Plant Protein *O*-Arabinosylation

Bent Larsen Petersen[1]*[†], Cora A. MacAlister[2†] and Peter Ulvskov[1]*

[1]*Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Copenhagen, Denmark,*
[2]*Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI, United States*

A wide range of proteins with diverse functions in development, defense, and stress responses are *O*-arabinosylated at hydroxyprolines (Hyps) within distinct amino acid motifs of continuous stretches of Hyps, as found in the structural cell wall extensins, or at non-continuous Hyps as, for example, found in small peptide hormones and a variety of plasma membrane proteins involved in signaling. Plant *O*-glycosylation relies on hydroxylation of Prolines to Hyps in the protein backbone, mediated by prolyl-4-hydroxylase (P4H) which is followed by *O*-glycosylation of the Hyp C_4-OH group by either galactosyltransferases (GalTs) or arabinofuranosyltranferases (Ara*f*Ts) yielding either Hyp-galactosylation or Hyp-arabinosylation. A subset of the P4H enzymes with putative preference to hydroxylation of continuous prolines and presumably all Ara*f*T enzymes needed for synthesis of the substituted arabinose chains of one to four arabinose units, have been identified and functionally characterized. Truncated root-hair phenotype is one common denominator of mutants of Hyp formation and Hyp-arabinosylation glycogenes, which act on diverse groups of *O*-glycosylated proteins, e.g., the small peptide hormones and cell wall extensins. Dissection of different substrate derived effects may not be regularly feasible and thus complicate translation from genotype to phenotype. Recently, lack of proper arabinosylation on arabinosylated proteins has been shown to influence their transport/fate in the secretory pathway, hinting to an additional layer of functionality of *O*-arabinosylation. Here, we provide an update on the prevalence and types of *O*-arabinosylated proteins and the enzymatic machinery responsible for their modifications.

Keywords: plant protein *O*-glycosylation, hydroxyproline-arabinosylation, secretory pathway, extensin, peptide hormone, plant allergens, arabinogalactan protein, hydroxyproline glycoprotein module

INTRODUCTION

Glycosylation of proteins is a common post-translational modification (PTM) on a large number of proteins across the domains of life. While some types of glycosylation, at least in part, are conserved broadly, others display more limited phylogenetic distributions. Specialization is particularly prominent for *O*-linked modifications where sugars are added to the oxygen present on mainly serine and threonine, or in the case of plants, hydroxyproline (Hyp). Hyp *O*-glycosylation takes place in the secretory pathway and occurs in two major forms, Hyp *O*-galactosylation and Hyp *O*-arabinosylation. Hyp-arabinosylation was first discovered in the structural cell wall glycoprotein family extensins (Lamport, 1963). It has since then been found in a number of unrelated families of proteins including small peptide hormones and (receptor) kinases. This review provides an update on recent insights in *O*-arabinosylation prevalence, functionality, and regulation. Some emphasis will be put on the role of glycosylation in correct protein processing and targeting

that takes place in the secretory pathway. The importance of arabinosylation to cell wall sensing and signaling will also be covered.

Hyp-*O*-arabinosylation radically sets plant *O*-glycosylation apart from that of mammalian cells. In fact, a number of allergies in man are caused by plant allergens that feature Hyp-*O*-arabinosylation. This has implications: firstly, the prospect of using mammalian cells as a clean slate system for constructing plant *O*-arabinosylation in a host cell; secondly, spurious Hyp-*O*-arabinosylation of therapeutic human proteins when expressed in plants must be prevented for plant cells to become useful cell factories for therapeutic proteins (Gomord et al., 2010).

HYP-*O*-GLYCOSYLATED PROTEINS

Contiguous Hyp-*O*-Arabinosylation

Contiguous Hyp-*O*-arabinosylation refers to the class of glycosylation motifs originally discovered in extensins, see **Figure 1A**, and demonstrated to lead to serine α-galactosylation and core arabinosylation of Hyps with β-1,2-linked arabinofuranosides (Ara$f$) of length 1-3 (Shpak et al., 1999; Kieliszewski, 2001). "Contiguous" alludes to the characteristic repetitive Ser-Hyp$_{3+}$ that defines extensins. Hyp-Araf_3 may be further elongated with an α-linked Ara$f$ and occasionally a fifth arabinosyl residue of unknown regio- and stereochemistry (Mazau and Esquerretugaye, 1986; Møller et al., 2017). Arabinoside profiles are characteristic of a species (Lamport and Miller, 1971), yet vary between tissues (Estevez et al., 2006; Møller et al., 2017) and during development as recently shown in a study of developing cotton fiber cells (Guo et al., 2019). The glycosylation machinery, which we will return to below, is ancient at least considering the three β-linked residues that we refer to as the core structure (**Figure 2**), whereas the genes encoding the extensin polypeptides appear to be streptophyte inventions (MacAlister et al., 2016; Møller et al., 2017) with considerable structural variation across the plant kingdom; particularly with notable differences between grasses and other flowering plants (Carpita, 1996). Classical extensins feature a cross-linking motif comprising two Tyr residues that may be coupled oxidatively, both intra- and inter-chain (Mnich et al., 2020). The most important inter-chain cross-link is pulcherosine (**Figure 1**). The YVY cross-linking motif common to canonical EXTs (**Figure 1**) appears to be present in commelinid monocot leucine-rich repeat extensins (LRXs) but missing in the extensins of this group. Arabinoxylan-linked ferulic and p-coumaric acids have been proposed to replace extensins as cross linkers in these plants (Francoz et al., 2015).

Extensins play roles in cell wall architecture (Hijazi et al., 2014). They are basic proteins that interact with acidic pectic polymers (Cannon et al., 2008). *In vitro* studies suggest that proper glycosylation of extensins is required for acquiring the rod-like extended structures which are required for sequestering cell wall polymers and extensin cross-linking (see **Figure 1**; Vanholst and Varner, 1984; Stafstrom and Staehelin, 1986). A hypothesis has been presented for their mode of insertion by reptation into the wall orthogonally to the plasma membrane (Kieliszewski et al., 2011). Extensins are used for wall repair, e.g., following pathogen attack (Showalter et al., 1991) but are also secreted in mucilage to play roles in non-pathogenic microbial interactions (Castilleux et al., 2018) and during cell plate formation.

Extensin cross-linking is catalyzed by peroxidases (Schnabelrauch et al., 1996) and positioning of the flanking Tyr residues, as well as the glycan structure, are integral to the process: the importance of extensin *O*-glycosylation (Gille et al., 2009), and in particular the fourth Ara$f$ residue, for cross-linking has been substantiated *via in vitro* studies, which demonstrated that the initial rate of cross-linking was primarily determined by the number of cross-linking motifs (**Figure 1**) in the protein backbone and by Hyp-Araf_4 (Chen et al., 2015). Complete removal of the fourth Ara$f$ through knockout of the *Extensin Arabinose Deficient* (*ExAD*) gene in *Arabidopsis*, however, displayed no visible phenotypic effects (Møller et al., 2017). Actual cross-linking at the molecular level was not examined.

Schnabelrauch et al. (1996) showed that de-glycosylated extensin monomers were not cross-linked *in vitro* by extensin peroxidase from tomato cell culture medium substantiating the role for arabinosylation in extensin self-assembly and cross-linking. Recently, two extensin peroxidases have been identified and demonstrated to be essential to tapetum and pollen development (Jacobowitz et al., 2019). Interestingly, the two enzymes belong to the related F and C clades of peroxidases and are not very similar to previously known extensin peroxidases from *Vitis* and tomato (M clade), and French bean and lupin (A clade) (Mnich et al., 2020).

Covalent linkages between the different polymer classes, hemicelluloses, pectins, and hydroxyproline-rich glycoproteins (HRGP), were central to the earliest cell wall models (Keegstra et al., 1973), and then less appreciated (Talbott and Ray, 1992; Carpita and Gibeaut, 1993) but growing evidence for the occurrence of covalent cross-links between polymers has to be taken into account in cell wall analysis (Fry, 2011). Cross-links between pectic polymers and extensin have been demonstrated in sugarbeet (Nunez et al., 2009) and cotton (Qi et al., 1995) but the type of linkage is unknown.

Arabinosylation of extensins is essential for root hair integrity and the root hair phenotypes played a key role in the characterization of the arabinosylation mutants (Velasquez et al., 2011), see below. Whether similar defects occur in less easily observable cells featuring diffuse growth is an open question. Generalizing from root hairs to the larger class of tip growing cells should be done with caution as their cell wall compositions differ considerably. Root hairs contain high
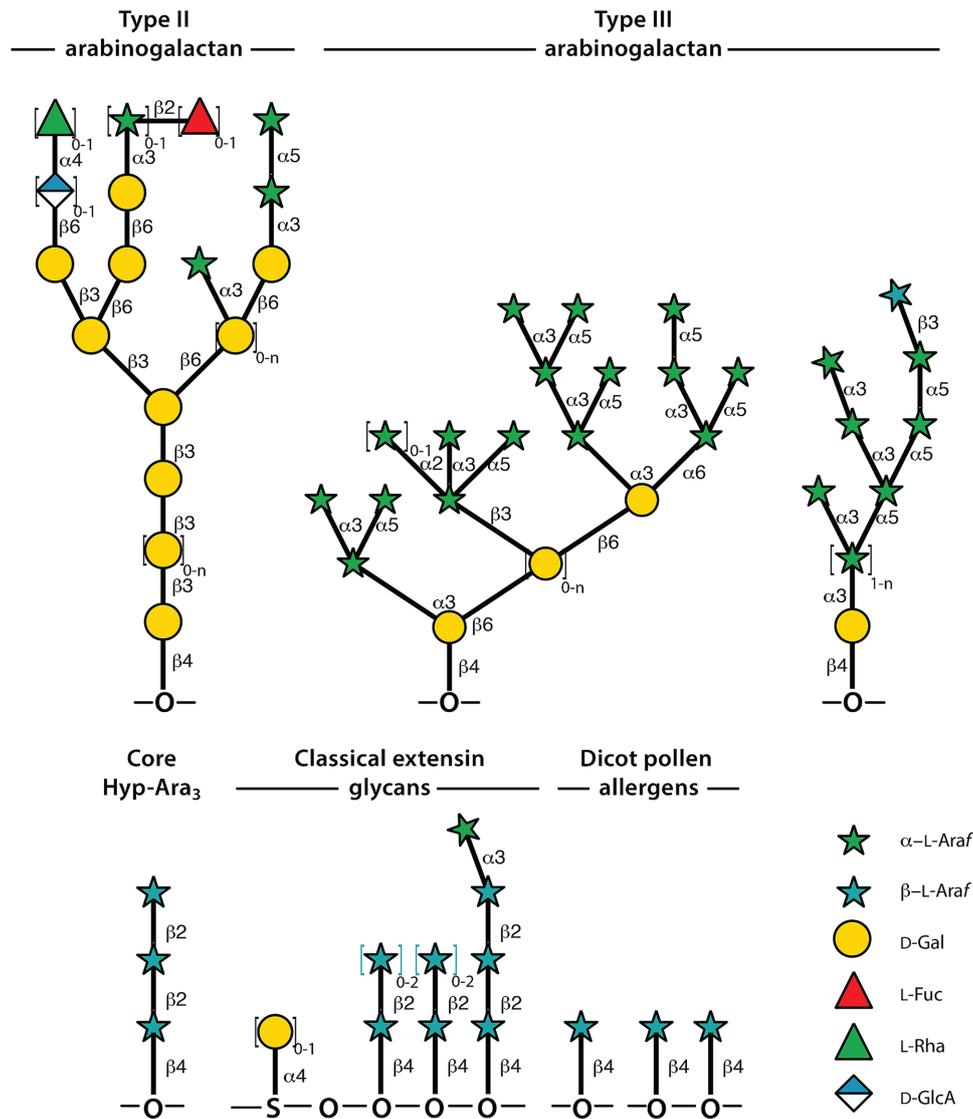
FIGURE 1 | Interchain phenolic cross-links in extensin. **(A)** Staggered partial alignment of *Arabidopsis* EXT3 with potential pulcherosine cross-links shaded in pink. **(B)** Chemical structure of the pulcherosine cross-link. The blue frame in **(A)** indicates the smallest extensin motif with experimentally validated arabinosylation (Beuder et al., 2020).

levels of cellulose, while pollen tubes produce little cellulose, but abundant callose (Park et al., 2011; Akkerman et al., 2012; Chebli et al., 2012). Root hairs also contain an acidic xyloglucan that is important for normal development (Pena et al., 2012) that does not occur in pollen tubes (Dardelle et al., 2010). In general, xyloglucan mutants show root hair growth defects, while pollen tubes appear to be unaffected (Cavalier et al., 2008), even though pollen tubes do contain xyloglucan (Lampugnani et al., 2013).

Despite the differences in their cell wall organization (Chebli et al., 2012), pollen tubes are also dependent on Hyp-arabinosylation (Ogawa-Ohnishi et al., 2013; MacAlister et al., 2016). Double mutants of *Arabidopsis hpat1* and *hpat3* (Hyp-arabinosyltransferases, see below) display severe male fertility defects due to compromised pollen tube cell wall polarity (Beuder et al., 2020). A recent genetic screen for suppressors of the *hpat* pollen phenotype identified mutants in the late stages of the secretory pathway, specifically in the vesicle-tethering exocyst complex. Beuder et al. (2020) found that *hpat* pollen tubes had increased rates of HPAT-modified protein secretion compared to WT pollen tubes and the rate of secretion was reduced in the suppressed line to near WT levels. This observation is consistent with a "toxic" effect of increased secretion of un-arabinosylated proteins. Which specific HPAT-target protein(s) may be responsible is unknown, but candidates include the canonical extensins or other proteins carrying Hyp-*O*-arabinosylation sites (see below).

Extensins are members of the HRGP superfamily which also comprises the lightly glycosylated proline-rich proteins (PRPs) and the often very heavily glycosylated arabinogalactan proteins (AGPs, see **Figure 2**). Early studies revealed that AGPs to be differentially and transiently expressed during embryogenesis in oilseed rape (Pennell et al., 1991) and are also implicated in somatic embryogenesis (Kreuger and Vanholst, 1993) suggesting roles in differentiation, cell identity, and cell-cell interaction. Examination of AGPs in carrot root similarly points to roles in differentiation (Knox et al., 1991). The glycan structure of AGPs is important as documented by the requirement of a fucosylated AGP for root cell expansion (van Hengel and Roberts, 2002). AGPs are important to growth of pollen tubes and more precisely in the endosomal transport that is essential to tip growth, reviewed in Dehors et al. (2019). Glycosylation sites in AGPs are of the clustered **non-contiguous Hyp *O*-galactosylation** type. The glycans are β-1,3-galactans featuring β-1,6-linked galactan side-chains and further decorated with rhamnose, (Me-)glucuronic acid, arabinose, and fucose (**Figure 2**; Tryfona et al., 2012). AGPs were reviewed recently (Seifert, 2020). One of the subfamilies, referred to as *hybrid* in a newly updated HRGP classification scheme (Liu et al., 2020), comprises both sites for clustered **non-contiguous Hyp *O*-galactosylation** and sites for **contiguous Hyp *O*-arabinosylation** and is thus included here.

While hybrid HRGPs refer to proteins featuring domains from more than one HRGP sub-family, chimeric HRGPs comprise one type of HRGP domain plus domains from another gene family (Liu et al., 2020). A wide variety of protein families fall into this class.

The updated HRGP classification comprises the following sub-classes of chimeric proteins with extensin domains: LRXs, proline-rich extensin-like receptor kinases (PERKs), formin-homolog EXTs (FH EXTs), and other chimeric EXTs.

FIGURE 2 | Representative structures of Hyp-glycans. Some of the structural variation is indicated by enclosing repeat structures in square brackets followed by indices representing the range of replicates, where *n* means that the maximum is not known. Not all structural variations of arabinogalactans are shown. The representative Type-II structures were assembled from Tryfona et al. (2012) and Seifert (2020). Type-III structures and the dicot pollen allergens were derived from Leonard et al. (2005) and Leonard et al. (2010). The ordering of arabinoside side-chains in extensins is not known; hence, the choice of placing Hyp-Ara*f*$_4$ toward the C-terminus of the motif is arbitrary. The structures shown are from Kieliszewski (2001).

The *Arabidopsis* genome encodes four LRXs involved in pollen and pollen tube development and seven that are expressed in vegetative tissues (Fabrice et al., 2018). LRX proteins function *via* the binding of their LRR domain to rapid alkalinization factors (RALF) signaling peptides and *Catharanthus roseus* receptor-like kinase1-like (CrRLK1L) proteins to monitor cell wall integrity (Blackburn et al., 2020; Herger et al., 2020; Moussu et al., 2020). The EXT-like domain of *Arabidopsis* LRX1 is required for its insolubilization in the cell wall and is essential for its function in root hair elongation (Ringli, 2010).

The molecular function of the PERK family is less well understood. *AtPERK4* is involved in the response to the plant

hormone abscisic acid and its kinase activity is activated by abscisic acid and calcium (Bai et al., 2009). Many PERKs are expressed primarily or exclusively in pollen and pollen tubes, but not all PERKs play roles in pollen development and in cell wall sensing (Chen et al., 2020), i.e., monitoring cell wall stresses during cell expansion, for example, but roles in responding to wounding and pathogens are also documented (Qanmber et al., 2019).

The intercellular part of formin-homolog EXTs interacts with actin and is thus another good candidate for coordinating the cell wall to the cell interior (Borassi et al., 2016). The EXT-like extracellular domain of *Arabidopsis* FH1 has been shown to physically interact with the cell wall, limiting its lateral mobility

at the plasma membrane (Martiniere et al., 2011). Other members of this family are required for proper root hair and pollen tube elongation (Cheung et al., 2010; Huang et al., 2013; Lan et al., 2018), suggesting a possible role in coordinating a polarized actin cytoskeleton and a polarized cell wall during tip growth.
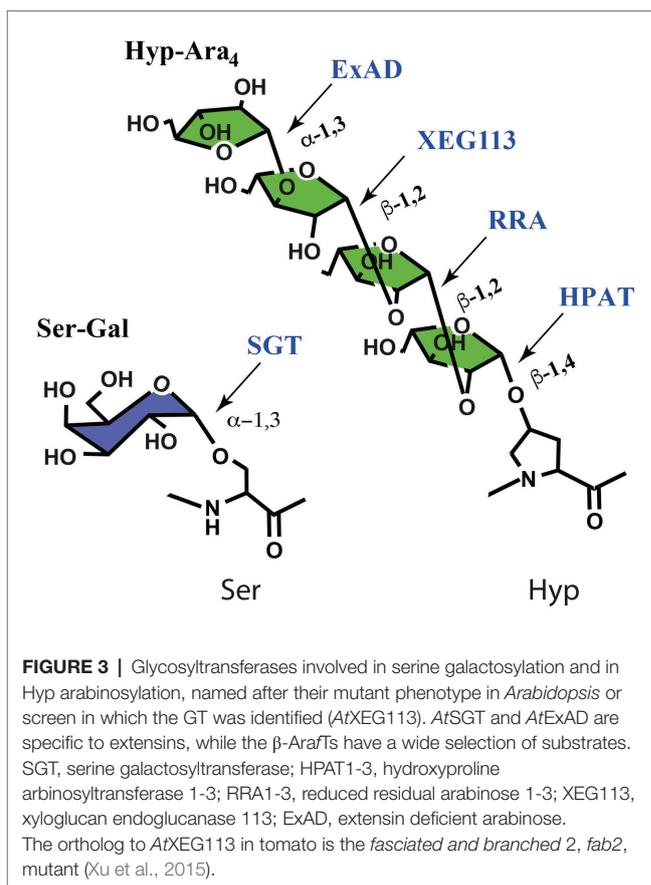
It is broadly accepted that the occurrence of the SP_{3+} motif is sufficient for Pro hydroxylation and *O*-arabinosylation but glycosylation of these domains in EXT chimeras has not been validated experimentally as far as we are aware.

Solanaceous lectins fall in the category of *other chimeric*. They feature chitin-binding domains interspaced with heavily glycosylated extensin-like regions (Kieliszewski et al., 1994). Roles in pathogen defense are inferred from the chitin affinity and in a few cases also documented (Chen et al., 2018). Their evolutionary origin is a conundrum: tomato (Oguri et al., 2008) and potato lectin (Van Damme et al., 2004) are closely related but their domain organization is quite different. Smaller differences in lectin specificity are inferred from *in silico* analyses (Jain et al., 2020).

Finally, two pollen allergens are considered which also fall in the category of *other chimeric* due to the presence of a defensin domain. Defensins are evolutionary ancient peptides involved in innate immunity. These allergens are also hybrid as they feature both **non-contiguous Hyp *O*-galactosylation** and sites for **contiguous Hyp *O*-arabinosylation**. The structures to be discussed are shown in **Figure 2**. Mugwort, *Artemisia vulgaris* Art v 1 (Himly et al., 2002), and short ragweed, *Ambrosia artemisiifolia* amb a 4 (Leonard et al., 2010), are both modular proteins featuring an N-terminal defensin-like and C-terminal hydroxyproline-rich domain. C-terminal SP_3 motifs in Art v 1 are β-arabinosylated, often on adjacent Hyps while some non-contiguous Hyps carry Type-III arabinogalactans (see **Figure 2**; Leonard et al., 2005). Type III arabinogalactans feature a galactan backbone that carry branched α-linked arabinans. The backbone is β-1,6-linked in contrast to Type-II arabinogalactans (**Figure 2**) of standard AGPs, yet, it is Yariv binding. If the Hyp-$Araf_1$ side-chains are synthesized as such, it raises the question how they evade being elongated. Processing is an alternative (Guo et al., 2019), but while β-arabinofuranoside degradation is known from bacteria (Miyake et al., 2020; Saito et al., 2020), it remains to be discovered in plants. The ragweed allergen is related but differs in that the P_3 motifs are preceded by an Ala rather than Ser, yet, the Hyps carry single, β-linked Araf just as the mugwort allergen. The Type III arabinogalactans are also different in that the galactan backbone is of length 1 and some of the α-arabinan side-chains are capped with a β-linked Araf. If α-arabinan capping is more wide-spread, and then the Ray1 gene (Gille et al., 2013), characterized in *Arabidopsis*, may encode a candidate GT for transferring the terminal β-linked Araf.

Non-contiguous Hyp *O*-Arabinosylation

Non-contiguous Hyp *O*-arabinosylation refers to synthesis of the $Araf_3$ core structure onto Hyps that are not derived from the typical SP_{3+} motif. The core structure is never elongated

FIGURE 3 | Glycosyltransferases involved in serine galactosylation and in Hyp arabinosylation, named after their mutant phenotype in *Arabidopsis* or screen in which the GT was identified (*At*XEG113). *At*SGT and *At*ExAD are specific to extensins, while the β-ArafTs have a wide selection of substrates. SGT, serine galactosyltransferase; HPAT1-3, hydroxyproline arbinosyltransferase 1-3; RRA1-3, reduced residual arabinose 1-3; XEG113, xyloglucan endoglucanase 113; ExAD, extensin deficient arabinose. The ortholog to *At*XEG113 in tomato is the *fasciated and branched* 2, *fab2*, mutant (Xu et al., 2015).

with a fourth α-linked Araf (**Figure 3**). Hyp-$Araf_3$ is found on a number of diverse proteins.

A number of small peptide hormones and certain other allergens fall under this heading. Genome and transcriptome analyses have identified more than 1,000 potential peptide hormones in *Arabidopsis* (Lease and Walker, 2006; Matsubayashi, 2014). Several PTMs are regularly required for peptide maturation and function. Most peptide hormones thus undergo PTM modifications and hydroxyproline arabinosylation and sulfation of tyrosine residues, for example, has been documented in a few instances. CLAVATA3 (CLV3) is part of a network involved in stem cell maintenance and differentiation in meristems (Clark et al., 1996; Schoof et al., 2000) and it features a single Hyp-$Araf_3$ side-chain (Ohyama et al., 2009). Loss of CLV3 results in disruption of the meristem size control pathway leading to a progressive increase in meristem size referred to as fasciation (Clark et al., 1995). In one study in tomato where the initiating and elongating arabinosyltransferases acting on CLV3 were knocked out by gene editing, the most severe fasciation resulted from the knockout of the initiating arabinosyltransferase, i.e., plants lacking the entire tri-arabinoside (Xu et al., 2015). The phenotype could only be rescued by arabinosylated CLV3, demonstrating that arabinosylation of CLV3 is important for its function. "Weaker" meristem phenotypes were observed in mutants with truncated arabinosides, indicating that CLV3 must be fully arabinosylated for full activity (Xu et al., 2015). This conclusion is corroborated

by *in vitro* studies using chemically synthesized differentially arabinosylated CLV3 peptides. CLV3/EMBRYO SURROUNDING REGION-related (CLE) peptides are related to CLV3 and several CLEs can complement *clv3* knockouts (Ni and Clark, 2006) and for CLE2 Hyp *O*-arabinosylation was demonstrated and also that it is essential to function (Ohyama et al., 2009).

The small peptide hormone PLANT PEPTIDE CONTAINING SULFATED TYROSINE 1 (PSY1) is a growth-promoting secreted tyrosine-sulfated glycopeptide whose receptor directly phosphorylates and activates the plasma membrane H$^+$-ATPase, which results in acidification of the apoplastic space facilitating cell wall loosening and is, therefore, a key step in cell expansion (Fuglsang et al., 2014; Mahmood et al., 2014). The mature PSY1 features a sulfated Tyr and a single Hyp-Araf_3 side-chain (Amano et al., 2007).

C-TERMINALLY ENCODED PEPTIDE (CEP) is Hyp-glycosylated at Asp-Ser-**Hyp**-Gly-Val (Patel et al., 2018), CLV3 and CLE2 share Gly-**Hyp**-Asp-Pro as glycosylation site while PSY1 is glycosylated at the first Hyp of Val-**Hyp**-Hyp-Ser. A large data-set is required to define a common motif or a set of motifs that lead non-contiguous Hyp *O*-arabinosylation. Structural features and biological functions of the small peptide hormones have been reviewed extensively (Matsubayashi, 2011, 2014; Tavormina et al., 2015; Stuhrwohldt and Schaller, 2019).

A number of pollen allergens belong under this heading but differ from the peptide hormones in two respects: arabinosylation is not confined to the core structure, Hyp-Araf_3; shorter side-chains are also found but not Hyp-Araf_4. The occurrence of a β-expansin domain in some of these suggests roles in pollen-pistil interactions (Wang et al., 2016). Major allergens from ryegrass and Timothy grass, *Lolium perenne* 1 (Lol p 1) and *Phleum pretense* 1 (Phl p 1), are β-expansins (Li et al., 2003) while Phl p 5 features a pollen RNase domain (Bufe et al., 1995). Phl p 1 and 5 were experimentally shown to carry Hyp and Hyp-Araf_{1-3} substitutions (Halim et al., 2015) and Lol p 1, a close homolog of Phl p 1, was shown to carry Hyp and likely Hyp substituted arabinosylation (Li et al., 2003).

BIOSYNTHETIC MACHINERY

The arabinosyltransferases (Ara$f$Ts) that catalyze Hyp-glycosylation are encoded by single genes or small gene families. There are three Ara$f$Ts in *Arabidopsis* that transfer the first and three that transfer the second β-linked Ara$f$ and only one Ara$f$T for each of the third and fourth Ara$f$, see **Figure 3**. This means that the diverse gene families considered above largely share the same glycosylation machinery. This raises questions regarding how Hyp-Ara$_1$ in allergens and Hyp-Ara$_3$ in peptide hormones evade further elongation; and it raises questions as to the recognition of glycosylation motifs. For the latter question the prolyl-4-hydroxylases (P4Hs), which insert the C$_4$-OH group onto prolyl residues in the protein backbone, thus enabling recognition of appropriate Hyp glycosylation sites for the initiating GTs should be considered. The P4Hs are type II membrane anchored proteins and form a family with 13 members in *Arabidopsis*. Some insights in the labor division, i.e., substrate specificities, among the P4Hs have been documented.

Consecutive Pro residues appear to be a favored motif for some P4Hs, and AtP4H5, -2, and -13 have been identified as involved in proline hydroxylation of cell wall extensins *in vivo* (Velasquez et al., 2011, 2015b). However, the widespread prevalence of Hyp residues without glycosylation necessitates that the initiating GTs also play roles in recognition of arabinosylation motifs.

Ara$f$Ts, both the enzymes that target the Hyps in the peptide backbone and those that elongate the glycan chains are Golgi localized type II membrane anchored proteins. The three hydroxyproline *O*-arabinosyltransferases, HPAT1-3 are founding members of CaZY-family GT95, identified and functionally characterized by Ogawa-Ohnishi et al. (2013). Family GT95 is related to family GT96 to which the extensin serine α-galactosyltransferase SGT1 belongs (Saito et al., 2014). Mutant data showed that HPAT1, HPAT2, and HPAT3 redundantly contribute to arabinosylation of the abundant extensin 3, EXT3, in *Arabidopsis*, and that at least subgroups of the CLE peptides are mainly arabinosylated by HPAT3. The three enzymes in *Arabidopsis* thus have overlapping but distinct target substrate preferences (Ogawa-Ohnishi et al., 2013). It would be tempting to guess that the role of the Ser-galactosylation would be to guide either the P4Hs or the arabinosyltransferases but the *sgt-1* knockout mutant is unaffected in arabinosylation (Møller et al., 2017).

The β-arabinosyltransferases that add the second residue, reduced residual arabinose 1-3 (RRA1-3; Egelund et al., 2007; Velasquez et al., 2011), and the third residue, XEG113 (Gille et al., 2009), are all GT77 family proteins, see **Figure 3**, and are conserved over large phylogenetic distances. We have detected Hyp-Araf_3 in the prasinophyte *Ostreococcus tauri* (unpublished observation), as well as gene orthologues in sequenced Chlorophytes but not in Rhodophyte genomes (Ulvskov et al., 2013). Extensin arabinose deficient (ExAD; Møller et al., 2017) belongs to family GT47 and is responsible for transferring the fourth α-linked arabinosyl residue, see **Figure 3**. ExAD is probably neofunctionalised in streptophytes, as probable homologs exist in chlorophytes which do not feature the fourth Ara$f$ (Møller et al., 2017).

DISCUSSION: OUTSTANDING QUESTIONS AND FUTURE RESEARCH IN *O*-ARABINOSYLATION

Hyp *O*-Glycosylation Prevalence in Plants

Non-contiguous Hyp *O*-arabinosylation appears to be present on various proteins encoded by several gene families. The prevalence of this PTM is probably significant and its detection dependent on whether or how these PTMs were included in the glycoprotein analysis. The *O*-glycoproteome is likely to be markedly under-annotated especially with respect to number of glycosylation sites and perhaps also with respect to described structures (for some indications of the latter, see, for example, Leonard et al., 2005, 2010; Halim et al., 2015). Some initial strategies for identification of proteins prone to Hyp-glycosylation in intrinsically disordered proteins (IDPs) to which the HRGPs

belong have been provided (Johnson et al., 2017), but general high throughput strategies have not been attempted. Recently high throughput system approaches, based on lectin glycopeptide/glycan enrichment and advanced Mass spec site specific glycosylation analysis, for assessing glycosylation in the mammalian proteome, have resulted in multiple doublings of the glycosylation sites, glycosylated proteins, and structures (reviewed in Schjoldager et al., 2020 and refs herein). For mammals, it is estimated that the vast majority of secretory proteins (>85%) are glycosylated (Zielinska et al., 2010; Steentoft et al., 2013) and most nuclear and cytoplasmic proteins undergo dynamic *O*-GlcNAcylation (Hart, 2019). Implementation of such strategies on the plant Hyp *O*-glycome is expected to yield similar insights, at least.

Protein Complexes in Biosynthesis

The sugar donor substrate for arabinosylation exists as UDP-Ara*p* and thus requires a sugar ring contraction to produce UDP-Ara*f* and eventually the structures discussed above. This reaction is catalyzed by neofunctionalized GT75s (Konishi et al., 2007). These proteins are cytosolic but associated with the Golgi membrane (Dhugga et al., 1991), while the GTs are located in the secretory pathway. This calls for organization of the mutase, a transporter and the Ara*f*Ts on the luminal side of the Golgi membrane as analyzed in detail recently (Saqib et al., 2019).

There are also indications of protein complexes playing a role in Pro hydroxylation. It has been demonstrated that *At*P4H5 may form homodimers and heterodimers with *At*P4H2 and *At*P4H13 in the Golgi and that P4Hs complexation may be required for prober localization and Pro hydroxylation (Velasquez et al., 2015a). Further studies are needed to validate these findings and potential complex-formation between initiating P4H and Hyp-glycosyltransferases though the reverse genetics and/or advanced protein-protein interaction techniques may be hampered by the high prevalence of isoenzyme redundancies *in vivo*. Defining sites for **non-contiguous Hyp *O*-arabinosylation** is a major research challenge and we expect that many more families of secreted proteins will be found to carry arabinosides.

Shared Glycosylation Pathways: Genotype to Phenotype

The degree to which the general Hyp-arabinosylation machinery (i.e., the P4Hs and Ara*f*Ts required for Hyp-arabinosylation) are shared between the classes of target proteins (i.e., contiguous and non-contiguous proline hydroxylation and Hyp-arabinosylation) is unclear. But, some cases of substrate preference have been reported. For example, when the CLE2 peptide was overexpressed in *Arabidopsis hpat3-1* mutants, the peptide was largely detected in an un-arabinosylated form while when overexpressed in the *hpat1-1 hpat2-1* double mutant, arabinosylation was maintained, suggesting that *HPAT3* is primarily responsible for modification of this peptide, and presumably other CLE peptides. However, even within the CLE family, Ara*f*Ts may have substrate preferences. A *Lotus japonicus* putative *HPAT*, *PLENTY* helps regulate the

number of nitrogen-fixing root nodules formed as part of the autoregulation of nodulation mechanism (Yoro et al., 2019). Other legume *HPAT*s, i.e., *ROOT DETERMINED NODULATION1* in *Medicago truncatula* and NOD3 in *Pisum sativum* serve the same role (Schnabel et al., 2011). Tri-arabinosylation is required for the nodule suppressing activity of pea CLE peptides, including suppression of the hypernodulation of *nod3* mutants (Hastwell et al., 2019) and in Medicago wild type and *rdn1* mutants. In *L. japonicus*, three CLE peptides (CLE-RS1, 2, and 3) are involved in autoregulation of nodulation (Okamoto et al., 2009; Nishida et al., 2016). Constitutive expression of CLE-RS1 and 2 could suppress the increased root nodule phenotype in *plenty* mutants, but overexpression of CLE-RS3 could not, suggesting that CLE-RS3 is dependent on PLENTY arabinosylation for function, but CLE-RS1 and 2 are not and thus may be arabinosylated by other HPATs (Yoro et al., 2019).

Despite potential target preferences, the Hyp-arabinosylation machinery is remarkably well conserved. Somewhat surprisingly, the phenotypic consequences of its disruption vary dramatically between species. In tomato, loss of the HPAT *fasciated inflorescence* (*fin*) results in severe meristem fasciation due to loss of arabinosylation of *Sl*Clv3. However, in *Arabidopsis*, fasciation does not result from loss of *hpat* activity (MacAlister et al., 2016). Although the fully glycosylated form of Clv3 (Ara*f*₃-Clv3) has higher activity in *Arabidopsis* than the un-modified form when applied exogenously, the un-arabinosylated form is active (Shinohara and Matsubayashi, 2013; Kim et al., 2017). This difference between tomato and *Arabidopsis* might be a consequence of the selection for larger fruit and, therefore, larger meristems during tomato domestication leaving the size regulation mechanism already near its limit in tomato (Munos et al., 2011). Somewhat surprisingly, the total loss of Hyp-arabinoslyation through mutations of all three *Arabidopsis HPAT* genes is not lethal and does not result in a severe vegetative phenotype (MacAlister et al., 2016). The major phenotype of the triple mutants is a reduction in pollen fertility leading to reduced seed set (Beuder et al., 2020). Similarly, HPAT activity is not essential in the moss *Physcomitrella patens* where knockout of both HPATs resulted in increased biomass due to enhanced elongation of tip-growing vegetative filament cells (MacAlister et al., 2016). In these cases, which Hyp-arabinosylated proteins are responsible for the phenotype is difficult to determine.

Hyp *O*-Glycosylation in the Secretory Pathway: Novel Insights and Applicative Opportunities

Large-scale bioinformatics identification of *EXTs* across the plant kingdom indicates that most (76%) encode recognizable signal peptides and are, therefore, likely directed into the conventional secretory pathway (Liu et al., 2016). Signal peptides derived from EXT have been used to direct entry of heterologous proteins into the plant secretory pathway (De Loose et al., 1991; Beuder et al., 2020; Jiang et al., 2020). In addition to a signal peptide, the inclusion of plant *O*-glycosylation motifs can further enhance heterologous protein production in plant

cells, *via* an unknown mechanism. Recently, Zhang et al. (2019) demonstrated significantly increased secreted protein yield for enhanced green fluorescence protein (EGFP) when fused to a HypGP module consisting of either 18 tandem repeats of an EXT-like "Ser-Hyp-Hyp-Hyp-Hyp" motif $(SP_4)_{18}$ or an AGP-like sequence of 32 tandem "Ser-Hyp" repeats $(SP)_{32}$. When expressed in tobacco hairy root culture, the HypGP fused protein was recovered from the culture media at up to 56-fold greater levels compared to an EGFP control lacking a HypGP module. Similarly, Jiang et al. (2020) demonstrated increased secretion and improved solubility of human interferon gamma expressed in *Nicotiana benthamiana* when including a C-terminal AG-type glycomodule, $(SP)_{10}$. The secretion of such HypGP tagged proteins is also highly influenced by nutrient availability, particularly the availability of nitrogen (Zhang et al., 2016). Interestingly, the lack of detectable partially glycosylated intermediates of the $(SP_4)_{18}$ glycomodule suggests that the rate-limiting step in arabinosylation is either transport to the Golgi or the initiation of arabinosylation with later steps proceeding quickly (Zhang et al., 2019). Work in tomato has demonstrated that the sequentially acting Ara*f*Ts required to modify the *Sl*CLV3 peptide are spatial separated in the *cis*, *medial*, and *trans* Golgi in an order reflecting their order of action (Xu et al., 2015). This suggests a simple linear progression of modification as target proteins move through the Golgi. The observation of the increased rates of Hyp-arabinosylated protein secretion in *hpat1 hpat3* double mutant pollen tubes (Beuder et al., 2020), suggests the existence of either as an active retention mechanism for partially glycosylated species (mutants fail to initiate glycosylation and thus lack the retention signal) or passive retention through the physical interaction between glycosylated proteins and the Ara*f*Ts modifying them until glycosylation is completed. Within the ER, N-linked glycosylation serves as a well-described, conserved mechanism to monitor protein folding, targeting misfolded proteins for degradation (Nagashima et al., 2018; Shenkman and Lederkremer, 2019). How or if glycosylation

status is monitored in the Golgi is unclear, however. Unlike the ER, to date, there is no evidence for a general protein glycosylation checkpoint or glycosylation-based quality control mechanism in the Golgi. In animal systems, disrupted Golgi stacking is reported to increase trafficking of several glycoproteins which are released in an under-glycosylated form, suggesting that the rate of glycoprotein secretion is not directly controlled by glycosylation status, but is a consequence of Golgi organization and the accessibility of proteins to the trafficking machinery (Zhang and Wang, 2016). With regard to glycoprotein movement through the secretory pathway, the unanswered questions are numerous and include how or if the rate of trafficking is regulated, the degree to which glycoproteins are selectively chosen for packaging into secretory vesicles, how secretory vesicles are directed to their target membrane, and how glycoprotein secretion relates to secretion of other Golgi products, particularly the carbohydrates destined for the cell wall. Use of synthetic glycosylated reporter proteins holds great promise to answer these and other questions.

AUTHOR CONTRIBUTIONS

BP: Hyp-*O*-glycosylation machinery, introduction, discussion, and abstract. PU: HYP-*O*-glycosylated proteins, introduction, discussion, and abstract. CM: discussion: *O*-glycosylation in the secretory pathway, significance of glycosylation on peptide hormones, and plant *O*-glycosylation modules for boosting of glycoprotein production in plants. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Akkerman, M., Franssen-Verheijen, M. A., Immerzeel, P., Hollander, L. D., Schel, J. H., and Emons, A. M. (2012). Texture of cellulose microfibrils of root hair cell walls of *Arabidopsis thaliana*, Medicago truncatula, and *Vicia sativa*. *J. Microsc.* 247, 60–67. doi: 10.1111/j.1365-2818.2012.03611.x

Amano, Y., Tsubouchi, H., Shinohara, H., Ogawa, M., and Matsubayashi, Y. (2007). Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 18333–18338. doi: 10.1073/pnas.0706403104

Bai, L., Zhang, G., Zhou, Y., Zhang, Z., Wang, W., Du, Y., et al. (2009). Plasma membrane-associated proline-rich extensin-like receptor kinase 4, a novel regulator of Ca signalling, is required for abscisic acid responses in *Arabidopsis thaliana*. *Plant J.* 60, 314–327. doi: 10.1111/j.1365-313X.2009.03956.x

Beuder, S., Dorchak, A., Bhide, A., Moeller, S. R., Petersen, B. L., and MacAlister, C. A. (2020). Exocyst mutants suppress pollen tube growth and cell wall structural defects of hydroxyproline O-arabinosyltransferase mutants. *Plant J.* 103, 1399–1419. doi: 10.1111/tpj.14808

Blackburn, M. R., Haruta, M., and Moura, D. S. (2020). Twenty years of progress in physiological and biochemical investigation of RALF peptides. *Plant Physiol.* 182, 1657–1666. doi: 10.1104/pp.19.01310

Borassi, C., Sede, A. R., Mecchia, M. A., Salgado Salter, J. D., Marzol, E., Muschietti, J. P., et al. (2016). An update on cell surface proteins containing extensin-motifs. *J. Exp. Bot.* 67, 477–487. doi: 10.1093/jxb/erv455

Bufe, A., Schramm, G., Keown, M. B., Schlaak, M., and Becker, W. M. (1995). Major allergen Phl-P-Vb in timothy-grass is a novel pollen RNase. *FEBS Lett.* 363, 6–12. doi: 10.1016/0014-5793(95)00259-C

Cannon, M. C., Terneus, K., Hall, Q., Tan, L., Wang, Y., Wegenhart, B. L., et al. (2008). Self-assembly of the plant cell wall requires an extensin scaffold. *Proc. Natl. Acad. Sci.* 105, 2226–2231. doi: 10.1073/pnas.0711980105

Carpita, N. C. (1996). Structure and biogenesis of the cell walls of grasses. *Annu. Rev. Plant Biol.* 47, 445–476. doi: 10.1146/annurev.arplant.47.1.445

Carpita, N. C., and 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. doi: 10.1111/j.1365-313X.1993.tb00007.x

Castilleux, R., Plancot, B., Ropitaux, M., Carreras, A., Leprince, J., Boulogne, I., et al. (2018). Cell wall extensins in root-microbe interactions and root secretions. *J. Exp. Bot.* 69, 4235–4247. doi: 10.1093/jxb/ery238

Cavalier, D. M., Lerouxel, O., Neumetzler, L., Yamauchi, K., Reinecke, A., Freshour, G., 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. doi: 10.1105/tpc.108.059873

Chebli, Y., Kaneda, M., Zerzour, R., and Geitmann, A. (2012). The cell wall of the *Arabidopsis* pollen tube—spatial distribution, recycling, and network formation of polysaccharides. *Plant Physiol.* 160, 1940–1955. doi: 10.1104/pp.112.199729

Chen, C. S., Chen, C. Y., Ravinath, D. M., Bungahot, A., Cheng, C. P., and You, R. I. (2018). Functional characterization of chitin-binding lectin from Solanum integrifolium containing anti-fungal and insecticidal activities. *BMC Plant Biol.* 18:3. doi: 10.1186/s12870-017-1222-0

Chen, Y., Dong, W., Tan, L., Held, M. A., and Kieliszewski, M. J. (2015). Arabinosylation plays a crucial role in extensin cross-linking in vitro. *Biochem. Insights* 8, 1–13. doi: 10.4137/BCI.S31353

Chen, G., Wang, J., Wang, H., Wang, C., Tang, X., Li, J., et al. (2020). Genome-wide analysis of proline-rich extension-like receptor protein kinase (PERK) in Brassica rapa and its association with the pollen development. *BMC Genomics* 21:401. doi: 10.1186/s12864-020-06802-9

Cheung, A. Y., Niroomand, S., Zou, Y., and Wu, H. M. (2010). A transmembrane formin nucleates subapical actin assembly and controls tip-focused growth in pollen tubes. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16390–16395. doi: 10.1073/pnas.1008527107

Clark, S. E., Jacobsen, S. E., Levin, J. Z., and Meyerowitz, E. M. (1996). The CLAVATA and SHOOT MERISTEMLESS loci competitively regulate meristem activity in *Arabidopsis*. *Development* 122, 1567–1575.

Clark, S. E., Running, M. P., and Meyerowitz, E. M. (1995). Clavata3 is a specific regulator of shoot and floral meristem development affecting the same processes as Clavata1. *Development* 121, 2057–2067.

Dardelle, F., Lehner, A., Ramdani, Y., Bardor, M., Lerouge, P., Driouich, A., et al. (2010). Biochemical and immunocytological characterizations of *Arabidopsis* pollen tube cell wall. *Plant Physiol.* 153, 1563–1576. doi: 10.1104/pp.110.158881

Dehors, J., Mareck, A., Kiefer-Meyer, M. C., Menu-Bouaouiche, L., Lehner, A., and Mollet, J. C. (2019). Evolution of cell wall polymers in tip-growing land plant gametophytes: composition, distribution, functional aspects and their remodeling. *Front. Plant Sci.* 10:441. doi: 10.3389/fpls.2019.00441

De Loose, M., Gheysen, G., Tire, C., Gielen, J., Villarroel, R., Genetello, C., et al. (1991). The extensin signal peptide allows secretion of a heterologous protein from protoplasts. *Gene* 99, 95–100. doi: 10.1016/0378-1119(91)90038-D

Dhugga, K. S., Ulvskov, P., Gallagher, S. R., and Ray, P. M. (1991). Plant polypeptides reversibly glycosylated by UDP-glucose. Possible components of Golgi beta-glucan synthase in pea cells. *J. Biol. Chem.* 266, 21977–21984. doi: 10.1016/S0021-9258(18)54733-5

Egelund, J., Obel, N., Ulvskov, P., Geshi, N., Pauly, M., Bacic, A., et al. (2007). Molecular characterization of two *Arabidopsis thaliana* glycosyltransferase mutants, rra1 and rra2, which have a reduced residual arabinose content in a polymer tightly associated with the cellulosic wall residue. *Plant Mol. Biol.* 64, 439–451. doi: 10.1007/s11103-007-9162-y

Estevez, J. M., Kieliszewski, M. J., Khitrov, N., and Somerville, C. (2006). Characterization of synthetic hydroxyproline-rich proteoglycans with arabinogalactan protein and extensin motifs in arabidopsis. *Plant Physiol.* 142, 458–470. doi: 10.1104/pp.106.084244

Fabrice, T. N., Vogler, H., Draeger, C., Munglani, G., Gupta, S., Herger, A. G., et al. (2018). LRX proteins play a crucial role in pollen grain and pollen tube cell wall development. *Plant Physiol.* 176, 1981–1992. doi: 10.1104/pp.17.01374

Francoz, E., Ranocha, P., Nguyen-Kim, H., Jamet, E., Burlat, V., and Dunand, C. (2015). Roles of cell wall peroxidases in plant development. *Phytochemistry* 112, 15–21. doi: 10.1016/j.phytochem.2014.07.020

Fry, S. C. (2011). "Cell wall polysaccharide composition and covalent crosslinking" in *Annual plant reviews: Plant polysaccharides, biosynthesis and bioengineering. Vol. 41.* ed. P. Ulvskov (Blackwell), 1–42.

Fuglsang, A. T., Kristensen, A., Cuin, T. A., Schulze, W. X., Persson, J., Thuesen, K. H., et al. (2014). Receptor kinase-mediated control of primary active proton pumping at the plasma membrane. *Plant J.* 80, 951–964. doi: 10.1111/tpj.12680

Gille, S., Hansel, U., Ziemann, M., and Pauly, M. (2009). Identification of plant cell wall mutants by means of a forward chemical genetic approach using hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* 106, 14699–14704. doi: 10.1073/pnas.0905434106

Gille, S., Sharma, V., Baidoo, E. E. K., Keasling, J. D., Scheller, H. V., and Pauly, M. (2013). Arabinosylation of a Yariv-precipitable cell wall polymer

impacts plant growth as exemplified by the *Arabidopsis* glycosyltransferase mutant ray1. *Mol. Plant* 6, 1369–1372. doi: 10.1093/mp/sst029

Gomord, V., Fitchette, A. C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D., et al. (2010). Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol. J.* 8, 564–587. doi: 10.1111/j.1467-7652.2009.00497.x

Guo, X., Hansen, B. O., Moeller, S. R., Harholt, J., Mravec, J., Willats, W., et al. (2019). Extensin arabinoside chain length is modulated in elongating cotton fibre. *Cell Surf.* 5:100033. doi: 10.1016/j.tcsw.2019.100033

Halim, A., Carlsson, M. C., Madsen, C. B., Brand, S., Moller, S. R., Olsen, C. E., et al. (2015). Glycoproteomic analysis of seven major allergenic proteins reveals novel post-translational modifications. *Mol. Cell. Proteomics* 14, 191–204. doi: 10.1074/mcp.M114.042614

Hart, G. W. (2019). Nutrient regulation of signaling and transcription. *J. Biol. Chem.* 294, 2211–2231. doi: 10.1074/jbc.AW119.003226

Hastwell, A. H., Corcilius, L., Williams, J. T., Gresshoff, P. M., Payne, R. J., and Ferguson, B. J. (2019). Triarabinosylation is required for nodulation-suppressive CLE peptides to systemically inhibit nodulation in *Pisum sativum*. *Plant Cell Environ.* 42, 188–197. doi: 10.1111/pce.13325

Herger, A., Gupta, S., Kadler, G., Franck, C. M., Boisson-Dernier, A., and Ringli, C. (2020). Overlapping functions and protein-protein interactions of LRR-extensins in *Arabidopsis*. *PLoS Genet.* 16:e1008847. doi: 10.1371/journal.pgen.1008847

Hijazi, M., Velasquez, S. M., Jamet, E., Estevez, J. M., and Albenne, C. (2014). An update on post-translational modifications of hydroxyproline-rich glycoproteins: toward a model highlighting their contribution to plant cell wall architecture. *Front. Plant Sci.* 5:395. doi: 10.3389/fpls.2014.00395

Himly, M., Jahn-Schmid, B., Dedic, A., Kelemen, P., Wopfner, N., Altmann, F., et al. (2002). Art v 1, the major allergen of mugwort pollen, is a modular glycoprotein with a defensin-like and a hydroxyproline-rich domain. *FASEB J.* 16, 106–108. doi: 10.1096/fj.02-0472fje

Huang, J., Kim, C. M., Xuan, Y. H., Liu, J., Kim, T. H., Kim, B. K., et al. (2013). Formin homology 1 (OsFH1) regulates root-hair elongation in rice (*Oryza sativa*). *Planta* 237, 1227–1239. doi: 10.1007/s00425-013-1838-8

Jacobowitz, J. R., Doyle, W. C., and Weng, J. K. (2019). PRX9 and PRX40 are extensin peroxidases essential for maintaining tapetum and microspore cell wall integrity during *Arabidopsis* anther development. *Plant Cell* 31, 848–861. doi: 10.1105/tpc.18.00907

Jain, M., Muthukumaran, J., and Singh, A. K. (2020). Comparative structural and functional analysis of STL and SLL, chitin-binding lectins from Solanum spp. *J. Biomol. Struct. Dyn.* 1–16. doi: 10.1080/07391102.2020.1781693 [Epub ahead of print]

Jiang, M. C., Hu, C. C., Hsu, W. L., Hsu, T. L., Lin, N. S., and Hsu, Y. H. (2020). Fusion of a novel native signal peptide enhanced the secretion and solubility of bioactive human interferon gamma glycoproteins in nicotiana benthamiana using the bamboo mosaic virus-based expression system. *Front. Plant Sci.* 11:594758. doi: 10.3389/fpls.2020.594758

Johnson, K. L., Cassin, A. M., Lonsdale, A., Bacic, A., Doblin, M. S., and Schultz, C. J. (2017). Pipeline to identify hydroxyproline-rich glycoproteins. *Plant Physiol.* 174, 886–903. doi: 10.1104/pp.17.00294

Keegstra, K., Talmadge, K. W., Bauer, W. D., and Albersheim, P. (1973). The structure of plant cell walls: III. A model of the walls of suspension-cultured sycamore cells based on the interconnections of the macromolecular components. *Plant Physiol.* 51, 188–197. doi: 10.1104/pp.51.1.188

Kieliszewski, M. J. (2001). The latest hype on Hyp-O-glycosylation codes. *Phytochemistry* 57, 319–323. doi: 10.1016/S0031-9422(01)00029-2

Kieliszewski, M. J., Lamport, D. T. A., Tan, L., and Cannon, M. C. (2011). "Hydroxyproline-rich glycoproteins: form and function" in *Annual plant reviews: plant polysaccharides, biosynthesis and bioengineering. Vol. 41.* ed. P. Ulvskov (Blackwell), 321–342.

Kieliszewski, M. J., Showalter, A. M., and Leykam, J. F. (1994). Potato lectin: a modular protein sharing sequence similarities with the extensin family, the hevein lectin family, and snake venom disintegrins (platelet aggregation inhibitors). *Plant J.* 5, 849–861. doi: 10.1046/j.1365-313X.1994.5060849.x

Kim, H., Wu, C. Y., Yu, H. M., Sheen, J., and Lee, H. (2017). Dual CLAVATA3 peptides in *Arabidopsis* shoot stem cell signaling. *J. Plant Biol.* 60, 506–512. doi: 10.1007/s12374-017-0083-2

Knox, J. P., Linstead, P. J., Peart, J., Cooper, C., and Roberts, K. (1991). Developmentally regulated epitopes of cell-surface arabinogalactan proteins

and their relation to root-tissue pattern-formation. *Plant J.* 1, 317–326. doi: 10.1046/j.1365-313X.1991.t01-9-00999.x

Konishi, T., Takeda, T., Miyazaki, Y., Ohnishi-Kameyama, M., Hayashi, T., O'neill, M. A., et al. (2007). A plant mutase that interconverts UDP-arabinofuranose and UDP-arabinopyranose. *Glycobiology* 17, 345–354. doi: 10.1093/glycob/cwl081

Kreuger, M., and Vanholst, G. J. (1993). Arabinogalactan proteins are essential in somatic embryogenesis of Daucus-carota L. *Planta* 189, 243–248. doi: 10.1007/BF00195083

Lamport, D. T. A. (1963). Oxygen fixation into hydroxyproline of plant cell wall protein. *J. Biol. Chem.* 238, 1438–1440. doi: 10.1016/S0021-9258(18)81202-9

Lamport, D. T., and Miller, D. H. (1971). Hydroxyproline arabinosides in the plant kingdom. *Plant Physiol.* 48, 454–456. doi: 10.1104/pp.48.4.454

Lampugnani, E. R., Moller, I. E., Cassin, A., Jones, D. F., Koh, P. L., Ratnayake, S., et al. (2013). In vitro grown pollen tubes of *Nicotiana alata* actively synthesise a fucosylated xyloglucan. *PLoS One* 8:e77140. doi: 10.1371/journal.pone.0077140

Lan, Y. X., Liu, X. N., Fu, Y., and Huang, S. J. (2018). *Arabidopsis* class I formins control membrane-originated actin polymerization at pollen tube tips. *PLoS Genet.* 14:e1007789. doi: 10.1371/journal.pgen.1007789

Lease, K. A., and Walker, J. C. (2006). The *Arabidopsis* unannotated secreted peptide database, a resource for plant peptidomics. *Plant Physiol.* 142, 831–838. doi: 10.1104/pp.106.086041

Leonard, R., Petersen, B. O., Himly, M., Kaar, W., Wopfner, N., Kolarich, D., et al. (2005). Two novel types of O-glycans on the mugwort pollen allergen art v 1 and their role in antibody binding. *J. Biol. Chem.* 280, 7932–7940. doi: 10.1074/jbc.M410407200

Leonard, R., Wopfner, N., Pabst, M., Stadlmann, J., Petersen, B. O., Duus, J. O., et al. (2010). A new allergen from ragweed (*Ambrosia artemisiifolia*) with homology to art v 1 from mugwort. *J. Biol. Chem.* 285, 27192–27200. doi: 10.1074/jbc.M110.127118

Li, L. C., Bedinger, P. A., Volk, C., Jones, A. D., and Cosgrove, D. J. (2003). Purification and characterization of four beta-expansins (Zea m 1 isoforms) from maize pollen. *Plant Physiol.* 132, 2073–2085. doi: 10.1104/pp.103.020024

Liu, X., Mckenna, S., Welch, L. R., and Showalter, A. M. (2020). Bioinformatic identification of plant hydroxyproline-rich glycoproteins. *Methods Mol. Biol.* 2149, 463–481. doi: 10.1007/978-1-0716-0621-6_26

Liu, X., Wolfe, R., Welch, L. R., Domozych, D. S., Popper, Z. A., and Showalter, A. M. (2016). Bioinformatic identification and analysis of extensins in the plant kingdom. *PLoS One* 11:e0150177. doi: 10.1371/journal.pone.0168743

MacAlister, C. A., Ortiz-Ramirez, C., Becker, J. D., Feijo, J. A., and Lippman, Z. B. (2016). Hydroxyproline O-arabinosyltransferase mutants oppositely alter tip growth in *Arabidopsis* thaliana and *Physcomitrella patens*. *Plant J.* 85, 193–208. doi: 10.1111/tpj.13079

Mahmood, K., Kannangara, R., Jorgensen, K., and Fuglsang, A. T. (2014). Analysis of peptide PSY1 responding transcripts in the two *Arabidopsis* plant lines: wild type and psy1r receptor mutant. *BMC Genomics* 15:441. doi: 10.1186/1471-2164-15-441

Martiniere, A., Gayral, P., Hawes, C., and Runions, J. (2011). Building bridges: formin1 of *Arabidopsis* forms a connection between the cell wall and the actin cytoskeleton. *Plant J.* 66, 354–365. doi: 10.1111/j.1365-313X.2011.04497.x

Matsubayashi, Y. (2011). Post-translational modifications in secreted peptide hormones in plants. *Plant Cell Physiol.* 52, 5–13. doi: 10.1093/pcp/pcq169

Matsubayashi, Y. (2014). Posttranslationally modified small-peptide signals in plants. *Annu. Rev. Plant Biol.* 65, 385–413. doi: 10.1146/annurev-arplant-050312-120122

Mazau, D., and Esquerretugaye, M. T. (1986). Hydroxyproline-rich glycoprotein accumulation in the cell-walls of plants infected by various pathogens. *Physiol. Mol. Plant Pathol.* 29, 147–157. doi: 10.1016/S0048-4059(86)80017-0

Miyake, M., Terada, T., Shimokawa, M., Sugimoto, N., Arakawa, T., Shimizu, K., et al. (2020). Structural analysis of beta-L-arabinobiose-binding protein in the metabolic pathway of hydroxyproline-rich glycoproteins in Bifidobacterium longum. *FEBS J.* 287, 5114–5129. doi: 10.1111/febs.15315

Mnich, E., Bjarnholt, N., Eudes, A., Harholt, J., Holland, C., Jorgensen, B., et al. (2020). Phenolic cross-links: building and de-constructing the plant cell wall. *Nat. Prod. Rep.* 37, 919–961. doi: 10.1039/C9NP00028C

Møller, S. R., Yi, X., Velásquez, S. M., Gille, S., Hansen, P. L. M., Poulsen, C. P., et al. (2017). Identification and evolution of a plant cell wall specific glycoprotein glycosyl transferase, ExAD. *Sci. Rep.* 7:45341. doi: 10.1038/srep45341

Moussu, S., Broyart, C., Santos-Fernandez, G., Augustin, S., Wehrle, S., Grossniklaus, U., et al. (2020). Structural basis for recognition of RALF peptides by LRX proteins during pollen tube growth. *Proc. Natl. Acad. Sci. U. S. A.* 117, 7494–7503. doi: 10.1073/pnas.2000100117

Munos, S., Ranc, N., Botton, E., Berard, A., Rolland, S., Duffe, P., et al. (2011). Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL. *Plant Physiol.* 156, 2244–2254. doi: 10.1104/pp.111.173997

Nagashima, Y., Von Schaewen, A., and Koiwa, H. (2018). Function of N-glycosylation in plants. *Plant Sci.* 274, 70–79. doi: 10.1016/j.plantsci.2018.05.007

Ni, J., and Clark, S. E. (2006). Evidence for functional conservation, sufficiency, and proteolytic processing of the CLAVATA3 CLE domain. *Plant Physiol.* 140, 726–733. doi: 10.1104/pp.105.072678

Nishida, H., Handa, Y., Tanaka, S., Suzaki, T., and Kawaguchi, M. (2016). Expression of the CLE-RS3 gene suppresses root nodulation in Lotus japonicus. *J. Plant Res.* 129, 909–919. doi: 10.1007/s10265-016-0842-z

Nunez, A., Fishman, M. L., Fortis, L. L., Cooke, P. H., and Hotchkiss, A. T. Jr. (2009). Identification of extensin protein associated with sugar beet pectin. *J. Agric. Food Chem.* 57, 10951–10958. doi: 10.1021/jf902162t

Ogawa-Ohnishi, M., Matsushita, W., and Matsubayashi, Y. (2013). Identification of three hydroxyproline O-arabinosyltransferases in *Arabidopsis thaliana*. *Nat. Chem. Biol.* 9, 726–730. doi: 10.1038/nchembio.1351

Oguri, S., Amano, K., Nakashita, H., Nagata, Y., and Momonoki, Y. S. (2008). Molecular structure and properties of lectin from tomato fruit. *Biosci. Biotechnol. Biochem.* 72, 2640–2650. doi: 10.1271/bbb.80310

Ohyama, K., Shinohara, H., Ogawa-Ohnishi, M., and Matsubayashi, Y. (2009). A glycopeptide regulating stem cell fate in *Arabidopsis thaliana*. *Nat. Chem. Biol.* 5, 578–580. doi: 10.1038/nchembio.182

Okamoto, S., Ohnishi, E., Sato, S., Takahashi, H., Nakazono, M., Tabata, S., et al. (2009). Nod factor/nitrate-induced CLE genes that drive HAR1-mediated systemic regulation of nodulation. *Plant Cell Physiol.* 50, 67–77. doi: 10.1093/pcp/pcn194

Park, S., Szumlanski, A. L., Gu, F., Guo, F., and Nielsen, E. (2011). A role for CSLD3 during cell-wall synthesis in apical plasma membranes of tip-growing root-hair cells. *Nat. Cell Biol.* 13, 973–980. doi: 10.1038/ncb2294

Patel, N., Mohd-Radzman, N. A., Corcilius, L., Crossett, B., Connolly, A., Cordwell, S. J., et al. (2018). Diverse peptide hormones affecting root growth identified in the *Medicago truncatula* secreted peptidome. *Mol. Cell. Proteomics* 17, 160–174. doi: 10.1074/mcp.RA117.000168

Pena, M. J., Kong, Y. Z., York, W. S., and O'neill, M. A. (2012). A galacturonic acid-containing xyloglucan is involved in *Arabidopsis* root hair tip growth. *Plant Cell* 24, 4511–4524. doi: 10.1105/tpc.112.103390

Pennell, R. I., Janniche, L., Kjellbom, P., Scofield, G. N., Peart, J. M., and Roberts, K. (1991). Developmental regulation of a plasma membrane arabinogalactan protein epitope in oilseed rape flowers. *Plant Cell* 3, 1317–1326. doi: 10.2307/3869311

Qanmber, G., Liu, J., Yu, D., Liu, Z., Lu, L., Mo, H., et al. (2019). Genome-wide identification and characterization of the PERK gene family in *Gossypium hirsutum* reveals gene duplication and functional divergence. *Int. J. Mol. Sci.* 20:1750. doi: 10.3390/ijms20071750

Qi, X., Behrens, B. X., West, P. R., and Mort, A. J. (1995). Solubilization and partial characterization of extensin fragments from cell walls of cotton suspension cultures. Evidence for a covalent cross-link between extensin and pectin. *Plant Physiol.* 108, 1691–1701. doi: 10.1104/pp.108.4.1691

Ringli, C. (2010). The hydroxyproline-rich glycoprotein domain of the *Arabidopsis* LRX1 requires Tyr for function but not for insolubilization in the cell wall. *Plant J.* 63, 662–669. doi: 10.1111/j.1365-313X.2010.04270.x

Saito, F., Suyama, A., Oka, T., Yoko, O. T., Matsuoka, K., Jigami, Y., et al. (2014). Identification of novel peptidyl serine alpha-galactosyltransferase gene family in plants. *J. Biol. Chem.* 289, 20405–20420. doi: 10.1074/jbc.M114.553933

Saito, K., Viborg, A. H., Sakamoto, S., Arakawa, T., Yamada, C., Fujita, K., et al. (2020). Crystal structure of beta-L-arabinobiosidase belonging to glycoside hydrolase family 121. *PLoS One* 15:e0231513. doi: 10.1371/journal.pone.0231513

Saqib, A., Scheller, H. V., Fredslund, F., and Welner, D. H. (2019). Molecular characteristics of plant UDP-arabinopyranose mutases. *Glycobiology* 29, 839–846. doi: 10.1093/glycob/cwz067

Schjoldager, K. T., Narimatsu, Y., Joshi, H. J., and Clausen, H. (2020). Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* 21, 729–749. doi: 10.1038/s41580-020-00294-x

Schnabel, E. L., Kassaw, T. K., Smith, L. S., Marsh, J. F., Oldroyd, G. E., Long, S. R., et al. (2011). The ROOT DETERMINED NODULATION1 gene regulates nodule number in roots of Medicago truncatula and defines a highly conserved, uncharacterized plant gene family. *Plant Physiol.* 157, 328–340. doi: 10.1104/pp.111.178756

Schnabelrauch, L. S., Kieliszewski, M., Upham, B. L., Alizedeh, H., and Lamport, D. T. (1996). Isolation of pI 4.6 extensin peroxidase from tomato cell suspension cultures and identification of Val-Tyr-Lys as putative intermolecular cross-link site. *Plant J.* 9, 477–489. doi: 10.1046/j.1365-313X.1996.09040477.x

Schoof, H., Lenhard, M., Haecker, A., Mayer, K. F. X., Jurgens, G., and Laux, T. (2000). The stem cell population of *Arabidopsis* shoot meristems is maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. *Cell* 100, 635–644. doi: 10.1016/S0092-8674(00)80700-X

Seifert, G. J. (2020). On the potential function of type II arabinogalactan O-glycosylation in regulating the fate of plant secretory proteins. *Front. Plant Sci.* 11:563735. doi: 10.3389/fpls.2020.563735

Shenkman, M., and Lederkremer, G. Z. (2019). Compartmentalization and selective tagging for disposal of misfolded glycoproteins. *Trends Biochem. Sci.* 44, 827–836. doi: 10.1016/j.tibs.2019.04.012

Shinohara, H., and Matsubayashi, Y. (2013). Chemical synthesis of *Arabidopsis* CLV3 glycopeptide reveals the impact of hydroxyproline arabinosylation on peptide conformation and activity. *Plant Cell Physiol.* 54, 369–374. doi: 10.1093/pcp/pcs174

Showalter, A. M., Zhou, J., Rumeau, D., Worst, S. G., and Varner, J. E. (1991). Tomato extensin and extensin-like cDNAs: structure and expression in response to wounding. *Plant Mol. Biol.* 16, 547–565. doi: 10.1007/BF00023421

Shpak, E., Leykam, J. F., and Kieliszewski, M. J. (1999). Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes. *Proc. Natl. Acad. Sci. U. S. A.* 96, 14736–14741. doi: 10.1073/pnas.96.26.14736

Stafstrom, J. P., and Staehelin, L. A. (1986). The role of carbohydrate in maintaining extensin in an extended conformation. *Plant Physiol.* 81, 242–246. doi: 10.1104/pp.81.1.242

Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T. B. G., et al. (2013). Precision mapping of the human O-GalNAc glycoproteome through simple cell technology. *EMBO J.* 32, 1478–1488. doi: 10.1038/emboj.2013.79

Stuhrwohldt, N., and Schaller, A. (2019). Regulation of plant peptide hormones and growth factors by post-translational modification. *Plant Biol.* 21(Suppl. 1), 49–63. doi: 10.1111/plb.12881

Talbott, L. D., and Ray, P. M. (1992). Molecular size and separability features of pea cell wall polysaccharides: implications for models of primary wall structure. *Plant Physiol.* 98, 357–368. doi: 10.1104/pp.98.1.357

Tavormina, P., De Coninck, B., Nikonorova, N., De Smet, I., and Cammue, B. P. (2015). The plant peptidome: an expanding repertoire of structural features and biological functions. *Plant Cell* 27, 2095–2118. doi: 10.1105/tpc.15.00440

Tryfona, T., Liang, H.-C., Kotake, T., Tsumuraya, Y., Stephens, E., and Dupree, P. (2012). Structural characterization of *Arabidopsis* leaf arabinogalactan polysaccharides. *Plant Physiol.* 160, 653–666. doi: 10.1104/pp.112.202309

Ulvskov, P., Paiva, D. S., Domozych, D., and Harholt, J. (2013). Classification, naming and evolutionary history of glycosyltransferases from sequenced green and red algal genomes. *PLoS One* 8:e76511. doi: 10.1371/journal.pone.0076511

Van Damme, E. J., Barre, A., Rouge, P., and Peumans, W. J. (2004). Potato lectin: an updated model of a unique chimeric plant protein. *Plant J.* 37, 34–45. doi: 10.1046/j.1365-313X.2003.01929.x

Van Hengel, A. J., and Roberts, K. (2002). Fucosylated arabinogalactan-proteins are required for full root cell elongation in *Arabidopsis*. *Plant J.* 32, 105–113. doi: 10.1046/j.1365-313X.2002.01406.x

Vanholst, G. J., and Varner, J. E. (1984). Reinforced polyproline-II conformation in a hydroxyproline-rich cell-wall glycoprotein from carrot root. *Plant Physiol.* 74, 247–251. doi: 10.1104/pp.74.2.247

Velasquez, S. M., Marzol, E., Borassi, C., Pol-Fachin, L., Ricardi, M. M., Mangano, S., et al. (2015a). Low sugar is not always good: impact of specific O-glycan defects on tip growth in *Arabidopsis*. *Plant Physiol.* 168, 808–813. doi: 10.1104/pp.114.255521

Velasquez, S. M., Ricardi, M. M., Dorosz, J. G., Fernandez, P. V., Nadra, A. D., Pol-Fachin, L., et al. (2011). O-glycosylated cell wall proteins are essential in root hair growth. *Science* 332, 1401–1403. doi: 10.1126/science.1206657

Velasquez, S. M., Ricardi, M. M., Poulsen, C. P., Oikawa, A., Dilokpimol, A., Halim, A., et al. (2015b). Complex regulation of prolyl-4-hydroxylases impacts root hair expansion. *Mol. Plant* 8, 734–746. doi: 10.1016/j.molp.2014.11.017

Wang, T., Chen, Y., Tabuchi, A., Cosgrove, D., and Hong, M. (2016). The target of β-expansin EXPB1 in maize cell walls from binding and solid-state NMR studies. *Plant Physiol.* 172, 2107–2119. doi: 10.1104/pp.16.01311

Xu, C., Liberatore, K. L., MacAlister, C. A., Huang, Z. J., Chu, Y. H., Jiang, K., et al. (2015). A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nat. Genet.* 47, 784–792. doi: 10.1038/ng.3309

Yoro, E., Nishida, H., Ogawa-Ohnishi, M., Yoshida, C., Suzaki, T., Matsubayashi, Y., et al. (2019). PLENTY, a hydroxyproline O-arabinosyltransferase, negatively regulates root nodule symbiosis in Lotus japonicus. *J. Exp. Bot.* 70, 507–517. doi: 10.1093/jxb/ery364

Zhang, N., Dolan, M., Wu, D., Phillips, G. C., and Xu, J. (2016). Dramatic secretion of recombinant protein expressed in tobacco cells with a designer glycopeptide tag is highly impacted by medium composition. *Plant Cell Rep.* 35, 2513–2522. doi: 10.1007/s00299-016-2051-6

Zhang, X., and Wang, Y. (2016). Glycosylation quality control by the Golgi structure. *J. Mol. Biol.* 428, 3183–3193. doi: 10.1016/j.jmb.2016.02.030

Zhang, N. N., Wright, T., Wang, X. T., Karki, U., Savary, B. J., and Xu, J. F. (2019). Engineering "designer" glycomodules for boosting recombinant protein secretion in tobacco hairy root culture and studying hydroxyproline-O-glycosylation process in plants. *Plant Biotechnol. J.* 17, 1130–1141. doi: 10.1111/pbi.13043

Zielinska, D. F., Gnad, F., Wisniewski, J. R., and Mann, M. (2010). Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* 141, 897–907. doi: 10.1016/j.cell.2010.04.012

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.