2013)

Glycosyltransferases

The numerous oligo- and polysaccharides found in nature entail an enormous complexity with specific linkages (Rademacher et al., 1988). The formation of these highly specific linkages requires the action of glycosyltransferases (GTs) (Lairson et al., 2008). GTs catalyze the formation of glycosidic bonds by transferring a sugar moiety from an activated donor to a variety of acceptors (Lairson et al., 2008). The activated donor is typically a nucleotide

sugar (e.g. UDP-glucose, UDP-xylose, GDP-mannose), but lipid phosphates are also used. The acceptor substrates can be any other molecule with the most common being other sugars, but can also be lipids, nucleic acids, proteins, or small molecules.

GTs are classified into 105 families based on their amino acid sequence in the Carbohydrate Active Enzyme Database (Campbell et al., 1997; Coutinho et al., 2003). Most GTs adopt one of two folds, GT-A or GT-B. GT-A fold has two Rossman-like domains that form a central β-sheet that is surrounded by α-helices. These are metal dependent enzymes that require a DxD motif for metal coordination (Busch et al., 1998; Wiggins and Munro, 1998). GT-B fold glycosyltransferases also have two Rossman-like domains, but they are less tightly associated with the active site between these two domains.

In addition to their classification by overall fold, GTs are also classified by their inversion or retention in stereochemistry in product compared to that of donor substrate. Inverting GTs follow a single displacement mechanism, a mechanism similar to the inverting glycosyl hydrolases (Davies and Henrissat, 1995; Lairson et al., 2008). This entails a protein side chain acting as catalytic base to deprotonate the acceptor hydroxyl, which then facilitates a S_N2 displacement of the anomeric carbon of the activated sugar (Lairson et al., 2008). The catalytic mechanism of retaining GTs is still under heavy debate. It was first proposed that these enzymes utilized a double displacement mechanism, similar to the retaining glycosyl hydrolases. This mechanism involves a protein side chain that performs a nucleophilic attack on the anomeric carbon of the activated sugar to form a glycosyl-enzyme intermediate with inverted stereochemistry. The acceptor hydroxyl would

then nucleophilically attack this glycosyl-enzyme intermediate, inverting the stereochemistry again, resulting in a net retention of anomeric carbon stereochemistry. This mechanism has fallen into disfavor due to the lack of a suitably placed catalytic nucleophile for the first SN_2 reaction and the failure to trap a glycosyl-enzyme intermediate. Instead, retaining GTs may utilize a S_Ni-like mechanism (Persson et al., 2001), which has gained support through kinetic isotope studies (Lee et al., 2011) and crystallographic snapshots along the pathway (Yu et al., 2015). S_Ni-like reactions are similar to S_N1 reactions but the bond collapses back to the original stereochemistry in the product. This mechanism is typically proposed due to the lack of supporting evidence for other mechanisms.

Xyloglucan Biosynthesis

Xyloglucan (XyG) is a ubiquitous plant polysaccharide that is found in all land plants (Pauly and Keegstra, 2016). XyG encompasses roughly 20-25% of the hemicellulose in the primary cell wall of dicots, while only encompassing 2-10% in grasses (Scheller and Ulvskov, 2010). XyG is believed to associate with cellulose through hydrogen bonds (Valent and Albersheim, 1974). Due to this interaction, it is believed that XyG may cross-links cellulose microfibrils or it may act as a spacer to prevent the cellulose aggregation (Somerville et al., 2004).

XyG contains a β-(1,4) linked glucan backbone that is substituted with numerous glycosyl residues (Pauly and Keegstra, 2016). The type and order of glycosyl substitution is dependent on numerous factors including species, tissue, cell type, and developmental state. Additionally, non-glycosyl substitutions can also be

present, such as acetyl groups. A nomenclature has been developed to describe the glucan backbone substitution (Fry et al., 1993). For example, G represents unsubstituted glucosyl residue, X represents glucosyl residue substituted at O-6 with α- D-xylose, L represents the xylose substituted with α-(1,2) linked D-Gal*p*, and F represents the xylose substituted with α-(1,2) linked L-Fuc*p*-(1,2)-β-D-Gal*p* (Fry et al., 1993). The composition of XyG in *Arabidopsis* primarily is composed of XXXG, XXGF, and XLFG subunits (Vanzin et al., 2002; Madson, 2003). Dicots typically contain XXXG-type xyloglucan, whereas grasses typically contain XXGG-type xyloglucan (Figure 1; Vincken et al., 1997).

XyG biosynthesis occurs in the Golgi apparatus (Chevalier et al., 2010) by action of multiple highly specific glycosyltransferases (decsribed in details below in the chapter 2). In this process, a glycan synthase synthesizes the glucan chain, transporting it into the Golgi, which is then decorated by numerous GTs (Pauly and Keegstra, 2016). Following biosynthesis, XyG is transported out of the cell via secretory vesicles. Cellulose synthase-like C4 (CSLC4) is the primary synthase responsible for XyG backbone synthesis (Cocuron et al., 2007). Interestingly, co-expression of CLSC4 with xyloglucan xylosyltransferase 1 (XXT1) increased glucan production. Additionally, it was shown that N-terminus, C-terminus, and active site loop are localized on the cytosolic side of the Golgi membrane (Davis et al., 2010).

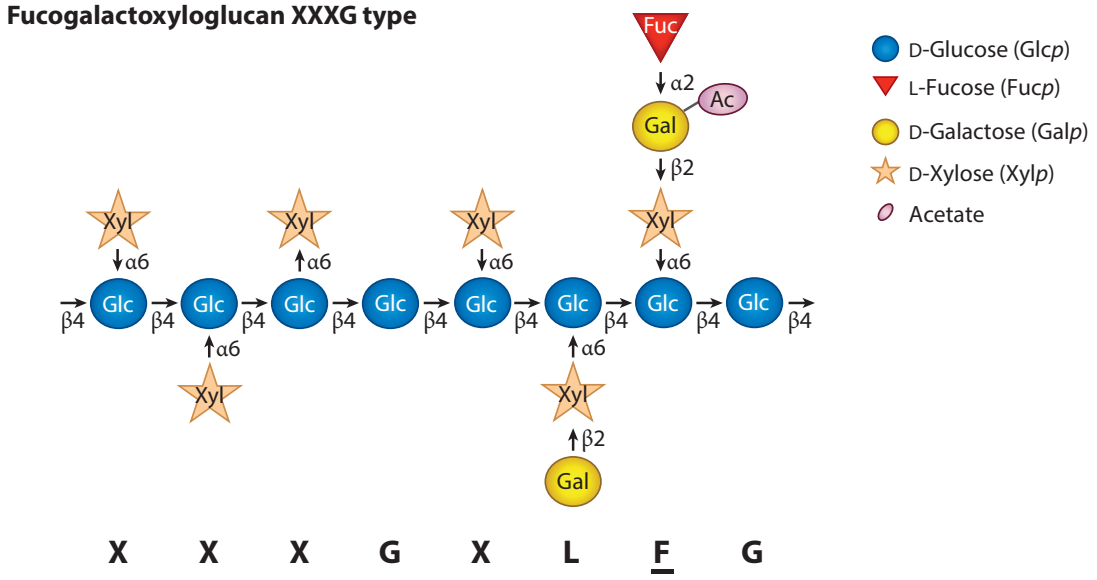


Figure 1. Schematic of the xyloglucan structure. Figure from Pauly and Keegstra, 2016.

To identify the enzymes involved in the xylosylation of the glucan backbone, Fiak *et al.* first tested enzymes that were in the same GT family as a characterized mannan GT (Faik et al., 2002). This is because the mannan GTs and XXTs both glycosylate a linear β-(1,4) glycan backbone at the O-6 position, thus it is likely that they would have high sequence similarity. This study identified XXT1 as being an enzyme that xylosylates a linear glucan backbone (Faik et al., 2002). Later, product characterization of XXT1 and XXT2 demonstrated that they both primarily xylosylate the fourth glucose from the reducing end of cellohexaose (Cavalier and Keegstra, 2006). Numerous reverse genetic studies have been investigated on the XXTs and it was found that a single knockout of either *xxt1* or *xxt2* had no change in XyG content, whereas a double knock-out of *xxt1xxt2* had no detectable XyG (Cavalier et al., 2008; Zabotina et al., 2012). This led to the proposal that XXT1 and XXT2 are at least partially redundant. The *xxt1xxt2* plants have minor morphological

phenotype changes, primarily in the root hairs (Cavalier et al., 2008; Zabotina et al., 2012). Later, XXT5 was identified to be another XXT with some distinct differences compared to XXT1 or XXT2. One difference was the fact that a single knock-out of *xxt5* resulted in a 50% reduction in xyloglucan (Zabotina et al., 2008; Zabotina et al., 2012). Additionally, attempts to demonstrate activity of XXT5 *in vitro* were all unsuccessful (Faik et al., 2002; Zabotina et al., 2008; Vuttipongchaikij et al., 2012), raising questions about the function of XXT5.

Galactosyltransferases were identified and were shown to have high specificity for the xylose they galactosylate (Madson, 2003; Jensen et al., 2012). XLT2 was shown to galactosylate the second xylose, and MUR3 galactosylates the third xylose in the XXXG subunit repeat (Madson, 2003; Jensen et al., 2012). A fucosyltransferase, FUT1, fucosylates the second galactose residue in the XLLG repeat, yielding an XLFG repeat (Perrin, 1999). All of these xyloglucan biosynthetic proteins were shown to form protein-protein interactions with each other and have been proposed to form a multi-enzyme protein complex (Chou et al., 2012; Chou et al., 2015).

Following its biosynthesis, XyG is transported to the plasma membrane via vesicles. Following incorporation in the cell wall, the structure remains highly dynamic as a result of numerous apoplastic enzymes (Frankova and Fry, 2013) which is highly correlated with cell elongation and plant growth (Pauly and Keegstra, 2016). Examples of these apoplastic enzymes include XyG endohydrolases, that hydrolyze the xyloglucan backbone releasing the XyG oligosaccharides into the

wall, and endotransglycosylases, which cleave and re-ligate the XyG to another XyG chain (Nishitani and Tominaga, 1992).

References

Atmodjo MA, Hao Z, Mohnen D (2013) Evolving views of pectin biosynthesis. Annu Rev Plant Biol **64:** 747-779

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annu Rev Plant Biol **54:** 519-546

Burton RA, Gidley MJ, Fincher GB (2010) Heterogeneity in the chemistry, structure and function of plant cell walls. Nat Chem Biol **6:** 724-732

Busch C, Hofmann F, Selzer J, Munro S, Jeckel D, Aktories K (1998) A Common Motif of Eukaryotic Glycosyltransferases Is Essential for the Enzyme Activity of Large Clostridial Cytotoxins. Journal of Biological Chemistry **273:** 19566-19572

Campbell J, Davies G, Bulone V, Henrissat B (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem. J **326:** 929-939

Carpita NC (2012) Progress in the biological synthesis of the plant cell wall: new ideas for improving biomass for bioenergy. Curr Opin Biotechnol **23:** 330-337

Carroll A, Somerville C (2009) Cellulosic biofuels. Annu Rev Plant Biol **60:** 165-182

Cavalier DM, Keegstra K (2006) Two xyloglucan xylosyltransferases catalyze the addition of multiple xylosyl residues to cellohexaose. J Biol Chem **281:** 34197-34207

Cavalier DM, Lerouxel O, Neumetzler L, Yamauchi K, Reinecke A, Freshour G, Zabotina OA, Hahn MG, Burgert I, Pauly M, Raikhel NV, Keegstra K (2008) Disrupting two Arabidopsis thaliana xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component. Plant Cell **20:** 1519-1537

Chevalier L, Bernard S, Ramdani Y, Lamour R, Bardor M, Lerouge P, Follet-Gueye ML, Driouich A (2010) Subcompartment localization of the side chain xyloglucan-synthesizing enzymes within Golgi stacks of tobacco suspension-cultured cells. Plant J **64:** 977-989

Chou Y, Pogorelko G, Young ZT, Zabotina OA (2015) Protein-Protein Interactions Among Xyloglucan-Synthesizing Enzymes and Formation of Golgi-Localized Multiprotein Complexes. Plant Cell Physiol

Chou YH, Pogorelko G, Zabotina OA (2012) Xyloglucan xylosyltransferases XXT1, XXT2, and XXT5 and the glucan synthase CSLC4 form Golgi-localized multiprotein complexes. Plant Physiol **159:** 1355-1366

Ciesielski PN, Resch MG, Hewetson B, Killgore JP, Curtin A, Anderson N, Chiaramonti AN, Hurley DC, Sanders A, Himmel ME, Chapple C, Mosier N, Donohoe BS (2014) Engineering plant cell walls: tuning lignin monomer composition for deconstructable biofuel feedstocks or resilient biomaterials. Green Chemistry **16:** 2627-2635

Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, Raikhel N, Wilkerson CG (2007) A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. Proc Natl Acad Sci U S A **104:** 8550-8555

Cosgrove DJ (2005) Growth of the plant cell wall. Nat Rev Mol Cell Biol **6:** 850-861

Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An Evolving Hierarchical Family Classification for Glycosyltransferases. Journal of Molecular Biology **328:** 307-317

Davies G, Henrissat B (1995) Structures and Mechanisms of Glycosyl Hydrolases. Structure **3:** 853-859

Davis J, Brandizzi F, Liepman AH, Keegstra K (2010) Arabidopsis mannan synthase CSLA9 and glucan synthase CSLC4 have opposite orientations in the Golgi membrane. Plant J **64:** 1028-1037

Faik A, Price NJ, Raikhel NV, Keegstra K (2002) An Arabidopsis gene encoding an alpha-xylosyltransferase involved in xyloglucan biosynthesis. Proc Natl Acad Sci U S A **99:** 7797-7802

Frankova L, Fry SC (2013) Biochemistry and physiological roles of enzymes that 'cut and paste' plant cell-wall polysaccharides. J Exp Bot **64:** 3519-3550

Fry SC, York WS, Albersheim P, Darvill A, Hayashi T, Joseleau JP, Kato Y, Lorences EP, Maclachlan GA, McNeil M (1993) An unambiguous nomenclature for xyloglucan-derived oligosaccharides. Physiologia Plantarum **89:** 1-3

Jensen JK, Schultink A, Keegstra K, Wilkerson CG, Pauly M (2012) RNA-Seq analysis of developing nasturtium seeds (Tropaeolum majus): identification and characterization of an additional galactosyltransferase involved in xyloglucan biosynthesis. Mol Plant **5:** 984-992

Keegstra K (2010) Plant Cell Walls. Plant Physiology **154:** 483-486

Lairson LL, Henrissat B, Davies GJ, Withers SG (2008) Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem **77:** 521-555

Lee SS, Hong SY, Errey JC, Izumi A, Davies GJ, Davis BG (2011) Mechanistic evidence for a front-side, SNi-type reaction in a retaining glycosyltransferase. Nat Chem Biol **7:** 631-638

Madson M (2003) The MUR3 Gene of Arabidopsis Encodes a Xyloglucan Galactosyltransferase That Is Evolutionarily Related to Animal Exostosins. The Plant Cell Online **15:** 1662-1670

McCann MC, Carpita NC (2008) Designing the deconstruction of plant cell walls. Curr Opin Plant Biol **11:** 314-320

McFarlane HE, Doring A, Persson S (2014) The cell biology of cellulose synthesis. Annu Rev Plant Biol **65:** 69-94

Morgan JL, Strumillo J, Zimmer J (2013) Crystallographic snapshot of cellulose synthesis and membrane translocation. Nature **493:** 181-186

Mueller SC, Brown RM, Jr. (1980) Evidence for an intramembrane component associated with a cellulose microfibril-synthesizing complex in higher plants. J Cell Biol **84:** 315-326

Nishitani K, Tominaga R (1992) Endo-xyloglucan transferase, a novel class of glycosyltransferase that catalyzes transfer of a segment of xyloglucan molecule to another xyloglucan molecule. J Biol Chem **267:** 21058-21064

Pauly M, Gille S, Liu L, Mansoori N, de Souza A, Schultink A, Xiong G (2013) Hemicellulose biosynthesis. Planta **238:** 627-642

Pauly M, Keegstra K (2010) Plant cell wall polymers as precursors for biofuels. Curr Opin Plant Biol **13:** 305-312

Pauly M, Keegstra K (2016) Biosynthesis of the Plant Cell Wall Matrix Polysaccharide Xyloglucan. Annu Rev Plant Biol

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. Proc Natl Acad Sci U S A **93:** 12637-12642

Perrin RM (1999) Xyloglucan Fucosyltransferase, an Enzyme Involved in Plant Cell Wall Biosynthesis. Science **284:** 1976-1979

Persson K, Ly HD, Dieckelmann M, Wakarchuk WW, Withers SG, Strynadka NC (2001) Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs. Nat Struct Biol **8:** 166-175

Rademacher TW, Parekh RB, Dwek RA (1988) Glycobiology. Annu Rev Biochem **57:** 785-838

Scheller HV, Ulvskov P (2010) Hemicelluloses. Annu Rev Plant Biol **61:** 263-289

Sethaphong L, Haigler CH, Kubicki JD, Zimmer J, Bonetta D, DeBolt S, Yingling YG (2013) Tertiary model of a plant cellulose synthase. Proc Natl Acad Sci U S A **110:** 7512-7517

Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, Milne J, Osborne E, Paredez A, Persson S, Raab T, Vorwerk S, Youngs H (2004) Toward a systems approach to understanding plant cell walls. Science **306:** 2206-2211

Valent BS, Albersheim P (1974) The structure of plant cell walls: v. On the binding of xyloglucan to cellulose fibers. Plant Physiol **54:** 105-108

Vanzin GF, Madson M, Carpita NC, Raikhel NV, Keegstra K, Reiter WD (2002) The mur2 mutant of Arabidopsis thaliana lacks fucosylated xyloglucan because of a lesion in fucosyltransferase AtFUT1. Proc Natl Acad Sci U S A **99:** 3340-3345

Vincken JP, York WS, Beldman G, Voragen AG (1997) Two general branching patterns of xyloglucan, XXXG and XXGG. Plant Physiol **114:** 9-13

Vuttipongchaikij S, Brocklehurst D, Steele-King C, Ashford DA, Gomez LD, McQueen-Mason SJ (2012) Arabidopsis GT34 family contains five xyloglucan alpha-1,6-xylosyltransferases. New Phytol **195:** 585-595

Wiggins CA, Munro S (1998) Activity of the yeast MNN1 α-1, 3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. Proceedings of the National Academy of Sciences **95:** 7945-7950

Wilson DB (2009) Cellulases and biofuels. Curr Opin Biotechnol **20:** 295-299

Yu H, Takeuchi M, LeBarron J, Kantharia J, London E, Bakker H, Haltiwanger RS, Li H, Takeuchi H (2015) Notch-modifying xylosyltransferase structures support an Si-like retaining mechanism. Nat Chem Biol

Zabotina OA, Avci U, Cavalier D, Pattathil S, Chou YH, Eberhard S, Danhof L, Keegstra K, Hahn MG (2012) Mutations in multiple XXT genes of Arabidopsis reveal the complexity of xyloglucan biosynthesis. Plant Physiol **159:** 1367-1384

Zabotina OA, van de Ven WT, Freshour G, Drakakaki G, Cavalier D, Mouille G, Hahn MG, Keegstra K, Raikhel NV (2008) Arabidopsis XXT5 gene encodes a putative alpha-1,6-xylosyltransferase that is involved in xyloglucan biosynthesis. Plant J **56:** 101-115

CHAPTER 2. THE GLYCOSYLTRANSFERASES INVOLVED IN SYNTHESIS OF PLANT CELL WALL POLYSACCHARIDES: PRESENT AND FUTURE

Modified from a paper published in *JSM Enzymology and Protein Science*

Alan T. Culbertson and Olga A. Zabotina*

Abstract

Glycosyltransferases are enzymes which transfer an activated sugar to an acceptor substrate such as polysaccharides, peptide, lipid or various small molecules. In the past 10-15 years, substantial progress has been made in the identification and cloning of genes that encode polysaccharide synthesizing glycosyltransferases. However, majority of these enzymes remain structurally and mechanistically uncharacterized. This short review will focus on the questions in biochemistry of polysaccharide synthesizing glycosyltransferases to be answered in coming years.

Introduction

Plant cell walls have been proposed to be a source of renewable energy in the form of lignocellulosic liquid biofuels. In the past 10-15 years, significant progress has been made in understanding of cell wall polysaccharide biosynthesis, particularly, in identifying and characterizing the numerous genes involved in this complex process. With respect to the challenges of revealing the genes required, molecular biology, reverse-genetics, and genomics has provided many powerful tools and significantly

advanced our understanding of plant cell wall formation. A significant body of recent reviews describes the advances and the current state in understanding of plant polysaccharide biosynthesis [1-5]. However, the progress in biochemical characterization of the gene products, glycosyltransferases, is being much slower and currently falls behind the successful genetic studies. In part, this is due to the low solubility of these enzymes and the lack of suitable enzyme assays. For example, the small stereochemical differences between sugar moieties and the multiple ways these moieties can be linked to each other, which were used by nature to achieve the wide diversity of oligo- and polysaccharide present in different types of plant cell walls, limits selection of suitable substrates and complicates characterization of products. Despite current limitations, the structural characterization and mechanisms of catalysis of plant glycosyltransferases will certainly be a subject of intensive research in the coming years due their essential function in plant cell wall biosynthesis. We present here the brief overview on the long-standing unanswered questions and directions we believe the field of cell wall polysaccharide synthesizing glycosyltransferases is headed.

Synthesis of the branched and heterogeneous polysaccharide structures requires action of multiple specific glycosyltransferases and synthases which transfer a donor sugar substrate to oligosaccharide acceptor. The donor substrates are typically activated sugars such as nucleotide sugars (UDP or GDP-bound) or, more rarely, phosphorylated sugars [6]. The structures solved for GTs from other organisms, the majority of which are involved in glycosylation of small lipophilic molecules, showed that the catalytic domains of most GTs have two types of fold,

GT-A and GT-B. However, the existence of the different type, GT-C was also proposed [6]. GT-A contains two β/α/β Rossmann-like folds tightly associated forming a continuous β-sheet and while GT-B also has two β/α/β Rossmann-like folds, they are not tightly intertwined but face each other with the active site residing between them. GT-A folds are metal dependent, which is coordinated by the well documented DxD motif [7], while GT-B are metal independent. In addition to these two structural folds, GTs are further characterized whether the chemical bond formed is an inversion or retention of stereochemistry with respect to the donor substrate. The most common donor substrate is nucleotide sugars where the sugar is linked via alpha bond. If a GT catalyzes the formation of the glycosidic bond attaching the sugar to the acceptor molecule via beta bond, the stereochemistry is inverted, while if the new glycosidic bond formed is alpha the stereochemistry is retained. The catalytic mechanism of inverting glycosyltransferases has been demonstrated to be a direct displacement S_N2-like reaction in which an active site residue acts as a general base to deprotonate the acceptor which performs a nucleophilic attack on the donor anomeric carbon [6]. The mechanisms of retaining glycosyltransferases has yet to be elucidated but numerous possibilities have been proposed such as a double displacement mechanism or a front-side single displacement (S_Ni) mechansism [8,9,10]. The plant cell wall biosynthetic enzymes are found in at least three of these classifications.

Glycosylsynthases

Cellulose Synthase (Ces) and Cellulose Synthase-Like (CSL) are integral, membrane proteins with multiple transmembrane domains (TD) which span the Golgi membrane (or plasma membrane in the case of Ces) multiple times and belong to CAZy family GT2 [11]. The first solved structure for a protein from GT2 family was the structure of the catalytic domain of polysaccharide synthesizing protein, SpsA, from B. subtilis [12], which demonstrated that catalytic domains of GT2 proteins adopt GT-A fold. The 3D structure of SpsA allowed prediction of the active site amino acids important for substrate binding and catalysis [13] and served as a prototype for the organization of other family GT2 synthases. More recently, the structure for *R. sphaeroides* Ces domains BcsA and BcsB demonstrated that TDs form a pore through which the synthesized glucan chain is translocated across the plasma membrane [14]. Another study resulted in a 3D computational model of the predicted cytosolic domain of cotton CESA (GhCESA1) [15], which showed good structural agreement between BcsA and GhCESA1. In both structures, the catalytic residues within GT-A fold included the matching motifs DDG, DCD and TED. The DDG and DCD motifs coordinate UDP and divalent cation and the D of the TED motif, most likely, acts as the catalytic base [14, 16]. The motif QRGRW in BcsA is positioned near the plasma membrane and was shown to interact with cellulose acceptor substrate, whereas in GhCESA1 a QVLRW motif was proposed to have similar function and similar positioning near plasma membrane [16]. These solved structures dismissed the long standing speculation about two active sites possibly present within the same peptide [17, 18] to explain cellulose synthesis, presenting

convincing model of how it is done by a single active site in concert with the pore for glucan translocation [14].

Although the β-glycan synthases in the gene family GT2 synthesize the polysaccharides with different linkages (1,3-β, 1,4-β or mixed 1,3-β;1,4-β), they have related sequences and conservative motifs implicated in substrate binding and catalysis. Therefore, it was proposed that they most likely have similar folding patterns and mechanisms of catalysis [19]. For example, the topologies of AtCSLC4 involved in synthesis of xyloglucan backbone [20] and BdCSLF6 involved in synthesis of mixed (1,3;1,4)-β-glucan [21] suggest that these two synthases translocate the corresponding glucan chain into Golgi lumen (Figure 1). However, AtCSLC4 has only six predicted TDs, so it is unclear whether six TDs are able to form wide enough pore to accommodate the glucan chain similar to BcsA. On the other hand, although BdCSLF6 has eight TDs, it was shown to synthesize mixed (1,3;1,4)-β-glucan which has a structure with a kink formed by 1,3-β-linkage. So, how can non-linear glucan chain with two different linkages be formed and accommodated inside the presumably tightly organized pore? The structural characterization of additional family GT2 members which have different number of predicted TDs and synthesize different products would aid in answering those questions.

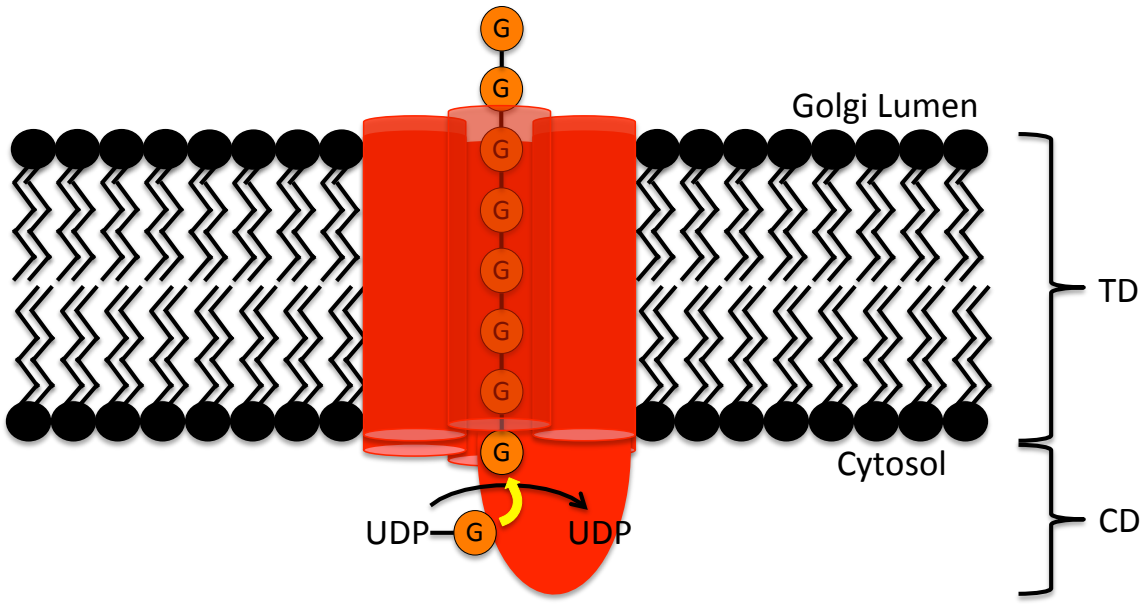


Figure 1. Proposed mode of action of CSLC4. Six predicted transmembrane domains (TD) and one catalytic domain (CD) residing in the cytosol are shown. The catalytic domain transfers glucose from UDP-glucose to a growing glucan chain which is channeled through the protein formed pore across the membrane into the Golgi lumen.

Glycosyltransferases

In addition to the synthases described above numerous other glycosyltransferases (GT) are involved in cell wall polysaccharide biosynthesis, which are type II membrane proteins residing in Golgi. They possess the short N-terminus domain localized in cytosol, one TD, flexible stem region and a catalytic domain localized in Golgi lumen [22], though the presence of GTs without predicted TD was also reported [23]. To date, none of the plant GTs involved in cell wall biosynthesis has been structurally characterized. The lack of structural data is a significant gap in our current knowledge about glycosyltransferases, which currently precludes further progress in understanding polysaccharide biosynthesis. In addition, determination of substrate binding or catalysis of each enzyme is

extremely difficult due to the lack of structural similarity and diversity in chemical bond formed, thus one cannot use a single well defined enzyme as a model for another. For example, according to classification system based on amino acid sequence [24, 6] two glycosyltransferases with diverse specificity, Xyloglucan Xylosyltransferase (XXT) involved in xyloglucan biosynthesis [25] and Galacturonosyltransferase 1 (GAUT1) involved in homogalacturonan biosynthesis [26] are retaining GTs with the GT-A fold. On the other hand, two xylosyltransferases, which are involved in xylan biosynthesis, are classified differently according to their sequences: Irregular Xylem 14 (IRX14) [27] is an inverting enzyme with the GT-A fold, whereas Irregular Xylem 10-Like (IRX10L), renamed to XYS1 [28], is an inverting enzyme with the GT-B fold. To date, none of plant cell wall synthesizing GTs is predicted to be a retaining enzyme with the GT-B fold [6]. Despite this difficulty, there are few well characterized enzymes from other organisms, which belong to the same gene families as plant cell wall biosynthetic enzymes and in some cases, like described above for Ces proteins, serve as prototypes for initial biochemical analysis.

In addition to the limitations imparted by the lack of structural data, it is also difficult to deduce any information in regards to acceptor substrate binding due to its complexity and diversity. This diversity of acceptor substrates requires the screening of various synthetic acceptor substrates, which are not readily available, for each specific glycosyltransferase. Cavalier and Keegstra [25] investigated the position and processivity of XXT1 and XXT2 catalysis of the xyloglucan backbone. It was found that both XXT1 and XXT2 are capable of xylosylation of cellohexaose at

two positions forming α-1,6 bonds and all of the acceptor substrate was first xylosylated at the fourth glucose from the reducing end and then once all acceptor was monoxylosylated, catalysis proceeds at the third xylose from the reducing end (Figure 2A). These results indicate that both UDP and the xylosylated acceptor are released and the xylosylated acceptor rebinds along with a fresh UDP-Xylose for another round of catalysis (Figure 2B). In addition, a oligoglucan with degree of polymerization (DP) of six was found to be the most suitable as the acceptor substrate, and enzyme activity decreased with DP was decreased. Increasing the DP further (e.g., celloheptaose) results in low solubility of the molecule and thus could not be analyzed, yet it cannot be ruled out that longer acceptor substrates may further increase XXT2 activity. In contrast to this, IRX10L/XYS1 was found to have an optimal DP of four and activity decreased with less or higher DP [28]. This is intriguing because XXT catalyze addition of sugar in the middle of the chain, while IRX10L/XYS1 glycosylated at the non-reducing end. Yet it is difficult to compare these two enzymes because XXTs are retaining enzymes with predicted GT-A fold, and IRX10/XYS1 is inverting enzyme with predicted GT-B fold GT.

There are also putative non-catalytic GTs involved in plant cell wall biosynthesis including GAUT7, IRX9, and IRX14 [26, 29]. XXT5 has also been proposed to be non-catalytic due to lack of *in vitro* activity, yet no evidence has confirmed this [30]. The function of these enzymes has been proposed to anchor other catalytic GTs to the golgi membrane which lack TDs or to channel the donor substrate to the active enzyme. For example, GAUT1 TD is post-translationally proteolytically cleaved *in vivo* yet is retained in the Golgi via physical interactions

with GAUT7 [26], and IRX14 was proposed to channel UDP-Xylose to IRX10 [29]. However, the hypothesis about IRX14 functioning has yet to be experimentally confirmed.

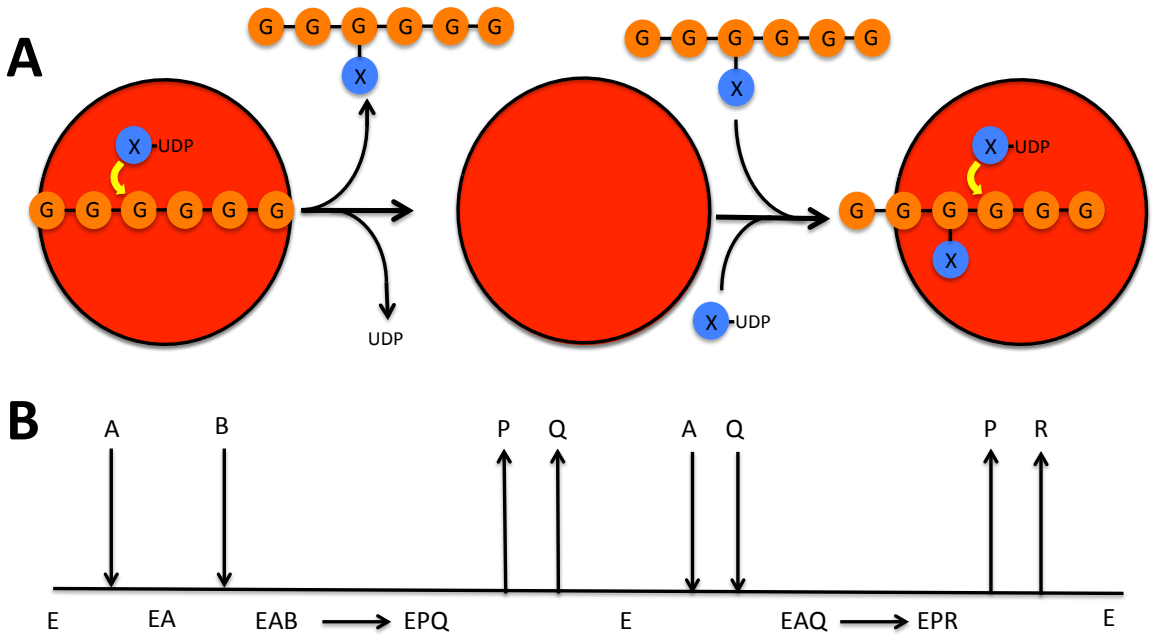


Figure 2. Proposed mode of action of XXT1 and XXT2. A) Positions of xylosylation and the main product formed, GGXGGG, which is then released and rebound for the second xylosylation which primarily forms GGXXGG. B) Kinetic mechanism of XXTs showing that both products are released before the second xylosylation occurs. Based on previous reports for other retaining GT-A fold glycosyltransferases, it is plausible that UDP-xylose binds first followed by acceptor binding, yet no data has been reported to confirm this. E: Enzyme; A: UDP-xylose; B: Cellohexaose; P: UDP; Q: Monoxylosylatedcellohexaose; R: Bixylosylatedcellohexaose

Concluding Remarks

Genetic studies have significantly advanced the field of plant cell wall biology, and currently, most of GT gene families, members of which are predicted to be involved in cell wall polysaccharide biosynthesis, possess at least one functionally characterized plant GTs. However, a simple gene family assignment is not sufficient

for the accurate prediction of GT's substrate specificities and mechanism of catalysis and there are a number of outstanding questions outlined in this brief overview. In order to engineer a plant cell wall for a specific purpose we must first understand the enzymes involved. This requires knowledge of not only the phenotypical changes of transgenic knock-out plants, but also the structure and mechanism of the specific enzyme. Characterization of these enzymes *in vitro* will aid in targeting specific enzymes for engineering of a bioenergy dedicated plant, or to engineer an enzymes with new specific functions to obtain the desired characteristics within the plant molecular framework. Although donor substrate binding can be predicted based on previous work on other glycosyltransferases in the same GT family, modes of acceptor binding are is still yet to be characterized due to the complexity and diversity of these substrates. This work will be aided by recently emerged new expression systems for recombinant protein production [28, 31] and advanced technologies in structural biology, enzymology, and computational simulations. Future advances in the field of polysaccharide synthesizing glycosyltransferases will aid in the biotechnological production of specific recombinant glycosyltransferases or plants with modified or even novel cell wall components which are engineered for specific applications such as biofuels [32], biomaterials [33], or pharmaceuticals [34].

References

1. Ulvskov P. (Editor) Plant polysaccharides, biosynthesis and bioengineering. Annual Plant Reviews. 2011; vol 41, 465 p, Blackwell Publishing, UK.

2. Atmodjo MA, Hao Z, Mohnen D. Evolving views of pectin biosynthesis. Annu Rev Plant Biol. 2013; 64:747-79.

3. Pauly M, Gille S, Liu L, Mansoori N, de Souza A, Schultink A, Xiong G. Hemicellulose biosynthesis. Planta. 2013; 238:627-42.

4. Hao Z., Mohnen D. A review of xylan and lignin biosynthesis: foundation for studying Arabidopsis irregular xylem mutants with pleiotropic phenotypes. Crit Rev Biochem Mol Biol. 2014; 49:212-41.

5. Rennie EA, Scheller HV. Xylan biosynthesis. Curr Opin Biotechnol. 2014; 26:100-107.

6. Lairson LL., Henrissat B., Davies GJ., Withers SG. Glycosyltransferases: Structures, Functions and Mechanisms. Annu. Rev. Biochem. 2008; 77:521-555.

7. Wiggins CA, Munro S. Activity of the yeast MNN1 α-1, 3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. Proceedings of the National Academy of Sciences. 1998; 95:7945-7950.

8. Schuman, B., Evans, S. V., and Fyles, T. M. (2013) Geometric Attributes of Retaining

Glycosyltransferase Enzymes Favor an Orthogonal Mechanism. PloS one 8, e71077.

9. Gomez, H., Polyak, I., Thiel, W., Lluch, J. M., and Masgrau, L. (2012) Retaining glycosyltransferase mechanism studied by QM/MM methods: lipopolysaccharyl-alpha-1,4-

galactosyltransferase C transfers alpha-galactose via an oxocarbenium ion-like transition state. Journal of
the American Chemical Society 134, 4743-4752.

10. Bobovska, A., Tvaroska, I., and Kona, J. (2014) A theoretical study on the catalytic mechanism of
the retaining alpha-1,2-mannosyltransferase Kre2p/Mnt1p: the impact of different metal ions on catalysis.
Organic & biomolecular chemistry 12, 4201-4210.

11. Cantarel BL., Coutinho PM., Rancurel C., Bernard T., Lombard V., Henrissat, B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009; 37:D233-238.

12. Charnock SJ., Davies GJ. Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms. Biochemistry. 1999; 38:6380-6385.

13. Tarbouriech N., Charnock SJ., Davies GJ. Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: A comparison with related NDP-sugar glycosyltransferases. J. Mol Biol. 2001; 314:655-661.

14. Morgan JLW., Strumillo J., Zommer J. Crystallographic snapshot of cellulose synthesis and membrane translocation. Nature. 2013; 493:181-186.

15. Sethaphong L., Haigler CH., Kubicki JD., Zimmer J., Bonetta D., DeBolt S., Yingling YG. Tertiary model of a plant cellulose synthase. Proceedings of the National Academy of Sciences. 2013; 110: 7512-7517.

16. Slabaugh E., Davis JK., Haigler CH., Yingling YG., Zimmer J. Cellulose synthases: new insights from crystallography and modeling. Trends in Plant Science. 2014; 19:99-106.

17. Saxena IM., Brown Jr RM., Dandekar T. Structure–function characterization of cellulose synthase: relationship to other glycosyltransferases. Phytochemistry. 2001; 57:1135-1148.

18. Carpita NC. Update on Mechanisms of Plant Cell Wall Biosynthesis: How Plants Make Cellulose and Other (1→4)-β-d-Glycans. Plant Physiol. 2011; 155:171-184.

19. Stone B., Jacobs AK., Hrmova M., Burton RA, Fincher GB. Biosynthesis of plant cell wall and related polysaccharides by enzymes of the GT2 and GT48 families. In: Ulvskov P. (Ed) Plant polysaccharides, biosynthesis and bioengineering. Annual Plant Reviews. 2011; vol 41, 109-165, Blackwell Publishing, UK.

20. Davis J., Brandizzi F., Liepman AH., Keegstra K. Arabidopsis mannan synthase CSLA9 and glucan synthase CSLC4 have opposite orientations in the Golgi membrane. Plant J. 2010; 64:1028-1037.

21. Kim SJ, Zemelis S, Keegstra K, Brandizzi F. The cytoplasmic localization of the catalytic site of CSLF6 supports a channeling model for the biosynthesis of mixed-linkage glucan. Plant J. 2015; 81(4):537-47.

22. Keegstra K., Raikhel N. Plant glycosyltransferases. Curr Opin Plant Biol. 2001; 4:219-224.

23. Kong Y, Zhou G, Yin Y, Xu Y, Pattathil S, Hahn MG. Molecular analysis of a family of Arabidopsis genes related to galacturonosyltransferases. Plant Physiol. 2011; 155:1791-805.

24. Coutinho PM., Deleury E., Davies GJ., Henrissat B. An evolving hierarchical family classification for glycosyltransferases. J. Mol Biol. 2003; 328:307-317.

25. Cavalier DM., Keegstra K. Two xyloglucan xylosyltransferases catalyze the addition of multiple xylosyl residues to cellohexaose. J. Biol Chem. 2006; 281:34197-34207.

26. Atmodjo MA, Sakuragi Y, Zhu X, Burrell AJ, Mohanty SS, Atwood JA, 3rd, Orlando R, Scheller HV, Mohnen D. Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:20225-20230.

27. Wu AM., Hornblad E., Voxeur A., Gerber L., Lorouge 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.

28. Urbanowicz BR., Pena MJ., Moniz HA., Moremen KW., York WS. Two Arabidopsis proteins synthesize acetylated xylan in vitro. Plant J. 2014; 80:197-206.

29. Ren Y, Hansen SF, Ebert B, Lau J, Scheller HV. 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:e105014.

30. Zabotina OA, van de Ven WT, Freshour G, Drakakaki G, Cavalier D, Mouille G, Hahn MG, Keegstra K, Raikhel NV. Arabidopsis XXT5 gene encodes a putative alpha-1,6-xylosyltransferase that is involved in xyloglucan biosynthesis. Plant J. 2008; 56:101-115.

31. Thomas P, Smart TG. HEK293 cell line: a vehicle for the expression of recombinant proteins. J Pharmacol Toxicol Methods. 2005; 51:187-200.

32. Phitsuwan P, Sakka K, Ratanakhanokchai K. Improvement of lignocellulosic biomass in planta: A review of feedstocks, biomass recalcitrance, and strategic manipulation of ideal plants designed for ethanol production and processability. Biomass and Bioenergy. 2013; 58:390-405.

33. Ciesielski PN, Resch MG, Hewetson B, Killgore JP, Curtin A, Anderson N, Chiaramonti AN, Hurley DC, Sanders A, Himmel ME, et al. Engineering plant cell walls: tuning lignin monomer composition for deconstructable biofuel feedstocks or resilient biomaterials. Green Chemistry. 2014; 16:2627-2635.

34. Seeberger PH, Werz DB. Synthesis and medical applications of oligosaccharides. Nature. 2007; 446:1046-1051.

CHAPTER 3. ENZYMATIC ACTIVITY OF ARABIDOPSIS XYLOGLUCAN XYLOSYLTRANSFERASE 5

Modified from a paper published in *Plant Physiology*

Alan T. Culbertson, Yi-Hsiang Chou, Adrienne L. Smith, Zachary T. Young, Alesia A. Tietze, Sylvain Cottaz[4], Régis Fauré, and Olga A. Zabotina

Abstract

Xyloglucan, the most abundant hemicellulosic component of the primary cell wall of flowering plants, is composed of a β-(1,4) glucan backbone decorated with D-xylosyl residues. Three Xyloglucan Xylosyltransferases (XXTs) participate in xyloglucan biosynthesis in *Arabidopsis thaliana*. Two of these, XXT1 and XXT2, have been shown to be active *in vitro,* whereas the catalytic activity of XXT5 has yet to be demonstrated. By optimizing XXT2 expression in a prokaryotic system and *in vitro* activity assay conditions, we demonstrate that non-glycosylated XXT2 lacking its cytosolic N-terminal and transmembrane domain displays high catalytic activity. Using this optimized procedure for expression of XXT5, we report for the first time that recombinant XXT5 shows enzymatic activity *in vitro*, although at a significantly slower rate than XXT1 and XXT2. Kinetic analysis showed that XXT5 has a 7-fold higher K_m and 9-fold lower k_{cat} compared to XXT1 and XXT2. Activity assays using XXT5 in combination with XXT1 or XXT2 indicate that XXT5 is not specific for their products. In addition, mutagenesis experiments showed that the *in vivo* function

and *in vitro* catalytic activity of XXT5 require the Asp-Ser-Asp motif. These results demonstrate that XXT5 is a catalytically active xylosyltransferase involved in xylosylation of the XyG backbone.

Introduction

Plant cell walls consist of cellulose (Endler and Persson, 2011), hemicellulose (Scheller and Ulvskov, 2010), pectin (Atmodjo et al., 2013), lignin (Vanholme et al., 2010), and glycoproteins (Josè-Estanyol and Puigdomènech, 2000). In plants, the cell wall strengthens the cell to resist turgor pressure and acts in intracellular communication and defense against biotic and abiotic stresses (Somerville et al., 2004; Keegstra, 2010). In addition, the plant cell wall has many potential and current industrial applications in biomaterials or biofuels (Somerville et al., 2004; McCann and Carpita, 2008; Abramson et al., 2010; Carpita, 2012)

Xyloglucan (XyG), the most abundant hemicellulose in the primary cell wall of dicotyledonous plants, has many proposed functions (Hayashi and Kaida, 2011; Park and Cosgrove, 2015) including cross-linking adjacent cellulose microfibrils (Hayashi, 1989; Carpita and Gibeaut, 1993; Somerville et al., 2004) and acting as a spacer to prevent the aggregation of cellulose microfibrils (Thompson, 2005). In Arabidopsis, XyG is composed of a β-(1,4) glucan backbone with 50-75% of the D-glucose residues substituted with α-(1,6) linked D-xylose residues. The side chain xylose residues can be further decorated with α-(1,2) linked D-Gal or L-Fuc*p*-(1,2)-β-D-Gal*p*. According to the established nomenclature for description of glucan backbone substitution patterns, G represents unsubstituted glucosyl residue, X

represents glucosyl residue substituted at O-6 with α- D-xylose, L represents the xylose substituted with α-(1,2) linked D-Gal*p*, and F represents the xylose substituted with α-(1,2) linked L-Fuc*p*-(1,2)-β-D-Gal*p* (Fry et al., 1993). Arabidopsis XyG is mainly composed of XXXG, XXFG, and XLFG subunits (Vanzin et al., 2002; Madson, 2003).

Enzymes involved in XyG biosynthesis in Arabidopsis have been identified through reverse genetics and biochemical approaches. All of the enzymes that decorate the glucan backbone of XyG are type II transmembrane proteins containing a short N-terminus localized in the cytosol, a transmembrane domain, a stem region, and a large catalytic domain localized in the Golgi lumen. Xylosylation of the glucan backbone is catalyzed by at least three glycosyltransferases, XXT1, XXT2, and XXT5, which are members of the GT34 family (Faik et al., 2002; Cavalier and Keegstra, 2006; Cavalier et al., 2008; Zabotina et al., 2008; Vuttipongchaikij et al., 2012; Zabotina et al., 2012) in the Carbohydrate Active enZyme database (CAZy; Campbell et al., 1997; Coutinho et al., 2003). Analysis of a phylogenetic tree of GT34 glycosyltransferases revealed three major clades (Mansoori et al., 2015). Clade A included AtXXT1-2, Clade B included AtXXT3-5, and Clade C included other glycosyltranferases that are involved in mannan biosynthesis (Mansoori et al., 2015). Glycosyltransferases are enzymes that transfer a sugar moiety from an activated donor substrate to an acceptor substrate (Lairson et al., 2008). XXTs are α-1,6 xylosyltranferases that transfer a xylose residue from UDP-Xylose onto the glucan backbone of XyG synthesizing XXXG type oligosaccharide (Supplemental Fig. S1A)

Predictions based on their primary amino acid sequences indicate that XXTs contain a GT-A fold and catalyze the formation of a glycosidic bond in which the product retains the anomeric configuration of UDP-Xylose. XXT1 and XXT2 have α-xylosyltransferase activity *in vitro* when cellohexaose was used as an acceptor (Faik et al., 2002; Cavalier and Keegstra, 2006). XXT1 and XXT2 primarily xylosylate the fourth D-glucosyl unit from the reducing end of cellohexaose, producing GGXGGG. Catalysis mainly continues on the third D-glucosyl unit from the reducing end, producing GGXXGG with small amounts of GXXGGG (Cavalier and Keegstra, 2006). Generation of knockout mutants in Arabidopsis showed that these enzymes, along with the putative xylosyltransferase XXT5, function to synthesize XyG (Cavalier et al., 2008; Zabotina et al., 2008; Zabotina et al., 2012). Another xylosyltransferase, XXT4, was later demonstrated to be catalytically active *in vitro*, although *xxt4* mutants had no detectable decrease in XyG content (Vuttipongchaikij et al., 2012). The *xxt1 xxt2* double mutant, *xxt1 xxt2 xxt5* triple mutant, and *xxt5* single mutant all show a severe root hair phenotype, displaying root hairs that were short and thick, formed bubble-like extrusions near the tips, and were lower in abundance compared with wild-type roots (Cavalier et al., 2008; Zabotina et al., 2008; Zabotina et al., 2012). By contrast, the *xxt1* and *xxt2* single mutants had normal root hairs, indicating that these enzymes are partially redundant (Cavalier et al., 2008). Analysis of the XyG content demonstrated that the *xxt5* single mutant plants had a 50% reduction in total XyG content, while the *xxt1 xxt2* double mutants had no detectable XyG (Cavalier et al., 2008; Zabotina et al., 2008). Although XXT5 clearly plays a significant role in XyG biosynthesis, its enzymatic activity has yet to be

demonstrated. XXTs, along with all other GTs involved in XyG biosynthesis, form homo- and heterodimers and have been proposed to form multi-protein complexes localized in the Golgi; these interactions may be required for the high efficiency of these enzymes *in vivo* (Chou et al., 2012; Chou et al., 2014; Lund et al., 2015).

In this study, we optimized the conditions for expression of XXTs in *E. coli* through the systematic investigation of several of the most commonly used *E. coli* strains, various solubility tags, and storage conditions. We demonstrate that XXT1, XXT2, and XXT5 can be expressed in a prokaryotic system, producing catalytically active recombinant proteins. Using this protocol, we demonstrated that XXT5 has enzymatic activity and transfers a D-xylosyl residue from UDP-Xylose onto cellohexaose at a similar position as XXT2. However, XXT5 has significantly lower activity compared with XXT1 and XXT2. For the first time, we show that XXT5 is catalytically active *in vitro* and we demonstrate that its activity does not depend on the products of XXT1 and XXT2. We also identified an Asp-Ser-Asp motif that is essential for XXT5 function as a xylosyltransferase.

Materials and Methods

XXT cloning. The pET20b expression vectors were modified as previously described (Boyken et al., 2012) to incorporate an N-terminal His-GB1, His-MBP, or His-GFP tag using *NdeI* and *XhoI* restriction sites. Overhangs of the forward primers were designed to contain the *NdeI* restriction site and an N-terminal 6xHis tag, and the reverse primers were designed to contain the *BamHI* and the *XhoI* restriction sites. Coding sequences of the *Arabidopsis thaliana* xylosyltransferase genes *XXT1*

(At3g62720), *XXT2* (At4g02500), and *XXT5* (At1g74380) were PCR amplified with N-terminal truncations of 135 base pairs, 126 base pairs, and 213 base pairs, respectively, and were cloned into the pET20b plasmid. *XXT2* was cloned using *BamHI* and *XhoI* restriction sites, whereas the compatible restriction enzymes *SalI* and *BglII* were used to clone *XXT1* and *XXT5*, respectively, due to the presence of an *XhoI* restriction site in *XXT1* and a *BamHI* restriction site in *XXT5* (Supplemental Table S1). All forward primers were designed to produce the Tobacco Etch Virus protease site (ENLYFQG) at the N-terminus of XXT. The resulting pET20b constructs were then transformed into *E. coli* DH10b for plasmid amplification and the sequence was verified by sequencing. Confirmed plasmids were then transformed into the various *E. coli* transgenic cells that were used for protein expression: BL21 (DE3) (Invitrogen), CodonPlus (Agilent Technologies), SoluBL21 (AMS Biotechnology), Rosetta 2 (DE3)(EMD Millipore), Arctic Express (DE3) CodonPlus (Agilent Technologies), and BL21 (DE3) pLysS (Agilent Technologies, Waldbronn, Germany). To generate the XXT5 D228A:D230A mutant, two fragments for each mutant were PCR amplified using primers designed to contain the mutated base pairs (Supplemental Table S1). The XXT5 D228A:D230A mutant fragment 1 was generated with XXT5 forward and XXT5 D228A:D230A reverse primers, and fragment 2 was generated using XXT5 D228A:D230A forward and XXT5 reverse primers. The resulting fragments containing the mutated base pairs were then used as the template in a fusion PCR reaction using XXT5 forward and XXT5 reverse primers to generate the final mutated gene fragment.

Protein expression. *E. coli* cells harboring the plasmids were grown at 37°C with shaking at 200 rpm in 500 ml of Luria-Bertani broth. When the cell culture reached an OD_{600} of 0.5, the temperature was lowered to 18°C and protein expression was induced by adding IPTG to a final concentration of 0.5 mM. Cells were incubated for 18 h at 18°C and then harvested by centrifugation. Pelleted cells were re-suspended in 12.5 mL of 25 mM Tris-HCl pH 7.4, 300 mM NaCl, 0.1 mM EDTA and rapidly frozen in liquid nitrogen. Cells were lysed by thawing and incubating for 30 min with 1 mg/mL of lysozyme and then sonicated for 15 s a total of five times. Solubilized proteins were collected by centrifugation at 20,000 xg for 30 min.

Protein purification. Crude lysate was loaded onto a Ni-NTA column with a lysate:resin ratio of 20:1. Affinity resin was incubated on a shaker for 1 h at 4°C. Unbound proteins were removed as a flow-through fraction and the resin was washed four times with washing buffer containing 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, and 20 mM imidazole. The protein of interest was eluted in three fractions with a total volume of 6 mL using elution buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl, and 300 mM imidazole. Buffer switch into 50mM Tris-HCl pH 7.4 and 150 mM NaCl was performed by concentrating the 6 mL elution fraction to 1 mL, then diluting it to 15 mL with buffer, and repeating this a total of 4 four times. Glycerol was added to the purified protein to a final concentration of 20% and proteins were stored at -80°C.

SDS-PAGE and Western Blots. Proteins were analyzed by SDS-PAGE in reducing conditions and stained with Coomassie Blue G250. Western blot analysis was performed as previously described (Chou et al., 2014). Proteins were electrophoretically transferred to a nitrocellulose membrane and blocked with non-fat milk. The GB1 tag was probed with horseradish peroxidase conjugated goat anti-rabbit antibody (Invitrogen) at a 10,000-fold dilution, developed using HyGLO quick spray, and visualized by ChemiDoc XRS+ (Bio-Rad). Pre-stained size markers were visualized on the same membrane using visible light. To estimate the concentration of XXT1 protein, XXT1 purity was estimated by determining the intensity of bands in Coomassie stained SDS-PAGE gel using ImageJ and total protein concentration measured by Bradford assay (Quick Start Bradford Dye reagent 1X, Cat# 500-0205) according to the manufacturer's instructions. Next, quantitative western blotting was performed using several volumes of protein loaded on the gel and quantifying the intensities of XXT band with ImageJ (Supplemental Fig. S3). Only full-length bands were used for quantitating XXT concentration.

Enzyme activity assay and product precipitation .The enzymatic activity assay was performed in a total volume 25 μL, consisting of 2 mM UDP-Xylose (CarboSource Services), 0.5 mM cellohexaose (Megazyme), and 2 mM $MnCl_2$. The reaction mixture was incubated at 28°C for 20 h with shaking at 100 rpm. All reactions were performed in duplicate. The reaction was stopped by adding 1 mL of 100% ethanol and incubating at -22°C for 6 h; the precipitated reaction product

was collected by centrifugation at 21,000 xg for 30 min. Combination assays were all performed as described above with equal protein concentration for each XXT.

XXT kinetic parameter estimations. Kinetic experiments were performed with 3 µM XXT1 or XXT2 with reaction containing various UDP-Xylose concentrations, 1 mM cellohexaose, 2 mM $MnCl_2$, 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, and reaction time for 30 min at 28°C. XXT5 reactions contained 11.5 µM XXT5 and the reaction time was 60 min. Kinetic parameters K_m and k_{cat} were calculated by non-linear curve fitting with the software Dynafit (Kuzmic, 1996). The reaction times were selected based on minimum reaction time to allow for detection by HPAEC.

HPLC and MALDI-TOF MS characterization of catalysis products. Products obtained from the enzymatic activity assays were analyzed by high-performance anion-exchange chromatography with a pulse-amperometric detector (HPAEC-PAD, Dionex), as described by Cavalier and Keegstra (Cavalier and Keegstra, 2006). Briefly, the aqueous solution of the product was injected onto a CarboPac PA-20 column and eluted using the following gradient: 30 mM to 100 mM Na-Acetate for 25 min with 100 mM NaOH remaining constant through the entire run; after the 25 min run, the column was re-equilibrated for 15 min to initial conditions. Non-xylosylated, mono-xylosylated, di-xylosylated, and tri-xylosylated cellohexaose eluted from the column with retention times of approximately 16.4, 17.8, 18.9, and 20.2 min, respectively. Quantitative estimation of xylose transfer was performed by peak integration of xylosylated cellohexaose to determine the relative percentage of each product (non-, mono-, di-, or tri-xylosated

cellohexaose). For example, if 40% of total cellohexaose (12.5 nmol) is mono-xylosylated, then 5 nmol of xylose was transferred to cellohexaose. Response factors for non-, mono-, di-, and tri-xylosylated glucan might vary when using PAD detection; nevertheless, this enabled estimation of the amount of xylose transfer in the reaction. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry analyses of the products was performed as described earlier (Zabotina et al., 2012). Briefly, one microliter of the product solution was spotted onto a MALDI-TOF sample plate using a 2,5-dihydroxybenzoic acid matrix at a sample:matrix ratio of 1:1. The sample spectra were analyzed on an Applied Biosystems VOYAGER-DE Pro MALDI-TOF mass spectrometry instrument in positive reflection mode with an acceleration voltage of 20 kV and extraction delay time of 350 ns.

Characterization of enzyme assay products. Enzyme assays and product precipitations were performed as described above, except that the product was resuspended in 25 mM sodium-acetate buffer (pH 5.0). Three μg of xyloglucan-specific *endo*-β-1,4-glucanase (XEG)(Megazyme) was added to the reaction and incubated at 37°C for 12 h. The sample was heated for 10 min at 100°C and centrifuged at 21,000 xg for 10 min. The resulting supernatant was used for analysis by HPAEC and MALDI-TOF.

Production of transgenic plants. PCR-amplified coding sequences were ligated into the pCR8/GW/TOPO entry vector (Invitrogen) as described in the manufacturer's instructions. The ligation product was transformed into *E. coli* DH10b and selected on LB agar plates with 50 μg/mL spectinomycin. The mutated gene in the

entry construct was transferred into the binary destination vector pEarleyGate 201 (Basta resistant) (TAIR). The binary destination vector was transformed into Agrobacterium GV3101 by electroporation. Arabidopsis was transformed using the floral-dip method (Clough and Bent, 1998). Plants were grown under long day conditions (16 hr light / 8 hr dark) at 22°C in a growth chamber.

Examination of protein expression in transgenic plants. Total membrane protein was extracted from 40 to 80 seedlings. The seedlings were ground in 10 mL protein extraction buffer (40 mM HEPES, 0.45 M sucrose, 1 mM EDTA, 1 mM $MgCl_2$, 1 mM KCl, 1mM dithiothreitol, and protease inhibitor cocktail [Roche], pH 8.0). The extract was homogenized using a Polytron homogenizer three times, 10 s each, at 10,000 rpm. The extract was filtered through three layers of miracloth and centrifuged for 30 min at 10,000 rpm at 4°C. The supernatant was transferred to a polycarbonate tube with aluminum cap assembly and ultracentrifuged at 37,000 rpm (100,000 xg) for 45 min with a 70Ti fixed rotor. After centrifugation, the supernatant was removed and the pellet was resuspended in 200 µL suspension buffer (40 mM HEPES, 0.2 M sucrose pH 7.3). The membrane protein fraction was treated with 1% Triton X-100 for 30 min at 4°C to solubilize membrane-bound proteins. After solubilization, proteins were ultracentrifugated at 37,000 rpm for 45 min to precipitate non-soluble membrane fragments. The protein concentration of the supernatant was measured using the Bradford protein assay. The supernatant was concentrated by precipitating with 10% trichloroacetic acid and the precipitated protein was resuspended in loading buffer with β-mercaptoethanol. Proteins were separated using SDS-PAGE and transferred to a

nitrocellulose membrane (0.2 μm, Bio-Rad) for immunodetection. Polyclonal anti-HA antibodies were used (1:500 dilution) to detect HA-fused proteins.

Root hair phenotype analysis and measurement of xyloglucan. Analysis of plant root hairs was done by plating sterilized seeds on half-strength MS with 0.3% gelrite. After the seeds germinated and the roots grew into the media, the plates were placed at a 45° angle. Pictures of 10-day-old roots were taken using a Leica DMIRE2 light microscope with a Retiga 1300 camera.

Alcohol-insoluble residues (AIR) from wild type Col-0, *xxt5,* and *xxt5* complemented either with HA-XXT5 or HA-XXT5(D228A:D230A) seedlings were prepared and digested with Driselase as described by (Zabotina et al., 2012). 5 mg of each AIR was homogenized in 0.5 ml of 20 mM Na acetate buffer (pH 5.0), ~0.5 U of Driselase was added, and the mixtures were incubated at 37°C for 48 h. Supernatants were separated by centrifugation and analyzed by HPAEC.

Results

Optimization of XXT2 Expression. Due to the hydrophobic nature of transmembrane domains, to increase protein solubility and expression levels, we produced truncated versions of XXT1, XXT2, and XXT5 that lacked the transmembrane domain and contained only the protein stem region and catalytic domain (Supplemental Fig. S1B). A previous study reported expression of truncated XXTs with an N-terminal GST tag in *E. coli* BL21 cells (Vuttipongchaikij et al., 2012). Here, we tested different *E. coli* strains and various N-terminal tags to determine the optimal conditions for expressing XXTs (Supplemental Fig. S2)

We investigated *E. coli* BL21 (DE3)-derived cell lines, which were engineered for different purposes (Francis and Page, 2010). We examined BL21 (DE3), CodonPlus, SoluBL21, Rosetta 2 (DE3), Arctic Express (DE3) codon plus, and BL21 (DE3) pLysS. SoluBL21 was engineered for expression of insoluble proteins and BL21 CodonPlus and Rosetta contain extra copies of tRNA genes, allowing for expression of proteins with codons rarely used in *E. coli* (Francis and Page, 2010). The pET20b-His-GB1-XXT2 plasmid was transformed into each cell line and protein expression was carried out in parallel. Cells were lysed and XXT2 was purified using Ni-NTA resin. We estimated XXT2 levels by analyzing the intensity of the corresponding band observed after immunoblotting using antibody against the GB1 tag (Supplemental Fig. S2A). The highest expression level of XXT2 was obtained in SoluBL21, and a slightly lower expression level was obtained in BL21 CodonPlus and Rosetta (Supplemental Fig. S2A).

Glycerol can improve protein stability and increase the specific activity of enzymes (Bradbury and Jakoby, 1972; Leandro et al., 2001). Therefore, we investigated the effect of the addition of glycerol following protein purification. Glycerol was added to the proteins eluted from Ni-NTA columns to a final concentration of 0%, 20%, or 50%, and the samples were frozen overnight at -80°C. The following day, the proteins were analyzed by activity assays using equal protein concentrations in all reactions. XXT2 stored in 20% glycerol had 1.4-fold higher activity compared to XXT2 stored in 0% or 50% glycerol (Supplemental Fig. S2B).

To further improve expression levels and affinity chromatography purification efficiency, we fused various solubility tags, along with a 6xHis tag, to the N-terminus of XXT2 (Supplemental Fig. S1C). We tested the following tags: the immunoglobulin-binding domain of streptococcal protein G (GB1) (Huth et al., 1997), maltose binding protein (MBP) (Nallamsetty and Waugh, 2006), green fluorescent protein (GFP) (Chalfie, 1995), and the 6xHis tag alone. Constructs containing XXT2 and the N-terminal tags were expressed in SoluBL21 cells and purified using Ni-NTA resin (Supplemental Fig. S2C). Activity assays of these XXT2 constructs showed that the 6xHis-GB1 and 6xHis-MBP tagged proteins have the highest activity (Supplemental Fig. S2D). Although XXT2 with only the 6xHis tag had a very large band at the expected size, the activity was significantly lower in comparison with XXT2 tagged with 6xHis-GB1 or 6xHis-MBP. The 6xHis-GB1 and 6xHis-MBP tagged proteins produced similar xylosylated products but, as observed by SDS-PAGE, the 6xHis-GB1 tagged protein had a much higher purity (Supplemental Fig. S2C). Based on these results, all further experiments were performed using 6xHis-GB1 fusion proteins expressed in SoluBL21 cells and stored in 20% glycerol until analysis.

Optimization of activity assay conditions. A previous study showed that maximum activity of XXT1 and XXT2 requires the presence of Mn^{2+} and an acceptor glucan chain with a degree of polymerization of at least six (Cavalier and Keegstra, 2006). Here, we investigated the effect of various salt concentrations, pH, temperature, and two buffers on XXT2 activity (Fig. 1). Increasing the sodium chloride concentration in the assay reaction mixture decreased XXT2 activity (Fig.

1A). Low pH in the assay reaction also decreased XXT2 activity, while pH between 7.4 and 9.0 did not significantly affect XXT2 activity (Fig 1B). Activity assay reactions incubated at 22°C or 28°C showed no difference in XXT2 activity, but increasing the incubation temperature to 37°C decreased XXT2 activity 4-fold (Fig. 1C).

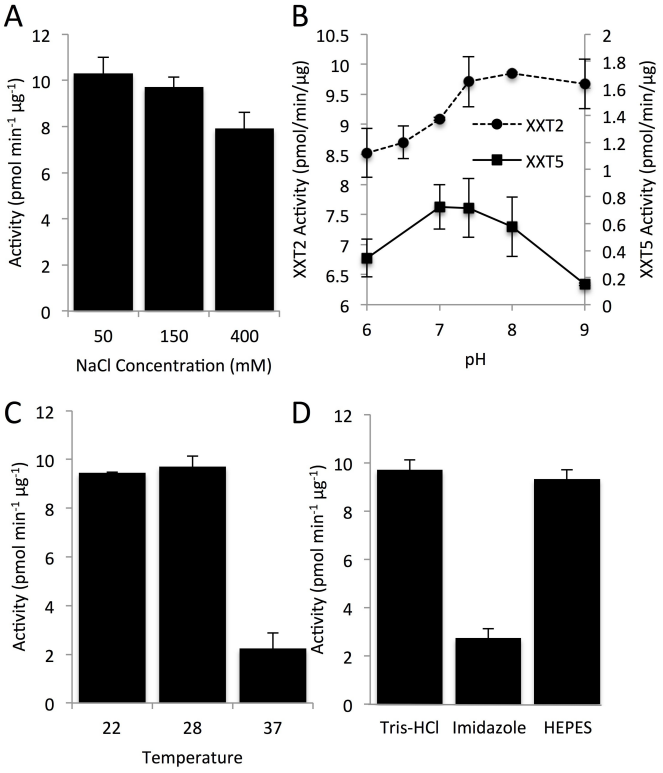


Figure 1. Optimization of assay conditions for XXT2 and XXT5. A. Effect of sodium chloride concentration (A), pH (B), and temperature (C) on His-GB1-XXT2 or His-GB1-XXT5 activity. (D) Effect of imidazole and HEPES on His-GB1-XXT2 activity. Imidazole sample in figure (D) were in Tris-HCl buffer. All assays were with XXT2 unless indicated. All buffers were adjusted to pH 7.4 unless indicated. All constructs were expressed in SoluBL21 cells and were purified with Ni-NTA columns. Assays contained 25 µL with concentrated His-GB1-XXT2 (3 µM; 2.5 µL) or His-GB1-XXT5 (5.5 µM; 8.25 µL), 12.5 µL of 2x buffer, 2 mM UDP-Xylose, 0.5 mM Cellohexaose, 2 mM $MnCl_2$ and were incubated for 4 hr. 2X buffer is a buffer that contains a concentration two times higher that of the desired final concentration of all components in the assay reaction. Final concentration of buffers were 50 mM for His-GB1-XXT2 reactions. Buffer concentration of His-GB1-XXT5 was 250 mM to reduce the effect of the larger volume of His-GB1-XXT5 protein that was added to the reaction on the overall reaction conditions. Products of the enzyme assays were analyzed by HPAEC, quantified by peak integration, and are presented as pmol of UDP-xylose per min per µg of XXT protein. All enzyme assays were performed in duplicate.

Next, two types of buffer (Tris-HCl or HEPES) and the presence of imidazole in the reaction mixture were investigated. A previous study reported that Tris-type buffers can inhibit enzymes by forming a complex with the nucleotide-sugar (Mukerjea et al., 2012). To investigate the effect of Tris-HCl buffer on XXT2 activity, we performed activity assays in Tris-HCl (pH 7.4) or HEPES (pH 7.4) buffers and found that reactions in Tris-HCl or HEPES buffer showed similar XXT2 activity (Fig. 1D). In addition, XXT2 was eluted from the Ni-NTA column using a buffer containing imidazole, which has been shown to affect activity of other enzymes (Cotovio et al., 1996; McGee et al., 2004; Zhang et al., 2011). To investigate the effect of imidazole on XXT2, activity assays were carried out in the presence of 250 mM imidazole and without imidazole in the reaction mixture. Activity of XXT2 with 250 mM imidazole was 4-fold lower compared to XXT2 activity without imidazole (Fig. 1D). Based on these results, all further activity assays were performed in 50 mM Tris-HCl buffer (pH 7.4) and 150 mM NaCl at 28°C, without imidazole.

***Expression and Catalytic Activity of XXT1, XXT2, and XXT5*.** XXT1, XXT2, and XXT5 lacking their cytosolic N-termini and transmembrane domains were fused with a 6xHis-GB1 tag and expressed in SoluBL21. When expressed in parallel using the same conditions, XXT1 showed approximately 2.5-fold higher expression level compared to XXT2; expression of XXT5 was slightly lower than that of XXT2 (Fig. 2A). SDS-PAGE and quantitative western blots were used to estimate the concentration of XXT proteins obtained after affinity purification (Supplemental Fig. S3). The purified XXT proteins were then assayed for enzymatic activity in a

reaction containing 2 mM UDP-xylose, 0.5 mM cellohexaose, and 2 mM $MnCl_2$. Enzyme assays contained maximum volume of XXT solution (maximum protein) to determine if XXT1, XXT2, and XXT5 can xylosylate cellohexaose at one, two, or three positions. Cellohexaose from reactions with XXT1 (6.2 μM) and XXT2 (2.5μM) was completely xylosylated and XXT1 was capable of xylosylating cellohexaose at three positions (Fig. 2B). XXT5 had very low activity under these conditions, yet when concentrated six-fold (6 μM), XXT5 was capable of xylosylating cellohexaose (Fig. 2B). All product sizes were verified with MALDI-TOF MS (Fig. 2C). Negative controls were assayed in parallel with XXT reactions (Fig. 2C; Supplemental Fig. S4). Localization of XXT5 in the cell or the XyG biosynthetic complex might differ from the localization XXT2 and XXT might have a different optimal pH compared to XXT2. Thus, we investigated various pH values in the XXT5 activity assay. XXT5 showed highest activity between pH 7 and 8 (Fig. 1B). XXT5 also showed higher sensitivity to increased pH, compared to XXT2. To the best of our knowledge, this is the first demonstration of XXT5 enzymatic activity *in vitro*.

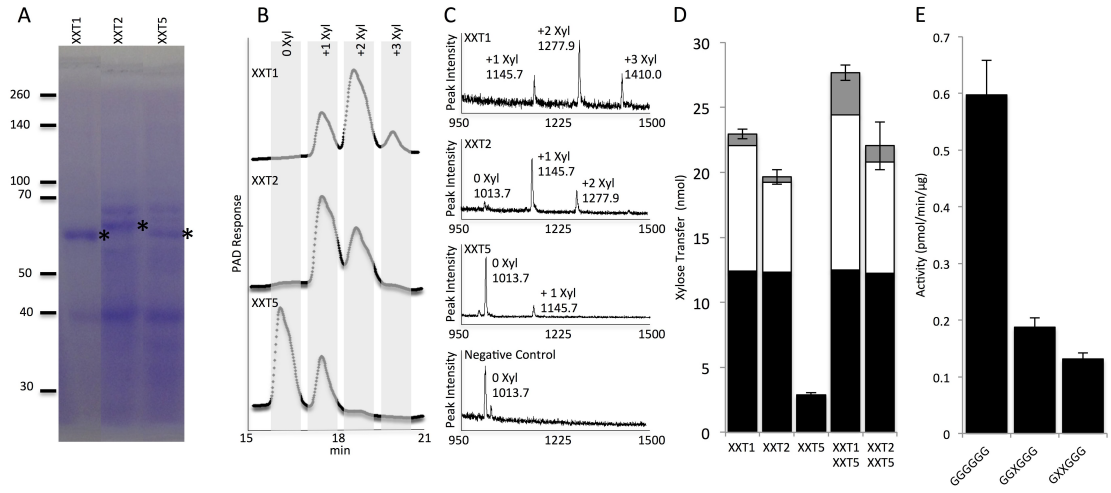


Figure 2. Protein expression and enzyme activity of XXT1, XXT2, and XXT5. All His-GB1-XXT proteins were expressed in SoluBL21 cells and purified on Ni-NTA resin. A. SDS-PAGE of XXT1, XXT2, and XXT5 with 10 µL (5.6 µg), 20 µL (12 µg), and 25 µL (12.5 µg) of elution loaded in each well, respectively. Positions of XXT proteins are shown with asterisks. B. HPAEC of XXT enzyme assay products. XXT1 and XXT2 concentration in assay reactions were 3 µM. XXT5 was concentrated 6-fold and the concentrated protein was assayed at 11.5 µM. C. MALDI-TOF chromatograms of enzyme assay products. Mass in MALDI-TOF chromatograms represents the mass of the oligosaccharide plus a sodium ion. Negative control consists of activity assay reaction with no enzyme. D. Enzyme assay of XXT1, XXT2, and XXT5 in combination. All XXT concentrations were 8.5 µM with reaction time of 20 hours. Black indicates mono-xylosylated cellohexaose, white indicates di-xylosylated cellohexaose, and grey indicates tri-xylosylated cellohexaose. E. XXT5 activity with various acceptors. Acceptor substrates were synthesized as described previously (Fauré et al., 2007). All reactions contained 2 mM UDP-xylose, 2 mM $MnCl_2$, 0.5 mM acceptor substrate, and 14 µM XXT5 with reaction time of 4 hr. Products of the enzyme assay were analyzed by HPAEC, quantified by peak integration, and presented as pmol of UDP-xylose per min per µg of XXT protein. All enzyme assays were performed in duplicate.

To determine whether XXT5's preferred substrate is the product of XXT1 or XXT2 (GGXXGG), we studied XXT5 activity in combination with either XXT1 or XXT2. Single assays were performed with only one XXT (8.5 µM) and combination assays were performed with XXT5 (8.5 µM) and XXT1 or XXT2 (8.5 µM). In the

single-enzyme reactions, XXT1 showed the highest activity, followed by XXT2. XXT5 showed the lowest catalytic activity, transferring only 2.9 nmol of D-xylose to cellohexaose (Fig. 2D). The combination of XXT1+XXT5 transferred 27.7 nmol of D-xylose to cellohexaose, slightly higher xylosylation than the sum of the single-enzyme reactions of XXT1 and XXT5 (25.8 nmol). Similarly, the combination of XXT2+XXT5 showed a slight decrease in xylosylation (22.0 nmol) compared to the single-enzyme reactions, transferring 22.0 nmol of xylose to cellohexaose, which was lower than the sum of the single-enzyme reactions of XXT2 and XXT5 (22.5 nmol). To further investigate if XXT5 is specific for xylosylated glucan, we ran XXT5 activity assays with different acceptor substrates, including GGXGGG and GXXGGG. Acceptor substrates were chemically synthesized as described previously (Fauré et al., 2007). XXT5 showed a 3-fold higher activity when cellohexaose (GGGGGG) was used as an acceptor, compared with the reactions when mono- or di-xylosylated cellohexaose were used (Fig. 2E).

Previous studies on XXTs used activity assays that were run for the maximum time-period to determine the ability of the XXTs to xylosylate cellohexaose at one, two, or three glucoses. Here, we performed kinetic studies to investigate the initial velocities of XXTs with various UDP-xylose concentrations and a constant cellohexaose concentration (1 mM). XXT1 and XXT2 assays required 30 min of reaction time and an enzyme concentration of 3 μM to produce detectable levels of xylosylated product, using high-performance anion-exchange chromatography (HPAEC) for detection. XXT5 assays required 60 min of reaction time and an enzyme concentration of 11.5 μM to produce detectable levels of

xylosylated product. In these conditions, XXT1 showed the lowest K_m and highest k_{cat} and XXT5 showed the highest K_m and lowest k_{cat}. XXT5 showed a 7-fold higher K_m and 9-fold lower k_{cat} compared to XXT1 and XXT2 (Fig. 3).

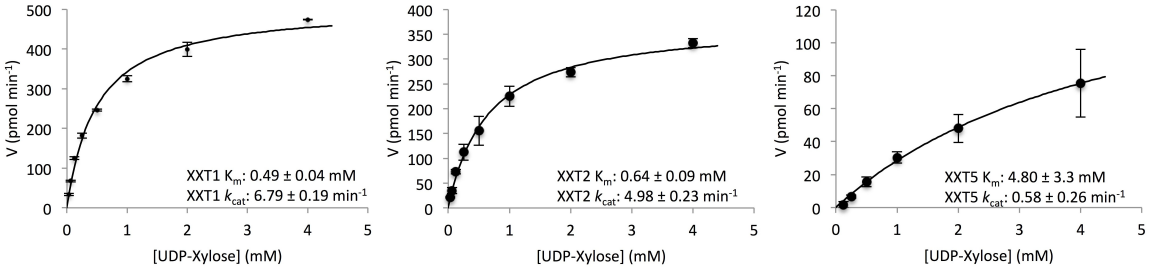


Figure 3. Kinetic analysis of XXT1, XXT2, and XXT5. XXT1 and XXT2 assays contained 3 µM of XXT and were performed for 30 min; XXT5 assays contained 11.5 µM of XXT5 and were performed for 60 min. All reactions contained 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM cellohexaose, 2 mM $MnCl_2$, and were incubated at 28°C. Kinetic constants of K_m and k_{cat} were calculated by fitting initial velocities as a function of UDP-Xylose concentration using non-linear curve fitting with Dynafit (Kuzmic, 1996). Products of the enzyme assay were analyzed by HPAEC, quantified by peak integration, and presented as pmol of UDP-xylose per min. All enzyme assays were performed in duplicate.

XXT5 and XXT2 produce similar xylosylation patterns. A previous study reported that both XXT1 and XXT2 link the first D-xylose to the fourth D-glucose from the reducing end of the cellohexaose oligosaccharide (Cavalier and Keegstra, 2006). Next, the enzymes link a second D-xylose preferentially to the third D-glucose from the reducing end; they also can add the second D-xylose to the fifth D-glucose, albeit at a lower frequency. Finally, the enzymes link a third xylose to the fifth or the third glucose, forming a tri-xylosylated product with three adjacent xyloses (Cavalier and Keegstra, 2006). To further investigate the activity of XXT5 *in vitro*, we determined the position of the D-xylose unit in the monoxylosylated product. Xylosylation positions were investigated by digesting the products of

XXT2 and XXT5 reactions with XyG-specific *endo*-β-1,4-glucanase (XEG), which cleaves the glucan backbone at the non-reducing end of a glucose substituted with xylose (Pauly et al., 1999; Yaoi et al., 2009). Digestion of the products of XXT2 (mono- and di-xylosylated) mainly produced XGG and XGGG. Similarly, digestion of the XXT5 product (only mono-xylosylated) mainly produced XGG with minor amounts of XGGG (Fig. 4A). Product sizes were verified with MALDI-TOF MS (Fig. 4B). These results indicated that XXT2 and XXT5 add D-xylose to cellohexaose at a similar position.

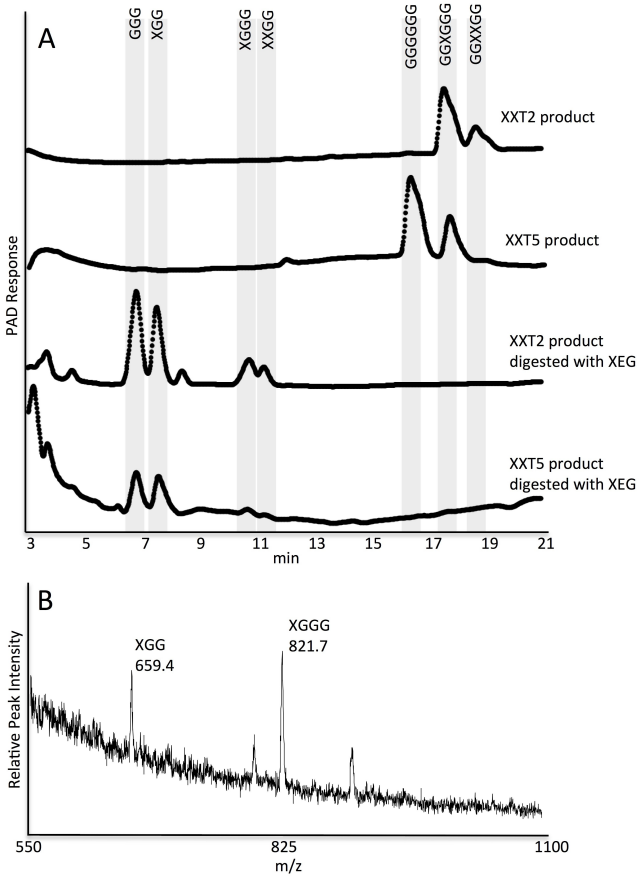


Figure 4. Analysis of the products of reactions catalyzed by XXT2 and XXT5 with HPAEC and MALDI-TOF. The products were precipitated (Materials and Methods) and treated with XEG for 12 hours. A. HPAEC chromatogram of enzyme assay products and products following XEG digestion. B. MALDI-TOF chromatogram of the XXT2 reaction to confirm product sizes.

XXT5 Activity Requires the DxD Motif. In GT-A fold glycosyltransferases, which include XXTs, the DxD motif (Asp-X-Asp) plays an important role in coordinating an essential divalent cation (Busch et al., 1998; Wiggins and Munro, 1998; Zhang et al., 2001; Boeggeman and Qasba, 2002; Gotting et al., 2004; Cavalier and Keegstra, 2006; Bobovska et al., 2014). Structural characterization of a putative glycosyltransferase from the Chlorella virus (PBVC-1), which is predicted to belong to the same GT 34 gene family as the XXTs, showed that the two Asp residues in the DxD motif participate in the coordination of the divalent cation (Zhang et al., 2007). Sequence alignment of the PBVC-1 glycosyltransferase with XXT1, XXT2, and XXT5 demonstrated that both Asp residues within the DxD motif are conserved in each of the XXTs (Supplemental Fig. S1D).

To test whether XXT5 catalytic activity requires the DxD motif, we mutated both Asp residues in the DxD motif of XXT5 (Asp-Ser-Asp) to Ala (Ala-Ser-Ala, D228A:D230A). The resulting protein was expressed, purified (Fig. 5A), and assayed for transferase activity. These activity assays showed no detectable xylosylated product for the XXT5 D228A:D230A mutant when analyzed by MALDI-TOF MS. To ensure that the lack of xylosylated product is not due to products below the sensitivity of MALDI-TOF MS, we concentrated both wild-type (WT) XXT5 and the XXT5 D228A:D230A mutant protein 15-fold, and performed activity assays with increased concentrations of UDP-Xylose and $MnCl_2$ (4 mM) and a reaction time of 30 h (Fig. 5B). Under these conditions, WT XXT5 added two D-xyloses,

producing di-xylosylated cellohexaose, while the XXT5 D228A:D230A mutant did not produce any xylosylated oligosaccharide (Fig. 5B).

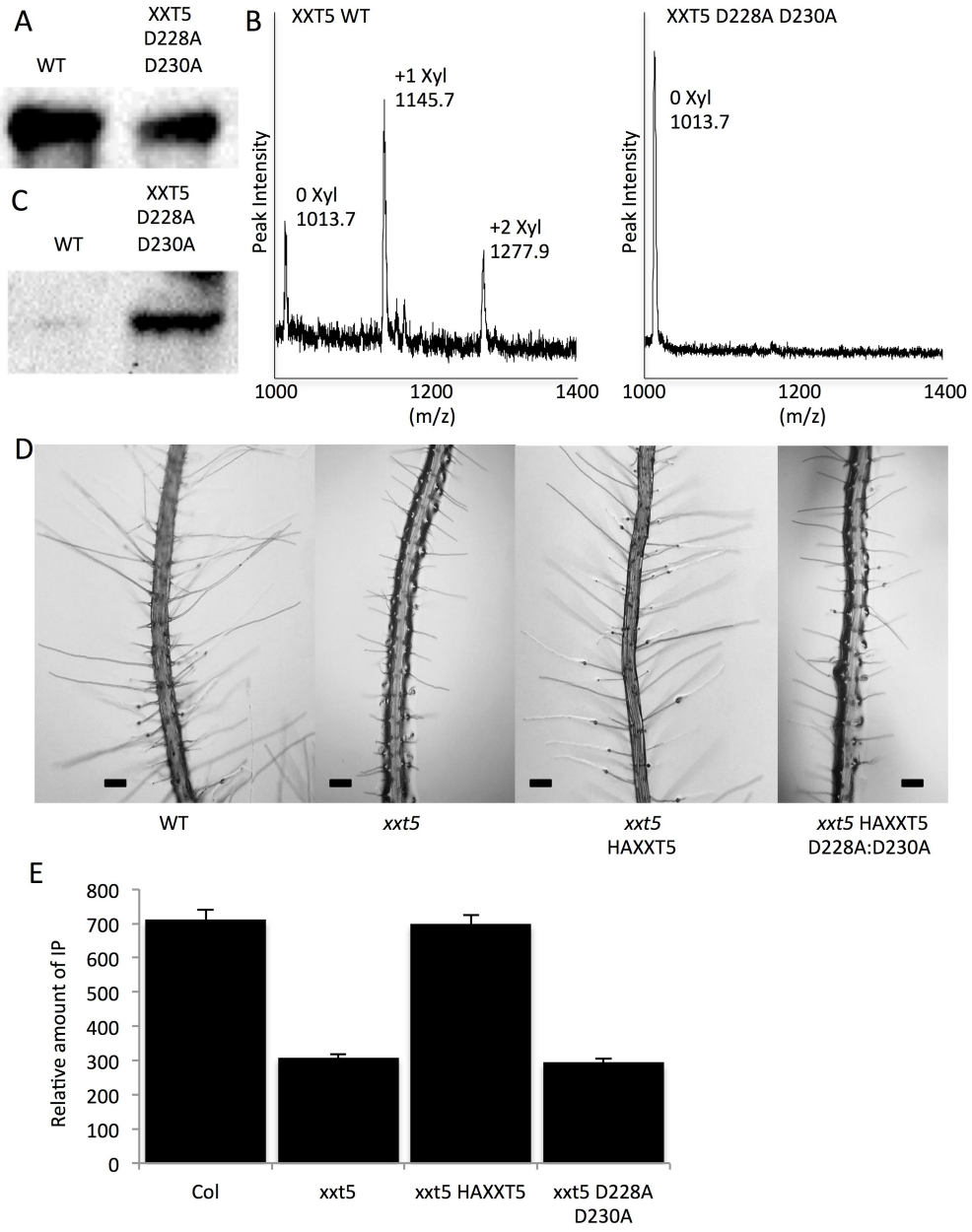


Figure 5. Characterization of the XXT5 D228A/D230A mutant. A. MALDI-TOF chromatograms of enzyme assay products of XXT5 D228A/D230A. The enzyme assay was slightly modified. Affinity purified protein was concentrated 15-fold; enzyme assays ran for 30 hr and contained 4 mM UDP-Xylose and 4 mM $MnCl_2$. B. Driselase digestion of plant cell wall extracts. C. Western blot of bacterially expressed XXT5 proteins. D. Western blot of proteins extracted from transgenic plants. E. Root hair phenotype of WT and transgenic plants.

To investigate the function of enzymatically inactive XXT5 *in vivo*, we overexpressed the XXT5 D228A:D230A mutant protein in the *xxt5* knockout mutant (Zabotina et al., 2008). The mutant cDNA of *XXT5 D228A:D230A* was fused with an HA-tag and introduced into the binary vector pEarleyGate 201. The resulting construct was transformed into *xxt5* homozygous mutant plants (Zabotina et al., 2008). *In planta* expression of XXT5 D228A:D230A was confirmed by immunoblot using HA antibodies (Fig. 5C). Over-expression of XXT5 D228A:D230A did not complement the root hair phenotype of the *xxt5* mutants (Fig. 5D). To measure XyG content in the transgenic plants, we used driselase to digest cell walls extracted from WT, *xxt5, xxt5* complemented with *HA-XXT5*, and *xxt5* complemented with *HA-XXT5 D228A:D230A*. Driselase is a mixture of hydrolases that digests most cell wall polysaccharides but lacks α-xylosidase activity. Hence, hydrolysis of XyG will produce the signature disaccharide isoprimeverose (D-Xyl*p*-α-(1-6)-D-Glc). Driselase digestion of the cell wall from *xxt5* plants complemented with *XXT5 D228A:D230A* resulted in isoprimeverose levels similar to those of the *xxt5* knockout mutant, indicating that XXT5 D228A:D230A cannot rescue XyG biosynthesis in the *xxt5* mutant (Fig. 5E).

Discussion

Three Arabidopsis xylosyltransferases, XXT1, XXT2, and XXT5, have been shown to be involved in XyG biosynthesis (Faik et al., 2002; Cavalier et al., 2008; Zabotina et al., 2008;). The *xxt1* and *xxt2* single knockout mutant plants had no

apparent phenotype, whereas the *xxt1 xxt2* double mutants had no detectable XyG and a strong root hair phenotype, indicating that XyG biosynthesis requires the presence of at least one of these enzymes (Zabotina et al., 2012). The *xxt5* mutant had a 50% reduction in XyG content as well as a distinct root hair phenotype. In addition, XXT1 and XXT2 have been confirmed to be catalytically active *in vitro* as α-1,6-xylosyltransferases (Cavalier and Keegstra, 2006), while the catalytic activity of XXT5 has not been shown despite several attempts (Faik et al., 2002; Zabotina et al., 2008; Vuttipongchaikij et al., 2012). It was therefore suggested that XXT5 might lack enzymatic activity and, most likely, perform other functions. Here, we demonstrate for the first time that XXT5 is enzymatically active *in vitro*, catalyzing the formation of a xylosylated oligosaccharide product, but at a significantly lower rate compared to XXT1 and XXT2.

Plant glycosyltransferases have mainly been heterologously expressed in eukaryotic cells for *in vitro* characterization (Faik et al., 2002; Cavalier and Keegstra, 2006; Jensen et al., 2014; Urbanowicz et al., 2014). Here, we express truncated versions of XXTs in prokaryotic cells, demonstrating that catalytic activity does not require the N-terminus and transmembrane domains or glycosylation; however glycosylation may improve protein folding and/or solubility. Negative controls of reactions lacking either UDP-Xylose or enzyme were done to confirm that the xylosylation is the product of XXTs (Fig. 2C; Supplemental Fig. S4). In addition, the XXT5 D228A:D230A mutant showed no xylosylation demonstrating that other proteins co-purified from *E. coli* cells do not xylosylate cellohexaose (Fig. 5A). Vuttipongchaikij *et al.* (2012) reported the only attempt to

produce recombinant XXTs in *E. coli* cells, expressing XXTs without transmembrane domains in BL21 cells, and demonstrated enzymatic activity for XXT1, XXT2, and XXT4. However, protein expression levels in this study were low, as indicated by faint bands SDS-PAGE and by reported yields of 40-90 μg of total protein per liter of cell culture. Vuttipongchaikij *et al.* showed no *in vitro* activity for XXT5 with an enzyme concentration of 0.3 μg of total protein per 25 μL reaction (Vuttipongchaikij *et al*. 2012). Here, increasing the XXT5 concentration 26-fold to 6 μM (7.8 μg) showed catalytic activity of XXT5 *in vitro* (Fig. 2). Thus, the lack of XXT5 activity described previously is most likely due to low protein levels in activity assay reactions. However, the demonstration of XXT5 catalytic activity *in vitro* does not exclude any of the non-catalytic functions proposed earlier for this protein *in vivo*.

XXT1 showed a higher degree of xylosylated product (Fig. 2), lower K_m, and higher k_{cat} (Fig. 3) compared to XXT2. This confirms previous reports showing higher activity for XXT1 compared to XXT2 when expressed in *Pichia pastoris* (Faik et al., 2002), *Drosophila S*2 cells (Cavalier and Keegstra, 2006), *Spodopstera frugiperda* 21 cells (Cavalier and Keegstra, 2006), although the kinetic parameters for XXTs have not been determined in those studies. Due to limitations in the product detection by HPAEC, the values reported here are only an estimation of kinetic parameters, yet allow for comparison between the XXTs. Development of more sensitive detection methods for more accurate determination of kinetic parameters remains a topic for future research. In the conditions used in this study, XXT5 shows a 7-fold higher K_m and 9-fold lower k_{cat} in comparison with XXT1 or

XXT2 (Fig. 3). This slower rate of XXT5 activity might be the reason that XXT5 cannot fully xylosylate the glucan backbone *in vivo* in the *xxt1 xxt2* double mutant.

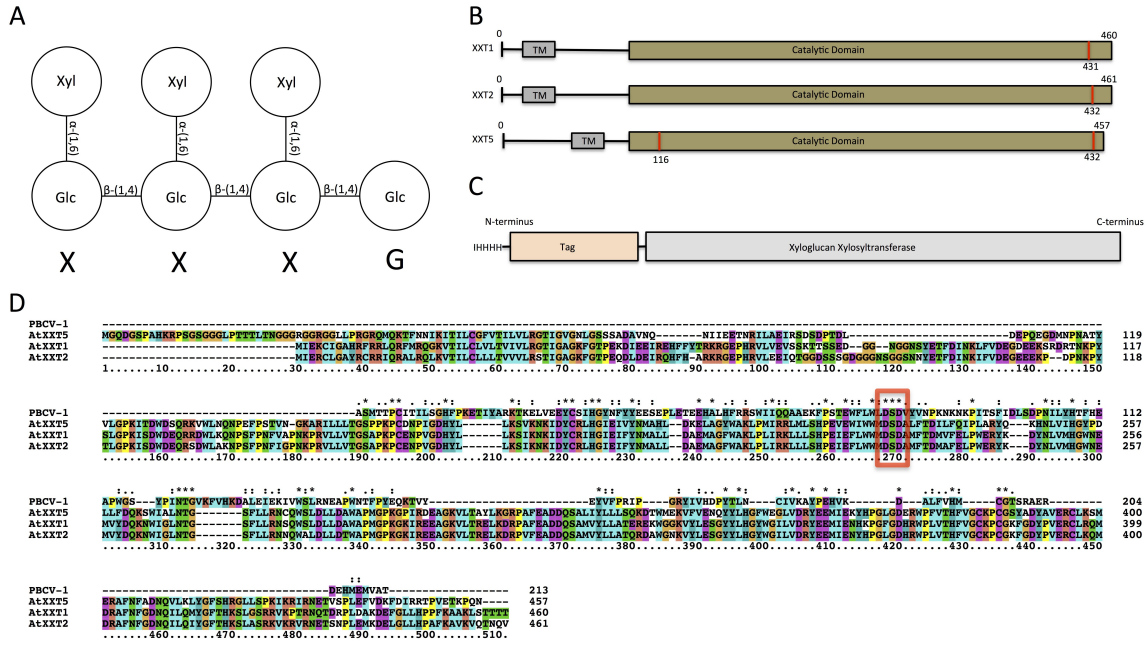
It was proposed that XXT5 might require pre-xylosylated product formed by XXT1 or XXT2 to add D-xylose onto the third D-glucose in the XXGG subunit to form an XXXG oligosaccharide (Cavalier et al., 2008). The results in our study do not support this hypothesis. First, both XXT1 and XXT2 can xylosylate cellohexaose at three positions, as shown here (Fig. 2) and reported earlier (Cavalier and Keegstra, 2006). Second, characterization of the reaction products by XEG digestion shows that both XXT2 and XXT5 add the first D-xylose at the same position on the cellohexaose acceptor (Fig. 4). XXT5 was also able to add the second D-xylose when the enzyme was present at high concentration and higher amount of donor substrate was provided (Fig. 5A). Third, in the conditions used in this study, XXT5 in combination with either XXT1 or XXT2 did not significantly increase the xylosylation of cellohexaose compared to the sum of the single reactions (Fig. 2D). Finally, XXT5 showed reduced activity when xylosylated cellohexaose was used as an acceptor compared to non-xylosylated cellohexaose (Fig. 2E) and this is similar to what was shown before for XXT1 and XXT2 (Faure et al., 2007). This suggests that XXT5 is not specific for the products of XXT1 or XXT2. However, it is possible that XXT5 might require the acceptor with higher degree of polymerization or one containing the XLGG subunit. Further work will be needed to investigate this possibility.

XXT5 can form strong heterocomplexes with XXT2 and CSLC4 (Chou et al., 2012, 2015; Lund et al., 2015). Protein-protein interaction between XXT5 and XXT2

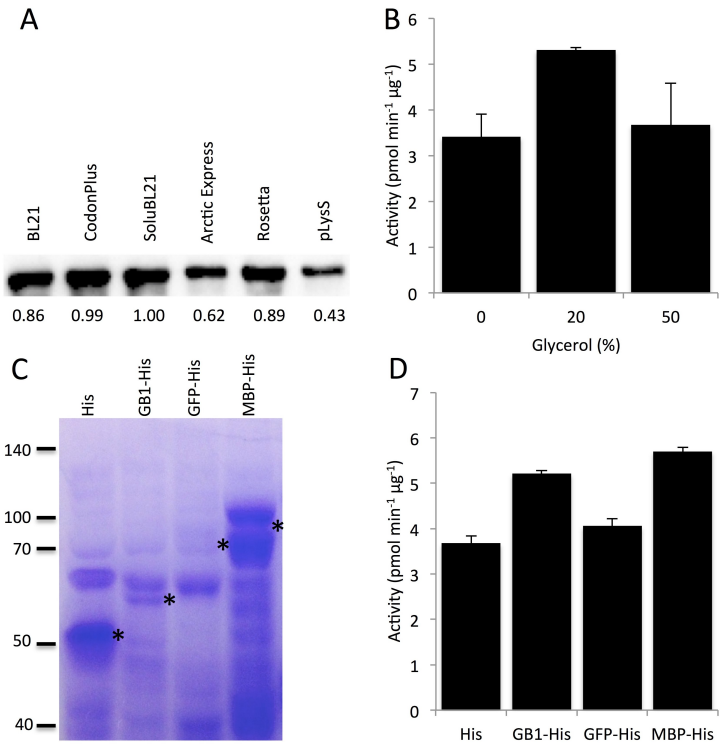
or CSLC4 could produce a conformational change in XXT5 to increase its activity. Combination assays with XXT2+XXT5 did not increase xylosylation compared to the sums of the XXT2 and XXT5 single-enzyme reactions (Fig. 2D). This indicates that the protein-protein interaction with XXT2 does not increase the activity of XXT5. Further work is needed to investigate if the interaction between XXT5 and CSLC4 increases XXT5 activity.

Our mutagenesis studies showed that the XXT5 D228A:D230A mutant has no detectable activity *in vitro*, demonstrating the importance of the DxD motif for the catalytic activity of XXT5. The DxD motif functions to coordinate the divalent cation that forms ionic interactions with the negatively charged diphosphate moiety in UDP-Xylose, and is critical for catalytic activity in numerous other glycosyltransferases (Busch et al., 1998; Wiggins and Munro, 1998; Li et al., 2001; Gotting et al., 2004). To determine the functional role of XXT5 *in vivo*, we over-expressed the XXT5 D228A:D230A mutant in the *xxt5* knockout mutant plant. Based on the XyG content and root-hair phenotype (Fig. 5), we conclude that the XXT5 D228A:D230A mutant is not functional *in vivo*, indicating that XXT5 function requires the DxD putative donor substrate binding motif. It is possible that XXT5 binds UDP-Xylose and functions to channel the substrate to other XXTs, such as XXT1 or XXT2. This has also been suggested for IRX14, which is involved in xylan biosynthesis, since the DxD motif is required for complementation of the *irx14* mutant phenotype and IRX14 lacks *in vitro* activity (Ren et al., 2014). Our demonstration of the importance of the DxD motif for the function of XXT5 *in vivo* does not exclude the possibility that XXT5 also has a similar channeling function.

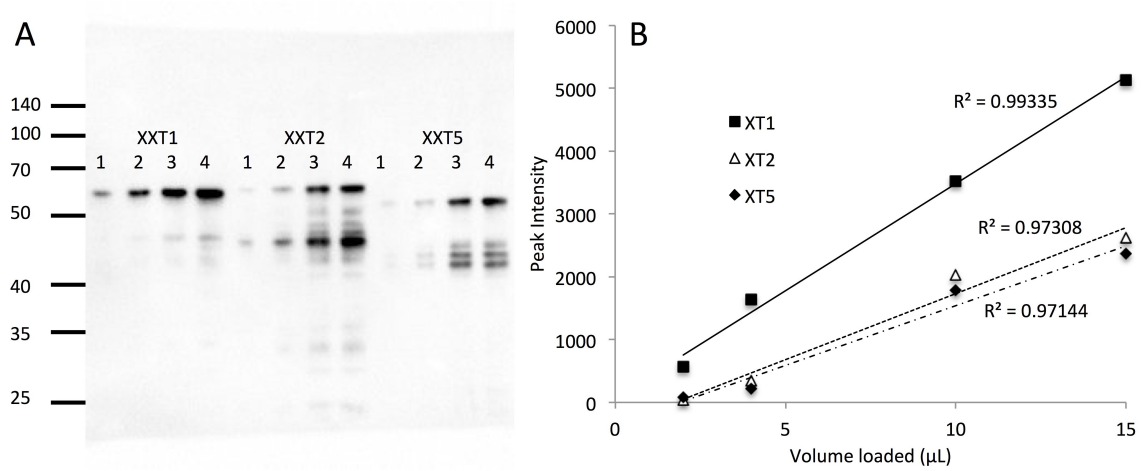
Supplementary Information



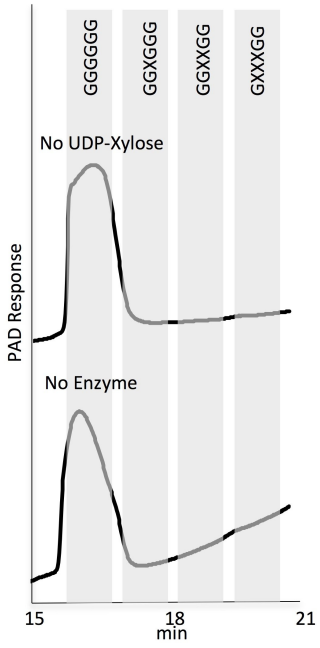
Supplemental Figure S1. Predicted topology of xyloglucan xylosyltransferases according to the Universal Protein Resource (http://www.uniprot.org). A. Diagram of xylosylated glucan subunit showing nomenclature used. B. Positions of the predicted transmembrane domain (TM) and catalytic domain based on primary amino acid sequence. The section between the TM and the catalytic domain corresponds to the stem region of the protein. Red lines indicate the predicted glycosylation sites (Asn-X-Ser/Thr). C. Position of the His-Tag, solubility tag, and cDNA of recombinant proteins. D. Sequence alignment of XXT1, XXT2, XXT5, and PBVC-1 GT (Zhang et al., 2007). Red boxes indicate the positions of mutated amino acids. Sequence alignment was created using Clustal Omega.



Supplementary Figure S2. Optimization of expression conditions for XXT2. A. Western blot of purified His-GB1-XXT2 expressed in various cell lines. Numbers below the bands indicate the relative intensity of the bands as analyzed with ImageJ. B. Effect of glycerol on His-GB1-XXT2 activity. All assay reactions contain equal amounts of total protein. C. Coomassie stained SDS-PAGE gel of XXT2 expressed with various solubility tags. All constructs were expressed in SoluBL21 cells and were purified with Ni-NTA columns. Proteins were loaded in equal volumes. XXT2 protein with the various tags is shown with asterisks. Expected sizes are His-XXT2: 49.5 kDa; His-GB1-XXT2: 56.9 kDa; His-GFP-XXT2: 76.4 kDa; His-MBP-XXT2: 91.2 kDa. D. Enzyme activity of purified XXT2 fused with various tags. All reactions contained an equal volume of the elution fraction collected from Ni-NTA resin. Products of the enzyme assays in B and D were analyzed by HPAEC, quantified by peak integration, and are presented as pmol xylose transferred per minute per µg of XXT2. All enzyme assays were performed in duplicate.



Supplemental Figure S3. Quantitative western of XXT1, XXT2, and XXT5. A. Western blot of XXT1, XXT2, and XXT5. Concentrated protein was diluted to the appropriate level for detection on western blot, XXT1 was diluted 70-fold, XXT2 was diluted 32-fold, and XXT5 was diluted 20-fold. All diluted protein samples were loaded at four volumes, 2, 4, 10, and 15 µL. B. Band intensity was plotted against volume loaded on the SDS-PAGE gel. Band intensity of full-length proteins was determined using ImageJ. The solid line indicates XXT1 line, the dashed line indicates XXT2 , and the dashed line with large spaces indicates XXT5.



Supplemental Figure S4. Negative Controls of XXT2 activity assay. Negative control reactions were done as described in Material and Methods with either no UDP-xylose or no XXT2 in the activity assay mixture.

Supplemental Table S1. Forward and reverse primers used for the amplification of the XXTs and the truncations of XXT2.

Gene	Direction	Organism	Sequence
XXT1	forward	Bacterial	ATA GGA TCC GAG AAC CTG TAC TTT CAG GGC ACG CCG GAG AAA GAT ATC
XXT1	reverse	Bacterial	ATA GTC GAC TCA CGT CGT CGT CG
XXT2	forward	Bacterial	ATA GGA TCC GAG AAC CTG TAC TTT CAG GGC AAA TTC GGA ACT CCG GAG
XXT2	reverse	Bacterial	GCA CTC GAG TCA AAC TTG ATT GGT
XXT5	forward	Bacterial	ATA AGA TCT GAG AAC CTG TAC TTT CAG GGC AAC CTA GGA AGC TCA AGC GCC GAT
XXT5	reverse	Bacterial	ATA CTC GAG CTA GTT CTG TGG TTT GGT
XXT5 D228A:D230A	forward	Bacterial	GATCTGGTGGATGGCTAGTGCTGCTTTGTTCACT
XXT5 D228A:D230A	reverse	Bacterial	AGTGAACAAAGCAGCACTAGCCATCCACCAGATC
XXT5	forward	Plant	AATAAGCTTATGGGTCAAGATGGTTC
XXT5	reverse	Plant	ACCATATGCTAGTTCTGTGGTTTGGTTTCCAC
XXT5 D228A:D230A	forward	Plant	GTTGATGTTGTCTCATCCAGAAGTTGAGTGGATCTGGTGGATGGCTAGTGCTGCTTTGTTC
XXT5 D228A:D230A	reverse	Plant	CCACCAGATCCACTCAACTTCTGGATGAGACAACATCAAC

Acknowledgements: We would like to thank Dr. Amy Andreotti for the pET20b-GB1 and pET20b-MBP plasmids. This research was supported by the NSF-MCB grant #1121163. RF and SC were supported by ICMG FR 2607, PolyNat Carnot Institute, and LabEx ARCANE (ANR-11-LABX-0003-01)

References

Abramson M, Shoseyov O, Shani Z (2010) Plant cell wall reconstruction toward improved lignocellulosic production and processability. Plant Sci **178:** 61-72

Atmodjo MA, Hao Z, Mohnen D (2013) Evolving views of pectin biosynthesis. Annu Rev Plant Biol **64:** 747-779

Bobovska A, Tvaroska I, Kona J (2014) A theoretical study on the catalytic mechanism of the retaining alpha-1,2-mannosyltransferase Kre2p/Mnt1p: the impact of different metal ions on catalysis. Org Biomol Chem **12:** 4201-4210

Boeggeman E, Qasba PK (2002) Studies on the metal binding sites in the catalytic domain of beta1,4-galactosyltransferase. Glycobiology **12:** 395-407

Boyken SE, Fulton DB, Andreotti AH (2012) Rescue of the aggregation prone Itk Pleckstrin Homology domain by two mutations derived from the related kinases, Btk and Tec. Protein Sci **21:** 1288-1297

Bradbury SL, Jakoby WB (1972) Glycerol as an Enzyme-Stabilizing Agent: Effects on Aldehyde Dehydrogenase. Proc Natl Acad Sci USA **69:** 2373-2376

Busch C, Hofmann F, Selzer J, Munro S, Jeckel D, Aktories K (1998) A Common Motif of Eukaryotic Glycosyltransferases Is Essential for the Enzyme Activity of Large Clostridial Cytotoxins. J Biol Chem **273:** 19566-19572

Carpita NC (2012) Progress in the biological synthesis of the plant cell wall: new ideas for improving biomass for bioenergy. Curr Opin Biotechnol **23:** 330-337

Carpita NC, Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. Plant J **3:** 1-30

Cavalier DM, Keegstra K (2006) Two xyloglucan xylosyltransferases catalyze the addition of multiple xylosyl residues to cellohexaose. J Biol Chem **281:** 34197-34207

Cavalier DM, Lerouxel O, Neumetzler L, Yamauchi K, Reinecke A, Freshour G, Zabotina OA, Hahn MG, Burgert I, Pauly M, Raikhel NV, Keegstra K (2008) Disrupting two Arabidopsis thaliana xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component. Plant Cell **20:** 1519-1537

Chalfie M (1995) Green Fluorescent Protein. Photochem and Photobiol **62:** 651-656

Campbell J, Davies G, Bulone V, Henrissat B (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem J **326:** 929-939

Chou Y, Pogorelko G, Young ZT, Zabotina OA (2015) Protein-Protein Interactions Among Xyloglucan-Synthesizing Enzymes and Formation of Golgi-Localized Multiprotein Complexes. Plant Cell Physiol, **56**: 255-267

Chou YH, Pogorelko G, Zabotina OA (2012) Xyloglucan xylosyltransferases XXT1, XXT2, and XXT5 and the glucan synthase CSLC4 form Golgi-localized multiprotein complexes. Plant Physiol **159:** 1355-1366

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16:** 735-743

Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, Raikhel N, Wilkerson CG (2007) A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. Proc Natl Acad Sci USA **104:** 8550-8555

Cotovio J, Roguet R, Pion FX, Rougier A, Leclaire J (1996) Effect of imidazole derivatives on cytochrome P-450 enzyme activities in a reconstructed human epidermis. Skin Pharmacol **9:** 242-249

Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An Evolving Hierarchical Family Classification for Glycosyltransferases. J Mol Biol **328:** 307-317

Endler A, Persson S (2011) Cellulose synthases and synthesis in Arabidopsis. Mol Plant **4:** 199-211

Faik A, Price NJ, Raikhel NV, Keegstra K (2002) An Arabidopsis gene encoding an alpha-xylosyltransferase involved in xyloglucan biosynthesis. Proc Natl Acad Sci USA **99:** 7797-7802

Fauré R, Cavalier D, Keegstra K, Cottaz S, Driguez H (2007) Glycosynthase-Assisted Synthesis of Xylo-Gluco-Oligosaccharide Probes for α-Xylosyltransferases. Eur J Org Chem **2007:** 4313-4319

Francis DM, Page R (2010) Strategies to optimize protein expression in E. coli. Curr Protoc Protein Sci **Chapter 5:** Unit 5 24 21-29

Fry SC, York WS, Albersheim P, Darvill A, Hayashi T, Joseleau JP, Kato Y, Lorences EP, Maclachlan GA, McNeil M (1993) An unambiguous nomenclature for xyloglucan‐derived oligosaccharides. Physiologia Plantarum **89:** 1-3

Gotting C, Muller S, Schottler M, Schon S, Prante C, Brinkmann T, Kuhn J, Kleesiek K (2004) Analysis of the DXD motifs in human xylosyltransferase I required for enzyme activity. J Biol Chem **279:** 42566-42573

Hayashi T (1989) Xyloglucans in the primary cell wall. Annu Rev Plant Biol **40:** 139-168

Hayashi T, Kaida R (2011) Functions of xyloglucan in plant cells. Mol Plant **4:** 17-24

Huth JR, Bewley CA, Jackson BM, Hinnebusch AG, Clore GM, Gronenborn AM (1997) Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. Protein Sci **6:** 2359-2364

Jensen JK, Johnson NR, Wilkerson CG (2014) Arabidopsis thaliana IRX10 and two related proteins from psyllium and Physcomitrella patens are xylan xylosyltransferases. Plant J **80:** 207-215

Josè-Estanyol M, Puigdomènech P (2000) Plant cell wall glycoproteins and their genes. Plant Physiol and Biochem **38:** 97-108

Keegstra K (2010) Plant Cell Walls. Plant Physiol **154:** 483-486

Kuzmic P (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. Anal Biochem **237:** 260-273

Lairson LL, Henrissat B, Davies GJ, Withers SG (2008) Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem **77:** 521-555

Leandro P, Lechner MC, Tavares de Almeida I, Konecki D (2001) Glycerol Increases the Yield and Activity of Human Phenylalanine Hydroxylase Mutant Enzymes Produced in a Prokaryotic Expression System. Mol Genet Metab **73:** 173-178

Li J, Rancour DM, Allende ML, Worth CA, Darling DS, Gilbert JB, Menon AK, Young WW, Jr. (2001) The DXD motif is required for GM2 synthase activity but is not critical for nucleotide binding. Glycobiol **11:** 217-229

Lund CH, Bromley JR, Stenbaek A, Rasmussen RE, Scheller HV, Sakuragi Y (2015) A reversible Renilla luciferase protein complementation assay for rapid identification of protein-protein interactions reveals the existence of an interaction network involved in xyloglucan biosynthesis in the plant Golgi apparatus. J Exp Bot **66:** 85-97

Madson M (2003) The MUR3 Gene of Arabidopsis Encodes a Xyloglucan Galactosyltransferase That Is Evolutionarily Related to Animal Exostosins. Plant Cell **15:** 1662-1670

Mansoori N, Schultink A, Schubert J, Pauly M (2015) Expression of heterologous xyloglucan xylosyltransferases in Arabidopsis to investigate their role in determining xyloglucan xylosylation substitution patterns. Planta **241**: 1145-1158

McCann MC, Carpita NC (2008) Designing the deconstruction of plant cell walls. Curr Opin Plant Biol **11:** 314-320

McGee DJ, Zabaleta J, Viator RJ, Testerman TL, Ochoa AC, Mendz GL (2004) Purification and characterization of Helicobacter pylori arginase, RocF: unique features among the arginase superfamily. Eur J Biochem **271:** 1952-1962

Mukerjea R, McIntyre AP, Robyt JF (2012) Potent inhibition of starch-synthase by Tris-type buffers is responsible for the perpetuation of the primer myth for starch biosynthesis. Carbohydr Res **355:** 28-34

Nallamsetty S, Waugh DS (2006) Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. Protein Expr Purif **45:** 175-182

Park YB, Cosgrove DJ (2015) Xyloglucan and its interactions with other components of the growing cell wall. Plant Cell Physiol **56:** 180-194

Pauly M, Andersen LN, Kauppinen S, Kofod LV, York WS, Albersheim P, Darvill A (1999) A xyloglucan-specific endo-beta-1,4-glucanase from Aspergillus aculeatus: expression cloning in yeast, purification and characterization of the recombinant enzyme. Glycobiol **9:** 93-100

Ren Y, Hansen SF, Ebert B, Lau J, Scheller HV (2014) 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 **9:** e105014

Scheller HV, Ulvskov P (2010) Hemicelluloses. Annu Rev Plant Biol **61:** 263-289

Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, Milne J, Osborne E, Paredez A, Persson S, Raab T, Vorwerk S, Youngs H (2004) Toward a systems approach to understanding plant cell walls. Science **306:** 2206-2211

Thompson DS (2005) How do cell walls regulate plant growth? J Exp Bot **56:** 2275-2285

Urbanowicz BR, Pena MJ, Moniz HA, Moremen KW, York WS (2014) Two Arabidopsis proteins synthesize acetylated xylan in vitro. Plant J **80:** 197-206

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. Plant Physiol **153:** 895-905

Vanzin GF, Madson M, Carpita NC, Raikhel NV, Keegstra K, Reiter WD (2002) The mur2 mutant of Arabidopsis thaliana lacks fucosylated xyloglucan because of a lesion in fucosyltransferase AtFUT1. Proc Natl Acad Sci USA **99:** 3340-3345

Vuttipongchaikij S, Brocklehurst D, Steele-King C, Ashford DA, Gomez LD, McQueen-Mason SJ (2012) Arabidopsis GT34 family contains five xyloglucan alpha-1,6-xylosyltransferases. New Phytol **195:** 585-595

Wiggins CA, Munro S (1998) Activity of the yeast MNN1 α-1, 3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. Proc Natl Acad Sci USA **95:** 7945-7950

Yaoi K, Kondo H, Hiyoshi A, Noro N, Sugimoto H, Tsuda S, Miyazaki K (2009) The crystal structure of a xyloglucan-specific endo-beta-1,4-glucanase from Geotrichum sp. M128 xyloglucanase reveals a key amino acid residue for substrate specificity. FEBS J **276:** 5094-5100

Zabotina OA, Avci U, Cavalier D, Pattathil S, Chou YH, Eberhard S, Danhof L, Keegstra K, Hahn MG (2012) Mutations in multiple XXT genes of Arabidopsis reveal the complexity of xyloglucan biosynthesis. Plant Physiol **159:** 1367-1384

Zabotina OA, van de Ven WT, Freshour G, Drakakaki G, Cavalier D, Mouille G, Hahn MG, Keegstra K, Raikhel NV (2008) Arabidopsis XXT5 gene encodes a putative alpha-1,6-xylosyltransferase that is involved in xyloglucan biosynthesis. Plant J **56:** 101-115

Zhang J, Zhang X, Wu C, Lu D, Guo G, Mao X, Zhang Y, Wang DC, Li D, Zou Q (2011) Expression, purification and characterization of arginase from Helicobacter pylori in its apo form. PLoS One **6:** e26205

Zhang Y, Wang PG, Brew K (2001) Specificity and mechanism of metal ion activation in UDP-galactose:beta -galactoside-alpha -1,3-galactosyltransferase. J Biol Chem **276:** 11567-11574

Zhang Y, Xiang Y, Van Etten JL, Rossmann MG (2007) Structure and function of a chlorella virus-encoded glycosyltransferase. Structure **15:** 1031-1039

CHAPTER 4. TRUNCATIONS OF XYLOGLUCAN XYLOSYLTRANSFERASE 2 PROVIDE INSIGHTS INTO THE ROLES OF THE N- AND C-TERMINUS

Modified from a paper published in *Phytochemistry*

Alan T. Culbertson, Adrienne L. Smith, Matthew D. Cook, Olga A. Zabotina

Abstract

Xyloglucan is the most abundant hemicellulose in the primary cell wall of dicotyledonous plants. In Arabidopsis, three xyloglucan xylosyltransferases, XXT1, XXT2, and XXT5, participate in xylosylation of the xyloglucan backbone. Despite the importance of these enzymes, we lack information on their structure and the critical residues required for substrate binding and transferase activity. In this study, we investigated the roles of different domains of XX2 in protein expression and catalytic activity by constructing a series of N- and C-terminal truncations. XXT2 with an N-terminal truncation of 31 amino acids after the predicted transmembrane domain showed the highest protein expression, but truncations of more than 31 residues decreased protein expression and catalytic activity. XXT2 constructs with C-terminal truncations showed increased protein expression but decreased activity, particularly for truncations of 44 or more amino acids. We also used site-directed mutagenesis to investigate six positively charged residues near the C-terminus and found that four of the mutants showed decreased enzymatic activity. We conclude that the N- and C-termini of XXT2 have important roles in protein folding and enzymatic activity: the stem region (particularly the N-terminus of the catalytic

domain) is critical for protein folding and the C-terminus is essential for enzymatic activity but not for protein folding.

Introduction

In plants, the cell wall strengthens the cell and helps it to resist turgor pressure but also can expand to allow cells to grow (Keegstra, 2010; Somerville et al., 2004). In addition, plant biomass has many uses as biofuels and biomaterials and manipulation of cell wall composition can alter the value of plant feedstocks for different industrial applications (Carpita, 2012; McCann and Carpita, 2008). Plant cell walls consist of cellulose, hemicellulose, and pectin; some cell walls also include lignin (Atmodjo et al., 2013; Endler and Persson, 2011; Scheller and Ulvskov, 2010; Vanholme et al., 2010). Xyloglucan (XyG), the most abundant hemicellulose in the primary cell wall of dicotyledonous plants, has been proposed to cross-link cellulose microfibrils or to act as a spacer to prevent aggregation of cellulose microfibrils (Carpita and Gibeaut, 1993; Hayashi, 1989; Thompson, 2005). In Arabidopsis, XyG is composed of a β-(1,4) linked glucan backbone with 50-75% of the glucose residues decorated with α-(1,6) linked xylose residues. These can be further decorated with α-(1,2) linked $_{D}$-Gal or $_{L}$-Fuc-(1,2)-β-$_{D}$-Gal.

Xyloglucan xylosyltransferases (XXTs) catalyze the transfer of xylose residues from UDP-xylose to the glucan backbone of XyG. Three XXTs in Arabidopsis, XXT1, XXT2, and XXT5, xylosylate the glucan backbone both *in vivo* (Cavalier et al., 2008; Zabotina et al., 2012; Zabotina et al., 2008) and *in vitro* (Cavalier and Keegstra, 2006; Faik et al., 2002; Vuttipongchaikij et al., 2012). XXTs catalyze the formation of an α-(1,6) glycosidic bond that retains the anomeric configuration of UDP-xylose. The XXTs

are type II transmembrane proteins consisting of a short cytosolic N-terminal, a transmembrane domain, a stem-region, and a large C-terminal catalytic domain localized in the lumen of the Golgi. All XXTs belong to the glycosyltransferase (GT) family 34, which are predicted to have a GT-A fold based on their primary amino acid sequence (Lairson et al., 2008). XXTs, along with all other enzymes involved in xyloglucan biosynthesis, physically interact, forming homo- and hetero-complexes (Chou et al., 2015; Chou et al., 2012; Lund et al., 2015).

Analysis of knock-out mutants in Arabidopsis demonstrated that either XXT1 or XXT2 must be present for xylosylation of the xyloglucan backbone. The *xxt1 xxt2* double knock-out mutant has no detectable xyloglucan; these mutants also have defects in root hairs (Cavalier et al., 2008; Zabotina et al., 2012). By contrast, the *xxt1* and *xxt2* single mutants have phenotypes similar to the wild type, indicating that XXT1 and XXT2 have partially redundant functions. Enzymatic assays showed that XXT1 and XXT2 have xylosyltransferase activity *in vitro* when cellohexaose is used as an acceptor and these enzymes primarily xylosylate the fourth glucose from the reducing end of cellohexaose, producing mono-xylosylated cellohexaose (Cavalier and Keegstra, 2006). XXT1 and XXT2 can also use the mono-xylosylated cellohexaose as an acceptor, adding a second xylose on the third glucose from the reducing end and producing di-xylosylated cellohexaose. Similarly, XXT1 and XXT2 can add a third xylose producing tri-xylosylated cellohexaose. The enzyme XXT5 also participates in xyloglucan biosynthesis. The *xxt5* knock-out mutant has a 50% reduction in xyloglucan content with a similar root hair phenotype to that of the *xxt1 xxt2* double mutant (Zabotina et al., 2012; Zabotina et al., 2008), but the catalytic activity of XXT5 has yet to be confirmed *in vitro*.

Plant cell wall biosynthetic enzymes have been extensively studied using reverse-genetics and heterologous expression to characterize their enzymatic activity. Despite this, we have little information about their structures or the residues involved in substrate binding and catalysis, mainly due to the low solubility of these enzymes and the lack of suitable acceptor substrates for *in vitro* activity assays. In this study, we used N- and C-terminal truncations of XXT2 to determine the minimal size required for full catalytic activity. In addition, examination of protein expression levels and activity of N- and C-terminally truncated XXT2 proteins provide insight to the roles of these protein domains in catalysis, protein folding, and substrate binding.

Results and Discussion

Sequence alignment of Arabidopsis XXTs and homologous enzymes. XXT2 is predicted to contain a short cytosolic N-terminus, a transmembrane domain, a stem region, and a large catalytic domain (Fig. 1 and Fig. 2). The hydropathy profile of XXT2 indicates that the transmembrane domain is between residues 21 and 42 (Fig. 1). The C-terminus has two hydrophobic regions, between residues 401-422 and 442-462 (Fig. 1). In addition, the XXT2 C-terminus contains a highly polar region between residues 423 and 440 (Fig. 1 and Fig. 2).

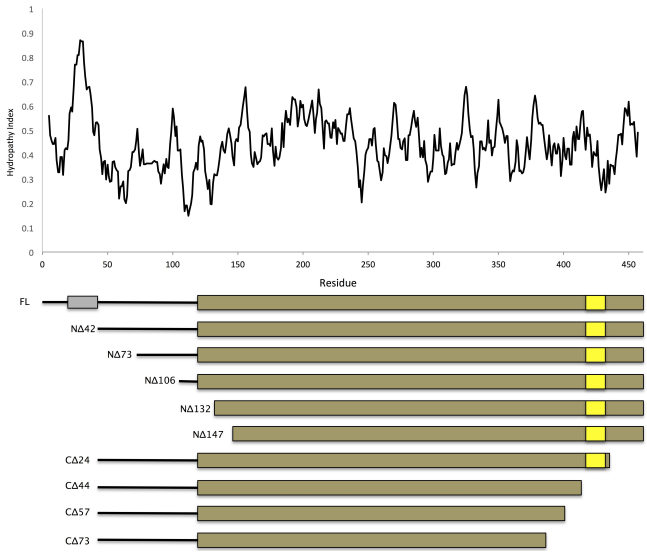


Figure 1. Hydropathy profile and predicted schematic topology of XXT2. (A) The hydropathy plot was prepared with the Kyte-Doolittle method (Kyte and Doolittle, 1982) with window setting of 9. (B) Schematic topology of full-length and truncated XXT2. FL: full length. NΔ42 indicates the truncation of the N-terminal domain and transmembrane domain. All C-terminal truncations also contain the NΔ42 truncation. The yellow box indicates the region with positively charged residues, the grey box indicates the transmembrane domain, and the brown region indicates the conserved catalytic region.

Most available plant genomes encode homologs of Arabidopsis XXT2, but only a few of these homologs have been cloned and functionally studied. For example, XXT2 homologs from *Pinus radiata* (Ade et al., 2014), *Oryza sativa* (Wang et al., 2014), and *Solanum lycopersicum* (Mansoori et al., 2015) were able to complement the *xxt1 xxt2* phenotype in Arabidopsis. To investigate the conservation of the predicted domains of XXT proteins, we aligned the sequences of all five Arabidopsis XXTs from GT family 34 (Faik et al., 2002) and their homologs from *P. radiata, O. sativa*, and *S. lycopersicum*. This alignment showed that the transmembrane and catalytic domains have high sequence similarity (Fig. 2), which is common for glycosyltransferases belonging to the same GT family (Hennet et al., 1998; Ihara et al., 1993). The cytosolic N-terminus and stem region of all XXTs have low sequence similarity and different lengths (Fig. 2). Stem

regions of different lengths have been reported previously for other glycosyltransferases, even for functionally homologous proteins (Hennet et al., 1998; Ihara et al., 1993). In addition, XXT5 homologs have a larger N-terminus, consisting of approximately 40-50 amino acids compared with the XXT1 and XXT2 N-terminus consisting of 22-25 amino acids. The sequence alignment also showed that the XXTs have numerous positively charged residues (Fig. 2) in the polar region near the C-terminus (Fig. 1). Numerous enzymes in GT family 8, such as WaaJ (Leipold et al., 2007) and α3GT (Henion et al., 1994), have positively charged residues located near the C-terminus and these residues are important for protein function.

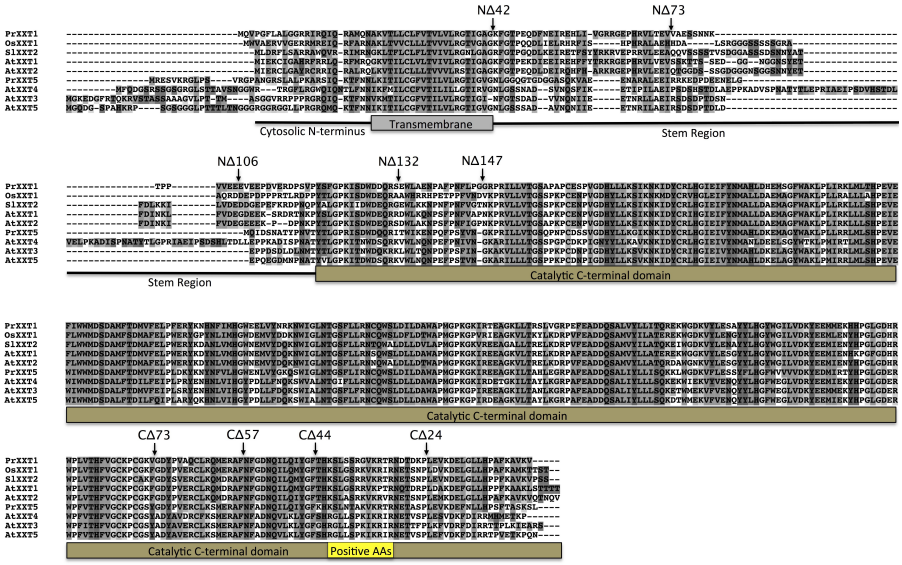


Figure 2. Sequence alignment of Arabidopsis XXTs and their functional homologs. Sequence alignment was done using Clustal Omega (Sievers et al., 2011). Domains shown below the sequence alignment are those predicted for XXT2. The yellow box indicates the positively charged amino acids described in the text. Genes used for sequence alignment were *AtXXT1* (At3g62720), *AtXXT2* (At4g02500), *AtXXT3* (At5g07720), *AtXXT4* (Atg18690), *AtXXT5* (At1g74380), *PrXXT1* (PrGT34B), *PrXXT5* (PrGT34C), *SiXXT2* (Si01g067930), and *OsXXT1* (Os03g18820).

Truncations of the XXT2 N-terminus. The transmembrane domain of XXT proteins is buried in the hydrophobic bilayer of the Golgi membrane and has low solubility in solution. To circumvent this, Vuttipongchaikij et al. (2012) used truncated proteins containing only the stem region and catalytic domain to express catalytically active recombinant proteins (Vuttipongchaikij et al., 2012). However, that study reported very low expression levels of recombinant proteins. Here, we examined the effect of N-terminal truncations into the stem region (NΔ73 and NΔ106) and catalytic domain (NΔ132 and NΔ147) on the expression and activity of recombinant XXT2 (Fig. 1). XXT2 was expressed with a streptococcal protein G immunoglobulin-binding domain (GB1) tag to improve solubility and a 6xHis tag to allow purification of the proteins by affinity chromatography.

XXT2 without the N-terminus and transmembrane domain (NΔ42) and the various truncated mutants were expressed in *E. coli* SoluBL21 cells and purified using Ni-NTA column as described in the Methods section. Protein expression of this construct showed high yield of soluble His-GB1-XXT2 (Fig. 3), yet most His-GB1-XXT2 was in the insoluble fraction following cell lysis (data not shown). When analyzed by western blot, the elution fraction of XXT2 contained two additional smaller bands that were approximately 8 and 9 kDa smaller than the full-length NΔ42 protein (Fig. 3A). These likely result from degradation of XXT2 because the different N-terminal truncations and the full-length proteins showed similar shifts in band size. This possibility was not explored further; however, these bands were not observed in the XXT2 C-terminal truncations, which can argue against the idea that these bands result from non-specific detection of bacterial proteins co-purified with XXT2.

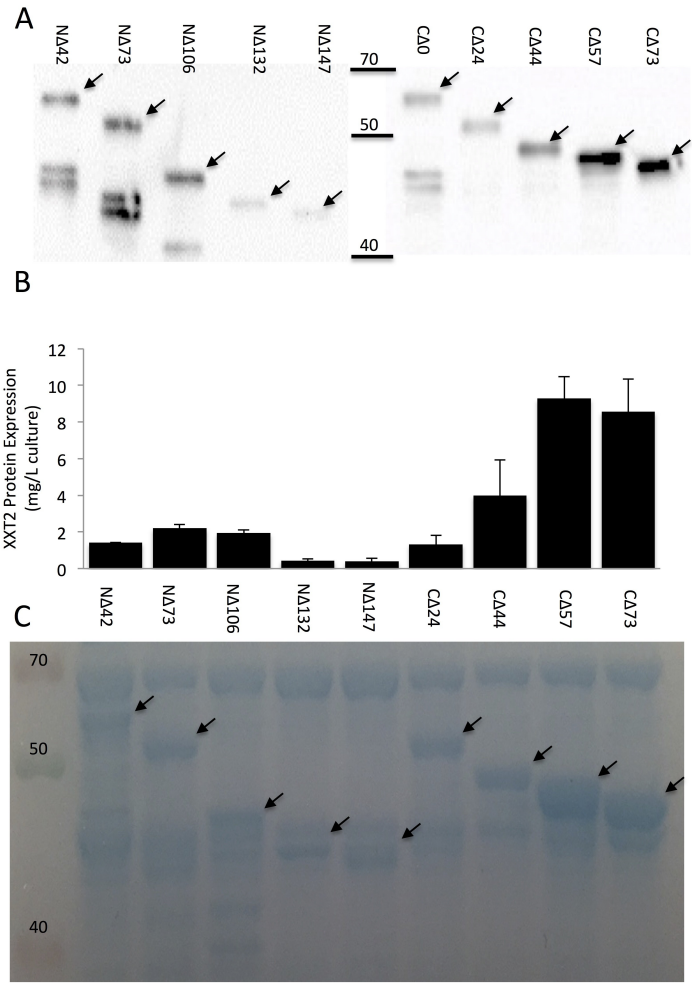


Figure 3. Protein expression levels of XXT2 truncations. Truncated proteins were expressed and purified in parallel using identical conditions. Proteins were loaded onto the SDS-PAGE gel in equal volumes of the corresponding purified elution factions. Quantification of protein expression was done using only the non-degraded XXT2 bands with expected sizes, as indicated by the black arrows. (A) Western blot of N- and C-terminal truncations detected using horseradish peroxidase-conjugated goat anti-rabbit antibody at a 10,000-fold dilution. NΔ42 in left blot is same construct as CΔ0 in right blot. (B) Protein expression of NΔ42 and N- and C-terminal truncations. Quantitation was preformed with ImageJ. The error bars represent the average ± S. D. of three replicates. (C) Coomassie Blue stained SDS-PAGE of NΔ42 and N- and C-terminal truncations. Black arrows indicate non-degraded XXT2 bands with expected sizes.

The amount of XXT2 NΔ42 protein in the elution fraction was estimated by measuring the amount of total protein by Bradford assay and the intensity of the band corresponding to XXT2 compared to the intensity of all bands on Coomassie-stained

SDS-PAGE gels using ImageJ software (Fig. 3C). Expression of XXT2 NΔ42 yielded 1.42 mg/L of cell culture (Fig. 3B). Quantitative western blots showed a linear relationship between XXT2 NΔ42 band intensity and total protein loaded in the gel, demonstrating that western blot can be used for protein quantification (Fig. S1). The expression levels of all truncated proteins were then determined by measuring the intensities of the upper bands (corresponding to the non-degraded proteins) on the western blot using ImageJ and comparing them to the intensity of the XXT2 NΔ42 band. Truncations into the stem region (NΔ73 and NΔ106) resulted in protein expression close to that of NΔ42 with maximal expression when 73 residues were truncated. Truncations into the conserved catalytic domain (NΔ132 and NΔ147) resulted in a severe reduction in protein expression with a 3.5-fold reduction for the both proteins (Fig. 3B).

SDS-PAGE analysis of XXT2 NΔ42 and all the XXT2 truncated mutants purified with Ni-NTA column showed that they co-purified with a 68 kDa protein (Fig. 3C). This protein was not removed even with multiple rounds of affinity chromatography or gel-filtration chromatography (data not shown). Previous work showed that the bacterial chaperonin GroEL can persistently co-purify with glycosyltransferases expressed in *E. coli* cells (Hou et al., 2004), including XXTs (Vuttipongchaikij et al., 2012). Here, the NΔ73 truncation resulted in lowest intensity of this 68 kDa band, and increasing truncation size further increased the intensity of the putative GroEL band (Fig. 3C). GroEL mediates protein folding by binding hydrophobic regions of misfolded proteins (Lin and Rye, 2006). Thus, it is possible that N-terminal truncations into the catalytic domain disrupt protein folding, resulting in more exposed hydrophobic residues, which

are then bound by GroEL. However, this speculation will need to be experimentally confirmed in the future.

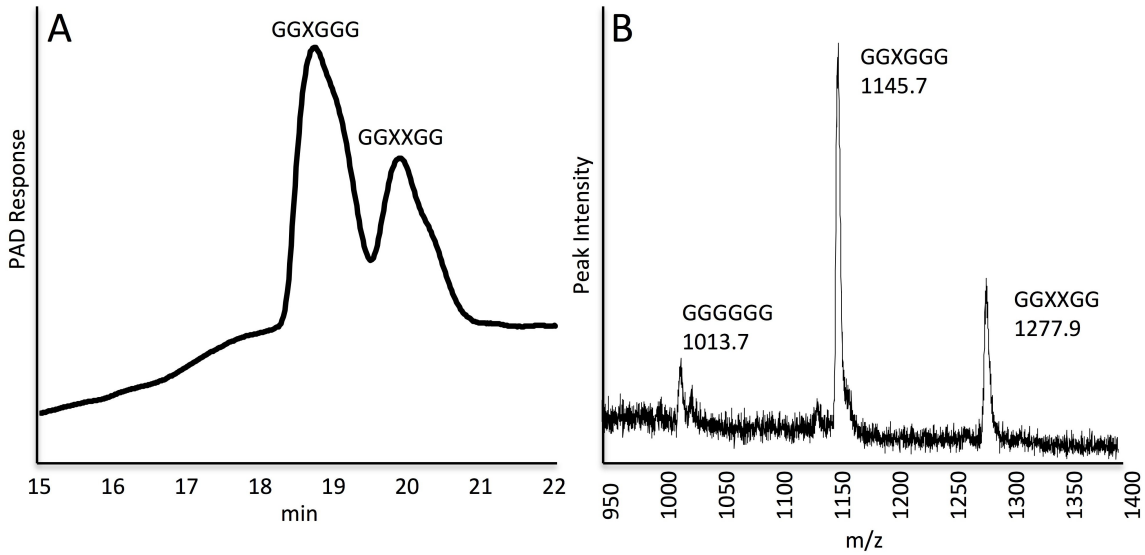


Figure 4. Activity of XXT2 NΔ42. Enzyme assays were performed in a 25 µL reaction for 20 h at 28°C and contained 2 mM UDP-Xylose, 0.5 mM cellohexaose, 2 mM $MnCl_2$, and 1.5 µM XXT2 NΔ42. (A) HPAEC profile of XXT2 NΔ42 reaction products. (B) MALDI-TOF analysis of the reaction product. GGGGGG: cellohexaose; GGXGGG: mono-xylosylated cellohexaose; GGXXGG: di-xylosylated cellohexaose.

Transferase activity of N-terminal truncated XXT2 proteins was assayed as described in the Methods section, using UDP-xylose and cellohexaose as donor and acceptor substrates, respectively. The products were quantified by peak integration on chromatograms obtained by HPAEC and their molecular weights were verified by MALDI-TOF analysis (Fig. 4A and Fig. 4B). Several control reactions were also performed to confirm that xylosylation of the acceptor glucan occurs only in the presence of XXT2 (data not shown). A previous study showed that XXT2 transfers the first xylose onto non-xylosylated cellohexaose; XXT2 only begins to transfer a second xylose onto mono-xylosylated glucan acceptors when it has xylosylated most of the cellohexaose (Cavalier and Keegstra, 2006). Performing the activity assays for different durations demonstrated that the maximum amount of xylose transferred onto glucan was after 20 h

of reaction (Fig. 5A). Addition of the second xylose onto glucan was initiated only after 4 hours of reaction and 33% of mono-xylosylated acceptor was converted into di-xylosylated during the 20-h reaction (Fig. 5A). Therefore, we measured the activity of truncated proteins with 20 h reactions to investigate the maximum amount of xylose transferred by the XXT2 truncated proteins. We choose these conditions because many of the XXT2 truncated proteins resulted in undetectable xylosylated product with shorter reaction times due to their severely reduced activity. In addition, these conditions demonstrating the effect of the truncations on the ability of XXT2 to xylosylate cellohexaose at one or two positions. Longer reactions times are required for investigation of xylosylation of mono-xylosylated cellohexaose by XXT2 due to the slower rate of catalysis on this acceptor substrate (Fauré et al., 2007).

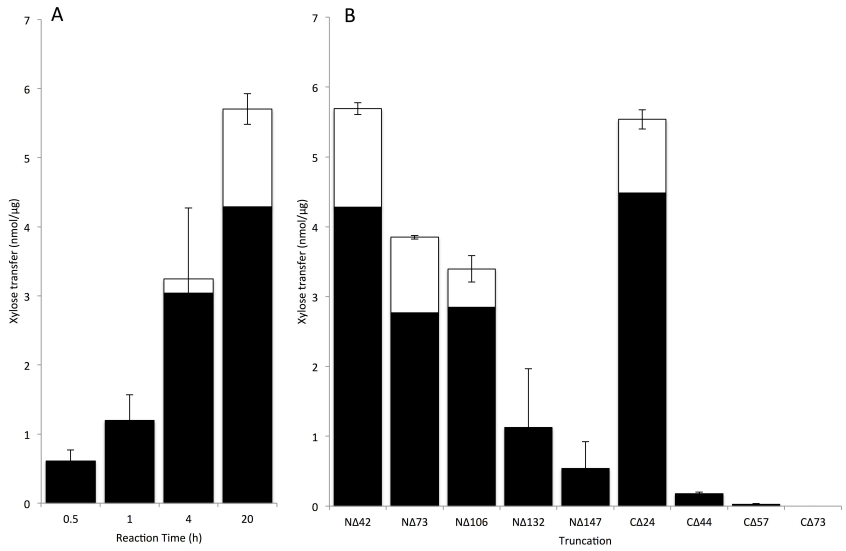


Figure 5. Activity of truncated XXT2 proteins and XXT2 NΔ42 at various time-points. The activity assays were performed as described in the Materials and methods section. (A) Amount of xylose transferred from UDP-Xylose onto cellohexaose by XXT2 in reactions of different durations. All assays contained 1.5 µM XXT2 NΔ42. (B) Activity of XXT2 truncations. Enzyme activity is shown as the amount of xylose transferred to cellohexaose by 1 µg of XXT2 protein after 20 h of incubation at 28°C. All proteins were expressed, purified, and assayed with identical conditions. Peak corresponding to xylosylated cellohexaose was quantified by peak integration on the HPAEC profile. The error bars represent the average ± S. D. of two replicates.

The N-terminal truncation mutants of XXT2 showed a decrease in activity with increasing truncation size (Fig. 5B). In the case of NΔ132 and NΔ147, this drop of activity was particularly strong and correlated with low expression of these mutants. To investigate whether the small amount of reaction product formed by NΔ132 and NΔ147 proteins was due to low protein expression or reduction of their activity, both proteins were concentrated approximately 6-fold and assayed for activity. A slight increase in activity was observed when the truncated proteins were concentrated (Fig. S2). These results suggest that N-terminal truncations of the stem region do not affect protein expression but to some extent affect enzymatic activity, whereas truncation into the catalytic domain severely affects both protein expression and catalytic activity. This is similar to the observations reported for the GT-A fold glycosyltransferase, α3GT (Henion et al., 1994). N-terminal truncations into the stem region of α3GT (residues 23-89) did not result in a significant change in enzymatic activity, but truncations into the catalytic domain severely reduced protein activity. In addition, in the crystal structure of the GT-A fold glycosyltransferase, XXYLT1, the stem region was disordered (Yu et al., 2015). This indicates that the XXYLT1 stem region might not participate in enzymatic activity, but rather might allow the enzyme to move close to or away from the membrane (Yu et al., 2015). From the results obtained here for N-terminal truncations of XXT2, we propose that the stem region of XXT2 also likely does not participate in catalysis, but assists in mobility of the protein catalytic domain or possibly in the accessibility of the acceptor substrate to the catalytic center of the protein (Fig. 6). Furthermore, N-terminal truncation into the catalytic domain likely affects protein folding, which results in both a

significant decrease in expression of the soluble protein in *E. coli* cells and in enzymatic activity. Further studies will be needed to reveal particular residues that are critical for proper protein folding and activity.

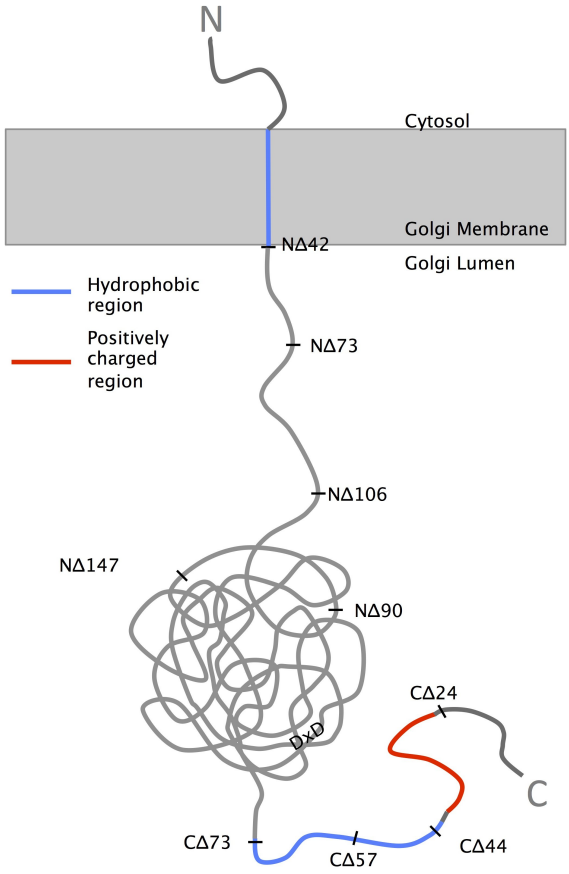


Figure 6. Schematic representation of the proposed topology of XXT2. The N-terminus of XXT2 localizes in the cytosol and the C-terminus localizes in the Golgi lumen. Hydrophobic regions are shown in blue and the positively charged region is shown in red.

Truncations of the XXT2 C-terminus. The C-terminal truncations of XXT2 were designed based on the predicted regions of hydrophobicity (Fig. 1) and the highly polar region with numerous positively charged residues (Fig. 2). Thus, the CΔ24 truncation removed the first hydrophobic region from the C-terminus of XXT2, CΔ44 removed the

region rich in positively charged residues, CΔ57 removed the second hydrophobic region from the C-terminus, and CΔ73 truncated further into the catalytic domain of XXT2. All C-terminal truncations also contained the NΔ42 truncation. Truncations of 44 or more amino acids from the C-terminus increased protein expression compared to the NΔ42 protein, particularly when 57 or 73 amino acids were truncated (Fig. 3). Despite the increase in protein expression, the CΔ44, CΔ57, and CΔ73 truncations showed reduced catalytic activity of XXT2 (Fig. 5B). It is unlikely that the reduction of activity for these truncations is due to errors in protein folding because partially folded or misfolded proteins usually aggregate, forming inclusion bodies, or are degraded in *E. coli* cells (Baneyx and Mujacic, 2004). Hence, it is more likely that the C-terminus of XXT2 contains critical amino acids that are involved in substrate binding or ordering of the active site.

The results obtained here for C-terminal truncations are very similar to the results reported for the GT-A fold glycosyltransferase WaaJ, which is involved in lipopolysaccharide biosynthesis in *E. coli* (Leipold et al., 2007). Truncations from the C-terminal of WaaJ resulted in increased protein expression and decreased catalytic activity (Leipold et al., 2007). The subcellular localization of the truncated WaaJ proteins was investigated by centrifugation of the whole-cell lysate to separate the soluble and membrane-associated protein fractions. Increases of the size of the C-terminal truncation did not affect protein expression in the whole-cell lysate, but increased the amount of protein in the soluble fraction while decreasing the amount of protein associated with the membrane fraction (Leipold et al., 2007). The C-terminal of WaaJ, like the C-termini of other enzymes in GT family 8, is rich in positively charged residues. Thus, these

positively charged residues could participate in association with the membrane (Leipold et al., 2007). XXT2 also has a region that is rich in positively charged residues, but it is unlikely they are involved in membrane association because, in contrast to WaaJ, XXT2 possesses a transmembrane domain near the N-terminus. In addition, truncation of the region with numerous positively charged residues (CΔ44) resulted in only a slight change in protein expression but resulted in a significant decrease in protein activity (Fig. 3 and Fig. 5).

Another possible explanation for the negative relationship between protein expression and activity is that the XXT2 C-terminus is disordered (Fig. 6). Thus, a truncation that removes this disordered region increases protein expression by removing the solvent-exposed hydrophobic residues predicted in the hydropathy index plot (Fig. 1). This notion is supported by the increase in protein expression when the hydrophobic region near the C-terminus (CΔ57) was removed (Fig. 3). Crystal structure analysis of the apo form of α3GT showed that this glycosyltransferase has a disordered C-terminus (Jamaluddin et al., 2007). UDP binding by α3GT induces a conformational change and ordering of the C-terminus, positioning two positively charged residues, Arg^{365} and Lys^{359}, near the active site to form ionic interactions with the negatively charged oxygen groups of UDP (Boix et al., 2001; Jamaluddin et al., 2007). Mutating either of these positively charged residues severely reduced catalytic activity of α3GT, demonstrating the importance of these residues (Jamaluddin et al., 2007). The six positively charged residues in the region between 24 and 44 amino acids from the C-terminus of XXT2 might have similar roles to those observed for α3GT.

To investigate these six positively charged residues in XXT2, each of the residues was individually mutated to alanine and the resulting mutant proteins were assayed for xylosyltransferase activity. Assays were performed for 1 h (Fig. 7) to compare the activity transferring the first xylose onto the glucan acceptor. Two mutations, $R^{425}A$ and $R^{431}A$ had no effect on enzymatic activity (Fig. 7). Four of the mutants, $K^{420}A$, $K^{426}A$, $K^{428}A$, and $R^{429}A$, had decreased enzymatic activity in transferring first xylose onto non-xylosylated cellohexaose (Fig. 7).

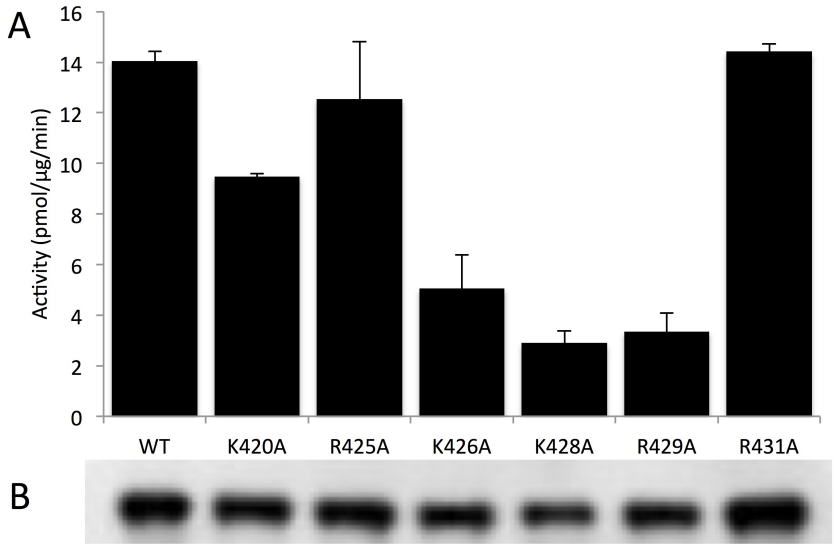


Figure 7. Enzyme activity of XXT2 mutants. All mutant proteins were expressed, purified, and assayed using identical conditions. (A) Enzyme activity of XXT2 NΔ42 and the mutants expressed as an amount of xylose transferred to cellohexaose by 1 μg of protein per min of incubation. The peak corresponding to xylosylated cellohexaose was quantified by peak integration on the HPAEC profile. The error bars represent the average ± S. D. of two replicates. (B) Western blot of NΔ42 and XXT2 mutants.

It is possible that these positively charged residues participate in substrate binding, similar to what was observed for α3GT (Boix et al., 2001; Jamaluddin et al., 2007). Another possibility is these positively charged residues may be involved in the

formation of open and closed states of XXT2. For example, crystal structures of the ABO(H) blood group glycosyltransferases showed that these enzymes are in the open state when no substrate is bound, whereas substrate binding induces a conformational change to the closed state (Johal et al., 2014). It was postulated that the mutual repulsion of the positively charged residues in these disordered loops were responsible for the open state of these enzymes (Johal et al., 2014). Thus, binding of UDP-Donor substrate introduces a negative charge that offsets these repulsions and forms ionic interactions to stabilize the closed state. The positively charged residues on the C-terminus of XXT2 may have a similar role. Further work on XXT2, including structural and kinetic studies, will be needed to verify this possibility and to clarify the involvement of these residues in catalytic activity.

Conclusions

Cell wall polysaccharide synthesizing GTs, including XXTs, have been studied extensively using reverse-genetics and heterologous expression to characterize their function. Despite this, there is no information on their atomic structures, modes of substrate binding, or critical residues involved in catalysis. Here, N- and C-terminal truncations of XXT2 were investigated to determine their role on protein folding and catalysis. N-terminal truncations into the stem-region resulted in minor changes in protein expression or activity, while truncation into the catalytic domain severely reduced protein expression and activity. This indicates that the N-terminal region of the catalytic domain may be critical for proper folding of active recombinant protein. In contrast, C-terminal truncations resulted in an increase in XXT2 expression but a dramatic decrease in its

enzymatic activity. This suggests that removal of amino acids from the C-terminus affects protein activity without impacting its folding. A hydropathy plot and sequence alignment showed the presence of hydrophobic regions and positively charged amino acids on the C-terminus of XXT2. Removal of these regions resulted in a decrease of enzymatic activity. This indicates that the C-terminus, most likely including these positively charged residues, is involved in substrate binding or catalysis. These results suggest that the stem region of XXT2 functions to link the transmembrane and catalytic domain and does not significantly contribute to enzymatic activity, whereas the N-terminal sequence of the catalytic domain is, mostly likely, required for proper folding of active recombinant protein. The C-terminus is likely involved in catalysis or substrate binding and might be disordered due to the increase in expression levels upon its removal (Fig. 5).

Materials and Methods

XXT2 Cloning. The pET20b-GB1 expression vector containing N-terminal 6xHis-GB1 tag and the NdeI and Xho1 restriction sites was obtained from Dr. Andreotti's laboratory at Iowa State University (Boyken et al., 2012). Coding sequences of *Arabidopsis thaliana* xylosyltransferase *XXT2* (At4g02500) were PCR amplified with N- or C-terminal truncations and cloned into the pET20b-GB1 plasmid (Table S1) with a 126 base pair N-terminal truncation to remove the cytoplasmic N-terminus and transmembrane domain. All forward primers were designed to contain the Tobacco Etch Virus protease site (ENLYFQG) on the N-terminus of XXT2. To generate the constructs

encoding the XXT2 mutants, two fragments for each mutant were PCR amplified using primers designed to contain the mutated base pairs (Table S1). The first XXT2 mutant fragment was generated with XXT2 NΔ42 forward and XXT2 mutant reverse primers and the second fragment was generated using XXT2 mutant forward and XXT2 C-term reverse primers. The resulting two fragments containing the mutated base pairs were then used as the template in a fusion PCR reaction using XXT2 NΔ42 forward and XXT2 C-term reverse primers to generate the final mutated gene fragment. The resulting pET20b construct was then transformed into *E. coli* DH10b for amplification of the plasmid and was verified by sequencing. Verified plasmids were then transformed into *E. coli* SoluBL21 for protein expression.

Protein expression. *E. coli* cells harboring the plasmids were grown at 37°C with shaking at 200 rpm in 500 ml of Luria-Bertani broth. When the cell culture reached an OD_{600} of 0.5, the temperature was lowered to 18°C and protein expression was induced by adding IPTG to a final concentration of 0.5 mM. Cells were incubated for 18 h at 18°C and then harvested by centrifugation. Pelleted cells were re-suspended in 12.5 mL of 25 mM Tris-HCl pH 7.4, 300 mM NaCl, 0.1 mM EDTA and rapidly frozen in liquid nitrogen. Cells were lysed by thawing and incubating for 30 min with 1 mg/mL of lysozyme and then sonicated for 15 s a total of five times. Solubilized proteins were collected by centrifugation at 20,000 xg for 30 min.

Protein purification. Crude lysate was loaded onto a Ni-NTA column with a lysate:resin ratio of 20:1. Affinity resin was incubated on a shaker for 1 h at 4°C.

Unbound proteins were removed as a flow-through fraction and the resin was washed four times with washing buffer containing 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, and 20 mM imidazole. The protein of interest was eluted in three fractions with a total volume of 6 mL using elution buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl, and 300 mM imidazole. Glycerol was added to the elution fractions to a final concentration of 20% and proteins were stored at -80°C. Total protein concentration was measured using a Bio-Rad kit (Quick Start Bradford Dye reagent 1X, Cat# 500-0205) according to the manufacturer's instructions.

Enzymatic activity assay and product precipitation. The protein was concentrated from the elution buffer using Amicon Ultra centrifuge filter units (10,000 Da cutoff, Millipore). The enzymatic activity assay was performed in a total volume of 25 µL consisting of 2 mM UDP-xylose (CarboSource Services), 0.5 mM cellohexaose (Megazyme), 2 mM $MnCl_2$, and 1.5 µM XXT2 protein. The reaction mixture was incubated at 28°C for 20 h with shaking at 100 rpm. All reactions were performed in duplicate. The reaction was stopped by adding 900 µL of 100% ethanol and incubating at -22°C for 6 h. The precipitated reaction product was collected by centrifugation at 21,000 xg for 30 min.

HPLC and MALDI-TOF MS characterization of reaction products. Products obtained from the enzymatic activity assays were analyzed by high-performance anion-exchange chromatography with a pulse-amperometric detector (HPAEC-PAD, Dionex),

as described (Cavalier and Keegstra, 2006). Briefly, the aqueous solution of the product was injected onto a CarboPac PA-20 column and eluted using the gradient 30 mM to 100 mM Na-Acetate for 25 min with 100 mM NaOH remaining constant through the entire run; after the 25-min run, the column was re-equilibrated for 15 min to initial conditions. Non-xylosylated, mono-xylosylated, and di-xylosylated cellohexaose eluted from the column with retention times of approximately 17.5, 19, and 20 min, respectively. Quantitation of xylose transfer was performed by peak integration of xylosylated cellohexaose to determine the relative percentage of each product (non-, mono-, or di-xylosated cellohexaose), which allowed us to determine the amount of xylose transfer in the reaction. For example, if 40% of total cellohexaose (12.5 nmol) is mono-xylosylated, then 5 nmol of xylose was transferred to cellohexaose. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry analyses of the products was performed as described (Zabotina et al., 2012). Briefly, one microliter of the product solution was spotted onto a MALDI-TOF sample plate using a 2,5-dihydroxybenzoic acid matrix at a sample:matrix ratio of 1:1. The sample spectra were analyzed on an Applied Biosystems VOYAGER-DE Pro MALDI-TOF mass spectrometry instrument in positive reflection mode with an acceleration voltage of 20 kV and extraction delay time of 350 ns.

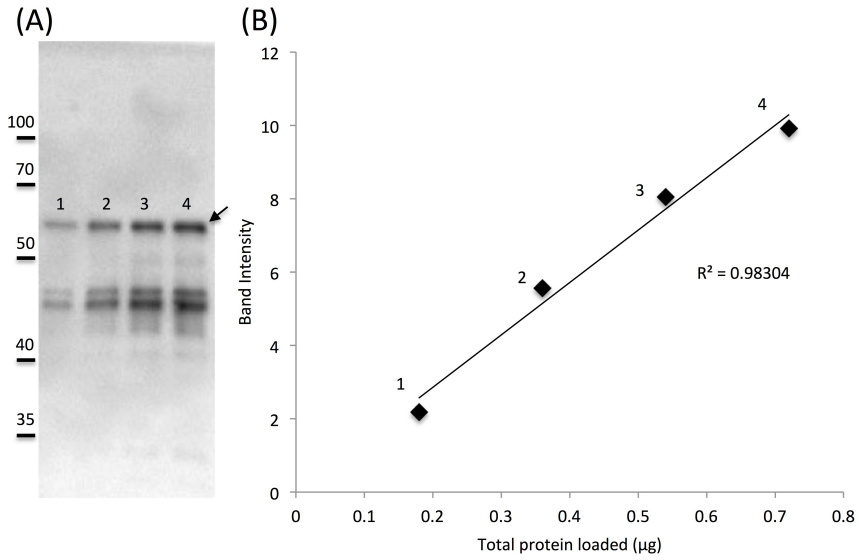
SDS-PAGE and Western Blots. Proteins were analyzed by SDS-PAGE in reducing conditions and stained with Coomassie Blue G250. Western blot analysis was performed as previously described (Chou et al., 2015). Proteins were electrophoretically transferred to a nitrocellulose membrane and blocked with non-fat milk. The GB1 tag

was recognized with horseradish peroxidase-conjugated goat anti-rabbit antibody (Invitrogen) at a 10,000-fold dilution, developed using HyGLO quick spray, and visualized with a ChemiDoc XRS+ (Bio-Rad). Pre-stained size markers were visualized on the same membrane using visible light. Analysis of protein expression levels was done using the bands with the expected size for NΔ42 or truncated versions of XXT2. Protein expression and purification for all constructs was done in parallel with identical conditions as described above. The purity of XXT2 NΔ42 was estimated by determining the intensity of bands in Coomassie stained SDS-PAGE gel using ImageJ, to determine the concentration of XXT2 NΔ42 protein based on the total protein concentration measured by Bradford assay (Quick Start Bradford Dye reagent 1X, Cat# 500-0205) according to the manufacturer's instructions. Next, the band intensity of the wild type and each truncated XXT2 band on the western blot was determined with ImageJ to estimate the concentrations of each of the truncated XXT2 proteins. Quantitative western blotting was performed using several dilutions of protein loaded on the gel and quantifying the intensities of XXT2 NΔ42 band with ImageJ.

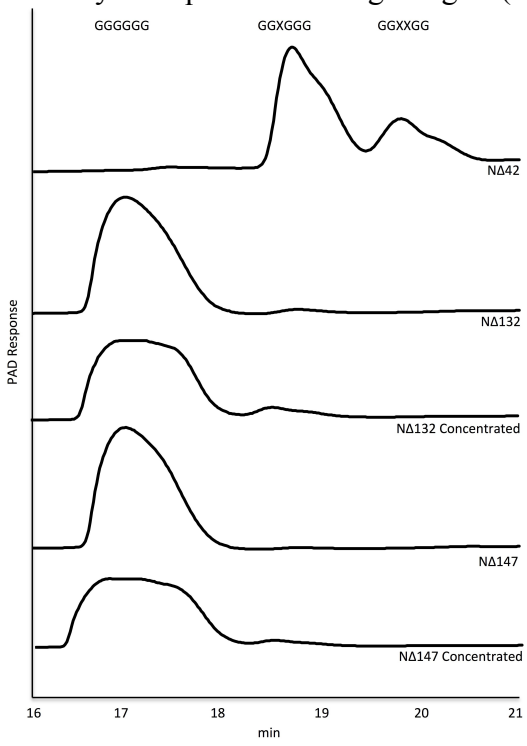
Supplementary Information

Supplementary Table 1. Primer sequences used for site-directed mutagenesis and protein cloning

Construct	Direction	Sequence
NΔ42	Forward	ATA GGA TCC AGC GAG AAC CTG TAC TTT CAG GGC ATG AAA TTC GGA ACT CCG GAG
C-term	Reverse	GCA CTC GAG TCA AAC TTG ATT GGT
NΔ73	Forward	ATA GGA TCC CAG ACC GGA GGA
NΔ106	Forward	CTA GGA TCC GGT GAA GAA GAG
NΔ132	Forward	GTC GGA TCC GAT TGG TTA GCT
NΔ147	Forward	CTA GGA TCC AAC AAG CCA CGT
CΔ24	Reverse	GCA CTC GAG TCA CGG ATT GCT AGT
CΔ44	Reverse	GCA CTC GAG TCA GAA ACC ATA GAT
CΔ57	Reverse	GCA CTC GAG TCA AAA GGC TCT GTC
CΔ73	Reverse	GCA CTC GAG TCA AAA TTT CCC ACA
K420A	Forward	GGT TTC ACT CAC GCA TCT TTG GCT AGT
K420A	Reverse	ACT AGC CAA AGA TGC GTG AGT GAA ACC
R425A	Forward	TCT TTG GCT AGT GCT AAA GTC AAG AGA GTG
R425A	Reverse	CAC TCT CTT GAC TTT AGC ACT AGC CAA AGA
K426A	Forward	TTG GCT AGT CGT GCA GTC AAG AGA GTG
K426A	Reverse	CAC TCT CTT GAC TGC ACG ACT AGC CAA
K428A	Forward	AGT CGT AAA GTC GCG AGA GTG CGG AA
K428A	Reverse	TTC CGC ACT CTC GCG ACT TTA CGA CT
R429A	Forward	GTA AAG TCA AGG CAG TGC GGA ACG A
R429A	Reverse	TCG TTC CGC ACT GCC TTG ACT TTA C
R431A	Forward	AGT CAA GAG AGT GGC GAA CGA GAC TAG CA
R431A	Reverse	TGC TAG TCT CGT TCG CCA CTC TCT TGA CT



Supplementary Figure 1. Quantification of XXT2 NΔ42. (A) Western blot of XXT2 NΔ42 with various amounts of protein loaded in each lane. The total protein loaded in each lane was: 0.18 µg, 0.36 µg, 0.54 µg, and 0.72 µg for lanes 1-4, respectively. Band intensity was quantified using ImageJ. (B) Plot of band intensity vs total protein loaded.



Supplementary Figure 2. Enzyme assay of concentrated N-terminal truncation proteins. The NΔ132 and NΔ147 protein samples were concentrated roughly six-fold. Enzyme assays were performed with 25 µL reaction containing 2 mM UDP-Xylose, 0.5 mM Cellohexaose, 2 mM $MnCl_2$, and 1.5 µM XXT2. GGGGGG: cellohexaose; GGXGGG: mono-xylosylated cellohexaose; GGXXGG: di-xylosylated cellohexaose

References

Ade, C. P., Bemm, F., Dickson, J. M., Walter, C., Harris, P. J., 2014. Family 34 glycosyltransferase (GT34) genes and proteins in *Pinus radiata* (radiata pine) and *Pinus taeda* (loblolly pine). Plant J. 78, 305-318.

Atmodjo, M. A., Hao, Z., Mohnen, D., 2013. Evolving views of pectin biosynthesis. Annu. Rev. Plant Biol. 64, 747-779.

Baneyx, F., Mujacic, M, 2004. Recombinant protein folding and misfolding in *Escherichia coli.* Nat. Biotechnol. 22, 1399-1408

Boix, E., Swaminathan, G. J., Zhang, Y., Natesh, R., Brew, K., Acharya, K. R., 2001. Structure of UDP complex of UDP-galactose: β-galactoside-α-1,3-galactosyltransferase at 1.53 Å resolution reveals a conformational change in the catalytically important C terminus. J. Biol. Chem. 276, 48608-48614.

Boyken, S. E., Fulton, D. B., Andreotti, A. H., 2012. Rescue of the aggregation prone Itk Pleckstrin Homology domain by two mutations derived from the related kinases, Btk and Tec. Protein Sci. 21, 1288-1297.

Carpita, N. C., 2012. Progress in the biological synthesis of the plant cell wall: new ideas for improving biomass for bioenergy. Curr. Opin. Biotechnol. 23, 330-337.

Carpita, N. C., Gibeaut, D. M., 1993. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. Plant J. 3, 1-30.

Cavalier, D. M., Keegstra, K., 2006. Two xyloglucan xylosyltransferases catalyze the addition of multiple xylosyl residues to cellohexaose. J. Biol. Chem. 281, 34197-34207.

Cavalier, D. M., Lerouxel, O., Neumetzler, L., Yamauchi, K., Reinecke, A., Freshour, G., Zabotina, O. A., Hahn, M. G., Burgert, I., Pauly, M., Raikhel, N. V., Keegstra, K., 2008. Disrupting two Arabidopsis thaliana xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component. Plant Cell 20, 1519-1537.

Chou, Y., Pogorelko, G., Young, Z. T., Zabotina, O. A., 2015. Protein-Protein Interactions Among Xyloglucan-Synthesizing Enzymes and Formation of Golgi-Localized Multiprotein Complexes. Plant Cell Physiol. 56, 255-267.

Chou, Y. H., Pogorelko, G., Zabotina, O. A., 2012. Xyloglucan xylosyltransferases XXT1, XXT2, and XXT5 and the glucan synthase CSLC4 form Golgi-localized multiprotein complexes. Plant Physiol. 159, 1355-1366.

Endler, A., Persson, S., 2011. Cellulose synthases and synthesis in Arabidopsis. Mol. Plant 4, 199-211.

Faik, A., Price, N. J., Raikhel, N. V., Keegstra, K., 2002. An *Arabidopsis* gene encoding an α-xylosyltransferase involved in xyloglucan biosynthesis. PNAS 99, 7797-7802.

Fauré, R., Cavalier, D., Keegstra, K., Cottaz, S., Driguez, H., 2007. Glycosynthase-Assisted Synthesis of Xylo-Gluco-Oligosaccharide Probes for α-Xylosyltransferases. European J Org Chem. 26, 4313-4319.

Hayashi, T., 1989. Xyloglucans in the primary cell wall. Annu. Rev. Plant Biol. 40, 139-168.

Henion, T. R., Macher, B. A., Anaraki, F., Galili, U., 1994. Defining the minimal size of catalytically active primate α1,3 galactosyltransferase: structure-function studies on the recombinant truncated enzyme. Glycobiology 4, 193-201.

Hennet, T., Dinter, A., Kuhnert, P., Mattu, T. S., Rudd, P. M., Berger, E. G., 1998. Genomic cloning and expression of three murine UDP-galactose: β-N-acetylglucosamine β1,3-galactosyltransferase genes. J. Biol. Chem. 273, 58-65.

Hou, B., Lim, E. K., Higgins, G. S., Bowles, D. J., 2004. N-glucosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana*. J. Biol. Chem. 279, 47822-47832.

Ihara, Y., Nishikawa, A., Tohma, T., Soejima, H., Niikawa, N., Taniguchi, N., 1993. cDNA cloning, expression, and chromosomal localization of human N-acetylglucosaminyltransferase III (GnT-III). J. Biochem. 113, 692-698.

Jamaluddin, H., Tumbale, P., Withers, S. G., Acharya, K. R., Brew, K., 2007. Conformational changes induced by binding UDP-2F-galactose to α-1,3 galactosyltransferase- implications for catalysis. J. Mol. Biol. 369, 1270-1281.

Johal, A. R., Blackler, R. J., Alfaro, J. A., Schuman, B., Borisova, S., Evans, S. V., 2014. pH-induced conformational changes in human ABO(H) blood group glycosyltransferases confirm the importance of electrostatic interactions in the formation of the semi-closed state. Glycobiology 24, 237-246.

Keegstra, K., 2010. Plant Cell Walls. Plant Physiol. 154, 483-486.

Kyte, J., Doolittle, R. F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132.

Lairson, L. L., Henrissat, B., Davies, G. J., Withers, S. G., 2008. Glycosyltransferases: structures, functions, and mechanisms. Annu. Rev. Biochem. 77, 521-555.

Leipold, M. D., Kaniuk, N. A., Whitfield, C., 2007. The C-terminal Domain of the *Escherichia coli* WaaJ glycosyltransferase is important for catalytic activity and membrane association. J. Biol. Chem. 282, 1257-1264.

Lin, Z., Rye, H. S., 2006. GroEL-mediated protein folding: making the impossible, possible. Crit. Rev. Biochem. Mol. Biol. 41, 211-239.

Lund, C. H., Bromley, J. R., Stenbaek, A., Rasmussen, R. E., Scheller, H. V., Sakuragi, Y., 2015. A reversible Renilla luciferase protein complementation assay for rapid identification of protein-protein interactions reveals the existence of an interaction network involved in xyloglucan biosynthesis in the plant Golgi apparatus. J. Exp. Bot. 66, 85-97.

Mansoori, N., Schultink, A., Schubert, J., Pauly, M., 2015. Expression of heterologous xyloglucan xylosyltransferases in *Arabidopsis* to investigate their role in determining xyloglucan xylosylation substitution patterns. Planta 241, 1145-1158.

McCann, M. C., Carpita, N. C., 2008. Designing the deconstruction of plant cell walls. Curr. Opin. Plant Biol. 11, 314-320.

Scheller, H. V., Ulvskov, P., 2010. Hemicelluloses. Annu. Rev. Plant. Biol. 61, 263-289.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., Higgins, D. G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.

Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredez, A., Persson, S., Raab, T., Vorwerk, S., Youngs, H., 2004. Toward a systems approach to understanding plant cell walls. Science 306, 2206-2211.

Thompson, D. S., 2005. How do cell walls regulate plant growth? J. Exp. Bot. 56, 2275-2285.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J., Boerjan, W., 2010. Lignin biosynthesis and structure. Plant Physiol. 153, 895-905.

Vuttipongchaikij, S., Brocklehurst, D., Steele-King, C., Ashford, D. A., Gomez, L. D., McQueen-Mason, S. J., 2012. *Arabidopsis* GT34 family contains five xyloglucan α-1,6-xylosyltransferases. New Phytol. 195, 585-595.

Wang, C., Li, S., Ng, S., Zhang, B., Zhou, Y., Whelan, J., Wu, P., Shou, H., 2014. Mutation in xyloglucan 6-xylosytransferase results in abnormal root hair development in *Oryza sativa*. J. Exp. Bot. 65, 4149-4157.

Yu, H., Takeuchi, M., LeBarron, J., Kantharia, J., London, E., Bakker, H., Haltiwanger, R. S., Li, H., Takeuchi, H., 2015. Notch-modifying xylosyltransferase structures support an Si-like retaining mechanism. Nat. Chem. Biol. 11, 847-854

Zabotina, O. A., Avci, U., Cavalier, D., Pattathil, S., Chou, Y. H., Eberhard, S., Danhof, L., Keegstra, K., Hahn, M. G., 2012. Mutations in multiple XXT genes of *Arabidopsis* reveal the complexity of xyloglucan biosynthesis. Plant Physiol. 159, 1367-1384.

Zabotina, O. A., van de Ven, W. T., Freshour, G., Drakakaki, G., Cavalier, D., Mouille, G., Hahn, M. G., Keegstra, K., Raikhel, N. V., 2008. *Arabidopsis* XXT5 gene encodes a putative α-1,6-xylosyltransferase that is involved in xyloglucan biosynthesis. Plant J. 56, 101-115.

CHAPTER 5. A HOMOLOGY MODEL OF XYLOGLUCAN XYLOSYLTRANSFERASE 2 REVEALS CRITICAL AMINO ACIDS INVOLVED IN SUBSTRATE BINDING

Modified from a paper published in *Glycobiology*

Alan T. Culbertson, Alesia A. Tietze, Daniel Tietze, Yi-Hsiang Chou, Adrienne L. Smith, Zachary T. Young, and Olga A. Zabotina

Abstract

In dicotyledonous plants, xyloglucan is the most abundant hemicellulose of the primary cell wall. The enzymes involved in xyloglucan biosynthesis have been identified through reverse genetics and activity was characterized by heterologous expression. Currently, there is no information on the atomic structures or amino acids involved in activity or substrate binding of any of the Golgi localized xyloglucan biosynthetic enzymes. A homology model of the Xyloglucan Xylosyltransferase 2 (XXT2) catalytic domain was built on the basis of the crystal structure of A64Rp. Molecular dynamics simulations revealed that the homology model retains the GT-A fold of the template structure used to build the homology model indicating that XXT2 likely has a GT-A fold. According to the XXT2 homology model, six amino acids (Phe204, Lys207, Asp228, Ser229, Asp230, His378) were selected and their contribution in catalytic activity was investigated. Site-directed mutagenesis studies show that Asp228, Asp230, and His378 are critical for XXT2

activity and are predicted to be involved in coordination of manganese ion. Lys207 was also found to be critical for protein activity and the homology model indicates a critical role in substrate binding. Additionally, Phe204 mutants have less of an impact on XXT2 activity with the largest effect when replaced with a polar residue. This is the first study that investigates the amino acids involved in substrate binding of the xyloglucan synthesizing xylosyltransferases and contributes to the understanding of the mechanisms of polysaccharide synthesizing glycosyltransferases and xyloglucan biosynthesis.

Introduction

The plant cell wall consists of different types of polymers including cellulose, hemicellulose, pectin, and lignin, and has many industrial applications such as biofuels and biomaterials (Carpita et al. 2008; Pauly and Keegstra 2008). Xyloglucan (XyG), the most abundant hemicellulose of dicotyledonous plants, has been proposed to interact with cellulose microfibrils by providing cross-links that prevent their aggregation (Carpita and Gibeaut 1993; Hayashi 1989; Park and Cosgrove 2015). XyG is composed of a β-(1,4) linked glucan backbone that is decorated with α-(1,6) linked xylose residues, which can be further decorated with various mono- or disaccharides, such as α-(1,2) linked $_{D}$-Gal, $_{L}$-Ara, or $_{L}$-Fuc-(1,2)-β-$_{D}$-Gal. The biosynthesis of XyG involves a diversity of glycosyltransferases (GTs) including glucan synthases, xylosyltransferases, galactosyltransferases, arabinosyltransferases, and a fucosyltransferase (Cavalier and Keegstra 2006;

Cocuron et al. 2007; Faik et al. 2002; Madson 2003; Schultink et al. 2013; Zabotina 2012; Zabotina et al. 2012).

In *Arabidopsis*, at least three Xyloglucan Xylosyltransferases (XXTs), XXT1, XXT2, and XXT5, are responsible for xylosylation of the XyG glucan backbone. Single knock-out mutants of *xxt1* or *xxt2* have only minor effects on plant phenotype and cell wall composition, while the *xxt5* single mutant results in a 50% reduction of total XyG content (Cavalier et al. 2008; Zabotina et al. 2008). The *xxt1xxt2* double or *xxt1xxt2xxt5* triple mutants have no detectable XyG and have root hairs that are shorter, thicker, and lower in abundance compared to wild type plants (Cavalier et al. 2008; Zabotina et al. 2012). XXT1 and XXT2 have been shown to be catalytically active *in vitro* and are able to xylosylate the acceptor substrate, cellohexaose, at three positions, producing tri-xylosylated cellohexaose (Cavalier and Keegstra 2006). The catalytic activity of XXT5 has not been demonstrated.

GTs are enzymes that transfer a sugar moiety from an activated donor substrate, typically a nucleotide sugar, to an acceptor substrate (Lairson et al. 2008). The acceptor substrates can be proteins, lipids, nucleic acids, or carbohydrates. Modes of acceptor substrate binding by GTs are difficult to predict because of the large diversity of substrates, which typically have unique binding motifs. In contrast, donor substrates are typically activated nucleotides, thus, donor substrate binding motifs can be predicted from previous reported atomic structures (Lairson et al. 2008).

GTs are classified into 98 different gene families in the Carbohydrate Active enZyme (CAZy) database based on their primary amino acid sequence (Campbell et

al. 1997; Coutinho et al. 2003). It is believed that all GTs have one of two structural folds referred to as GT-A or GT-B. The GT-A fold contains two Rossmann-like domains that form a continuous β-sheet that is surrounded by α-helices. The GT-B fold also contains two Rossmann-like domains, yet unlike the GT-A fold, are not as tightly associated, and the active site is localized between these two Rossmann-like domains. Furthermore, GTs with a GT-A fold are typically divalent cation dependent, which is coordinated by the well characterized DXD motif (Busch et al. 1998; Wiggins and Munro 1998). There is another structural fold, GT-C, but this fold is less commonly observed compared to GT-A or GT-B folds (Liu and Mushegian 2003). In addition to the different folds, GTs are also classified by whether the stereochemistry of the product's glycosidic bond is retained or inverted compared to the stereochemistry of the donor substrate.

Although there are numerous crystal structures solved for different GTs (Gloster 2014), there is no structural information for any of the Golgi localized GTs involved in plant cell wall polysaccharide biosynthesis. Structural information revealing the amino acids involved in substrate binding and catalysis would aid in the understanding of polysaccharide biosynthesis, which is critical for engineering plants with a cell wall that is more suitable for the production of biofuels or biomaterials. Here, we use a computational approach to produce a homology model of the XXT2 catalytic domain. A viral GT A64R (A64Rp), predicted to be in the same GT34 family as the XXTs, was used as a template (Faik et al. 2002; Zhang et al. 2007). The XXT2 homology model was used to predict amino acids localized in the active site of XXT2. The function of these amino acids was then investigated by site-

directed mutagenesis to confirm their role in donor substrate binding. We demonstrated that the DSD motif of XXT2 is critical for catalytic activity *in vitro* and function *in vivo*. In addition, we identified three other amino acids, His378, Phe204, and Lys207, which are also important for the catalytic activity of XXT2.

Materials and methods

Homology models and molecular dynamics simulations. The YASARA molecular modeling program (YASARA "structure" Vers. 15.11.18 and 16.2.23)(Krieger et al. 2009; Krieger et al. 2014) was used to build the homology model of XXT2 and for the molecular dynamics simulations. Possible templates were identified by YASARA by running PSI-BLAST iterations (Altschul et al. 1997) on the basis of the amino acid sequence of Xyloglucan Xylosyltransferase 2, to extract a position specific scoring matrix (PSSM) from UniRef90, and then searching the Protein Data Bank (PDB) for a match (hits with an E-value below the cutoff 0.5). Only a single hit (PDB 2P6W) was identified that could be used as a template for homology modeling.

Multiple models were built according to alternative alignments of the target and template protein sequence. Side chains were added using YASARA's implementation of SCWRL3 (Canutescu et al. 2003) and fine-tuned by considering electrostatic, knowledge-based packing iterations, and solvation effects. Hydrogen bond network was optimized (Krieger et al. 2012) and each model was then subjected to an unrestrained energy minimization with explicit water molecules by

simulated annealing employing the YASARA2 force field (Krieger et al. 2009). In addition, a hybrid model was built by combination of the best regions of the individual models. The models were rated according to a quality Z-score, the best scoring model was used as the final protein model. A more detailed description of YASARA's homology modeling protocol can be found online (http://yasara.org/homologymodeling.htm). UDP-xylose was then manually modeled into the final homology model according to the position of UDP-glucose in the crystal structure of the template (PDB 2P72; Zhang et al. 2007). Refinement of the UDP-xylose-XXT2 structure was achieved in two steps. First, a 10 ns molecular dynamics simulation in water was performed, restraining the residues involved in coordination of Mn^{2+} and restrained UDP-Xyl to the metal ion. Distances and angles used for the restraints of specific atoms are shown in the Supplementary data, Table S2. Next, an unrestrained, 90 ns, all-atom molecular dynamics simulation in water was performed. For comparison, an unrestrained, 100 ns, all-atom, molecular dynamics simulation was performed for the template structure of A64Rp using identical simulation parameters. UDP-Xyl was derived from the A64Rp UDP-Glc (PDB 2p72) and geometry was optimized using a semi-empirical quantum mechanics approach, which is implemented in YASARA's YAPAC module (Stewart 2000).

Unless otherwise stated TIP3P was used as the water model and the Amber14 force field was used for energy minimization and molecular dynamics simulations (Hornak et al. 2006; Maier et al. 2015). Energy minimization was done by simulated annealing in water including optimization of the hydrogen bond

network (Krieger et al. 2012) and equilibration of the water shell. Restrained and unrestrained all-atom molecular dynamics simulation were performed in water using the PME method (Essmann et al. 1995) to describe long-range electrostatic interactions at a cut off distance of 8 Å at physiological conditions (298 K, pH 7.4, 0.9% NaCl)(Krieger et al. 2006). Energy minimization was done for all structures before the molecular dynamics simulations. Charged amino acids were assigned according to the predicted pKa of the amino acid side chains by Ewald summation (Krieger et al. 2006) and were neutralized by adding counter ions (NaCl). A multiple time step algorithm together with a simulation time step interval of 1.25 fs was chosen (Grubmueller et al. 1998; Krieger and Vriend 2015). Molecular dynamics simulations were performed at constant temperature using a Berendsen thermostat and constant pressure. Mn^{2+} was modeled according to the procedure described by Li et al. (Li et al. 2013). Structural alignments were conducted using the Mustang algorithm (Konagurthu et al. 2006). Molecular graphics were created with YASARA (www.yasara.org) and POVRay (www.povray.org)

***XXT2 cloning*.** The pET20b expression vector was modified as previously described (Boyken et al. 2012) to incorporate an N-terminal His-immunoglobulin-binding domain of streptococcal protein G (GB1) tag using *NdeI* and *XhoI* restriction sites. Overhangs of the forward XXT specific primers were designed to contain the *NdeI* restriction site and an N-terminal 6xHis tag, and the reverse primers were designed to contain the *BamHI* and the *XhoI* restriction sites. The coding sequence of the *Arabidopsis thaliana* xylosyltransferase gene, *XXT2* (At4g02500), was PCR amplified with N-terminal truncations of 126 base pairs (42 amino acids) and was

cloned into the pET20b plasmid. XXT2 was cloned using *BamHI* and *XhoI* restriction sites. All forward primers were designed to produce the Tobacco Etch Virus protease site at the N-terminus of XXT2. The resulting pET20b constructs were then transformed into *E. coli* DH10b for plasmid amplification and the sequence was verified by sequencing. Confirmed plasmids were then transformed into *E. coli* SoluBL21. To generate the XXT2 mutants, two fragments for each mutant were PCR amplified using primers designed to contain the mutated base pairs (Supplementary data, Table SI).). The XXT2 mutant fragment 1 was generated with XXT2 forward and XXT2 mutant reverse primers, and fragment 2 was generated using XXT2 mutant forward and XXT2 reverse primers. The resulting fragments containing the mutated base pairs were then used as the template in a fusion PCR reaction using XXT2 forward and XXT2 reverse primers to generate the final mutated gene fragment.

Protein expression. *E. coli* SoluBL21 cells harboring the plasmids were grown at 37°C with shaking at 200 rpm in 500 mL of Luria-Bertani broth. When the cell culture reached an OD_{600} of 0.5, the temperature was lowered to 18°C and protein expression was induced by adding IPTG to a final concentration of 0.5 mM. Cells were incubated for 18 h at 18°C and then harvested by centrifugation. Pelleted cells were re-suspended in 12.5 mL of 25 mM Tris-HCl pH 7.4, 300 mM NaCl, 0.1 mM EDTA and rapidly frozen in liquid nitrogen. Cells were lysed by thawing and incubating for 30 min with 1 mg/mL of lysozyme and then sonicated

for 15 s a total of five times. Solubilized proteins were collected by centrifugation at 20,000 xg for 30 min.

Protein purification. The crude lysate was loaded onto a Ni-NTA column with a lysate:resin ratio of 20:1. Affinity resin was incubated on a shaker for 1 h at 4°C. Unbound proteins were removed as a flow-through fraction and the resin was washed four times with washing buffer containing 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, and 20 mM imidazole. The protein of interest was eluted in three fractions with a total volume of 6 mL using elution buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl, and 300 mM imidazole. Glycerol was added to the elution fractions to a final concentration of 20% and proteins were stored at -80°C. Protein concentration was measured using a Bio-Rad kit (Quick Start Bradford Dye reagent 1X, Cat# 500-0205) according to the manufacturer's instructions.

Enzyme activity assay and product precipitation. The enzymatic activity assay was performed in a total volume 25 µL, consisting of 2 mM UDP-Xyl (CarboSource Services), 0.5 mM cellohexaose (Megazyme), 2 mM $MnCl_2$, and 1.5 µM purified protein in elution buffer. The reaction mixture was incubated at 28°C for 20 h with shaking at 100 rpm. The reaction was stopped by adding 900 µL of absolute ethanol and incubating at -22°C for 6 h; the precipitated reaction product was collected by centrifugation at 21,000 xg for 30 min. All activity assays were performed in triplicates using two independent protein preparations; all XXT2

mutants were expressed and purified two separate times in parallel with wild type XXT2 as a positive control.

HPLC and MALDI-TOF MS characterization of catalysis products.

Products obtained from the enzymatic activity assays were analyzed by high-performance anion-exchange chromatography with a pulse-amperometric detector (HPAEC-PAD, Dionex), as described by Cavalier and Keegstra (Cavalier and Keegstra 2006). The aqueous solution of the product was injected onto a CarboPac PA-20 column and eluted using the following gradient: 30 mM to 100 mM Na-Acetate for 25 min with 100 mM NaOH remaining constant through the entire run; after the 25 min run, the column was re-equilibrated for 15 min to initial conditions. Non-xylosylated, mono-xylosylated, and di-xylosylated cellohexaose eluted from the column with retention times of approximately 16.4, 17.8, and 18.9 min, respectively. Quantitation of xylose transfer was performed by peak integration of xylosylated cellohexaose to determine the relative percentage of each product (non-, mono-, or di-xylosylated cellohexaose). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) analysis of the products was performed as described earlier (Zabotina et al. 2012). One µL of the product solution was spotted onto a MALDI-TOF sample plate using a 2,5-dihydroxybenzoic acid matrix at a sample:matrix ratio of 1:1. The sample spectra were analyzed on an Applied Biosystems VOYAGER-DE Pro MALDI-TOF mass spectrometry instrument in positive reflection mode with an acceleration voltage of 20 kV and extraction delay time of 350 ns.

SDS-PAGE and Western Blots. Proteins were analyzed by SDS-PAGE in reducing conditions and stained with Coomassie Blue G250. Western blot analysis was performed as previously described (Chou et al. 2012). Proteins were electrophoretically transferred to a nitrocellulose membrane and blocked with non-fat milk. The GB1 tag was probed with horseradish peroxidase conjugated goat anti-rabbit antibody (Invitrogen) at a 10,000-fold dilution, developed using HyGLO quick spray, and visualized by ChemiDoc XRS+ (Bio-Rad). Pre-stained size markers were visualized on the same membrane using visible light.

Production of transgenic plants. The PCR-amplified coding sequence of the full-length XXT2 D228A/D230A protein including Human influenza hemagglutinin (HA) tag was ligated into the pCR8/GW/TOPO entry vector (Invitrogen) as described in the manufacturer's instructions. The ligation product was transformed into *E. coli* DH10b and selected on LB agar plates with 50 µg/mL spectinomycin. The mutated gene in the entry construct was transferred into the binary destination vector pEarleyGate 201 containing Cauliflower mosaic virus 35S promoter and *bar* selectable marker gene (TAIR). The binary destination vector was transformed into Agrobacterium GV3101 by electroporation. Arabidopsis was transformed using the floral-dip method (Clough and Bent 1998). Plants were harvested and selected using 10mg L^{-1} of Basta. A total of four independent homozygous lines were prepared. For analysis, the homozygous plants (T3) were grown under long day conditions (16 h light / 8 h dark) at 22°C in a growth chamber.

Examination of protein expression in transgenic plants. The total membrane protein was extracted from 40 to 80 2-week-old seedlings of independent homozygous lines. The seedlings were ground in 10 mL protein extraction buffer (40 mM HEPES, 0.45 M sucrose, 1 mM EDTA, 1 mM $MgCl_2$, 1 mM KCl, 1mM dithiothreitol, protease inhibitor cocktail [Roche], pH 8.0). The extract was homogenized three times, 10 s each, at 10,000 rpm, using a Polytron homogenizer. The extract was filtered through three layers of miracloth and centrifuged for 30 min at 10,000 rpm at 4°C. The supernatant was transferred to a polycarbonate tube with aluminum cap assembly and ultracentrifuged at 37,000 rpm (100,000 xg) for 45 min with a 70Ti fixed rotor. After centrifugation, the supernatant was removed and the pellet was resuspended in 200 µL suspension buffer (40 mM HEPES, 0.2 M sucrose pH 7.3). The membrane protein fraction was treated with 1% Triton X-100 for 30 min at 4°C to solubilize membrane-bound proteins. After solubilization, proteins were ultracentrifugated at 37,000 rpm for 45 min to precipitate non-soluble membrane fragments. The protein concentration of the supernatant was measured using the Bradford protein assay. The supernatant was concentrated by precipitating with 10% trichloroacetic acid and the precipitated protein was resuspended in loading buffer with β-mercaptoethanol. Proteins were separated using SDS-PAGE and transferred to a nitrocellulose membrane (0.2 µm, Bio-Rad) for immunodetection. Polyclonal anti-HA antibodies were used (1:500 dilution) to detect HA-fused proteins.

Root hair phenotype and xyloglucan content analysis. Analysis of plant root hairs was done by plating sterilized seeds on half-strength MS with 0.3% gelrite. Plants were grown in a growth chamber with long day conditions (16 h light/ 8 h dark) at 22°C. After the seeds germinated and the roots grew into the media, the plates were placed at a 45° angle. Pictures of 7-day-old roots were taken using a Leica DMIRE2 light microscope with a Retiga 1300 camera.

Driselase digestion was performed as previously described (Zabotina et al. 2012). The alcohol insoluble residues (AIR) were prepared from wild type, *xxt1xxt2* mutant and complemented with HA-XXT2 D228A/D230A construct *xxt1xxt2* mutant plants. 5 mg of AIR were incubated with 10 μg of Driselase in citrate buffer (pH 5) overnight at 37°C. Reaction was stopped by boiling for 15 min and supernatant was separated by centrifugation. The amount of isoprimeverose produced by Driselase was measured by analyzing the supernatants on HPAEC and calculating the peak areas corresponding to isoprimeverose.

Results

Homology models of XXT2. A homology model of Arabidopsis XXT2 was built using the structure of the protein A64Rp (Figure 1A), a putative GT involved in glycosylation of the Chlorella virus major capsid protein, that belongs to the same gene family GT34 as XXT2 (Zhang et al. 2007). A64Rp was identified as a template by running 3 PSI-BLAST iterations to extract a position specific scoring matrix (PSSM) from UniRef90 and then searching the PDB for a match. Sequence alignment of full-length XXT2 (Uniprot ID: O22775) with A64Rp (Uniprot ID: Q89399) using

Clustal Omega (Sievers et al. 2011, Goujon et al. 2010, Drozdetskiy et al 2015) resulted in 7.7% and 25.6% sequence identity and sequence similarity, respectively. The alignment used to build the homology model contained 163 out of 419 XXT2 residues (38.9%), lacking the putative transmembrane domain, and was aligned to the template residues resulting in 23.3% and 45.4% sequence identity and similarity, respectively ('similar' means that the BLOSUM62 score is > 0). Although the sequence identity is rather low to build a high quality model, the structure of A64Rp has been rated as appropriate for homology modeling by YASARA internal quality score of 0.51 (the quality score ranges from 0.000 = bad to 1.000 = good). In addition, GT-A fold GTs, particularly ones in the same GT family, typically have high structural homology despite having low sequence similarity (Zhang et al. 2007, Gibbons et al. 2002).

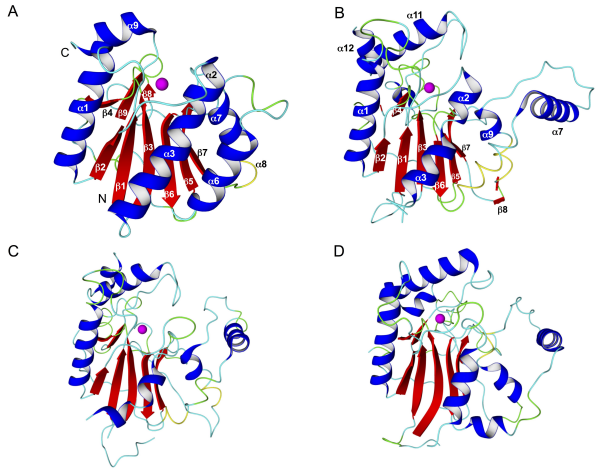


Figure 1. (**A**) The structure of A64Rp (PDB 2P6W) used as a template. (**B**) The XXT2 homology model before refinement simulations (all-atom RMSD to A64Rp is 1.63 Å) (**C**) The XXT2 homology model after 10 ns restrained simulation (all-atom RMSD to A64Rp is 3.48 Å), (**D**) The XXT2 homology model after 10 ns of restrained followed by 90 ns of unrestrained simulation (all-atom RMSD is 3.66 Å). **B**-**D** structures are energy minimized. The β-strands and α-helices are numbered. Colors of secondary structure elements are as follows: β-strand (red); α-helix (blue); Turn (green); Random coil (cyan), 310 helix (yellow). Mn2+ (magenta).

The final homology model included 276 amino acids of XXT2 (amino acids 140-415). The N-terminal residues 1-139 (containing the cytosolic N-terminus, transmembrane domain, and stem region) and the C-terminal residues 416-461 of XXT2 were not included in the homology model since no adequate template structure is available for these regions. Five models were built that are based on alternative structural alignments of the target and template protein sequence. Following refinement and scoring of the structural quality of the models, a hybrid model was built by combining the best parts of the five individual models. The hybrid model scored significantly better than the other modeled structures and was therefore selected as the final homology model for XXT2 (Figure 1B and 2B).

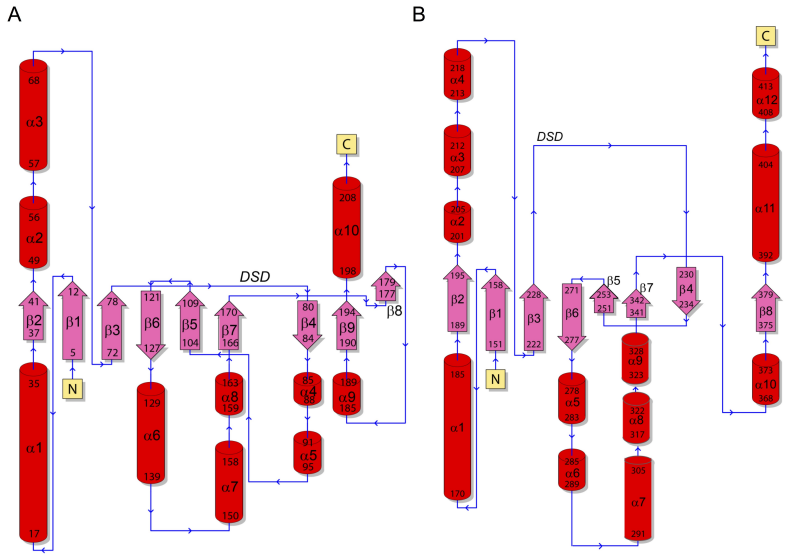


Figure 2. Topology diagrams (**A**) of the A64Rp crystal structure (PDB 2P6W; Zhang et al. 2007) and (**B**) of the XXT2 homology model after 10 ns of restrained simulation. The diagrams were generated by PDBsum (http://www.ebi.ac.uk/pdbsum; Laskowski 2009).

UDP-Xyl was then manually modeled into the XXT2 homology model to identify residues involved in substrate binding. The position of UDP-Xyl in the homology model was adapted from the substrate position of UDP-Glc in the crystal structure of A64Rp (PDB 2P72; Zhang et al. 2007). After the respective glucose moiety of UDP-Glc was changed to xylose, the energy of the UDP-Xyl-XXT2 structure was minimized in water before it was docked and refined in a restrained 10 ns molecular dynamics simulation. The active site environment was restrained during refinement by applying distance and angle restraints in order to allow the protein structure to adapt to the UDP-Xyl molecule (Supplementary data, Table S2). Additionally, the side chain atoms of Asp228, Asp230, His378, and an oxygen atom of the UDP-xylose phosphate group were restrained to the Mn^{2+} ion according to the distances and angles derived from the crystal structure of A64R with UDP-glucose bound (PDB 2P72 and 2P6W, Zhang et al. 2007).

The XXT2 secondary structure obtained from the final homology model (Figure 1B and 2B) was compared to the predicted secondary structure from the XXT2 primary amino acid sequence using the software JNet Secondary structure predictor (Drozdetskiy et al. 2015; Figure 3). The predicted secondary structure from the primary amino acid sequence is similar to the homology model (Figure 3). This further supports the accuracy of the predicted secondary structure of XXT2 from YASARA's homology modeling, particularly the core region of the homology model, which harbors the active site of XXT2, including α-helices α1 and α11, and the β-strands β2, β1, β3 β6, β5 and β7 (Figure 1A and B). In contrast, the orientation of α-helices α2, α7 and α9, as well as some loops and random coils between the β-

strand β7 and α-helix α11 do not resemble the template structure (Figure 1A and B). It is therefore assumed that the modeled structure of these regions is still undefined.

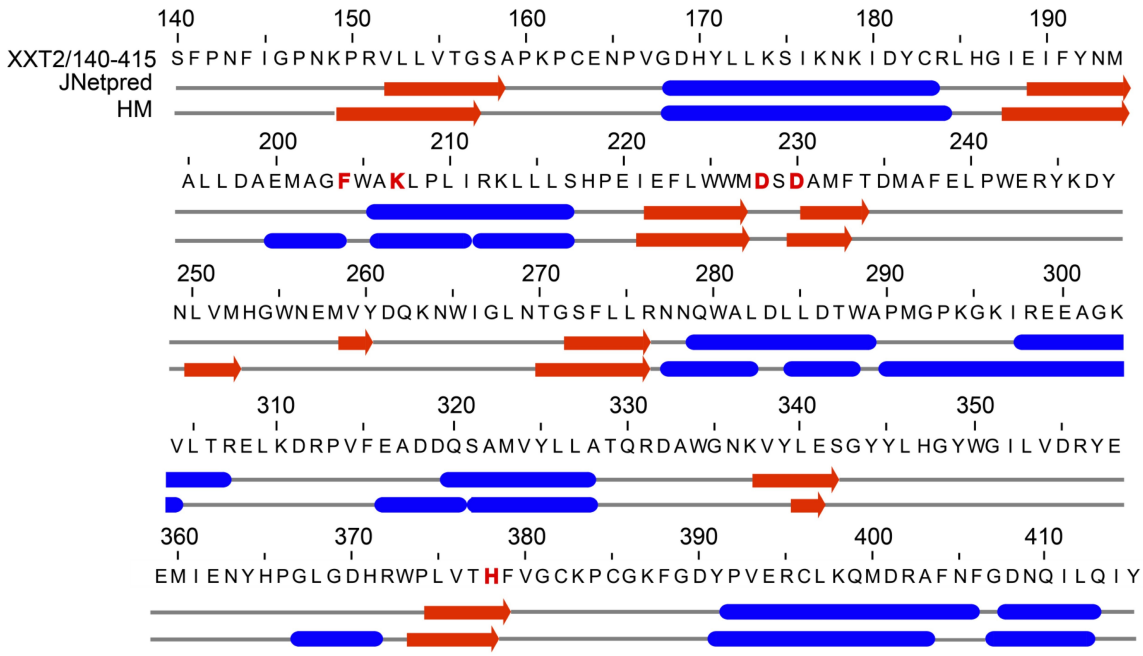


Figure 3. Secondary structure alignment for the XXT2 protein sequence obtained from the homology model and JNet Secondary structure predictor (Drozdetskiy et al 2015).

An unrestrained 90 ns molecular dynamics simulation was performed to further assess the stability of the UDP-Xyl-XXT2 homology model. The modeled UDP-Xyl bound XXT2 structure reached a stable conformation after about 50 to 60 ns with a Cα RMSD of around 6Å (Figure 4), demonstrating some conformational stress in the original homology model. However, the core region of the XXT2 homology model retained its fold during the restrained (10 ns) and unrestrained (90 ns) simulations indicating that this part of the protein is structurally stable (Figure 1B-D, Supplementary data, Figure S1 and S4B). The stable core region is also

characterized by low average per residue deviation of Cα atoms (Supplementary data, Fig. S4B) as well as low per residue flexibilities, given as average per residue RMSF (Supplementary data, Fig. S4B) compared to the high flexibility of other protein regions, such as α-helices α2-4, α7, α9, α10 and the disordered region between the β strand β7 and α-helix α11 (Figure 1B-D, Supplementary data, Fig. S4B). Average per residue RMSFs were calculated from the last 50 ns of the refinement simulation for the homology model and the template structure. Furthermore, the displacement of the UDP-Xyl substrate molecule with respect to its initial position follows the displacement of Cα atoms of XXT2 (Figure 4) during the simulations and shows only very little positional fluctuation (RMSF 1.43 Å, RMSD 1.36 Å, Supplementary data, Figure S4B) revealing that the substrate remains tightly bound in the active site pocket (Supplementary data, Figure S3). Similarly, the substrate molecule UDP-Glc of the template structure A64Rp also remains tightly bound in the active site pocket during the 100 ns simulation. The structure of the template is much more stable compared to the XXT2 homology model as shown by the calculated Cα RMSD (Figure 4) and per residue deviations of Cα atoms and per residue fluctuations (Supplementary data, Figure S4A), indicating very little conformational changes during the simulation. Both the template and homology model exhibit the lowest per residue flexibility in the active site pocket (Figure 5, Supplementary data, Figure S4).

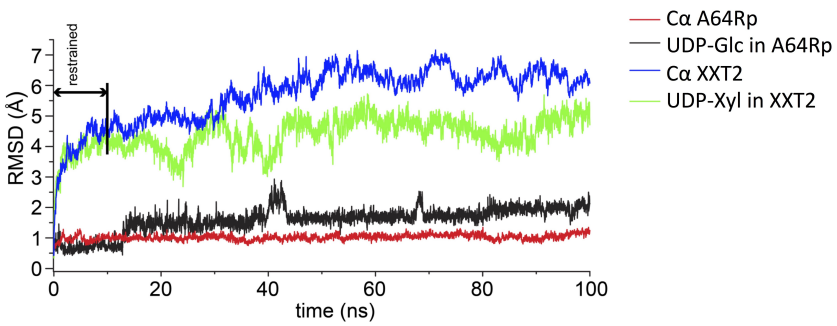


Figure 4. Cα root-mean-square deviation (RMSD) of the XXT2 homology model equilibrated for 10 ns restrained followed by 90 ns unrestrained equilibration (blue) and the template A64Rp (PDB ID: 2P6W) after 100 ns of unrestrained equilibration (red). Displacement of UDP-Xyl (green) in the XXT2 model and UDP-Glc (black) in the A64Rp template during molecular dynamic simulations.

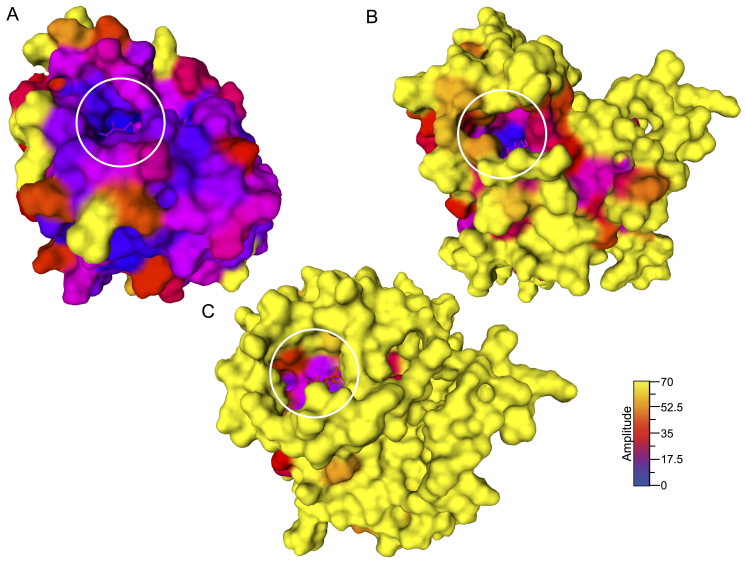


Figure 5. Molecular surfaces of A64Rp and XXT2. The proteins surfaces are colored according to the crystallographic B-factor which is calculated from average per residue root-mean-square fluctuations. The amplitude diagram represents the B-factor where blue corresponds to rigid and yellow to flexible residues. The orientations of the proteins are kept the same as for Figure 1. White circles show positions of active site. (**A**) The template A64Rp (PDB ID: 2P6W), (**B**) XXT2 after 10 ns restrained simulation. (**C**) The model shown in **B** after 90 ns unrestrained simulation.

Sequence alignment of A64Rp with XXT2 and XXT2 homologs. To predict amino acids that may be involved in UDP-Xyl binding by XXT2, a sequence alignment with A64Rp, XXT2, and XXT2 homologous proteins was investigated (Supplementary data, Figure S2). The XXT2 homologs from *Pinus radiata* (Ade et al. 2014), *Oryza sativa* (Wang et al. 2014), and *Solanum lycopersicum* (Mansoori et al. 2015) were included in the sequence alignment. Although there are many other XXT2 homologs in most available plant genomes, these were chosen because all were shown to be able to restore the *Arabidopsis* wild type phenotype when over-expressed in the *xxt1xxt2* mutant plant. Four conserved residues were identified from the sequence alignment (Supplementary data, Figure S2) that are predicted to reside in the XXT2 active site, including Asp 228, Ser229, Asp230, and His378. Asp228, Ser229 and Asp230 are predicted to be part of the DXD motif that is well characterized in GT-A fold GTs to be involved in the coordination of the divalent cation required for catalysis. In addition, His378 was also shown to be involved in the coordination of the divalent cation in the modeled XXT2 structure. Two additional amino acids, Phe204 and Lys207, that are predicted to be positioned near the active site pocket of XXT2 are conserved in all XXT2 homologs (Supplementary data, Figure S2). Lys207 is conserved in all XXT homologs except A64Rp, which has an Arg. Phe204 is also conserved in all XXT2 homologs, yet in the A64Rp sequence it is replaced with a His residue.

Active site of XXT2. The XXT2 predicted active site has numerous similarities with the A64Rp active site (Figure 6) and was remarkably stable during the molecular dynamics simulation. The Mn^{2+} coordinating residues in the XXT2

homology model, including Asp228, Asp230, and His378, are in similar positions to the aligned residues in the active site of the A64Rp crystal structure (Figure 6A, Supplementary data, Figure S5). As mentioned above, Phe204 and Lys207 are also predicted to be in the active site pocket of XXT2 (Figure 6A).

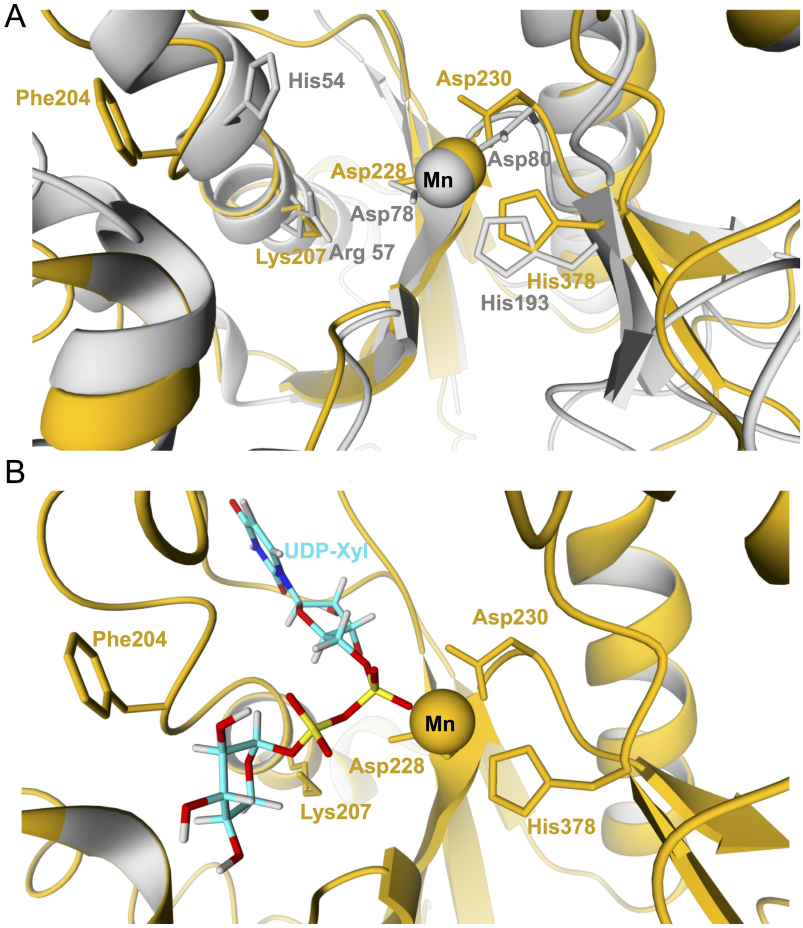


Figure 6. Structures of the active site of A64Rp and XXT2 (**A**) The alignment of the A64Rp (PDB 2P72) active site in gray and XXT2 homology model equilibrated for 10 ns in dark yellow. Relevant amino acids are highlighted. (**B**) The XXT2 homology model active site equilibrated for 10 ns with UDP-xylose bound. The colors are as follow: phosphorus (yellow), carbon (cyan); oxygen (red), nitrogen (blue), hydrogen (white).

The Mn^{2+} ion, which is coordinated by Asp228, Asp230, and His378, is also ligated by a single oxygen atom of the diphosphate in UDP-Xyl (Figure 6B and Supplementary data, Figure S5) resulting in an octahedral coordination environment in which two water molecules make the fifth and sixth ligand. His54 in A64Rp, which aligns with Phe204 in XXT2, forms a hydrogen bond to the hydroxyl group of carbon C6 of UDP-Glc (Zhang et al. 2007). In contrast, XXT2 is specific for UDP-Xyl as a donor substrate (Faik et al. 2002), which lacks the sixth carbon atom. Phe204 in XXT2 is roughly 7 Å away from UDP-Xyl (Figure 6B and Supplementary data, Figure S5). Arg57 in the crystal structure of A64Rp, which aligns with Lys207 in XXT2, forms a hydrogen bond network via a water molecule with Asp78 and the C4 hydroxyl group of the glucose moiety of UDP-Glc. In the XXT2 model, Lys207 interacts with the hemi-acetal oxygen of the xylose moiety of UDP-Xyl (Figure 6B and Supplementary data, Figure S5).

Mutagenesis of XXT2. The six amino acids revealed as important from the homology model and sequence alignment were individually mutated to various amino acids to investigate their role in XXT2 substrate binding and catalysis. All mutants were expressed without their N-termini and transmembrane domains and purified in parallel using identical conditions. Western blot analysis confirmed that all mutants have a similar level of expression (Figure 7B). Enzyme assays were performed for 20 h to determine the effect of these mutations on XXT2 ability to xylosylate cellohexaose at one or two glucoses as described in Materials and Methods. The products of the wild type XXT2 reaction analyzed by HPAEC (Figure 7A) were also verified using MALDI-TOF-MS (Figure 7C).

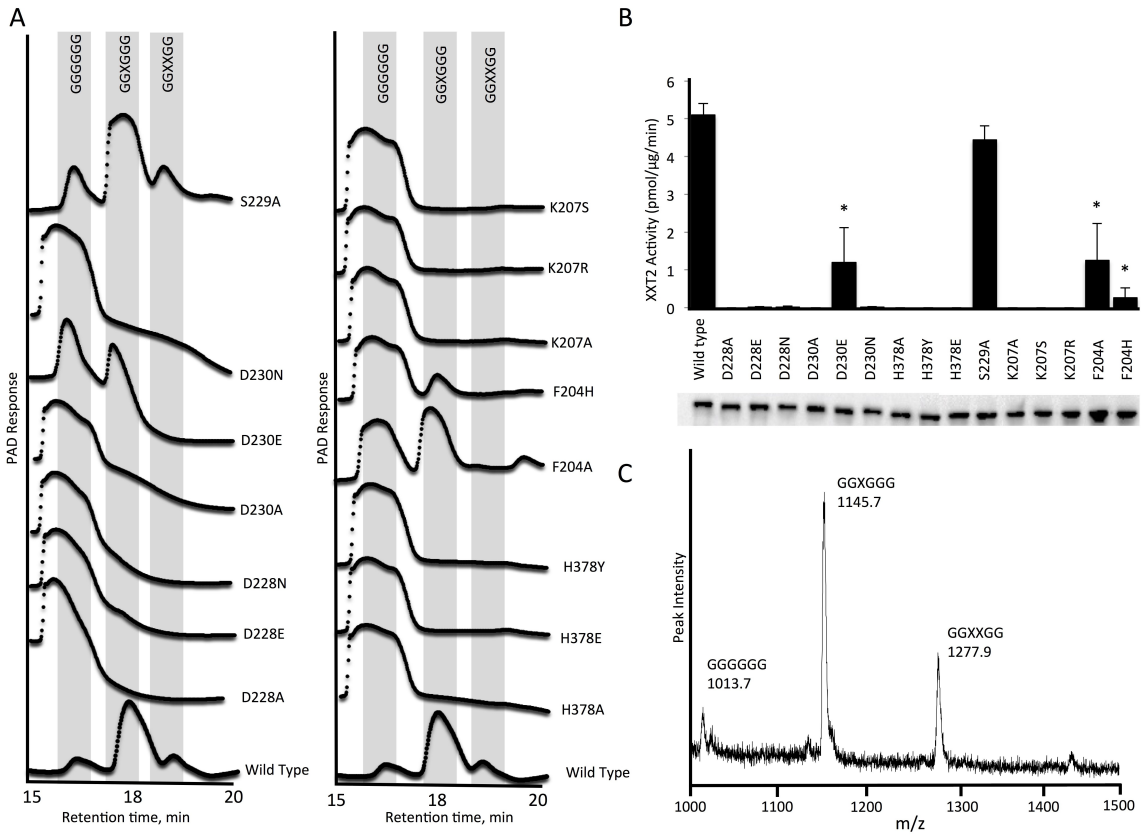


Figure 7. Site-directed mutagenesis of XXT2. (**A**) HPAEC chromatograms of the products produced in the reactions catalyzed by the XXT2 wild type and XXT2 mutants. GGGGGG: cellohexaose; GGXGGG: mono-xylosylated cellohexaose; GGXXGG: tri-xylosylated cellohexaose. (**B**) Enzyme activity (top) and the western blot of XXT2 wild type and mutant proteins after Ni-NTA column (bottom). Enzyme activity is shown in nmol according to xylose transferred to cellohexaose by 1 μg of XXT2 protein per minute. Asterisks indicate significantly different compared to wild type ($P<0.01$). (**C**) MALDI-TOF spectrum of the reaction product formed by wild type XXT2 protein correspond to the bottom chromatogram in (**A**). The mutants were produced as described in materials and methods. All mutant proteins were expressed without N-terminus and transmembrane domain, purified, and assayed for activity with identical conditions with equal protein concentration. The enzyme assay mixture consisted of 2mM UDP-Xyl, 0.5 mM cellohexaose, 2mM $MnCl_2$. Error bars represent standard deviation of three replicates. Wild type and mutant XXT2 proteins were expressed in parallel at least twice to verify activity.

Mutations of the Asp228 and Asp230 residues in the DXD motif resulted in a severe reduction in XXT2 activity for all mutants investigated. When either of these

residues was mutated to Ala or Asn, no activity was detected (Figure 7A and B). In contrast, mutation to Glu did not abolish activity completely, demonstrating that a carboxylate group at these positions is required for XXT2 activity (Figure 7A and B). The S229A mutation had no effect on XXT2 catalytic activity, indicating that only the Asp residues in the DXD motif are important for activity.

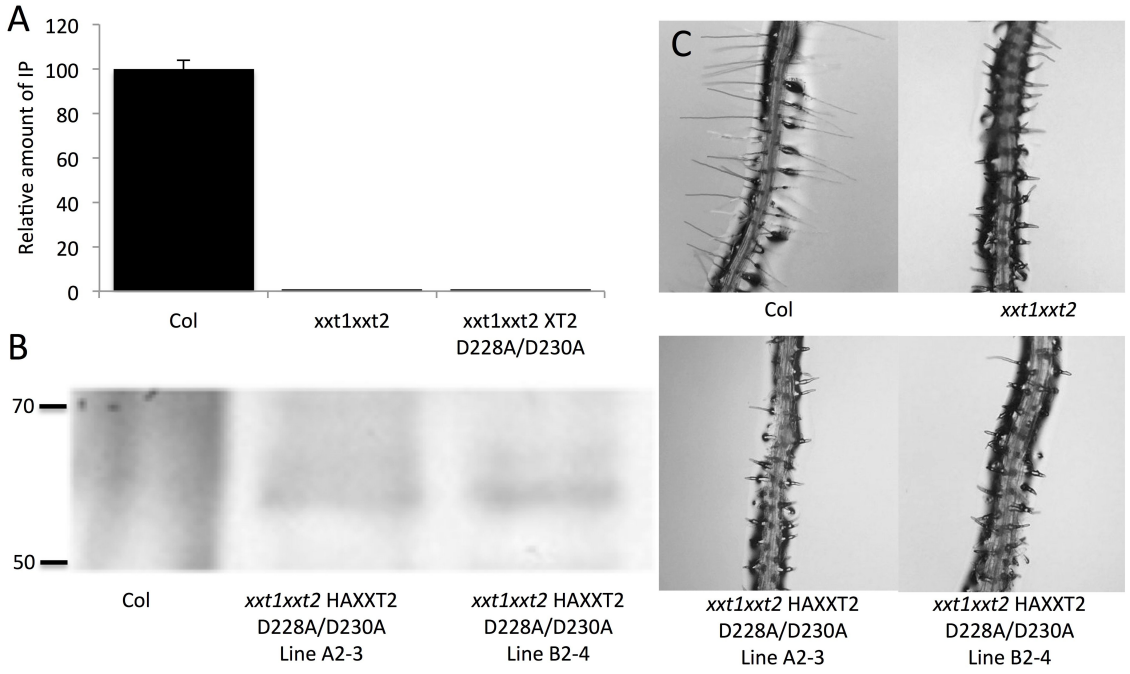


Figure 8. Over-expression of the XXT2 D228A/D230A mutant protein in *Arabidopsis xxt1xxt2* double mutant plants. (**A**) Driselase digestion of the cell wall extracted from the wild type Col-0, *xxt1xxt2* double mutant plants, and *xxt1xxt2* double mutant plants expressing the HAXXT2 D228A/D230A mutant protein. (**B**) Western blot of the total protein extracts from Arabidopsis wild type Col-0 and *xxt1xxt2* double mutant plants expressing the HAXXT2 D228A/D230A mutant protein. (**C**) The root hair phenotype of Col-0, *xxt1xxt2* double mutant, and *xxt1xxt2* double mutant over expressing HAXXT2 D228A/D230A mutant protein. Full-length XXT2 was expressed with 35S promotor.

To investigate the role of the DXD motif for XXT2 function *in vivo,* a full-length D228A/D230A double mutant protein was cloned and transformed into the *Arabidopsis xxt1xxt2* double mutant plants. The *xxt1xxt2* plants have no detectable XyG and root hairs that are shorter, thicker, and lower in abundance compared to the wild type plants. It was shown before that overexpression of non-mutated HA-XXT2 in the *xxt1xxt2* double mutant plants resulted in complementation of the mutant root hair phenotype (Cavalier et al. 2008). Here, overexpression of HA-XXT2 D228A/D230A mutant protein in *xxt1xxt2* double mutant plants was unable to restore the root hair phenotype and no XyG was detected (Figure 8A), indicating that the HA-XXT2 protein with both Asp residues mutated to Ala is unable to function *in vivo*. Analysis of total protein extract from transgenic plants by western blot using antibodies against HA epitope confirmed expression of the HA-XXT2 D228A/D230A mutant protein in the transgenic plants (Figure 8B). This confirmed that the *xxt1xxt2* phenotype was not complemented because of the lack of XXT2 activity and not because of the lack of protein itself.

His378 in the XXT2 model is predicted to be involved in the coordination of the manganese ion (Figure 6). XXT2 with His378 substituted to Ala was found to be inactive (Figure 7A and B). To investigate further, two additional mutants were cloned, H378Y and H378E. Mutation to Tyr was investigated because of its similarities to His, being a polar aromatic amino acid. Mutation to Glu was investigated because it contains a carboxylate group, similar to Asp228 and Asp230, which are also involved in the coordination of the divalent cation. Both H378Y and H378E mutant proteins were shown to be inactive (Figure 7A and B). These results

further demonstrate the importance of proper coordination of the divalent cation for XXT2 catalytic activity.

All mutations of Lys207 resulted in inactive proteins, including the K207R mutation, which would retain the positive charge of the amino acid at this position. On the contrary, mutations of Phe204 did not severely affect XXT2 activity (Figure 7B), however, the F204H mutant showed lower activity in comparison with the F204A mutant.

Discussion

Plant cell wall synthesizing GTs have been extensively studied using reverse-genetics and heterologous expression to confirm protein function. Despite this extensive effort to understand polysaccharide biosynthesis, there is little information about their atomic structures or amino acids involved in substrate binding of any of the XyG biosynthetic enzymes. Information about the atomic structure, catalytic mechanism, and amino acids involved in substrate binding is needed to fully understand the biosynthesis of the plant cell wall polysaccharides. The earlier reverse-genetics and biochemical studies demonstrated that XXT1 and XXT2 are the key proteins responsible for xylosylation of the glucan backbone in Arabidopsis xyloglucan (Cavalier and Keegstra, 2006; Cavalier et al., 2008). It was also demonstrated that XXT2 forms heterocomplexes with other GTs involved in xyloglucan biosynthesis, whereas XXT1 is less involved in complex formation (Chou et al., 2012, 2015). In addition, the *xxt2xxt5* double mutant plants had a significantly larger reduction of XyG content in comparison with the *xxt1xxt5* double mutant,

which indicates that XXT2 alone is more effective in xylosylation of the backbone compared to XXT1. These prior studies motivated us to choose XXT2 for the first investigation of molecular determinants of XXT activity. We built and utilized a homology model of XXT2 to predict amino acids that are involved in donor substrate binding and, thus, are critical for catalytic activity. Site-directed mutagenesis was used to validate the predicted 3-D structure of the XXT2 catalytic domain and to confirm the function of the selected amino acids in XXT2 activity.

Homology modeling and molecular dynamics simulations revealed that XXT2 likely adopts a GT-A fold, which consists of a central β-sheet surrounded by α-helices (Figure 1B). GT-A fold GTs have been shown to be metal-ion dependent. The metal ion is involved in coordination of donor substrate by forming a coordination bond with the negatively charged oxygen groups of the UDP-sugar. In addition, the metal ion may help to neutralize the newly formed negative charge on the phosphate group when the sugar moiety is transferred. Mutation of the two Asp residues (Asp228 and Asp230) in the DXD motif of XXT2 demonstrates that this motif is critical for XXT2 enzymatic activity. Only the conservative mutation to another acidic amino acid Glu, which retains the carboxylate functional group, produced catalytically active XXT2, yet activity was significantly reduced compared to the wild type protein, probably due to the longer side chain and additional Cγ atom in Glu that disturbs the active site geometry. In addition, over-expression of the XXT2 D228A/D230A mutant in the *xxt1xxt2* double mutant plant did not complement the root hair phenotype or restore the XyG content, suggesting that the DXD motif is critical for XXT2 function *in vivo* (Figure 8). Similar results have been

reported in numerous other studies, where mutations in DXD motifs resulted in either inactive GTs or proteins with severely reduced activity (Busch et al. 1998; Gotting et al. 2004; Li et al. 2001). For example, in the case of another xylosyltransferase, XT-I, it was demonstrated that mutation of the first Asp residue in the DXD motif to Gly resulted in significant reduction in activity, whereas mutation of the same Asp to the Glu residue had no effect on protein activity (Gotting et al. 2004). Mutations of the second Asp residue in the DXD motif of XT-I to either Gly or Glu residues resulted in an approximately 50% reduction in activity (Gotting et al. 2004). Mutations in XXT2 also showed the Asp228 residue being more sensitive to mutagenesis compared to Asp230. For example, the activity of XXT2 D228E mutant was severely reduced, resulting in only a minor amount of xylosylated product being produced. In contrast, the D230E mutant produced a significant amount of mono-xylosylated product, albeit in lower quantity in comparison with the wild type XXT2 (Figure 7). Previous studies demonstrated that mutations in the DXD motif did not impact GT's proper folding, thus it was concluded that the lack of enzymatic activity in the mutants was not due to misfolding (Li et al. 2001; Wiggins and Munro 1998). However, some local structural changes around the DXD motif have also been reported when it was mutated (Zhang et al. 2001). From these previous reports, and the similar levels of expression of the XXT2 mutants (Figure 7B), it is plausible that the mutations do not have an effect on XXT2 folding, however, further detailed studies are needed to confirm this.

Previous studies on GTs have also demonstrated that the divalent cation is additionally coordinated by a closely localized His residue (Lobsanov et al. 2004;

Zhang et al. 2007). In XXT2, mutation of the His378 residue to Ala, Tyr, or Glu resulted in complete inactivation of the enzyme. Glu was selected in an effort to introduce a negatively charged carboxylate group, which was anticipated to interact with the divalent cation similar to the Asp carboxylate groups in the DXD motif. However, this mutant showed no catalytic activity. These results, together with the results obtained with the mutants in the DXD motif, demonstrate that the proper coordination of the divalent cation is critical for XXT2 activity.

All Lys207 mutants of XXT2 analysed in this study were inactive. In the A64Rp structure, which was found to bind only UDP-Glc, the Arg57 residue of A64Rp was shown to interact with the carbon four hydroxyl group of glucose (Zhang et al. 2007). However, the model structure of the UDP-Xyl bound XXT2 protein indicates a possible interaction of Lys207 with the hemi-acetal oxygen between the first and the fifth carbon in the xylose ring (Figure 6B and Supplementary data, Figure S5). Positively charged residues in the active site have been shown to allow GTs to remain in the “open” conformation as a result of the mutual repulsion with other positively charged residues on a mobile loop. The repulsion of these positively charged residues in the active site and the mobile loop is neutralized by forming ionic interactions with the negatively charged diphosphate group of UDP bound to the active site, inducing a conformational change of the GT into the “closed state” (Johal et al. 2014). It is possible that the positively charged side chain of Lys207 which resides in the active site of XXT2 functions to maintain the “open state” much like in the ABO(H) GTs described above. As a result, the XXT2 Lys207 mutants may reside in “closed state”, disallowing UDP-Xyl to bind. However,

the K207R mutant did not have catalytic activity, even though the positive charge of the amino acid was retained (Figure 7). This indicates that the function of Lys207 is different from controlling the open-closed state of XXT2. The functional role of Lys207 will need further investigation in future studies.

The F204A mutation resulted in a slight reduction in catalytic activity compared to wild type XXT2, with protein activity further reduced in the F204H mutant. His54 in A64Rp, which aligns with Phe204 of XXT2, forms a hydrogen bond with the carbon six hydroxyl of glucose. In the XXT2 homology model, Phe204 is positioned about 7 Å away from carbon five of xylose. It is possible that mutating Phe204 to His results in unfavourable interactions with nearby amino acids, resulting in small local conformational changes, whereas mutating to Ala does not have this affect. Further work is needed to confirm this.

Conclusions

Plant polysaccharide synthesizing GTs have been extensively studied to elucidate their function, yet there is currently little information about their structures and the residues involved in substrate binding and catalysis. Here, we report a homology model built to predict a 3-D structure of the XXT2 catalytic domain and to reveal amino acids involved in catalytic activity, which were verified using site-directed mutagenesis. Molecular dynamics simulations revealed that the homology model is indeed able to retain the GT-A fold of the template structure used to build the homology model thus indicating that XXT2 likely has a GT-A fold. Six amino acids were identified to contribute to the *in vitro* activity of XXT2. The

amino acids predicted to be involved in the coordination of the manganese ion, Asp228, Asp230, and His378, are critical for XXT2 activity as demonstrated by obtaining mutants either inactive or with severely reduced activity for all mutations investigated. In addition, Lys207 was also found to be critical for activity and may be involved in the binding of UDP. Substitutions of Phe204 had less impact on XXT2 activity with the largest effect when a polar residue replaced the hydrophobic Phe. This is the first reported study investigating the amino acids involved in substrate binding by plant xylosyltransferase involved in XyG biosynthesis. This information will aid in future work in order to understand the mechanisms of plant polysaccharide biosynthesis, and will further aid in engineering plant cell walls for efficient biofuel production and the production of novel polysaccharide-based composites for industrial and medical applications.

References

Ade CP, Bemm F, Dickson JM, Walter C, Harris PJ. 2014. Family 34 glycosyltransferase (GT34) genes and proteins in *Pinus radiata* (radiata pine) and *Pinus taeda* (loblolly pine). *Plant J*. 78:305-318.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.

Benson NC, Daggett V. 2008. Dynameomics: large-scale assessment of native protein flexibility. *Protein Sci.* 17:2038-2050.

Boyken SE, Fulton DB, Andreotti AH. 2012. Rescue of the aggregation prone Itk Pleckstrin Homology domain by two mutations derived from the related kinases, Btk and Tec. *Protein Sci.* 21:1288-1297.

Busch C, Hofmann F, Selzer J, Munro S, Jeckel D, Aktories K. 1998. A Common Motif of Eukaryotic Glycosyltransferases Is Essential for the Enzyme Activity of Large Clostridial Cytotoxins. *J Biol Chem*. 273:19566-19572.

Campbell J, Davies G, Bulone V, Henrissat B. 1997. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *J Biochem.* 326:929-939.

Canutescu AA, Shelenkov AA, Dunbrack Jr RL. 2003. A graph theory algorithm for rapid protein side chain prediction. *Protein Sci.* 12:2001-2014.

Carpita NC. 2012. Progress in the biological synthesis of the plant cell wall: new ideas for improving biomass for bioenergy. *Curr Opin Biotechnol*. 23:330-337.

Carpita NC, Gibeaut DM. 1993. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* 3:1-30.

Cavalier DM, Keegstra K. 2006. Two xyloglucan xylosyltransferases catalyze the addition of multiple xylosyl residues to cellohexaose. *J Biol Chem*. 281:34197-34207.

Cavalier DM, Lerouxel O, Neumetzler L, Yamauchi K, Reinecke A, Freshour G, Zabotina OA, Hahn MG, Burgert I, Pauly M*, et al.* 2008. Disrupting two *Arabidopsis thaliana* xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component. *Plant cell*. 20:1519-1537.

Chou Y, Pogorelko G, Young ZT, Zabotina OA. 2015. Protein-Protein Interactions Among Xyloglucan-Synthesizing Enzymes and Formation of Golgi-Localized Multiprotein Complexes. *Plant Cell Physiol.* 56: 255-267

Chou YH, Pogorelko G, Zabotina OA. 2012. Xyloglucan xylosyltransferases XXT1, XXT2, and XXT5 and the glucan synthase CSLC4 form Golgi-localized multiprotein complexes. *Plant Physiol*. 159:1355-1366.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16:735-743.

Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, Raikhel N, Wilkerson CG. 2007. A gene from the cellulose synthase-like C family encodes a β-1,4 glucan synthase. *PNAS.* 104:8550-8555.

Coutinho PM, Deleury E, Davies GJ, Henrissat B. 2003. An Evolving Hierarchical Family Classification for Glycosyltransferases. *J Mol Biol.* 328:307-317.

Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res. 43*(W1), W389-W394.

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. 1995. A smooth particle mesh Ewald method. *J Chem Phys.* 103:8577-8593.

Faik A, Price NJ, Raikhel NV, Keegstra K. 2002. An *Arabidopsis* gene encoding an α-xylosyltransferase involved in xyloglucan biosynthesis. *PNAS.* 99:7797-7802.

Gibbons BJ, Roach PJ, Hurley TD. 2002. Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J Mol Biol*. 319:463-477.

Gloster TM. 2014. Advances in understanding glycosyltransferases from a structural perspective. *Curr Opin Struct Biol.* 28:131-141.

Gotting C, Muller S, Schottler M, Schon S, Prante C, Brinkmann T, Kuhn J, Kleesiek K. 2004. Analysis of the DXD motifs in human xylosyltransferase I required for enzyme activity. *J Biol Chem.* 279:42566-42573.

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res*. 38:W695-699.

Grubmüller H, Tavan P. 1998. Multiple time step algorithms for molecular dynamics simulations of proteins: How good are they?. *J Comput Chem*. *19*.13:1534-1552

Hayashi T. 1989. Xyloglucans in the primary cell wall. *Annu Rev Plant Biol.* 40:139-168.

Hooft RW, Vriend G, Sander C, Abola EE. 1996. Errors in protein structures. *Nature*. 381:272.

Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins. 65*:712-725.

Jamaluddin H, Tumbale P, Withers SG, Acharya KR, Brew K. 2007. Conformational changes induced by binding UDP-2F-galactose to α-1,3 galactosyltransferase-implications for catalysis. *J Mol Biol*. 369:1270-1281.

Johal AR, Blackler RJ, Alfaro JA, Schuman B, Borisova S, Evans SV. 2014. pH-induced conformational changes in human ABO(H) blood group glycosyltransferases confirm the importance of electrostatic interactions in the formation of the semi-closed state. *Glycobiology*. 24:237-246.

Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. 2006. MUSTANG: a multiple structural alignment algorithm. *Proteins*. 64:559-574.

Krieger E, Dunbrack RL, Jr., Hooft RW, Krieger B. 2012. Assignment of protonation states in proteins and ligands: combining pKa prediction with hydrogen bonding network optimization. *Methods Mol Biol*. 819:405-421.

Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K. 2009. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins*. 77:114-122.

Krieger E, Koraimann G, Vriend G. 2002. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins*. 47:393-402.

Krieger E, Nielsen JE, Spronk CA, Vriend G. 2006. Fast empirical pKa prediction by Ewald summation. *J Mol Graph Model*. 25:481-486.

Krieger E, Vriend G. 2014. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics*. 30:2981-2982.

Krieger E, Vriend G. 2015. New ways to boost molecular dynamics simulations. *J Comput Chem*, 36:996-1007.

Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem.* 77:521-555.

Laskowski RA. 2009. PDBsum new things. *Nucleic Acids Res*. 37:D355-359.

Li J, Rancour DM, Allende ML, Worth CA, Darling DS, Gilbert JB, Menon AK, Young WW, Jr. 2001. The DXD motif is required for GM2 synthase activity but is not critical for nucleotide binding. *Glycobiology*. 11:217-229.

Li P, Roberts BP, Chakravorty DK, Merz KM, Jr. 2013. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J Chem Theory Comput*. 9:2733-2748.

Linding R, Russell RB, Neduva V, Gibson TJ. 2003. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* 31:3701-3708.

Liu J, Mushegian A. 2003. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein science.* 12:1418-1431.

Lobsanov YD, Romero PA, Sleno B, Yu B, Yip P, Herscovics A, Howell PL. 2004. Structure of Kre2p/Mnt1p: a yeast alpha1,2-mannosyltransferase involved in mannoprotein biosynthesis. *J Biol Chem.* 279:17921-17931.

Madson M. 2003. The MUR3 Gene of Arabidopsis Encodes a Xyloglucan Galactosyltransferase That Is Evolutionarily Related to Animal Exostosins. *Plant Cell.* 15:1662-1670.

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. 2015. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput.* 11: 3696-3713.

Mansoori N, Schultink A, Schubert J, Pauly M. 2015. Expression of heterologous xyloglucan xylosyltransferases in Arabidopsis to investigate their role in determining xyloglucan xylosylation substitution patterns. *Planta*. 241: 1145-1158

McCann MC, Carpita NC. 2008. Designing the deconstruction of plant cell walls. *Curr Opin Plant Biol.* 11:314-320.

McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* 41:W597-600.

Park YB, Cosgrove DJ. 2015. Xyloglucan and its interactions with other components of the growing cell wall. *Plant Cell Physiol.* 56:180-194.

Pauly M, Keegstra K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J.* 54:559-568.

Schultink A, Cheng K, Park YB, Cosgrove DJ, Pauly M. 2013. The identification of two arabinosyltransferases from tomato reveals functional equivalency of xyloglucan side chain substituents. *Plant Physiol.* 163:86-94.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J*, et al.* 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 7:539.

Stewart JJ. 1990. MOPAC: a semiempirical molecular orbital program. *J Comput Aided Mol Des*. 4:1-105.

Wang C, Li S, Ng S, Zhang B, Zhou Y, Whelan J, Wu P, Shou H. 2014. Mutation in xyloglucan 6-xylosytransferase results in abnormal root hair development in Oryza sativa. *J Exp Bot*. 65:4149-4157.

Waterman MS, Smith TF, Beyer WA. 1976. Some biological sequence metrics. *Adv Math.* 20:367-387.

Wiggins CA, Munro S. 1998. Activity of the yeast MNN1 α-1, 3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *PNAS.* 95:7945-7950.

Yin Y, Chen H, Hahn MG, Mohnen D, Xu Y. 2010. Evolution and function of the plant cell wall synthesis-related glycosyltransferase family 8. *Plant Physiol.* 153:1729-1746.

Zabotina OA. 2012. Xyloglucan and its biosynthesis. *Front Plant Sci.* 3:134.

Zabotina OA, Avci U, Cavalier D, Pattathil S, Chou YH, Eberhard S, Danhof L, Keegstra K, Hahn MG. 2012. Mutations in multiple XXT genes of *Arabidopsis* reveal the complexity of xyloglucan biosynthesis. *Plant Physiol.* 159:1367-1384.

Zabotina OA, van de Ven WT, Freshour G, Drakakaki G, Cavalier D, Mouille G, Hahn MG, Keegstra K, Raikhel NV. 2008. *Arabidopsis* XXT5 gene encodes a putative α-1,6-xylosyltransferase that is involved in xyloglucan biosynthesis. *Plant J.* 56:101-115.

Zhang Y, Wang PG, Brew K. 2001. Specificity and mechanism of metal ion activation in UDP-galactose:β-galactoside-α-1,3-galactosyltransferase. *J Biol Chem.* 276:11567-11574.

Zhang Y, Xiang Y, Van Etten JL, Rossmann MG. 2007. Structure and function of a chlorella virus-encoded glycosyltransferase. *Structure.* 15:1031-1039.

CHAPTER 6. STRUCTURE OF XYLOGLUCAN XYLOSYLTRANSFERASE 1: HOW SIMPLE STERIC RULES DEFINE BIOLOGICAL PATTERNS OF XYLOGLUCAN POLYMERS

Modified from a paper submitted to *PNAS*

Alan T. Culbertson, Jacqueline J. Ehrlich, Jun-Yong Choe, Richard B. Honzatko, Olga A. Zabotina

Abstract

The plant cell wall is a complex network composed mainly of polysaccharides, the most abundant biopolymers on earth and rich source of biorenewable materials. Biosynthesis of these biopolymers is poorly understood, largely due to difficulties in the structural characterization of glycyosyltransferases and lack of suitable substrates for in vitro analysis. The dearth of structural information for enzymes involved in plant cell-wall polysaccharide biosynthesis impedes the development of more resilient plants better suited for numerous industrial applications. Presented here are structures of Arabidopsis xyloglucan xylosyltransferase 1 (XXT1) without ligands and in complexes with UDP and cellohexaose. XXT1 initiates side-chain extensions from a linear glucan polymer by transferring the xylosyl group from UDP-xylose during xyloglucan biosynthesis. XXT1, a homodimer and member of the GT-A fold family of glycosyltransferases, binds UDP analogously to other GT-A fold enzymes. Structures here and the

properties of mutant XXT1s are consistent with a SN_i-like catalytic mechanism. Distinct from other systems is the recognition of cellohexaose by way of an extended cleft. XXT1 alone cannot produce xylosylation patterns observed for native xyloglucans because of steric constraints imposed by the acceptor binding cleft. Homology modeling of XXT2 and XXT5, the other two xylosyltransferases involved in xyloglucan biosynthesis, reveals a structurally altered cleft in XXT5 that could accommodate a partially xylosylated glucan chain produced by XXT1 and/or XXT2. An assembly of the three XXTs can produce the xylosylation patterns of native xyloglucans, suggesting the involvement of an organized multi-enzyme complex in the xyloglucan biosynthesis.

Introduction

Plant cell walls consist of cellulose, hemicellulose, pectin, and lignin, all of which confer mechanical properties to plant structures, and are important for shape and development. Plant cell walls represent the largest pool of renewable carbohydrate and the potential to support numerous industrial applications in bioenergy and biomaterials (1,2). Efficient processing of lignocellulosic biomass is limited by its complex structure, which resists enzymatic and microorganism degradation (3). Engineering a biologically viable plant susceptible to enzymatic or non-biological degradation requires a complete understanding of plant cell wall polysaccharide biosynthesis and structure.

Xyloglucan (XyG) is the most abundant hemicellulose in the primary cell wall of dicotyledonous plants, and has many proposed structural and regulatory functions (4-7). XyG consists of a β-(1,4)-linked glucan backbone branched with various glycosyl residues depending on species or tissue (8). The nomenclature for XyG structure is as follows: G represents an unbranched glucose unit, whereas X, L, and F are glucosyls units with Xyl, Gal-Xyl, or Fuc-Gal-Xyl side-chains , respectively (9). *Arabidopsis* XyG consists of a glucan backbone branched with α-(1,6)-linked D-Xyl residues, resulting in XXXG-type pattern (10,11). The xylosyl units of the XXXG-type xyloglucan can undergo attachment of β-D-Gal to the 2-hydroxyl groups of xylose, followed by the attachment of α-L-Fuc to the 2-hydroxyl groups of β-D-Gal units.

Glycosyltransferases (GTs) catalyze the formation of glycosidic bonds by transferring a sugar moiety from an activated donor, typically a nucleotide-sugar, to a variety of acceptor substrates, including, carbohydrates, proteins and lipids (12). Amino acid sequences in the Carbohydrate Active Enzyme Database fall into 105 families of GTs (13,14). Available structures indicate most GT families adopt one of two folds, GT-A or GT-B, although a rarer GT-C fold has been proposed (12). The GT-A fold has two Rossmann-like domains that form a central β-sheet, each face of which is covered by α-helices. These are typically metal-dependent enzymes that require an Asp-x-Asp motif for metal coordination (15,16). GT-B folds also have two Rossmann-like domains, less tightly associated than those of GT-A folds, with the active site located between domains. GTs are also classified by the stereochemistry of the glycosidic bond in the product (inverted or retained) relative to that of the substrate (12). The catalytic mechanism of inverting GTs likely follows the single

displacement mechanism of inverting glycosyl hydrolases (12,17). The catalytic mechanism of retaining GTs, first proposed as a double displacement mechanism similar to retaining glycosyl hydrolases, has fallen into disfavor due to the absence of a suitably-placed catalytic base and the failure to trap a glycosyl-enzyme intermediate. Instead, retaining GTs may employ a SN_i-like mechanism (18-20). Although structural information is abundant for glycosyltransferases (21), structural information specifically for GTs involved in plant cell wall polysaccharide biosynthesis is available only for xyloglucan fucosyltransferase 1 (FUT1), which adds fucose to the terminal position of XyG side chains (22,23).

Xylosylation of the 6-hydroxyl group of glucose, catalyzed by xyloglucan xylosyltransferases (XXTs), is the first step in building branches on the XyG backbone (8,24). XXTs are type II transmembrane enzymes, having a short cytosolic N-terminal region, a transmembrane domain, a stem-region, and a large C-terminal catalytic domain localized in the Golgi lumen (25-27). Three XXTs xylosylate the XyG backbone both in vivo (28-30) and in vitro (25, 31-34). Enzymatic action of XXTs in vitro employ short glucans, such as cellohexaose or cellopentaose, due to the low solubility of long glucan chains. XXT1 (and XXT2) primarily adds the first xylosyl residue to the fourth glucose residue from the reducing end of cellohexaose to produce mono-xylosylated cellohexaose. XXT1 (and XXT2) then adds a xylosyl residue to the third or fifth glucose from the reducing end of cellohexaose (32).

Homology modeling of XXT2 was consistent with a GT-A fold, indicating residues critical for donor substrate binding and the Asp-X-Asp motif (34). Low sequence similarity of XXT2 to the threading template in the binding region of the acceptor

molecule undermined confidence in the model. Moreover, the template for the homology model, lacking N- and C-terminal regions of XXT2, covered only residues 140-415. Subsequent studies revealed that C-terminal residues omitted in the model are critical for catalysis (25).

Presented here are crystal structures of XXT1 without ligands, a binary complex with UDP, and ternary complex with UDP and cellohexaose. This is the first structure from the GT Family 34 plant protein and only the second crystal structure of a plant cell wall glycosyltransferase (the first being of FUT-1). XXT1 is a homodimer in solution and crystal. Dimer interface residues are conserved amongst XXT1, XXT2 and XXT5, indicating possible heterodimer formation. Sequence-conserved surface charges on XXT dimers suggest possible interactions with other proteins involved in biosynthesis of XyG. Structural comparisons with retaining GT-A fold glycosyltransferases and directed mutations support a SN_i-like catalytic mechanism. Moreover, the crystal structure and homology models indicate similar steric requirements for glucan interactions with XXT1 and XXT2, but different requirements for XXT5, suggesting the participation of XXT1 (or XXT2) with XXT5 in achieving the in vivo pattern of xylosyl transfer to the linear glucan.

Results

Overall structure of XXT1. Protein expression of the XXT1 stem region and catalytic domain (residues 45-460) was performed in human embryonic kidney cells (35). The structure of the enzyme (in crystals of spacegroup $P2_12_12_1$) was solved by single-wavelength anomalous diffraction (SAD), using data from a crystal

derivatized with K_2HgI_4. The complex of UDP/Mn^{2+} (hereafter, the binary complex) and the complex of cellohexaose/UDP/Mn^{2+} (hereafter the ternary complex) were formed by ligand soaks. A second crystal form of the apo-enzyme (spacegroup $C222_1$), diffracting to 1.5 Å resolution, was solved by molecular replacement. Given its superior resolution, the apo-enzyme in spacegroup $C222_1$ is reported in Table S1. Regardless of crystal form, the asymmetric unit has two subunits of XXT1, and for each, residues 45-115 from the N terminus and residues 454-460 from the C terminus are without electron density. The purified protein has an observed mass consistent with that expected for residues 45-460. Finally, a substantial void exists in both crystal forms that could accommodate 70 additional residues.

XXT1 adopts a GT-A fold with a central β-sheet having both faces covered by α-helices (Fig. 1*A* and *C*). The central β-sheet contains strands β2, β1, β3, β6, β5, and β7, all of which are parallel except for strand β6 (Fig. 1*A* and *C*). The loops extending from this central β-sheet and surrounding α-helices define the active site of XXT1, containing the Asp-x-Asp motif (site of Mn^{2+} binding), donor substrate binding site, and acceptor binding site (Fig. 1*A*). The active site is a cleft roughly 13 Å wide and 30 Å long (Fig. 1*A*).

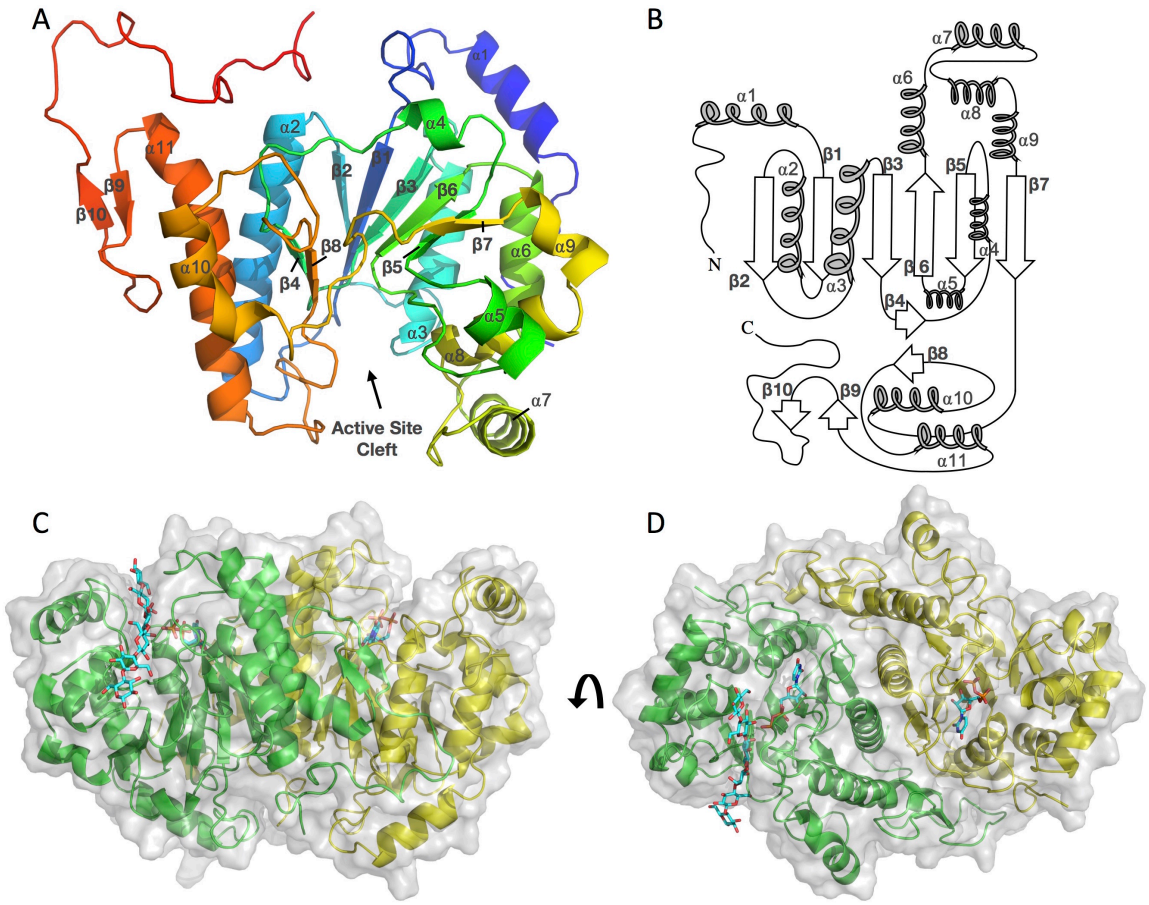


Fig. 1. Structure overview of XXT1. (*A*) XXT1 Chain A colored blue to red from N- to C-terminus, respectively. Chain B not shown. (*B*) Secondary and tertiary structure of XXT1. Names of α-helices and β-strands correspond to those in *panel a*. (*C*) XXT1 dimer with UDP and cellohexaose. Separate monomers of the dimer are colored green and yellow. UDP binds to both monomers, whereas cellohexaose binds only to one monomer. (*D*) Image of *panel C* rotated 90° to view down the symmetry axis of the dimer.

XXT1 Forms a Dimer. XXT1 forms a dimer with an interface of roughly 2900 Å^2 (PDBe PISA; http://www.ebi.ac.uk/pdbe/pisa/). XXT1 elutes from a gel filtration column with an expected size of roughly 101 kDa, consistent with a dimer subunit structure (Fig. S1*A* and *B*). The interface between subunits involves the side chains of Leu171, Leu172, Ile175, Ile179 and Tyr191 from helix α2 and strand β2 from each subunit (Fig. 1*D*). Surrounding this hydrophobic core are electrostatic

interactions and hydrogen bonds obeying the twofold symmetry of the dimer: Gln432-NE2 to the backbone carbonyl of Leu119; Asp168-OD1 to His169-ND1; the backbone amide of Tyr191 to Lys176-NZ; Tyr414-OH to His217-NE2; Glu188-O1 to Lys451-NZ; and Glu219-OE1 to His447-NE2. Interacting residues at the subunit interface of the dimer are identical in XXT1 and XXT2 suggesting the possibility of hetero-dimer formation (Fig. S2). Although twofold molecular symmetry is obeyed across the subunit interface, cellohexaose binds to only one of two active sites of the UDP/cellohexaose complex. The basis for this asymmetry in the binding of cellohexaose is provided below.

Donor Substrate Binding. Binary (UDP) and ternary (UDP and cellohexaose) structures of XXT1 were obtained by soaking XXT1 crystals with the substrates, along with $MnCl_2$, for two days; the structures were refined against data to a resolution of 2.1 Å and 2.4 Å, respectively (Table S1). In both the binary and ternary complexes, Mn^{2+} is coordinated through a monodentate interaction with Asp227, bidentate interaction with Asp229, and NE2 of His377. The UDP molecules have different conformations in the presence and absence of cellohexaose, distinguished by the torsional angle defined by atoms O4', C4', C5'and O5' (the rotational angle γ). In the absence of cellohexaose, $\gamma=-62°$ (hereafter the bent conformer), whereas in the presence of cellohexaose $\gamma=175°$ (extended conformer) (Fig. 2 and Fig. S3). The extended conformer coordinates Mn^{2+} through one oxygen atom from each of the α- and β-phosphoryl groups. The bent conformer coordinates Mn^{2+} through one oxygen atom of the β-phosphoryl group and a nitrate anion. In both the bent and

extended conformations of UDP, the Mn^{2+} coordination is octahedral. Interactions are identical for the bent conformer of UDP in the active site of the ternary complex that lacks cellohexaose and the active sites of the binary complex. The assignment of separate molecules of nitrate and UDP (bent conformer) rather than an equal-occupancy combination of bent and extended conformers of UDP rests on negative difference electron density at the oxygen atom bridging the α- and β-phosphoryl groups of the extended conformer (eliminated by the replacement of the extended conformer with nitrate) and strong positive difference density associated with the α-phosphoryl group of the bent conformer (eliminated by assigning full occupancy to the bent conformer) (Fig. S3). The β-phosphoryl group of the extended UDP conformer hydrogen bonds with the 6-hydroxyl group and the 2-hydroxyl group of the fourth and fifth glucose units, respectively, from the reducing end of bound cellohexaose. In addition, atom NZ of Lys382 hydrogen bonds with oxygen atoms of the α- and β-phosphoryl groups of the extended conformer (Fig. 2).

The base and ribosyl moieties of UDP are positioned and interact similarly in bent and extended conformers. Atoms O2'and O3'of the ribosyl group hydrogen bond with the backbone amide and side chain, respectively, of Ser229, (Fig. 2). In addition, the backbone carbonyl of Gly156 hydrogen bonds with atom O2’ of the ribosyl residue. The base of UDP does not hydrogen bond with XXT1, making instead hydrophobic contacts with Ala158 and Gly202.

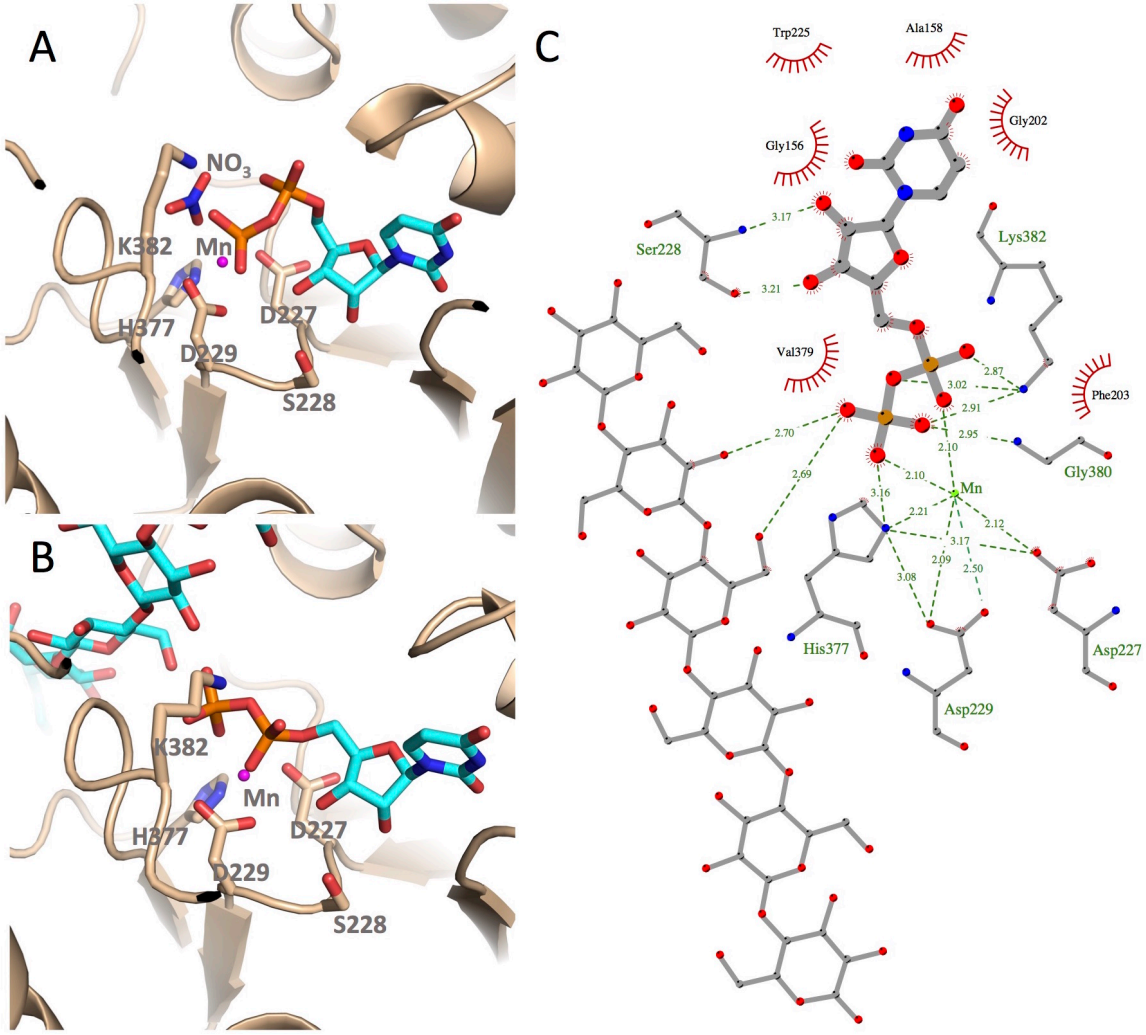


Fig. 2. Bound UDP in two conformations. (*A*) Bent conformer, cellohexaose absent. (*B*) Extended conformer, cellohexaose present. Colors in *panels a* and *b* are as follows: Tan: XXT1 secondary structure; blue: nitrogen atoms; red: oxygen atoms, cyan: carbon atoms of ligands; orange: phosphate atoms; purple: Mn^{2+}. (*C*) Interaction map for the extended conformer of UDP from LigPlot+(50).

Modeling of UDP-Xylose. Attempts to co-crystallize UDP-xylose with XXT1 or soak UDP-xylose into preformed crystals did not reveal a bound donor substrate. Instead, UDP-xylose was modeled into the active site of the ternary structure. GT structures typically have similar bound UDP-sugar conformations in which the sugar residue folds over the phosphoryl groups. A analogous conformation for UDP-

xylose is present in the structure of XXYLT1 (PDB:4WNH). That model for UDP-xylose, when superimposed on the extended UDP molecule in XXT1, fits the active site, with the xylosyl residue occupying a pocket of uniform/unstructured electron density. Energy minimization led to the model in Fig. 3, which retains the conformations of UDP, cellohexaose and active-site side chains (Fig. S4). The 3-hydroxyl group of xylose hydrogen bonds with Gln319 and Thr269, and the 4-hydroxyl group hydrogen bonds with Lys207 and Asp318 (Fig. 3). Phe203, which has a high B-parameter in crystal structures, makes a favorable contact with atom C5 of the xylosyl residue, a contact that would discriminate against a hexose sugar.

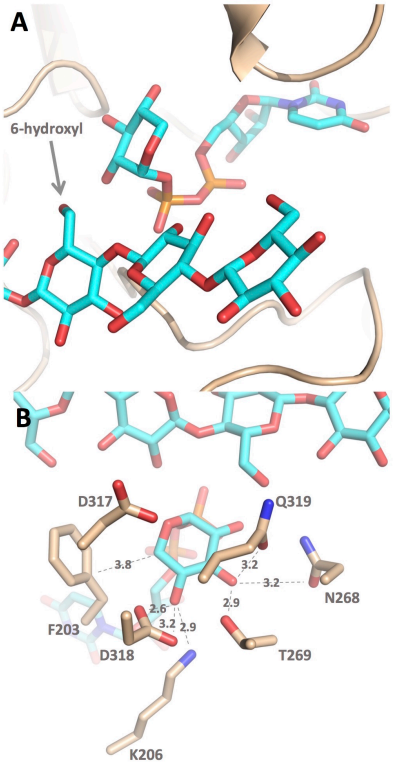


Fig. 3. Model of bound UDP-xylose. (*A*) Position of UDP-xylose with respect to cellohexaose. The arrow marks the 6-hydroxyl group of the fourth glucose from the reducing end. (*B*) Distances (in Å) between selected atoms of XXT1 and xylosyl group of UDP-xylose. Colors as define in Fig. 2

Acceptor substrate binding. The ternary complex of XXT1 has only one monomer of the dimer with bound cellohexaose (Fig. 1*C* and *D*). Access to the active site in the non-occupied monomer is limited due to packing contacts. The cellohexaose molecule is covered by strong electron density and has B-parameters significantly lower than those of UDP (Fig. S3*A*). The interactions of XXT1 with the glucosyl residue of cellohexaose defines six subsites (hereafter subsites 1-6). Subsite-1 binds the reducing end of cellohexaose. The 6-hydroxyl of the fourth glucosyl residue projects into the active site of XXT1 (Fig. 4), in agreement with the predominant modification of cellohexaose at its fourth glucosyl residue in solution (32). Glucose residues-3, -4 and -5 of cellohexaose make direct hydrogen bonds at subsites-3,-4 and -5, respectively. Conversely, the first, second, and sixth glucosyl residues have water-mediated or hydrophobic interactions at subsites-1, -2 and -6, respectively (Fig. 4). Glucosyl residues at the ends of cellohexaose have weak electron density, whereas residues 2-4 have strong electron density with clearly defined 6-hydroxyl groups that unambiguously define the orientation of cellohexaose and the placement of its reducing and non-reducing ends. The strong electron density of 6-hydroxyl groups of cellohexaose suggests a single allowed orientation of the bound glucan to XXT1 (Fig. 4*B* and Fig. S3*A*). The 6-hydroxyl group of glucosyl residue 4, positioned as an acceptor for the xylosyl residue, hydrogen bonds with an oxygen atom from the β-phosphoryl group of UDP (extended conformer). The glucosyl residue at subsite-3 hydrogen bonds with

His346 and stacks with Tyr348 (Fig. 4). Glucosyl residue 5 hydrogen bonds with Asp389, Lys382, and an oxygen atom of the β-phosphoryl group of UDP.

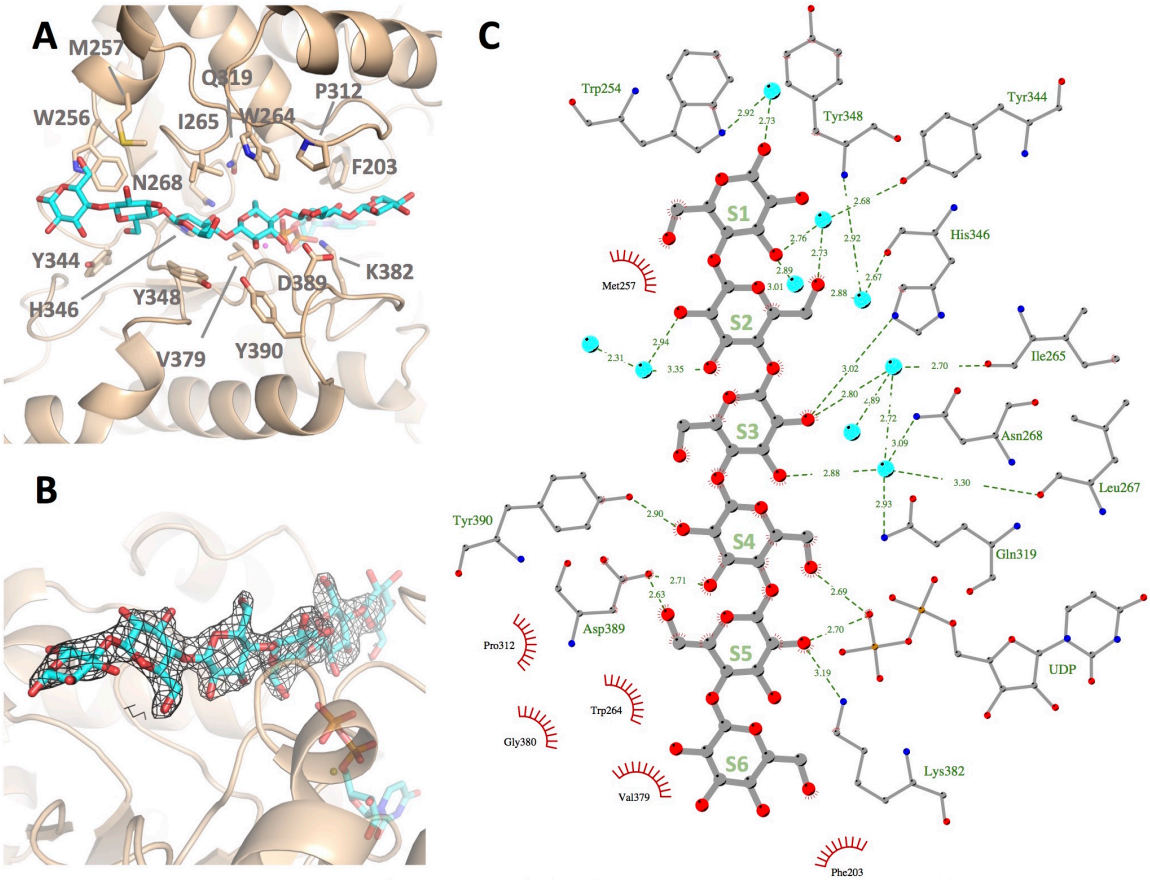


Fig. 4. Acceptor substrate binding in ternary complex. (*A*) Residues of XXT1 in proximity to cellohexaose. (*B*) Cellohexoase with electron density from an omit map contoured at 2σ. Colors as defined in Fig. 2. (*C*) Interaction map for cellohexaose from LigPlot+(50).

Xylosyl extensions from the 6-hydroxyl groups of glucosyl residues at subsites-1, -3, -5 or -6 should not introduce steric conflicts (Fig. S5). The 6-hydroxyl groups at subsites-3 and -5 are pointed away from the acceptor cleft and are solvent-exposed (Fig. S5*A*). The 6-hydroxyl groups of subsites-1 and -6 at opposite ends of the cleft and have space to accommodate a xylosyl adduct (Fig. S5*B* and *C*). Moreover, relatively low levels of electron density associated with glucose units at

subsites-1 and -6 suggests weaker binding at these subsites for an extended glucan chain (Fig. 4 and Fig. S3*A*). In contrast, a xylosyl extension at subsite-2 would overlap the side chains of His252, Ile351, and Tyr344, suggesting that xylose transfer to glucose unit N blocks subsequent binding necessary for xylose transfer to glucose unit N+2.

Mutations of active site residues. In 20-minute assays, mutant enzyme Ser228Ala generates UDP levels comparable to those of the wild-type enzyme. All other mutant enzymes produce UDP levels less than 25% of those of the wild-type enzyme (Fig. 5*A*). Assays were repeated for 20 hours, at which time the wild-type enzyme converts most of its UDP-xylose into UDP. Mutant enzymes Ser228Ala and Asn268Ala have UDP levels comparable to those of the wild-type enzyme. Mutations of residues that coordinate Mn^{2+} in the wild-type enzyme (single mutant enzymes Asp229Ala and His377Ala, and the double mutant enzyme Asp227Asn/Asp229Asn), produce 25-30% of the wild-type levels of UDP (Fig. 5). Mutant enzymes producing the lowest levels of UDP (less than 20% of wild-type) are Lys382Ala, Asp317Ala, Asp318Ala, and Gln319Ala. Lys382 hydrogen bonds with oxygen atoms of the α- and β-phosphoryl groups of UDP (extended conformer), whereas the remaining residues interact with the xylosyl group of modeled UDP-xylose. Evidently, mutations proximal to the xylosyl group of the donor substrate have the greatest impact on the rate of product formation.

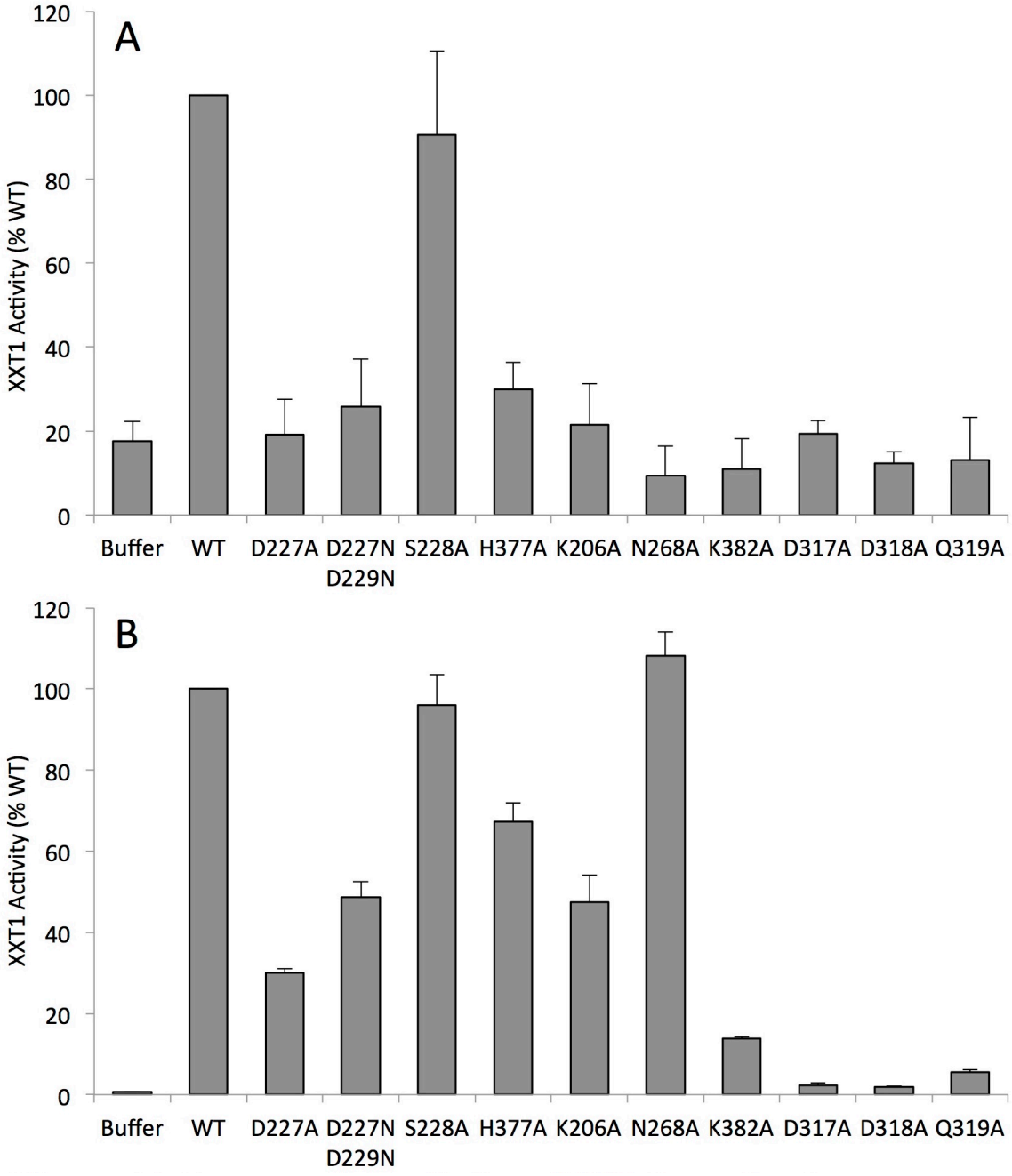


Fig. 5. Activity of XXT1 mutants and wild-type (WT) enzymes. (*A*) 20-min activity assays. (*B*) 20-hr activity assays. Each datum represents the average of three assays measuring the production of UDP from starting concentrations of 2 mM UDP-xylose, 1 mM cellohexaose and 2 mM $MnCl_2$, in 50 mM Tris pH 7.4 and 75 mM NaCl. Error bars drawn at one standard deviation.

Catalytic mechanism of XXT1. Inverting GTs utilize a single-displacement mechanism similar to that of inverting glycosyl hydrolases (12,17). The catalytic mechanism for retaining GTs, however, is unclear. Retaining glycosyl hydrolases utilize a double-displacement mechanism in which a catalytic base first inverts the anomeric stereochemistry of the sugar, forming a glycosyl-enzyme intermediate. A second nucleophilic attack at the anomeric carbon of the glycosyl-enzyme intermediate by the hydroxyl group of the acceptor restores the original anomer. For GTs however, a suitable catalytic base is not evident in structures (18,36,37) and efforts to trap a glycosyl-enzyme intermediate have been unsuccessful (38). Hence, some have suggested a SN_i-like mechanism for retaining GTs as an alternative (12,18). In the SN_i-like mechanism an oxocarbonium-ion intermediate forms; however, the acceptor approaches the same face of the oxocarbonium-ion intermediate as the leaving group. This mechanism has gained support from studies in quantum/molecular mechanics (39), kinetic isotope effects (20), and crystallography (snapshots along the reaction pathway) (19).

The XXT1 active site has no nucleophile positioned suitably for a double-displacement mechanism. Asp317 and Asp318 are 5.4 and 5.7 Å away, respectively from the anomeric carbon in the UDP-xylose model. Invariant conformations of Asp317 and Asp318 over all crystal structures and low B-parameters are not indicative of a region predisposed to conformational change, yet mutations at positions 317 and 318 have among the largest impacts on activity (Fig. 5). In the structure of XXYLT1 (19), a hydrogen bond between Gln330 and the 2-hydroxyl group of the xylosyl could in principle stabilize the positive charge of an

oxocarbonium ion during the SN_i-like reaction mechanism (19). Gln319 of XXT1 assumes a position analogous to Gln330 in XXYLT1 (Fig. S6), 4.7 Å from the anomeric carbon atom of modeled UDP-xylose, forming a hydrogen bond with the 3-hydroxyl group of the xylosyl residue. The Gln319Ala mutation has one of largest impacts on activity (Fig. 5).

Homology Modeling of XXT2 and XXT5. Efforts to purify and concentrate XXT2 and XXT5 to levels sufficient for crystallization have been unsuccessful; however, the structure of XXT1 leads to homology models of XXT2 and XXT5 that have no major departures from the backbone structure of XXT1 (Fig. S7*A*). XXT1, XXT2, and XXT5 have nearly identical positions for all residues in the active site with one notable difference: XXT1 puts an isoleucine residue at subsite-2 proximal to the 6-hydroxyl group of the second glucosyl from the reducing end of cellohexaose, whereas XXT5 has a glycine residue in the corresponding location. The glycine in XXT5 likely allows a xylosyl moiety at subsite-2, suggesting that XXT5 (in contrast to XXT1 and XXT2) can transfer a xylosyl moiety to glucose N when the N+2 glucosyl towards reducing end has been xylosylated (Fig. S7*B*-*E*). Additionally, the XXT1 surface contains a region rich in acidic residues (Glu357, Glu358, and Glu361) and one rich in basic residues (Lys419, Arg425, and Lys427; Fig. S8*A*-*B*). These residues are conserved in XXT1, XXT2, and XXT5 (Fig. S3). Matching the acidic region of one XXT1 dimer to the basic region of another XXT2 dimer results in a feasible dimer-to-dimer interface (Fig. S8*C*). A glucan chain would thread from one active site to that

of its neighboring dimer, suggesting the possibility of coordinated xylosyl transfers to the glucan chain.

Discussion

Of all plant cell wall GTs, only the structure of FUT1 is known (22,23). The crystal structure of XXT1 is the first instance of a GT family 34 member from plants and the second example of a GT involved in plant cell wall biosynthesis. XXT1 and FUT1 represent different folds, GT-A for XXT1 and GT-B for FUT1. In addition, XXT1 employs a large cleft in acceptor binding, whereas FUT1 does not (22,23).

XXT1 is catalytically active on cellopentaose and cellohexaose, but not on shorter glucan chains. Xylosylation is primarily on glucose residue four (31,32). The ternary complex (PDBID: 6BSW) directly supports xylosylation of cellohexaose at the fourth glucose residue from the reducing end (Fig 3). The addition of a second xylosyl moiety requires mono-xylosylated cellohexaose to disassociate from XXT1, rotate about the axis of the glucan by 180° and rebind, shifted by one glucosyl residue. Alternatively, cellohexaose can shift by two glucoses, placing glucose 2 at subsite 4. The two-subsite shift, however, would result in weak binding in which cellohexaose occupies subsites 3-6. Shifting (after rotation) by one glucosyl residue occupies subsites 1-5 or subsites 2-6. Binding of $_{6}$GGXGGG$_{1}$ at subsites 1-5 should generate the product $_{6}$GXXGGG$_{1}$, whereas binding at subsites 2-6 should generate the product $_{6}$GGXXGG$_{1}$ (Fig. S9). Cavalier et al. observed both products, with $_{6}$GGXXGG$_{1}$ preferred (32). Given that subsite-2 cannot accommodate a xylosyl

adduct, XXT1 cannot transfer a xylosyl moiety to $_6$GGXXGG$_1$ to produce $_6$GXXXGG$_1$, terminating further modification; however, XXT1 can transfer a xylosyl moiety to the third glucose of $_6$GXXGGG$_1$, forming $_6$GXXXGG$_1$ (Fig. S9). Hence, as long as $_6$GGXXGG$_1$ is the preferred product in the pool of doubly modified cellohexaoses, the addition of a third xylosyl moiety will never exceed 50%, as experimentally observed (33).

The steric limitation that prevents subsite 2 from accommodating a xylosyl adduct leads to the N+2 rule: XXT1 cannot transfer a xylosyl residue to the glucosyl residue N+2 when glucosyl residue N has a xylosyl adduct. This rule can be extended to XXT2 with some confidence. The homology model of XXT2 (and XXT5) and the crystal structures of XXT1 have nearly identical backbone positions (Fig. S7*A*), and the partial redundancy in function of XXT1 and XXT2 as demonstrated by in vivo (28,29) and in vitro (32) are consistent with similar substrate specificities, steric limitations, and modes of action. The residue that would be in steric conflict with a xylosylated glucose at subsite 2 is isoleucine in XXT1 and XXT2, but is glycine in XXT5 (Fig. S7). Hence, it is likely that the N+2 rule does not apply to XXT5 (Fig. 6).

A model for xylosylation of the glucan chain in vivo must satisfy the following assumptions and observations. First, cellulose synthase-like C4 (CSLC4) probably functions like cellulose synthase, adding β1-4 linked glucosyl residue onto the non-reducing end of the growing glucan chain (40,41). Hence, the reducing end of the glucan would emerge first from the synthase. Second, *xxt5* knock-out plants exhibit 2- to 4-fold increases in GXXG-type xyloglucan and a 3-fold decrease in XXXG-type xyloglucan (29). Third, the glucan chain binds in a specific orientation to XXT1, and

on the basis of homology models, the same orientation for XXT2 and XXT5 (this study). Fourth, XXT1, XXT2 and XXT5 have similar 6-subunit clefts (this study). Fifth, XXT1 and XXT2 adhere to the N+2 rule, whereas XXT5 does not (this study). Sixth, XXT1 interacts with XXT2, and XXT2 interacts with XXT5, but XXT1 does not interact strongly with XXT5 (42,43). Finally, the glucan is a helix of cellobiose repeats, with an angular displacement (rotation) of 51.4˚ per cellobiose unit (45). The second glucosyl residue of cellobiose has an additional angular offset of 180˚ relative to the first glucose. Figure 6 presents a plausible multi-enzyme system that accommodates the preceding observations/assumptions. Emerging glucosyl residues from the glucan synthase thread into XXT1 first, followed by XXT2, and then XXT5, the order determine by observed binding preferences of the XXTs (XXT1-XXT2 and XXT2-XXT5) and the recognition that XXT5 must come last in order to produce the XXXG repeat. The glucan chain in this model must advance by four glucosyl residues with each cycle. The active sites of the XXTs are "in phase" by virtue of the 51.4˚ helical repeat of the glucan in combination with the 180˚ offset of the second glucosyl residue of cellobiose (35). In the absence of XXT5, as in plant knock-outs, the product would have a GXXG repeat. The absence of steric limitations at subsite-2 of XXT5 allows xylosyl transfer to glucose N+2 when glucose N has a xylosyl adduct. Therefore, production of a glucan with a consistent XXXG repeat, as observed for the native xyloglucan in cell wall (11,29), requires the action of XXT1 and XXT2 to synthesize the GXXG repeat, followed by XXT5 to produce XXXG, all of which is organized in a specific manner as to prevent the xylosylation of all 6-

hydroxyl groups of the glucan chain (Fig. 6). This model accounts for all that we know or can be inferred in regard xylosylation of a glucan chain.

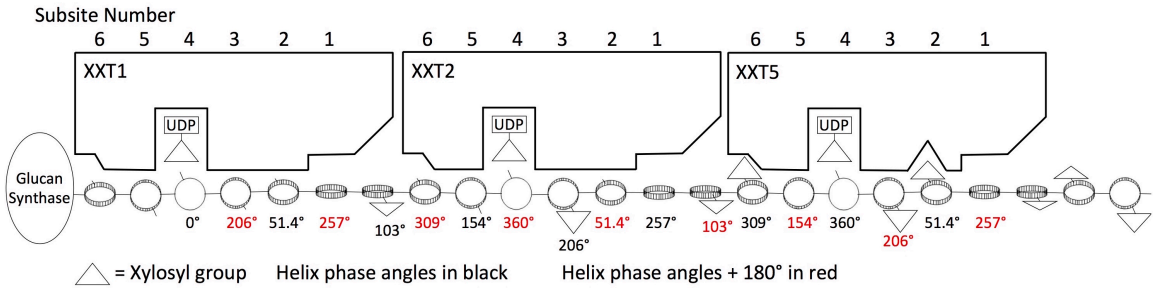


Fig. 6. Production of an XXXG-type xyloglucan. The model incorporates observed features of the xyloglucans elaborated by the wild-type organism (XXXG-repeats) and XXT5-knockout organism (GXXG-repeats), the emergence of the nonreducing end of the glucan from the synthase and the steric requirements imposed by the N+2 rule (for XXT1 and XXT2) and the lack of a N+2 rule (for XXT5). Angles in black represent the helical rotation of each cellobiose unit of cellulose, whereas angles in red combine a 180˚ offset angle with a half-cellobiose helical rotation (25.7˚). Circles and triangles represent glucose and xylose residues, respectively. The model, based on a 7-glucose separation of XXTs, enables a clear two-dimensional representation. Models based on other separations (11, and 15 glucose units for instance) are possible and may be necessary for allowable protein-protein contacts.

In summary, we demonstrate here that XXT1 and/or XXT2 are unlikely to produce the XXXG-type XyG in *Arabidopsis,* requiring the contribution of XXT5 to fully xylosylate the xyloglucan backbone. This limitation of XXT1 and XXT2 and the function of XXT5 were not apparent from reverse-genetics nor from in vitro studies. Steric constraints revealed by the structures, however, clarify the action of each XXT in xyloglucan biosynthesis. The model proposed here provides a satisfying explanation for all previous observations including the low activity of XXT5 on non-xylosylated cellohexaose (33), the reduction of XXXG-type xyloglucan in *xxt5* knock-out plants (29, 30), and why the *xxt1/xxt2* double knock-out has no XyG whereas the *xxt5* single knock-out has a significant reduction in XyG (29). Experiments to

determine the action of XXT5 on specifically xylosylated substrates and to understand the organization of the XyG biosynthetic complex will lead to a more certain understanding of XyG production in plants. The findings in this study will support strategies to create plant biomass with desired properties.

Methods

Protein expression and purification of XXT1. Arabidopsis xyloglucan xylosyltransferase I (residues 45-460) was cloned into the *EcoRI* and *BamHI* sites of the pGen2 vector containing a N terminal 8x histidine tag and a green fluorescent protein (GFP) tag. A Tobacco etch virus protease digestion site was between the XXT1 N terminus and GFP. Proteins were expressed during a transient transfection of HEK293F (Life technologies). HEK cells were grown in 90% FreeStyle293 and 10% Excell medium (Life Technologies) at 135 RPM with 8% CO_2 and 80% humidity at 37°C (35). Transfection was performed at a cell density of 3.0 x 10^6 cells per ml with 3.0 µg/ml DNA and 9.0 µg/ml PEI. Cells were diluted 1:1 after 24 hr of incubation with 90% FreeStyle293 and 10% Excell medium supplemented with 4.4 mM valproic acid. Protein was harvested after five days by centrifugation of the cells at 1000g for 10 min.

XXT1 proteins were purified using a Ni-NTA column and eluted with buffer containing 50mM Tris pH 7.4, 150mM NaCl, and 300mM imidazole. The eluent buffer was replaced by four rounds of volume reduction using an amicon centrifugal filter (Millipore) followed by additions of 50mM Tris pH 7.4, 150mM NaCl, and 40mM imidazole. Digestion of the GFP tag with the TEV protease was performed

overnight at 4˚C in a volume of 0.8 ml with an XXT1 concentration of 2 mg per ml and a TEV protease concentration of 0.1 mg per ml. The protein was purified with a Ni-NTA column to remove the TEV protease and His-GFP. Digested XXT1 protein was collected in the flow through. The flow through fraction was concentrated to less than 0.5 ml and injected into a Superdex 200 gel filtration column with buffer containing 50 mM Tris-HCl pH 7.4 and 150 mM NaCl. SDS-PAGE analysis of XXT1 confirms the expected size of roughly 48 kDa. Protein concentration was measured using a Bio-Rad kit (Quick Start Bradford Dye reagent 1X, Cat#500-0205) according to the manufacturer's instructions.

Crystallization, diffraction data collection, and processing. The XXT1 construct crystallized under two conditions by the hanging-drop, vapor-diffusion method, using a 1:1 ratio of XXT1 (10 mg/ml) and well solution. The first condition used a well solution of 0.1M Tris-HCl pH 8.0, 17% (w/v) PEG3350, 0.11M sodium nitrate, and 5% (w/v) trehalose. These crystals were used for soaking ligands and heavy atoms. The second condition (discovered more recently, and providing higher resolution data) employed a well solution of 0.1M Bis-Tris Propane pH 7.5, 160mM sodium formate, and 18% (w/v) PEG3350. Crystals appeared in 3-5 days. Crystals were harvested in a mixture of well solution supplemented with 15-20% (v/v) glycerol before flash-freezing in liquid nitrogen.

The heavy atom derivative was produced by adding 0.5 µl of K_2HgI_4 in well solution directly to the crystallization droplet to obtain an approximate final concentration of 0.4 mM K_2HgI_4. After 1.5 hr of incubation, crystals were transferred

to well solution supplemented with 15% (v/v) glycerol, but with no K_2HgI_4 and then frozen. Diffraction data were collected at the Advanced Photon Source. Initial phases were determined using Phenix Autosol (46). The initial model was built by Phenix Autobuild[46] and then refined against data to 2.4 Å resolution (not included in Table 1). This intermediate model was then used for the molecular replacement solutions (Phenix MR) of crystals used in ligand soaks and the 1.5-Å structure of the apo enzyme (PDB: 6BSU).

Binary (UDP and Mn^{2+}; PDB: 6BSV) and ternary (UDP, cellohexaose, and Mn^{2+}; PDB: 6BSW) complexes were formed by soaking crystals (spacegroup $P2_12_12_1$) in the well solution supplemented with ligands to concentrations of 10 mM $MnCl_2$ and 20 mM UDP (Sigma-Aldrich, >96%) for the binary complex and 10 mM $MnCl_2$, 8 mM UDP, and 2 mM cellohexaose (Megazyme) for ternary complex. The binary and ternary models were refined to resolutions of 2.1 Å and 2.4 Å, respectively. Manual adjustments to the model used COOT (47). Parameter refinement (x, y, z, B and occupancies) employed Phenix.refine (46). Distance restraints (2.1 Å) were applied to coordination bonds of Mn^{2+}. Evaluation of the stereochemistry of deposited models employed Molprobity (48) and all had scores less that 1.2. All structure related figures were made using PyMol (49).

Modeling of UDP-xylose. Co-crystallization of XXT1 in the presence of UDP-xylose resulted in no electron density in the active site of XXT1, necessitating the modeling of UDP-xylose in the active site ternary crystal structure. UDP-xylose from mouse xyloside xylosyltransferase 1 (PDB: 4WNH) was manually aligned with the

extended conformer of UDP in ternary model using atoms of the UDP moieties. Glycerol and water molecules were removed as necessary to avoid steric conflicts. Energy minimization was performed with YASARA by simulated annealing, employing the YASARA2 force field (44). First, all atoms were fixed except those of the xylosyl moiety, which were free for energy minimization. Second, the entire UDP-xylose molecule was free for energy minimization while keeping all other atoms fixed. Third, the XXT1 active site residues, UDP-xylose, and cellohexaose were free for energy minimization. The last step allowed the entire XXT1 protein along with all substrates to be free in energy minimization.

Activity assays. All activity assays were performed with 2 mM UDP-xylose, 1 mM cellohexaose, and 2 mM $MnCl_2$ in 50 mM Tris pH 7.4 and 75 mM NaCl. The 20 min reactions were in 10 µl volume with 200 ng of XXT1 enzyme, and the 20 hr reactions were in 15 µl volume with 1 µg of XXT1 enzyme. Product (UDP) formation employed the UDP-GLO glycosyltransferase Assay (Promega). The reaction was stopped by the addition of UDP-GLO detection reagent. The product solutons were then appropriately diluted to be within the range of standards.

Homology Models. The YASARA molecular modeling program (YASARA "structure") (44) was used to build the homology models as described previously (34), using standard protocols (http://yasara.org/homologymodeling.htm), but with the only the XXT1 ternary crystal structure as a template for both XXT2 and XXT5 homology models.

Accession codes. Protein Data Bank: The coordinates and structure factors of XXT1 apoform, the binary structure of XXT1 complexed with UDP, and the ternary form of XXT1 complexed with UDP and cellohexaose have been deposited at the PDB with accession codes 6BSU, 6BSV, and 6BSW, respectively

Acknowledgment

We would like to thank Dr. Adam Barb and Dr, Ganesh Subedi (Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University) for sharing the protocols and resources for transformation and growing HEK293 cells. This study was supported by NSF-MCB grant # 1121163, 2011-2017 (to OAZ), Roy J Carver trust, and Argonne National Laboratory Advanced Photon Source grant GUP-48455.

Author Contributions

A.T.C., O.A.Z., and R.B.H. designed the research. A.T.C., JY.C, and R.B.H. performed experimental phasing. A.T.C. performed protein expression, purification, and crystallization. A.T.C. and J.J.E. produced constructs of mutant variants of XXT1. A.T.C., O.A.Z., and R.B.H. wrote the manuscript.

Supplementary Information

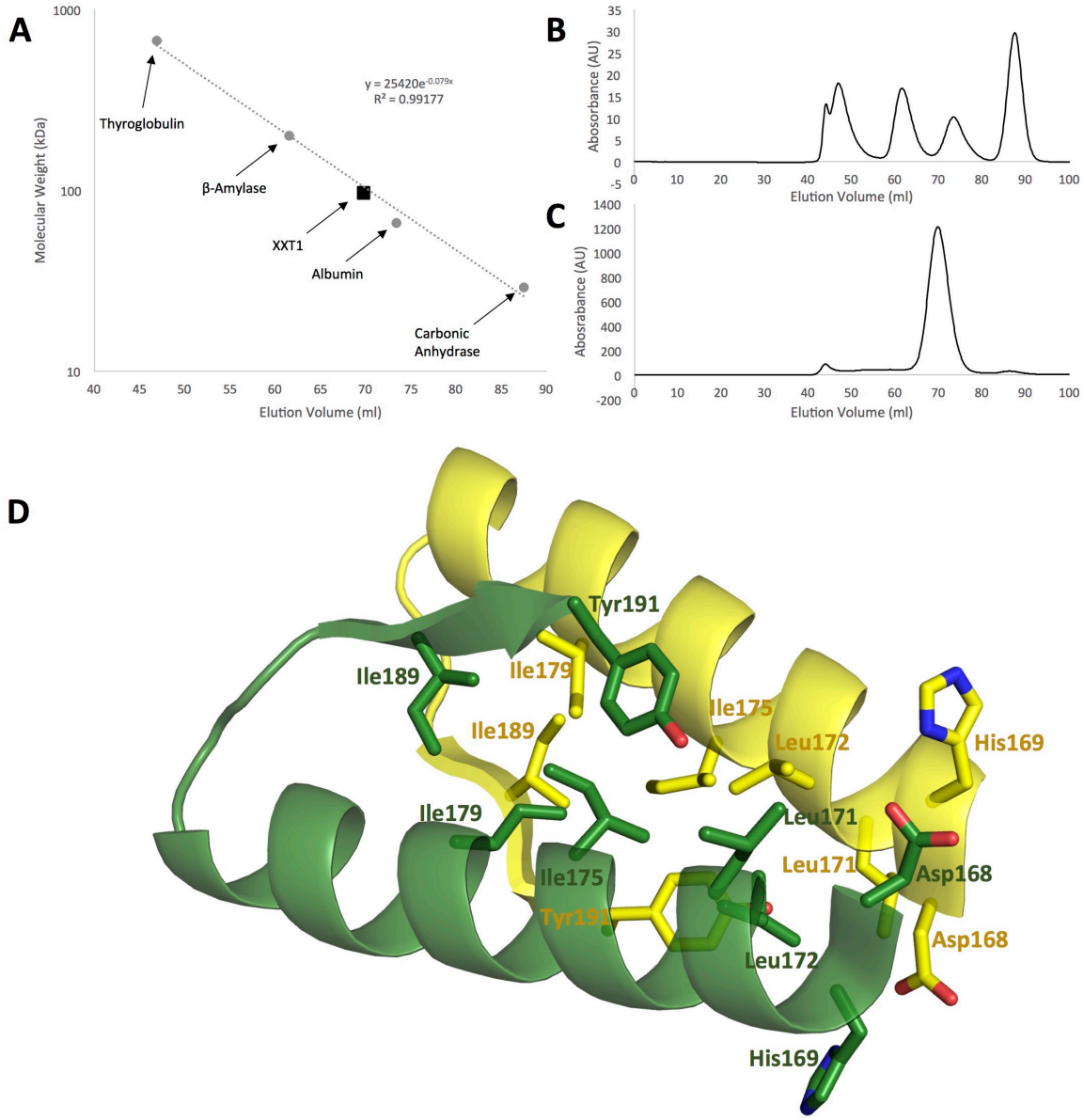


Fig. S1. Gel filtration and dimer interface of XXT1. (*A*) Calibration curve for the S200 column. Protein with known mass are shown with grey circles and XXT1 is shown with black square. Gel filtration chromatogram of (*A*) protein with known mass and (*B*) XXT1. The expected size of XXT1 monomer and dimer are 48.4 and 96.7 kDa, respectively. Based on standard curve, XXT1 molecular size was calculated to be 102.7 kDa. D. Dimer interface between α-helix α2 and β-strand β2. All other residues were removed for clarity. Chain A is green and chain B is yellow.

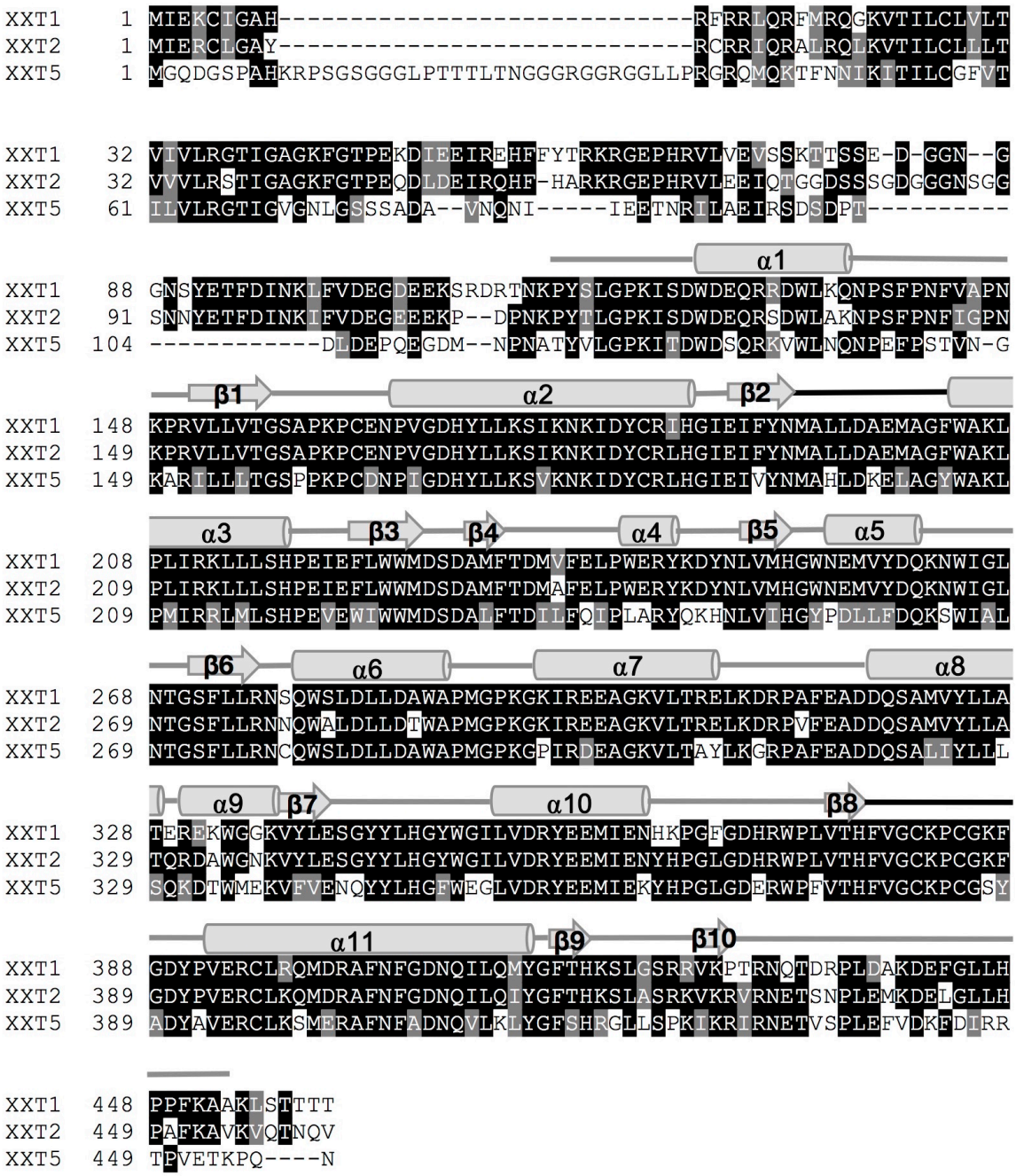


Fig. S2. Sequence alignment of XXT1, XXT2, and XXT5. Secondary structure of XXT1 is shown above sequences.

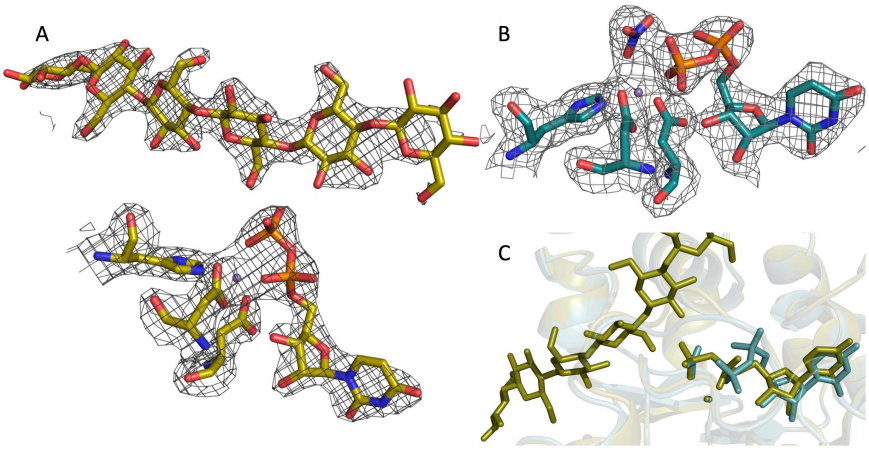


Fig. S3. Comparison between the binary (bent) and ternary (extended) UDP conformation. (*A*) Extended conformation of UDP in the ternary complex. (*B*) Bent conformation of UDP in binary structure. (*C*) Overlay comparing the bent and extended conformation. Bent conformation is in cyan and extended conformation is in tan. Electron density is contoured to 2.0 σ.

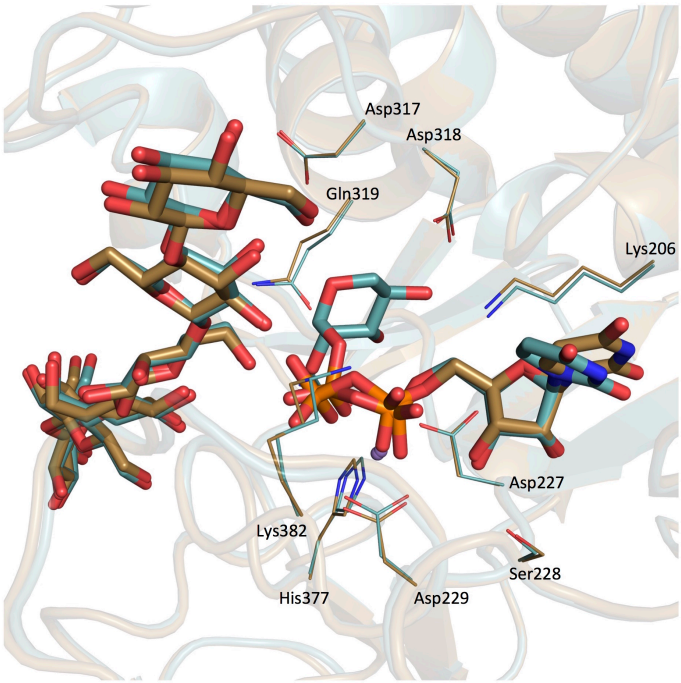


Fig. S4. Comparison UDP in ternary complex and modeled UDP-xylose. Tan represents the ternary crystal structure and cyan represents the modeled UDP-xylose. Modeling UDP xylose was performed as described in methods section. Conformations of the cellohexoase, UDP, and active site residues have no major pertubations between ternary structure and modeled UDP-xylose structure. Gln319 is localized in similar position compared to Gln330 in XXYLT1 suggesting it might also stabilize the oxocarbenium ion during the SN_i-like reaction mechanism.

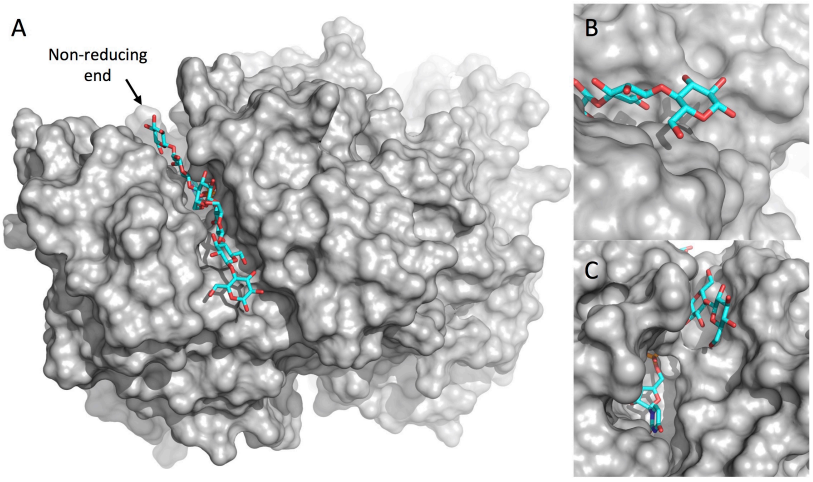


Fig. S5. Surface of XXT1 ternary complex showing acceptor substrate binding active site cleft. (*A*) View of cellohexaose entire active site cleft. Non-reducing end of the glucan chain is shown with arrow. Both terminal glucoses of cellohexaose partially hand out from active site cleft (*B*) Zoomed view of cellohexaose active site cleft near the reducing end. (*C*) Zoomed view of cellohexaose active site cleft near the non-reducing end.

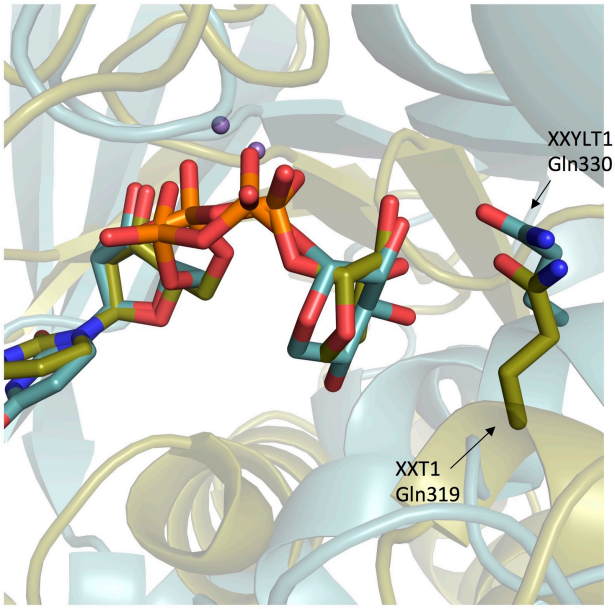


Fig. S6. Comparison of XXT1 and XXYLT1 Glutamine near anomeric carbon. This Gln330 in XXYLT1 was proposed in to be involved in stabilization of oxocarbenium ion intermediate in the SN_i-like reaction mechanism (19; PDB: 4WNH). Alignment of XXT1 and XXYLT1 was done by manually aligning the UDP-xylose substrates.

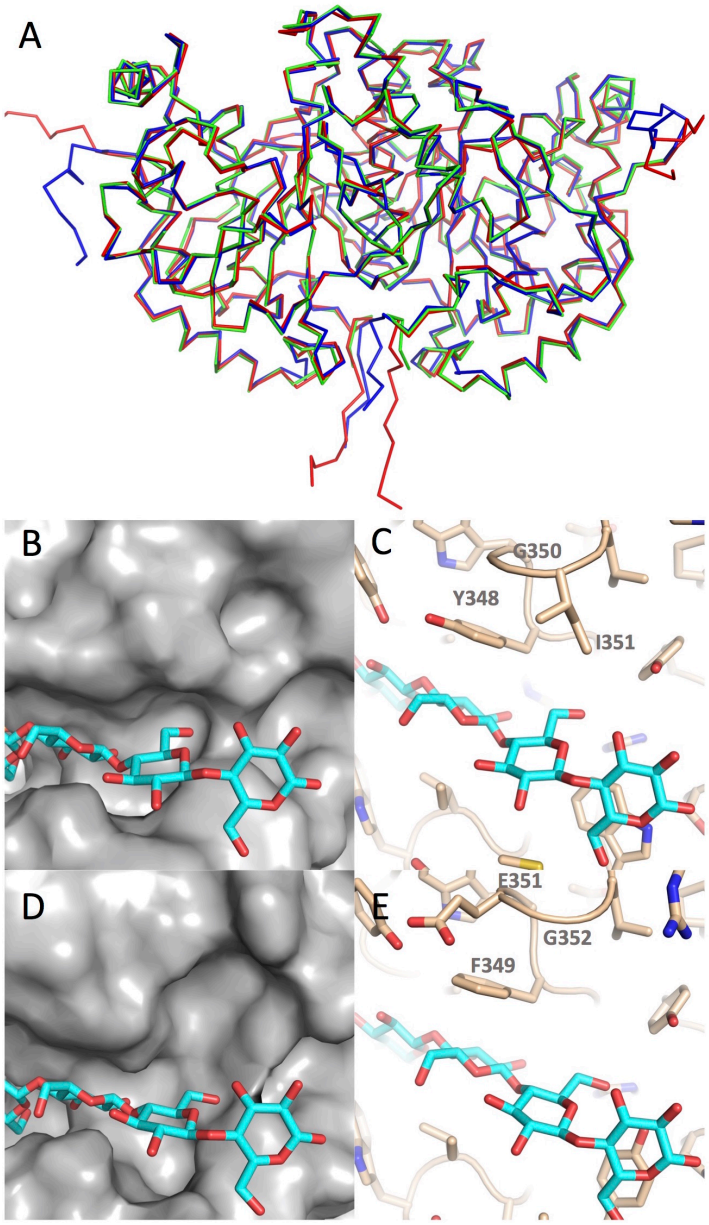


Fig. S7. Comparison of XXT1 crystal structure with XXT2 and XXT5 homology model. (*A*) Backbone alignment of XXT1 crystal structure with XXT2 and XXT5 homology model. XXT1 crystal structure is shown in green, XXT2 homology model in dark red, and XXT5 homology model in blue. Ten residues from the N- and C-terminus of XXT2 and XXT5 was built and resulted in extensions from the globular protein. (*B-E*) Comparison of the cellohexaose binding site near the second glucose from the reducing end bound to subsite 2. (*A*) XXT1 binding site surface. (*B*) XXT1 binding site residues at subsite 2 (*C*) XXT5 homology model binding site surface. (**d**) XXT5 homology model binding site residues at subsite 2. I351 in XXT1 aligns with G352 in the sequence alignment (Fig. S2).

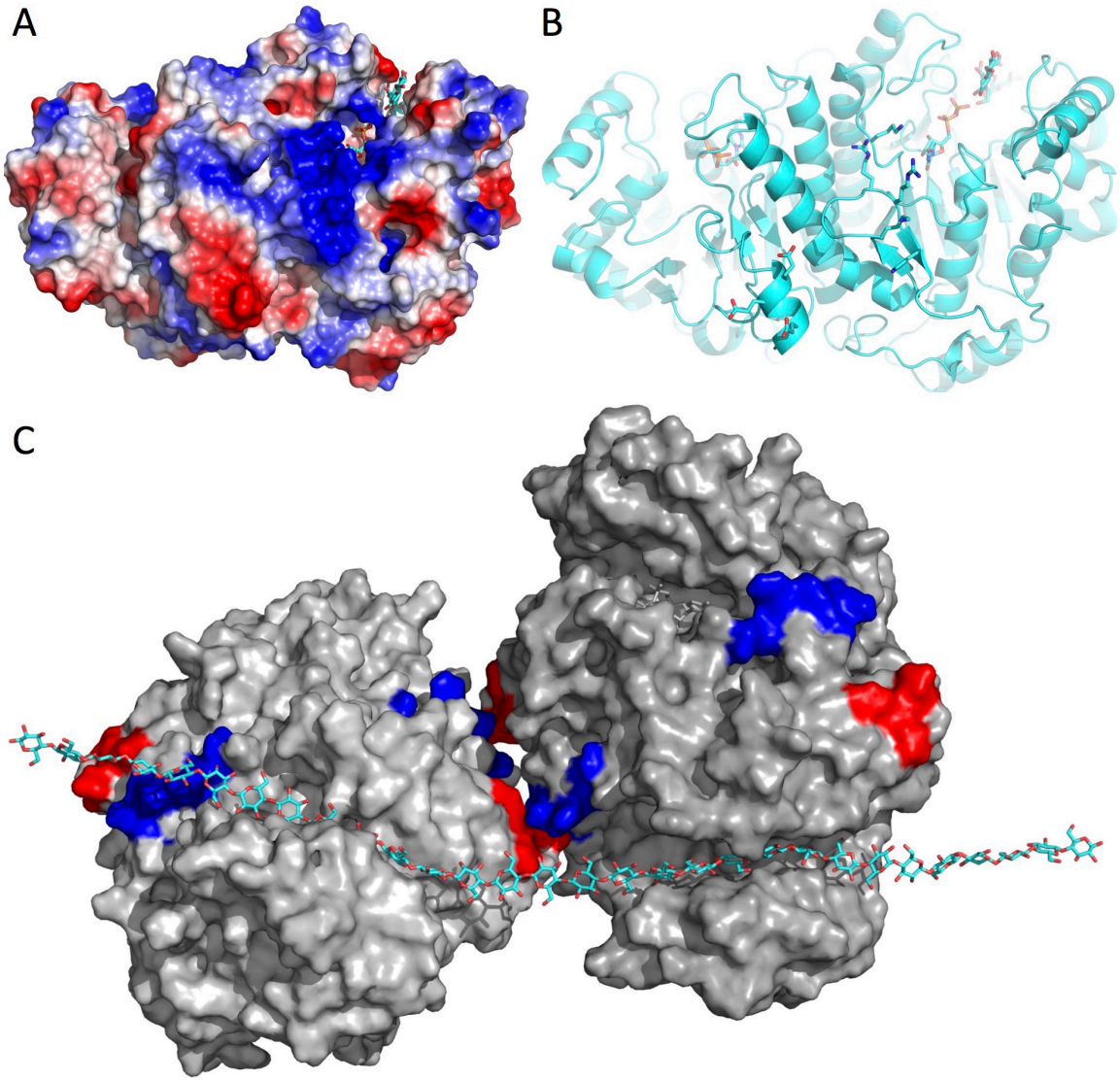


Fig. S8. XXT1 surface showing regions of concentrated charged residues. (*A*) Surface representation showing electrostatic potential of the XXT1 dimer surface. (*B*) Cartoon figure of XXT1 with same orientation as shown in *A* showing the acidic and basic residues in the region that is concentrated with charged amino acids. (*C*) Representation of manually aligned dimer-dimer of XXT1. Dimer-dimer was aligned so that the basic residues of one XXT1 dimer were interacting with the acidic residue of the same face of a identical XXT1 dimer. Glucan chain was manually directed from the original cellohexaose molecule in the ternary structure into the active site cleft of the other XXT1 dimer.

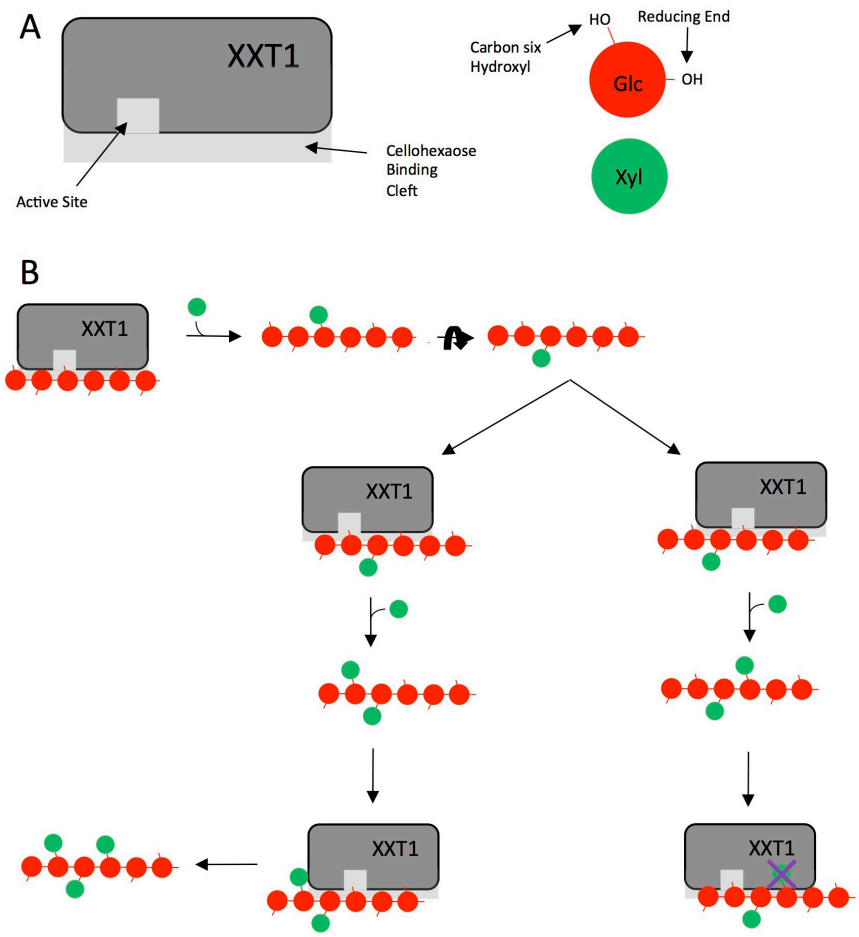


Fig. S9. Model of XXT1 Xylosylation pattern. XXT1 is predicted to be unable to xylosylate bi-xylosylated cellohexaose when it's in the form GGXXGG due to steric conflicts at subsite two of the XXT1 active site cleft. (*A*) Key showing regions of XXT1 and sugar schematics. (*B*) Xylosylation of cellohexaose. Celloehexaose is shown to be mainly xylosylated at the fourth glucose from the reducing end of cellohexaose (32). Following this initial xylosylation, cellohexaose is released from XXT1 and freely rotates in solution. To rebind XXT1, the mono-xylosylated cellohexaose would be required to rotate 180° along the axis of the glucan chain, and rebind XXT1. Rebinding of XXT1 to position the 6-hydroxyl into the XXT1 active site requires a one glucose shift towards either the reducing end to produce GXXGGG, or one glucose towards the non-reducing end to produce GGXXGG. If GGXXGG is made, XXT1 is unable to use this as a substrate to produce GXXXGG because of the steric limitation at the N+2 rule. Conversely, XXT1 would be able to utilize GXXGGG to produce GXXXGG.

Table S1. Data collection and structure refinement statistics of XXT1 structures.

Complex	Apo	Binary	Ternary
Ligand	Mn2+	Mn2+, UDP	Mn2+, UDP, Cellohexaose
PDB	6BSU	6BSV	6BSW
Data Collection			
Space group	C 2 2 21	P 21 21 21	P 21 21 21
Cell dimensions			
a, b, c (Å)	94.30, 135.68, 113.23	64.93, 94.55, 144.72	65.04, 94.89, 145.10
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	43 - 1.50 (1.55 - 1.50)	47 - 2.43 (2.52 - 2.43)	48 - 2.16 (2.23 - 2.16)
Rmerge (%)	7.9 (92.4)	12.7 (48.9)	11.4 (52.9)
I/σ(I)	14.31 (1.03)	9.80 (2.02)	10.39 (2.20)
Completeness (%)	96.36 (83.82)	93.75 (62.44)	92.14 (65.86)
Redundancy	7.8 (3.7)	6.1 (4.5)	5.5 (4.6)
Refinement			
Resolution (Å)	43.47 - 1.497	47.27 - 2.433	48.43 - 2.157
No. reflections	112693 (9911)	32045 (2083)	45315 (3194)
Rwork/Rfree (%)	14.2 (26.2) / 16.7 (27.9)	18.4 (26.0) / 21.6 (27.1)	18.2 (24.6) / 22.9 (30.6)
No. atoms			
Protein	5611	5566	5566
Ligands	2	76	145
Solvent	672	101	209
R.m.s deviations			
Bond lengths (Å)	0.006	0.01	0.009
Bond angles (°)	0.82	0.64	0.57
B-Parameter			
Protein	29.47	45.54	46.72
Ligands	20.43	70.35	57.82
Solvent	40.57	41.71	45.31

References

1. Pauly M & Keegstra K (2010) Plant cell wall polymers as precursors for biofuels. *Curr Opin Plant Biol* 13, 305-12.
2. Pauly M & Keegstra K (2008) Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J* 54, 559-68.
3. Gray KA, Zhao L & Emptage M (2006) Bioethanol. *Curr Opin Chem Biol* 10, 141-6.
4. Hayashi T & Kaida R (2011) Functions of xyloglucan in plant cells. *Mol Plant* 4, 17-24.
5. Carpita NC & Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J* 3, 1-30.
6. Somerville C et al (2004) Toward a systems approach to understanding plant cell walls. *Science* 306, 2206-11.
7. Thompson DS (2005) How do cell walls regulate plant growth? *Journal of Experimental Botany* 56, 2275-2285.

8. Pauly M & Keegstra K (2016) Biosynthesis of the Plant Cell Wall Matrix Polysaccharide Xyloglucan. *Annu Rev Plant Biol* 67, 235-259.
9. Fry SC, et al. (1993) An unambiguous nomenclature for xyloglucan-derived oligosaccharides. *Physiologia Plantarum* 89, 1-3.
10. Pena MJ, Darvill AG, Eberhard S, York WS & O'Neill MA (2008) Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants. *Glycobiology* 18, 891-904.
11. Hsieh YS & Harris PJ (2009) Xyloglucans of monocotyledons have diverse structures. *Mol Plant* 2, 943-65.
12. Lairson LL, Henrissat B, Davies GJ & Withers SG (2008) Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* 77, 521-55.
13. Campbell J, Davies G, Bulone V & Henrissat B (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J* 326, 929-939.
14. Coutinho PM, Deleury E, Davies GJ & Henrissat B (2003) An Evolving Hierarchical Family Classification for Glycosyltransferases. *Journal of Molecular Biology* 328, 307-317.
15. Busch C, et al. (1998) A Common Motif of Eukaryotic Glycosyltransferases Is Essential for the Enzyme Activity of Large Clostridial Cytotoxins. *Journal of Biological Chemistry* 273, 19566-19572.
16. Wiggins CA & Munro S (1998) Activity of the yeast MNN1 α-1, 3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proceedings of the National Academy of Sciences* 95, 7945-7950.
17. Davies G & Henrissat B (1995) Structures and Mechanisms of Glycosyl Hydrolases. *Structure* 3, 853-859.
18. Persson K, et al. (2001) Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs. *Nat Struct Biol* 8, 166-75.
19. Yu H, et al. (2015) Notch-modifying xylosyltransferase structures support an SNi-like retaining mechanism. *Nat Chem Biol* 11, 847-854.
20. Lee SS, et al. (2011) Mechanistic evidence for a front-side, SNi-type reaction in a retaining glycosyltransferase. *Nat Chem Biol* 7, 631-638.
21. Gloster TM (2014) Advances in understanding glycosyltransferases from a structural perspective. *Current Opinion in Structural Biology* 28, 131-141.
22. Rocha J, et al. (2016) Structure of Arabidopsis thaliana FUT1 Reveals a Variant of the GT-B Class Fold and Provides Insight into Xyloglucan Fucosylation. *Plant Cell* 28, 2352-2364.
23. Urbanowicz BR, et al. (2017) Structural, mutagenic and in silico studies of xyloglucan fucosylation in Arabidopsis thaliana suggest a water-mediated mechanism. *Plant J* 91, 931-949.
24. Zabotina OA (2012) Xyloglucan and its biosynthesis. *Front Plant Sci* 3, 134.
25. Culbertson AT, Smith AL, Cook MD & Zabotina OA (2016) Truncations of xyloglucan xylosyltransferase 2 provide insights into the roles of the N- and C-terminus. *Phytochemistry* 128, 12-9.

26. Oikawa A, Lund CH, Sakuragi Y & Scheller HV (2013) Golgi-localized enzyme complexes for plant cell wall biosynthesis. *Trends Plant Sci* 18, 49-58.
27. Sogaard C, et al. (2012) GO-PROMTO illuminates protein membrane topologies of glycan biosynthetic enzymes in the Golgi apparatus of living tissues. *PLoS One* 7, e31324.
28. Cavalier DM, et al. (2008) Disrupting two Arabidopsis thaliana xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component. *Plant Cell* 20, 1519-37.
29. Zabotina OA, et al. (2012) Mutations in multiple XXT genes of Arabidopsis reveal the complexity of xyloglucan biosynthesis. *Plant Physiol* 159, 1367-84.
30. Zabotina OA, et al. (2008) Arabidopsis XXT5 gene encodes a putative alpha-1,6-xylosyltransferase that is involved in xyloglucan biosynthesis. *Plant J* 56, 101-15.
31. Faik A, Price NJ, Raikhel NV & Keegstra K (2002) An Arabidopsis gene encoding an alpha-xylosyltransferase involved in xyloglucan biosynthesis. *Proc Natl Acad Sci U S A* 99, 7797-802.
32. Cavalier DM & Keegstra K (2006) Two xyloglucan xylosyltransferases catalyze the addition of multiple xylosyl residues to cellohexaose. *J Biol Chem* 281, 34197-207.
33. Culbertson A, et al. (2016) Enzymatic Activity of Arabidopsis Xyloglucan Xylosyltransferase 5. *Plant Physiol* 171, 1893-904.
34. Culbertson AT, et al. (2016) A Homology Model of Xyloglucan Xylosyltransferase 2 Reveals Critical Amino Acids Involved in Substrate Binding. *Glycobiology* 26, 961-972.
35. Subedi GP, Johnson RW, Moniz HA, Moremen KW & Barb A (2015) High yield expression of recombinant human proteins with the transient transfection of HEK293 cells in suspension. *Journal of visualized experiments* 106, e53568.
36. Gastinel LN, et al. (2001) Bovine alpha1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases. *Embo j* 20, 638-49.
37. Patenaude SI, et al. (2002) The structural basis for specificity in human ABO(H) blood group biosynthesis. *Nat Struct Biol* 9, 685-90.
38. Ly HD, Lougheed B, Wakarchuk WW & Withers SG (2002) Mechanistic studies of a retaining alpha-galactosyltransferase from Neisseria meningitidis. *Biochemistry* 41, 5075-85.
39. Gomez H, Polyak I, Thiel W, Lluch JM & Masgrau L (2012) Retaining glycosyltransferase mechanism studied by QM/MM methods: lipopolysaccharyl-alpha-1,4-galactosyltransferase C transfers alpha-galactose via an oxocarbenium ion-like transition state. *J Am Chem Soc* 134, 4743-52.
40. Slabaugh E, Davis JK, Haigler CH, Yingling YG & Zimmer J (2014) Cellulose synthases: new insights from crystallography and modeling. *Trends Plant Sci* 19, 99-106.
41. Morgan JL, Strumillo J & Zimmer J (2013) Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* 493, 181-6.

42. Chou YH, Pogorelko G & Zabotina OA(2012) Xyloglucan xylosyltransferases XXT1, XXT2, and XXT5 and the glucan synthase CSLC4 form Golgi-localized multiprotein complexes. *Plant Physiol* 159, 1355-66.
43. Chou Y, Pogorelko G, Young ZT & Zabotina OA (2015) Protein-Protein Interactions Among Xyloglucan-Synthesizing Enzymes and Formation of Golgi-Localized Multiprotein Complexes. *Plant Cell Physiol* 56, 255-67.
44. Krieger E, et al. (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* 77 Suppl 9, 114-22.
45. Viswanathan A & Shenouda SG (1971) The helical structure of cellulose I. *Journal of Applied Polymer Sci* 15, 519-535.
46. Adams PD, *et al*. (2010) PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213-221.
47. Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 486-501.
48. Chen VB, *et al.* (2010) MoleProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 12-21.
49. Schrodinger LLC (2010) The PyMol Molecular Graphics System, Version 1.7.4.5.
50. Wallace AC, Laskowski RA & Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams for protein-ligand interactions. *Protein Eng* 8, 127-34.

CHAPTER 7. CONCLUSIONS AND FUTURE DIRECTIONS

The plant cell wall is a complex and dynamic assembly of numerous biopolymers including polysaccharides, lignin, and glycoproteins (Keegstra, 2010; Hofte and Voxeur, 2017). Our understanding of the chemical structures of the polymers has only begun, and new techniques and technologies will surely reveal important aspects of this abundant resource. Despite the progress of understanding the composition of the plant cell wall polysaccharides, our understanding of the modes of biosynthesis of these polysaccharides is still very limited. This is mainly due to the lack of suitable substrates for *in vitro* analysis of the glycosyltransferases involved in this biosynthesis. Additionally, there are currently only two known structures of these glycosyltransferases, FUT1 (Rocha et al., 2016; Urbanowicz et al., 2017) and XXT1 (Chapter 6), both of which are involved in xyloglucan (XyG) biosynthesis. This accounts for only two structures out of the over 400 predicted cell wall biosynthetic GTs.

How can something such as the plant cell wall, which already has such a major influence on humanity, in addition to its potential use in biofuels and biomaterials, be so poorly understood? This is likely due to low economic pressure for its understanding due to the current low cost of fossil fuels (oil, natural gas, and coal) and the fact that the current yield of our agricultural system is sufficient for feeding most of the population. Surely, with reduction of available fossil fuels and the growing population, there will be increased pressure for engineering plants for the production of biomaterials, to increased food yields, or to increase abiotic or

biotic stress resistance. An understanding of the biosynthesis of the plant cell wall is essential for the above-mentioned engineering strategies.

One of the major bottlenecks in investigation of plant GTs is their poor recombinant protein expression *in vitro*. Thus, optimization of the expression conditions for the XXTs was performed. We optimized the expression conditions in XXT2 including various *E. coli* strain, N-terminal tags, and storage conditions (Chapter 3). We also optimized the assay conditions including pH, NaCl concentration, buffer, and presence of imidazole (Chapter 3). Using these conditions, various truncations from the N- and C-terminal were investigated (Chapter 4). This demonstrated that increasing truncation size from N-terminal decreased protein expression and activity, while increasing truncation size from C-terminal increased protein expression and decreased XXT2 activity (Chapter 4).

Using the optimized procedures described in Chapters 3 and 4, we expressed all three XXTs in *E. coli*. Using these recombinantly expressed proteins, along with concentrating to very high concentration, we show for the first time that XXT5 is catalytically active *in vitro* (Chapter 3). For many years, there was a debate about the function of XXT5 due to its clearly distinct function *in vivo* (Zabotina et al., 2008; Zabotina et al., 2012), but catalytic activity *in vitro* was not shown (Faik et al., 2002; Zabotina et al., 2008; Vuttipongchaikij et al., 2012). Thus, activity of XXT5 *in vitro* demonstrates that XXT5 is catalytically active and likely does not play a non-catalytic function (Chapter 3).

At the beginning of this work, there was no structural information for any plant cell wall GTs. Thus, we created a homology model of XXT2 using a

glucosyltransferase in the same GT family, GT-34 (Zhang et al., 2007), which was used to predict active site amino acid reisdues. Mutagenesis studies demonstrated that the DXD motif, including D228 and D230, is critical for catalysis, likely through coordination of the divalent cation. Additionally, K207 and F204 were also shown to be import for activity (Chapter 5).

We solved the crystal structure of XXT1 using K_2HgI_4 for experimental phasing (Chapter 6). The crystal structure of XXT1, a dimer both in crystal and solution, revealed a large active site cleft that accommodates the glucan chain (Chapter 6). Based on mutagenesis and previous studies of retaining GT-A fold GTs, it is likely that XXT1 utilizes a S_Ni-like catalytic mechanism. Interestingly, based on steric restraints, it is apparent that XXT1 would not be able to accommodate a glucan chain when the glucose two units towards the reducing end has previously be xylosylated (N+2 rule), thus disallowing XXT1 (or XXT2) to produce the biologically observed XXXG repeat. Analysis of the homology model of XXT5 showed that the site located at N+2 has a mutation of an Ile in XXT1/XXT2 to a Gly in XXT5. This would effectively open up this site which would likely allow XXT5 to accommodate a previously xylosylated glucose at the N+2 site. Thus, we propose that XXT1 and XXT2 xylosylate the glucan backbone to produce the GXXG repeat, which is then utilized by XXT5 to produce the biologically observed XXXG repeat.

This crystal structure also raises numerous questions. First, XyG biosynthetic enzymes have been shown to form protein-protein interactions with other XyG biosynthetic enzymes, thus it is likely that these enzymes form a large, multi-protein, XyG biosynthetic complex (Chou et al., 2012; Chou et al., 2015). For

example, XXT2 and XXT5 were shown to form a very strong interaction, whereas the interaction of XXT1 with other XyG proteins was weak (Chou et al., 2015). Thus, to identify residues involved in this protein-protein interaction, the surfaces of XXT1, XXT2, and XXT5 could be scanned to locate differences between XXT1 and XXT2/XXT5. These residues could then be mutated and binding affinity measured to confirm that these residues are at the protein-protein interaction interface. It is also interesting that the expression levels of XXT2 and XXT5 are significantly lower than XXT1, suggesting that these proteins may require each other to form soluble protein. Possibly, co-expression of XXT2 and XXT5 may result in higher net yield compared to when they are expressed individually.

Though the structure of XXT1 revealed numerous aspects of its activity and specificity, the ultimate goal is to obtain structural information of the XyG biosynthetic complex. One reasonable objective is to express these proteins (CSLC4, XXT1, XXT2, XXT5, XLT2, MUR3, FUT1) and place this biosynthetic complex in a nanodisc. With the recent advances in cryo-electron microscopy, this large complex could then be visualized at a near-atomic level. The structure of the XyG biosynthetic complex is the ultimate goal and would truly be a breakthrough in understanding of not only xyloglucan biosynthesis, but also other plant cell wall polysaccharide biosynthesis.

References

Chou Y, Pogorelko G, Young ZT, Zabotina OA (2015) Protein-Protein Interactions Among Xyloglucan-Synthesizing Enzymes and Formation of Golgi-Localized Multiprotein Complexes. Plant Cell Physiol

Chou YH, Pogorelko G, Zabotina OA (2012) Xyloglucan xylosyltransferases XXT1, XXT2, and XXT5 and the glucan synthase CSLC4 form Golgi-localized multiprotein complexes. Plant Physiol **159:** 1355-1366

Faik A, Price NJ, Raikhel NV, Keegstra K (2002) An Arabidopsis gene encoding an alpha-xylosyltransferase involved in xyloglucan biosynthesis. Proc Natl Acad Sci U S A **99:** 7797-7802

Hofte H, Voxeur A (2017) Plant cell walls. Curr Biol **27:** R865-R870

Keegstra K (2010) Plant Cell Walls. Plant Physiology **154:** 483-486

Rocha J, Ciceron F, de Sanctis D, Lelimousin M, Chazalet V, Lerouxel O, Breton C (2016) Structure of Arabidopsis thaliana FUT1 Reveals a Variant of the GT-B Class Fold and Provides Insight into Xyloglucan Fucosylation. Plant Cell **28:** 2352-2364

Urbanowicz BR, Bharadwaj VS, Alahuhta M, Pena MJ, Lunin VV, Bomble YJ, Wang S, Yang JY, Tuomivaara ST, Himmel ME, Moremen KW, York WS, Crowley MF (2017) Structural, mutagenic and in silico studies of xyloglucan fucosylation in Arabidopsis thaliana suggest a water-mediated mechanism. Plant J **91:** 931-949

Vuttipongchaikij S, Brocklehurst D, Steele-King C, Ashford DA, Gomez LD, McQueen-Mason SJ (2012) Arabidopsis GT34 family contains five xyloglucan alpha-1,6-xylosyltransferases. New Phytol **195:** 585-595

Zabotina OA, Avci U, Cavalier D, Pattathil S, Chou YH, Eberhard S, Danhof L, Keegstra K, Hahn MG (2012) Mutations in multiple XXT genes of Arabidopsis reveal the complexity of xyloglucan biosynthesis. Plant Physiol **159:** 1367-1384

Zabotina OA, van de Ven WT, Freshour G, Drakakaki G, Cavalier D, Mouille G, Hahn MG, Keegstra K, Raikhel NV (2008) Arabidopsis XXT5 gene encodes a putative alpha-1,6-xylosyltransferase that is involved in xyloglucan biosynthesis. Plant J **56:** 101-115

Zhang Y, Xiang Y, Van Etten JL, Rossmann MG (2007) Structure and function of a chlorella virus-encoded glycosyltransferase. Structure **15:** 1031-1039