Three Decades of Advances in Arabinogalactan-Protein Biosynthesis

Jessy Silva1,2, Ricardo Ferraz1,2, Paul Dupree3, Allan M. Showalter4 and Sílvia Coimbra1,2*

1Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal, 2LAQV Requimte, Sustainable Chemistry, Universidade do Porto, Porto, Portugal, 3Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom, 4Department of Environmental and Plant Biology, Molecular and Cellular Biology Program, Ohio University, Athens, OH, United States

Arabinogalactan-proteins (AGPs) are a large, complex, and highly diverse class of heavily glycosylated proteins that belong to the family of cell wall hydroxyproline-rich glycoproteins. Approximately 90% of the molecules consist of arabinogalactan polysaccharides, which are composed of arabinose and galactose as major sugars and minor sugars such as glucuronic acid, fucose, and rhamnose. About half of the AGP family members contain a glycosylphosphatidylinositol (GPI) lipid anchor, which allows for an association with the outer leaflet of the plasma membrane. The mysterious AGP family has captivated the attention of plant biologists for several decades. This diverse family of glycoproteins is widely distributed in the plant kingdom, including many algae, where they play fundamental roles in growth and development processes. The journey of AGP biosynthesis begins with the assembly of amino acids into peptide chains of proteins. An N-terminal signal peptide directs AGPs toward the endoplasmic reticulum, where proline hydroxylation occurs and a GPI anchor may be added. GPI-anchored AGPs, as well as unanchored AGPs, are then transferred to the Golgi apparatus, where extensive glycosylation occurs by the action of a variety glycosyltransferase enzymes. Following glycosylation, AGPs are transported by secretory vesicles to the cell wall or to the extracellular face of the plasma membrane (in the case of GPI-anchored AGPs). GPI-anchored proteins can be released from the plasma membrane into the cell wall by phospholipases. In this review, we present an overview of the accumulated knowledge on AGP biosynthesis over the past three decades. Particular emphasis is placed on the glycosylation of AGPs as the sugar moiety is essential to their function. Recent genetics and genomics approaches have significantly contributed to a broader knowledge of AGP biosynthesis. However, many questions remain to be elucidated in the decades ahead.

Keywords: arabinogalactan-proteins, arabinogalactan-protein biosynthesis, glypiation, proline hydroxylation, glycosylation, glycosyltransferases, hydroxyproline, cell wall

THE PLANT CELL WALL: THE PROTECTOR OF THE REALM

Plant growth and development are essential processes that are regulated and supported by the cell wall (Serpe and Nothnagel, 1999). The plant cell wall is a dynamic and complex structure composed of different components that act together and contribute to cell wall architecture and function (Somerville et al., 2004). The cell wall plays different key roles in plant growth
and development, cell differentiation, environmental sensing and signaling, intercellular communication, water movement, and defense against invading pathogens (Cosgrove, 2005; Jamet et al., 2008; Malinovsky et al., 2014; Bacete et al., 2018). Typically, the primary cell wall is a thin (0.1–1 μm) and flexible layer composed of three major classes of polysaccharides, namely cellulose, hemicelluloses, and pectins, along with cell wall proteins (CWPs; Somerville et al., 2004; Cosgrove, 2005; Atmodjo et al., 2013; Anderson and Kieber, 2020). Some specific cell types also include lignins (Jamet et al., 2006). Approximately 15% of the 27,000 Arabidopsis thaliana genes are involved in cell wall synthesis, modification, and turnover (Carpita et al., 2001; Wang et al., 2012).

Polysaccharides constitute up to 90–95% of the cell wall mass, whereas CWPs account for 5–10% (Cassab and Varner, 1988; Jamet et al., 2008). CWPs are dedicated to cell wall support, structure, signaling, and interactions with other proteins (Jamet et al., 2006). CWPs are classified into nine functional categories in Arabidopsis: (1) proteins acting on carbohydrates, which include glycoside hydrolases (GHs), glycosyltransferases (GTs), carbohydrate esterases, polysaccharide lyases, and expansins; (2) oxido-reductases such as peroxidases, multicomponent oxidases, berberine bridge enzyme, and blue copper binding proteins; (3) proteases like serine carboxypeptidases and aspartic, cysteine, and serine proteases; (4) proteins with interaction domains, for example, lectins, leucine-rich repeat domains, and enzyme inhibitors; (5) proteins possibly involved in signaling, which include arabinogalactan-proteins (AGPs) and receptors; (6) structural proteins like glycine-rich proteins, extensins (EXTs), leucine-rich-repeat extensins, and proline-rich proteins (PRPs); (7) proteins related to lipid metabolism, for instance, glycine, aspartic acid, serine, leucine (GDSL) lipases and lipid transfer proteins; (8) miscellaneous proteins such as purple acid phosphatases, phosphate-induced proteins, and germin; and (9) unknown function proteins with domains of unknown function (DUFs; Showalter, 1993; Jamet et al., 2008; Albenne et al., 2013).

The cell wall hydroxyproline-rich glycoproteins (HRGPs) superfamily is subdivided into three families: lightly glycosylated PRPs, moderately glycosylated EXTs, and highly glycosylated AGPs (Showalter, 1993; Nothnagel, 1997; Showalter et al., 2010; Hijazi et al., 2014; Johnson et al., 2017b). HRGPs are extracellular intrinsically disordered proteins (IDPs) as they are rich in Pro, the most disorder-promoting residue due to its rigid conformation and have motifs that direct posttranslational modifications (Johnson et al., 2017a).

**ARABINOGALACTAN-PROTEINS: ALL YOU NEED IS SUGAR**

AGPs are one of the most complex and diverse families of glycoproteins found in plants. AGPs may have a core-protein backbone rich in Pro/Hyp, Ala, Ser, and Thr (PAST) decorated by a diversity of carbohydrates (Showalter, 1993; Chasan, 1994; Schultz et al., 2002; Showalter et al., 2010). The amino acids of AGPs are often arranged in characteristic dipeptide repeats: Ala-Hyp, Ser-Hyp, Thr-Hyp, introduced as AG glycomodules (Tan et al., 2003; Ellis et al., 2010). The carbohydrate moiety of AGPs represents more than 90% of their total molecular mass, and it is composed mainly of arabinose (Ara) and galactose (Gal), as well as minor sugars such as glucuronic acid (GlcA), fucose (Fuc), rhamnose (Rha), and xylose (Xyl; Clarke et al., 1979; Fincher et al., 1983; Showalter, 1993; Chasan, 1994). AGPs have different molecular weights that reflect different extents of glycosylation of their specific protein cores. The extensive glycosylation of the protein backbone confers resistance to proteolysis (Showalter, 1993). Approximately half of the AGP family members are predicted to be tethered to the plasma membrane by a glycosylphosphatidylinositol (GPI) lipid anchor, being perfect candidates for signal perception and transduction (Youl et al., 1998; Oxley and Bacic, 1999; Sherrier et al., 1999; Svetek et al., 1999; Borner et al., 2002, 2003; Seifert and Roberts, 2007).

Positive reactions with β-Yariv reagent, a chemical reagent that specifically binds to the β-(1→3)-linked D-Galp backbone of AGPs (Yariv et al., 1967; Kitazawa et al., 2013), and immunolocalization studies with monoclonal antibodies that recognize AGP epitopes have shown that AGPs are ubiquitous in the plant kingdom, from bryophytes to angiosperms, and are also present in many algae (Clarke et al., 1979; Fincher et al., 1983; Showalter, 1993; Serpe and Nothnagel, 1999; Lee et al., 2005; Hervé et al., 2016). These glycoproteins were conserved during evolution, presumably because of their vital roles in plants (Serpe and Nothnagel, 1999). AGPs are found in plant cell walls, plasma membranes, apoplastic spaces, secretions, and intracellular multivesicular bodies (Herman and Lamb, 1992; Serpe and Nothnagel, 1999; Majewska-Sawka and Nothnagel, 2000; Ellis et al., 2010; Nguema-Ona et al., 2012).

AGPs are expressed in distinct cells and tissues and at particular stages of development. AGPs are a heterogenous family with their members, individually or collectively, implicated to function in many plant growth and development processes, such as cell proliferation and programed cell death, cell-cell signaling, embryo and postembryonic pattern formation, somatic embryogenesis, female and male gametophyte development, cell wall plasticizers, pollen tube growth and guidance, pollen incompatibility, root growth, xylem differentiation, secondary wall deposition, hormone signaling pathways, plant microbe interactions, and abiotic stress responses (reviewed in Showalter, 2001; Gaspar et al., 2004; Seifert and Roberts, 2007; Ellis et al., 2010; Nguema-Ona et al., 2013; Pereira et al., 2016). Recently, classical AGPs have been proposed to make a three-fold contribution: as a primary source of cytosolic Ca²⁺, as a pectic plasticizer and as Ca²⁺ signposts to the ovule (Lamport et al., 2018).

**FINDING AND CLASSIFYING AGPS: A HISTORY OVERVIEW**

AGPs were initially discovered as polysaccharides isolated from suspension-cultured sycamore (Acer pseudoplatanus L.) cells (Aspinall et al., 1969). Since then, AGPs have fascinated and challenged researchers with their huge diversity of protein...
backbones discovered by genome sequencing. Two decades ago, the term AGP was defined as a group of molecules that presents three criteria (Clarke et al., 1979; Du et al., 1996; Knox, 1999): a core-protein backbone rich in Hyp; type II arabino-3,6-galactan polysaccharides (AGs); and the ability to bind to Yariv reagents (Yariv et al., 1962, 1967). However, since then, investigators have discovered that several AGPs do not fit these criteria (Du et al., 1996), as it is the case of AG-peptides from wheat (Fincher et al., 1974) and two glycoproteins from Nicotiana alata style that do not bind the Yariv reagent (Lind et al., 1994; Sommer-Knudsen et al., 1996). Moreover, other AGPs are histidine-rich (Kieliszewski et al., 1992) or have short oligoarabinosides (Qi et al., 1991).

AGPs cDNA were first isolated, cloned, and sequenced in 1994 from cultured pear (Pyrus communis) cells (Chen et al., 1994) and from N. alata styles (Du et al., 1994). AGPs were first classified based on the amino acid sequence composition as classical AGPs and non-classical AGPs (Mau et al., 1995; Du et al., 1996). Classical AGPs are characterized by the presence of an N-terminal hydrophobic secretion signal sequence, a central domain rich in PAST residues (including sites for Hyp-O-glycosylation) and potentially a hydrophobic C-terminal anchor addition sequence that directs the attachment of a GPI anchor (Du et al., 1996). Non-classical AGPs contain an N-terminal hydrophobic secretion signal sequence followed by one or more PAST-rich regions, which can be Hyp-O-glycosylated, along with other non-PAST rich regions, such as hydrophilic C-terminal Asn-rich domains (Mau et al., 1995; Du et al., 1996).

After the sequencing of the Arabidopsis genome in 2000 by the Arabidopsis Genome Initiative (AGI; Kaul et al., 2000), Borner et al. (2002) identified on the basis of sequence analysis 210 predicted GPI-anchored proteins and over 40% of these proteins had putative AG glycomodules, including 13 classical AGPs, 9 AG peptides, 18 fasciclin-like proteins, 8 phytocyanin-like proteins, 8 early nodulin-like proteins, and 9 lipid transfer protein-like proteins. In the same year, Schultz et al. (2002) conducted the first bioinformatics identification of AGPs based on sequence analysis, the amino acid bias method identified and classified 47 candidate AGP genes in four different classes: classical AGPs (13), Lys-rich AGPs (3), AG peptides (10), and fasciclin-like AGPs (FLAs; 21). The other classes are structurally similar to the classical AGPs but have different lengths and domains in their polypeptide core. AG peptides are short classical AGPs with only 10–13 amino acids (Schultz et al., 2000). Lys-rich AGPs contain a Lys-rich domain of approximately 16 amino acid residues between a Pro-rich domain and the C-terminus (Schultz et al., 2002; Sun et al., 2005). FLAs can have one or two fasciclin-like (FAS1) cell adhesion domains with 110–150 amino acids and one or two AG domains (Gaspar et al., 2001; Schultz et al., 2002; Johnson et al., 2003).

Nowadays, with the improvement of technologies, whole genome sequences of several plant species have been annotated, which allowed bioinformatic identification of AGPs and consequently, an update on AGPs classification. A bioinformatics approach named as finding-AGP based on AG glycosylation has been performed in 47 plant species from Chlorophytta to Eudicot, enabling the identification of thousands of putative AGPs. The number of AGPs varied between 48 in Amborella trichopoda and 313 in Glycine max, with 151 AGPs being identified in Arabidopsis (Ma et al., 2017). According to their polypeptide core and based on the presence/absence of particular motif/domains, AGPs are now classified into classical AGPs, AG peptides, Lys-rich AGPs, FLAs, early nodulin-like AGPs (ENODLs) with plastocyanin-like domains, xylogen-like AGPs (XYLPs) with non-specific lipid transfer protein (nLTP) domains, other chimeric AGPs that do not belong in any of the mentioned groups and HAEs (Borner et al., 2002, 2003; Schultz et al., 2002; Johnson et al., 2003; Mashiguchi et al., 2009; Ma and Zhao, 2010; Showalter et al., 2010; Kobayashi et al., 2011; Ma et al., 2017). Therefore, we propose that AGPs should be defined as a large family of glycoproteins, which can share common features, such as the typical protein domain rich in PAST, the occurrence of Ala-Hyp, Ser-Hyp, and/or Thr-Hyp dipeptide repeats, the occurrence of an N-terminal hydrophobic secretion signal sequence, the presence of type II AGs attached to Hyp residues, the ability to interact with the synthetic chemical dye β-Yariv reagent (Yariv et al., 1962, 1967), which recognizes the β-1,3-Gal main chains of type II AGs attached to Hyp residues, the ability to interact with the synthetic chemical dye β-Yariv reagent (Yariv et al., 1962, 1967), which recognizes the β-1,3-Gal main chains of type II AGs (Kitazawa et al., 2002, 2013), the possibility to be attached to the plasma membrane by a C-terminal GPI anchor and the potential presence of other functional domains.

AGP BIOSYNTHESIS PATHWAY: THE SUGAR FACTORY

The biosynthesis of AGPs comprises the addition and removal of amino acids, lipids, and carbohydrates. As some AGPs are composed of approximately 90% sugar, the AGP biosynthetic pathway resembles a candy factory occurring inside the cells. The production line involves sequential operations and begins with the translation of the N-terminal signal sequence of AGPs on ribosomes, which allows entry into the endoplasmic reticulum (ER) and endomembrane system (Figure 1A). Translation continues of the AGPs with concomitant removal of the signal peptide (Figure 1A) and hydroxylation of Pro residues by prolyl hydroxylase enzymes located in the ER (Walter and Johnson, 1994; Schatz and Dobberstein, 1996; Faye et al., 2005; Figure 1B). The resulting AGP backbone can be further modified in the ER if it contains a GPI anchor addition sequence (Schultz et al., 1998; Yeats et al., 2018; Figure 1C). Subsequently, AGPs, either free in the ER lumen or tethered to the ER membrane by a GPI anchor (Figure 1C), are transported to the Golgi apparatus to allow for the successive addition of various monosaccharide units to the protein backbone (Figure 2A). Several enzymes constitute the required machinery for these
Glypiation: GPI Anchoring

The addition of GPI-anchors, also referred to as glypiation, is a post-translational modification allowing AGPs to be attached to the plasma membrane (Duruflé et al., 2017). In Arabidopsis, 55 of the 85 identified AGPs are predicted to contain a GPI-anchor addition motif (Showalter et al., 2010). GPI-anchors allow attachment of proteins to the cell surface and may increase their lateral mobility in the membrane, exclusion from clathrin-coated pits, targeting to membrane microdomains/lipid rafts, and function in signal transduction pathways (Schultz et al., 1998; Ellis et al., 2010; Desnoyer and Palanivelu, 2020). The GPI-attachment signal (GAS) in the C-terminal region of the protein consists of ~11 polar residues, followed by the ω region of ~4 small residues containing the ω site, a spacer region of ~6 moderately polar residues and a C-terminal hydrophobic region of variable length between 9 and 24 residues (Schultz et al., 1998; Eisenhaber et al., 2003; Ellis et al., 2010; Desnoyer and Palanivelu, 2020). The highly conserved GPI moiety is initially synthesized on the cytosolic surface of the ER via the sequential addition of glucosamine, three α-linked mannosyl (Man) residues, and phosphoethanolamine to phosphatidylinositol (Schultz et al., 1998; Yeats et al., 2018; Beihammer et al., 2020; Desnoyer and Palanivelu, 2020). Although GPI anchor biosynthesis has not been biochemically studied in plants, the proteins that catalyze this pathway have been well studied in mammalian cells, yeast, and protozoa (Schultz et al., 1998; Pittet and Conzelmann, 2007; Morotti et al., 2017; Desnoyer and Palanivelu, 2020; Kinoshita, 2020). Nevertheless, this process may be conserved as orthologs of GPI biosynthetic genes are found in plant genomes (Schultz et al., 1998; Ellis et al., 2010; Desnoyer and Palanivelu, 2020).

In mammals, the first step involves the transfer of β-N-acetylglucosamine (GlcNAc), from the nucleotide sugar UDP-GlcNAc, to phosphatidylinositol (PI) to generate GlcNAc-PI by the GPI-N acetylglucosaminyltransferase (GPI-GnT) complex, which contains seven subunits (PIG-A, PIG-C, PIG-H, PIG-Q, PIG-P, PIG-Y, and DPM2). The product formed is de-N-acetylated subsequently by PIG-L, a GPI deacylase, to yield GlcN-PI. At some point in the process, the synthesis switches from the cytoplasmic to the luminal face of the ER by an unknown flipase enzyme. In the ER lumen, GlcN-PI is acetylated by the acyltransferase PIG-W to generate GlcN-(acyl)PI. Then, three Man residues, donated by dolichol-phospho-mannose (Dol-P-Man), are sequentially added to GlcN-(acyl)PI by the GPI-mannosyltransferases PIG-M, PIG-V, and PIG-B.
The addition of phosphoethanolamine (PEtN) on the third Man residue is catalyzed by PIG-O and PIG-F (Schultz et al., 1998; Ellis et al., 2010; Beihammer et al., 2020; Desnoyer and Palanivelu, 2020; Kinoshita, 2020). The resulting GPI structure is ready to be transferred to proteins but other side chains may be added, such as phosphoethanolamines or sugars may be linked to Man and acyl groups linked to inositol (Luschnig and Seifert, 2010; Beihammer et al., 2020).
The GPI biosynthetic process converges with the cotranslational insertion of the protein backbone into the ER. The GAS is recognized and proteolytically cleaved between the ω and ω + 1 sites by the GPI transamidase (GPI-T) complex and the GPI anchor is attached to the ω site of the mature protein by a transamidation reaction (Schultz et al., 1998; Ellis et al., 2010; Yeats et al., 2018; Beihammer et al., 2020; Desnoyer and Palanivelu, 2020; Figure 1C). The mammal GPI-T consists of five subunits, PIG-K, GPAA1, PIG-S, PIG-T, and PIG-U (Kinoshita, 2020).

Some GPI biosynthesis enzymes in mammals have homologs already characterized in the Arabidopsis genome: SETH1 for PIG-C, SETH2 for PIG-A, PEANUT1 (PNT1) for PIG-M, PEANUT 5 (PNT5) for PIG-W, ABNORMAL POLLEN TUBE GUIDANCE1 (APTG1) for PIG-B, AtGPI8, and AtGPI-S (Lalanne et al., 2004; Gillmor et al., 2005; Dai et al., 2014; Bundy et al., 2016; Liu et al., 2016; Beihammer et al., 2020; Desnoyer et al., 2020; Desnoyer and Palanivelu, 2020). Although several different structures of GPI anchors have been observed in several kingdoms, to date, in plants, the only known structure was determined on PcAGPI isolated from pear (Oxley and Bacic, 1999). This structure contains a glycan moiety conserved in all eukaryotic GPI anchors, D-Manα(1-2)-D-Manα(1-6)-D-Manα(1-4)-D-GlcN-inositol with a specific β(1-4)-galactosyl substitution of the 6-linked Man residue, it is devoid of PEdN side chains and contains a phosphocheramide, composed of phytosphingosine and tetracosanoic acid, instead of the common glycerolipid (Oxley and Bacic, 1999; Ellis et al., 2010; Yeats et al., 2018; Desnoyer and Palanivelu, 2020; Figure 1C). Moreover, an AGP isolated from Rosa sp. cell suspension cultures also included a ceramide, composed of tetracosanoic acid and 4-hydroxysphinganine, as the GPI lipid component (Svetek et al., 1999). These results suggest the GPI glycerolipid moiety of some GPI-anchored proteins is remodeled to contain ceramide, as described in yeast (Bosson and Conzelmann, 2007).

Discovering other plant GPI structures will identify possible forms of GPI and determine the conservation of these structures in plants (Desnoyer and Palanivelu, 2020).

Glycosylation: Becoming Sweeter

AGPs free in the ER and AGPs anchored to the ER membrane are then transferred to the Golgi apparatus, where they undergo glycosylation (Figure 2A). Glycosylation is one of the major post-translational modifications found in almost every living organism. It is performed by GTs and includes N-glycosylation, O-glycosylation, and glypiation (Hurtado-Guerrero and Davies, 2012; Duruflé et al., 2017). Each of these post-translational modifications occurs on specific amino acid sequences (Duruflé et al., 2017). The addition of carbohydrates on a polypeptide backbone may affect the physico-chemical properties of a protein, including resistance to thermal denaturation, protection from proteolytic degradation, solubility, and it can alter essential biological functions (Faye et al., 2005).

N-glycosylation occurs on Asn residues in Asn-X-Ser/Thr-specific sequences, where X can be any amino acid except Pro (Faye et al., 2005; Duruflé et al., 2017). N-glycosylation starts in the ER by the co-translational transfer of Glc4Man9GlcNAc2, an oligosaccharide precursor, onto the amide nitrogen of Asn residues. During the transportation of the glycoprotein along the secretory pathway, the N-glycan undergoes a maturation process that involves the removal and the addition of sugar residues in the ER and the Golgi (Faye et al., 2005; Nguema-Ona et al., 2014). Classical AGPs do not contain this conserved sequence but many chimeric AGPs, including FLAs, contain the consensus sequence for N-glycosylation (Du et al., 1996; Johnson et al., 2003; Ellis et al., 2010).

O-glycosylation is the most complex type of glycosylation and, in plants, it occurs predominantly on Hyp residues and less often on Ser and Thr residues in the Golgi apparatus (Nothnagel, 1997; Faye et al., 2005; Duruflé et al., 2017). The Golgi apparatus, the central organelle in the secretory pathway, is responsible for glycosylation, protein sorting, and secretion; it contains a diverse group of membrane-bound GTs required for synthesis of a variety of linkage types (Nikolovski et al., 2012). O-glycosylation of Hyp is a complex mechanism unique to plants that involves the transfer of a glycan from the donor substrate to the acceptor hydroxyl group of Hyp residues (Faye et al., 2005). Gal can be linked to Ser and Hyp residues, whereas Ara can only be linked to Hyp residues (Duruflé et al., 2017). HRGPs are glycosylated by two types of O-glycosylation on their Hyp residues in the Golgi apparatus: Hyp arabinosylation and Hyp arabinogalactosylation (Kieliszewski, 2001; Nguema-Ona et al., 2014). The Hyp contiguity hypothesis (Kieliszewski and Lamport, 1994) predicts that contiguous Hyp residues are arabinosylated, adding short 4-6 residue long oligoarabinoside chains, as occurs in EXTs, whereas non-contiguous Hyp residues are arabinogalactosylated, adding larger 30–150 residue acidic or neutral AGs polysaccharides, as occurs in AGPs (Shpak et al., 1999, 2001; Kieliszewski, 2001; Faye et al., 2005).

AG glycans have a complex and heterogeneous structure that changes throughout development (Tryfona et al., 2014). AG polysaccharides may affect cell surface protein trafficking and stability, act as chaperoning polysaccharides or Ca2+ chelators and generate signaling molecules (Borner et al., 2003; Lamport and Várnai, 2013; Tryfona et al., 2014). In AGPs, type II AGs are O-glycosidically linked to Hyp residues (Ellis et al., 2010; Figure 2A). The structure of type II AGs is not fully resolved but it seems to consist of a β-(1→3)-linked D-galactopyranosyl (Galp) backbone substituted at O6 positions by side-chains of β-(1→6)-linked D-Galp, which are further decorated with α-(1→3)-l-arabinofuranosyl residues (Araf) and less frequently with other sugars, such as GlcA, Rha, Fuc, and Xyl (Clarke et al., 1979; Gaspar et al., 2001; Showalter, 2001; Tan et al., 2004, 2010; Tryfona et al., 2010, 2012; Figure 3). The β-1,6-galactan side chains often terminate with β-1,6-GlcA or 4-O-Me-GlcA, giving the molecules an overall negative charge (Ellis et al., 2010; Tryfona et al., 2012; Temple et al., 2019). Two other types of AG polysaccharides have been described in plants (Clarke et al., 1979; Hijazi et al., 2014). Type I AGs are formed by a linear chain of β-(1→4)-linked D-Galp with lateral chains of α-Araf and β-(1→4)-linked D-Galp (Clarke et al., 1979; Hijazi et al., 2014). Leonard et al. (2005) defined the Artemisia vulgaris
Art v 1 polysaccharide, formed by a short linear $\beta$-(1→6)-linked D-Galp chain, which contains Gal and Ara residues and large branched Ara chains, as a new AG type, the type III AG. However, this structure is inconsistent with the finding that it binds to $\beta$-Yariv, as chains of $\beta$-(1→3)-linked D-Galp are required for $\beta$-Yariv binding (Kitazawa et al., 2013). Therefore, this structure may need to be revised.

Three models have been proposed for the molecular structure of AGPs: the wattle-blossom, the twisted hairy rope, and the necklace (Fincher et al., 1983; Qi et al., 1991; Du et al., 1996; Lamport et al., 2014). The wattle blossom model predicts globular units of polysaccharide chains anchored to a protein core of a spheroidal molecule, and the twisted hairy rope model foresees an alignment of AG chains along the protein core.
backbone, whereas the recent necklace model compares AGP structure to an ancient gold necklace from Afghanistan with pendant glycomodules. Detailed analysis of different AGPs will determine the number, length, and sequence of polysaccharide chains, allowing the improvement of the existing models.

AGPs may be connected to other cell wall components such as pectins and hemicelluloses (Tan et al., 2013). Keegstra et al. (1973) hypothesized that type II AGs could be linked to rhamnogalacturonan (RGI) by the Rha residues. This linkage has been demonstrated in the complex arabinoxylan pectin arabinogalactan protein 1 (APAP1). This complex also has a link between arabinoxylan and an arabinose residue in the type II AGs (Tan et al., 2013). A FLA, SALT-OVERLY SENSITIVE 5 (SOS5) was proposed to mediate seed coat mucilage adherence by interacting with pectins (Griffiths et al., 2014). These results support the hypothesis that AGPs may serve as cross-linkers in the cell wall and act as polysaccharide plasticizers (Lampport, 2001; Lamport et al., 2006). The binding of the Yariv reagent specifically to AGs supports the view that AGs bind specific glycans (Kitazawa et al., 2013).

Glycosyltransferases: Meet the Sugar Workers

Glycoproteins have a complex structure and, therefore, plants require the action of numerous GTs families to assist in their biosynthesis. AGP glycosylation is catalyzed by a large number of GTs (EC 2.4.x.y) in the secretory pathway. GTs are enzymes that catalyze glycosidic bond formation between a sugar moiety and a specific acceptor molecule (sugars, proteins, lipids, or small molecules), creating a diverse collection of oligosaccharides and glycoconjugates in nature (Hansen et al., 2012; Gloster, 2014). GTs have been classified into 111 families in the CAZy database (Coutinho et al., 2003; Lombard et al., 2014). GTs are very specific and, thus, each different linkage may require a distinct GT. With AGP glycosylation, several GTs work together to regulate the density, length, and sequences of AG chains (Qu et al., 2008). Almost 500 putative GT families have been identified in the Arabidopsis genome and classified into 42 different CAZy families (Coutinho et al., 2003). In order to understand the structure of AGPs, the discovery of GTs involved in the initiation and elongation of AG chains is a priority. GTs can be type I membrane proteins located in the ER, type II membrane proteins located in the Golgi or integral membrane proteins such as cellulose synthases (Hansen et al., 2009). GTs generally are localized in the Golgi and have a type II membrane protein topology with a short N-terminal domain facing the cytosol, one helical transmembrane domain, and a hydrophilic C-terminal catalytic domain containing the active site attached to a flexible stem region facing the luminal side (Perrin et al., 2001; Hansen et al., 2009; Chou et al., 2012).

The biosynthesis of AGP glycans requires at least 10 functionally distinct enzymes, e.g., galactosyltransferases (GALTs), arabinosyltransferases, fucosyltransferases (FUTs), rhamnosyltransferases, xylosyltransferases, glucuronosyltransferases (GLCATs), and glucuronic acid methyltransferases. To date, 22 transferases responsible for AGP glycosylation have been identified in Arabidopsis (Tables 1 and 2): eight hydroxyproline-O-β-GALTs (HyP-O-GALTs; AtGALT2, AtGALT3, AtGALT4, AtGALT5, and AtGALT6, Basu et al., 2013, 2015a,b; Showalter and Basu, 2016; AtHPGT1, AtHPGT2, and AtHPGT3, Ogawa-Ohnishi and Matsubayashi, 2015), two β-1,3-GALTs (Atg77810, Qu et al., 2008 and AtKSN4, Suzuki et al., 2017), two β-1,6-GALTs (AtGALT31A, Geshi et al., 2013 and AtGALT29A; Dilokpimol et al., 2014), five β-1,6-GALTs (AtGLCAT14A, AtGLCAT14B, AtGLCAT14C, AtGLCAT14D, and AtGLCAT14E, Knoch et al., 2013; Dilokpimol et al., 2014; Lopez-Hernandez et al., 2020; Zhang et al., 2020), two α-1,2-FUTs (AtFUT4 and AtFUT6, Wu et al., 2010; Liang et al., 2013; Tryfona et al., 2014), a putative β-arabinosyltransferase (ATRAY1, Gille et al., 2013), and two glucuronic acid methyltransferases (AtAGM1 and AtAGM2, Temple et al., 2019; Figure 3). However, several enzymes remain to be identified, including α-arabinofuranosyltransferases, β-arabinopyranosyltransferases, α-rhamnosyltransferases, α-xylosyltransferases, an α-GALT, and other β-GALTs, β-GLCATs, α-FUTs, and glucuronic acid methyltransferases.

Galactosyltransferases: The First Ones to Arrive

AGP glycosylation is initiated by the action of HyP-O-GALTs that add the first Gal onto the hydroxyl group of HyP residues in the protein backbone (Ogawa-Ohnishi and Matsubayashi, 2015; Showalter and Basu, 2016). This process allows the subsequent addition of different sugars by other GTs. The eight GALTs specific for HyP identified in Arabidopsis belong to the CAZy GT31 family. AtGALT2, AtGALT3, AtGALT4, AtGALT5, and AtGALT6 encode a GALT domain as well as a GALECTIN domain, whereas AtHPGT1, AtHPGT2, and AtHPGT3 lack a GALECTIN domain (Showalter and Basu, 2016). Their activity was demonstrated by heterologous expression in Pichia pastoris (AtGALT2, AtGALT5; Basu et al., 2013, 2015b), in Nicotiana tabacum leaf epidermal cells (AtGALT2, AtGALT3, AtGALT4, AtGALT5, and AtGALT6; Basu et al., 2015a) and BY-2 cells (AtHPGT1, AtHPGT2, and AtHPGT3; Ogawa-Ohnishi and Matsubayashi, 2015). galt2, galt3, galt4, galt5, galt5, galt6, galt2galt5, and hpgt1hpgt2hpgt3 mutants demonstrated lower GALT activities and reduced β-Yariv-precipitable AGPs compared to wild-type plants. AtGLAT3-6 and AtHPGT1 transiently expressed in N. tabacum and Arabidopsis T87 protoplasts, respectively, localized to the Golgi, whereas AtGALT2 was found in both ER and Golgi when expressed in tobacco (Basu et al., 2013, 2015a,b; Ogawa-Ohnishi and Matsubayashi, 2015). These results may indicate that AGP glycosylation may start in the ER, but predominantly occurs in the Golgi (Basu et al., 2013, 2016). Nevertheless, further studies are required to explore this hypothesis.

Some physiological phenotypes were revealed under normal growth conditions in the GALT mutants: galt2, galt3, galt4, galt5, and galt2galt5 displayed reduced root hair length and density, galt4 and galt6 showed reduced seed production, galt3, galt6, and galt2galt5 presented reduced seed coat mucilage

1http://www.cazy.org
and galt6 revealed accelerated leaf senescence. Additionally, all GALT mutants (galt2, galt3, galt4, galt5, galt6, and galt2galt5) roots and pollen tubes exhibited less sensitivity to β-Yariv reagent, and root growth and root tip swelling were impaired under salt stress and in elevated levels of sucrose. galt2galt5 displayed a large number of rosette leaves, delayed flowering time, reduced silique length, and plant height (Basu et al., 2015a,b). The galt2galt5 double mutant phenocopies the root swelling, the reduced seed coat mucilage, and the reduced cellulose phenotypes of sos5 and fei1fei2, a double mutant of two cell wall-associated leucine-rich repeat receptor-like kinases, and sos5fei1fei2, indicating that the carbohydrate moiety of SOS5 is important for signaling in the cell (Basu et al., 2015b). The analysis of the quintuple mutant galt2galt5sos5fei1fei2 showed that these five genes act in a single and linear genetic pathway, and it was hypothesized that SOS5 glycosylation by GALT2 and GALT5 was required for its function in the SOS5/FEI1-FEI2 signaling pathway (Basu et al., 2016). In addition, HPGT mutants exhibited several pleiotropic phenotypes such as longer lateral roots, increased root hair length, and density, shorter inflorescence stems, shorter siliques (hpgt2hpgt3 and hpgt1hpgt2hpgt3), thicker primary roots, smaller rosette leaves, shorter petioles, and reduced fertility in the lower portion of the inflorescence (hpgt1hpgt2hpgt3; Ogawa-Ohnishi and Matsubayashi, 2015). Another GT31 member, AtGALT31A, is involved in elongation of β-1,6-galactan side chains. Its activity was demonstrated by heterologous expression in Escherichia coli and N. benthamiana. AtGALT31A accumulates in the Golgi apparatus and unidentified organelles in N. benthamiana. A mutation in this gene showed an abnormal asymmetric cell division in the hypocotyl causing the arrest of embryo development at the globular stage. This phenotype reveals the importance of AG glycans in embryo development (Geshi et al., 2013). AtGALT29A co-expresses with AtGALT31A and AtGLCAT14A. AtGALT29A resides in the GT29 family and this enzyme recombinantly expressed in N. benthamiana possesses β-1,6-GAL elongation and branch initiation activities. Dilokpimol et al. (2014) showed that AtGALT29A was localized in the Golgi in N. benthamiana, where it interacted with AtGALT31A, as indicated by Förster resonance energy transfer. AtGALT31A was also targeted in tobacco to uncharacterized small compartments, which are not part of the trans-Golgi network, cis-Golgi network, or endosomes and that colocalized with EXO70E2, a marker for exocyst-positive

| CAZy GT family | Enzyme | Locus | Enzyme Activity | Subcellular localization | References |
|----------------|--------|-------|-----------------|-------------------------|------------|
| GT14           | GLCAT14A | At5g39990 | β-1,6-GLCAT     | Golgi                   | Knoch et al., 2013; Dilokpimol and Geshi, 2014; Lopez-Hernandez et al., 2020; Zhang et al., 2020 |
|                | GLCAT14B | At5g15050 | Unique subcellular compartments | -                       | -          |
|                | GLCAT14C | At2g37585 | -               | -                       | -          |
|                | GLCAT14D | At3g24040 | -               | -                       | -          |
|                | GLCAT14E | At3g15350 | -               | -                       | -          |
| GT29           | GALT29A | At1g08280 | β-1,6-GALT      |戈尔吉                   | Dilokpimol et al., 2014 |
| GT31           | GALT2   | At4g21060 | Hyp-O-β-GALT    | ER and Golgi             | Basu et al., 2013, 2015a,b |
|                | GALT3   | At3g06440 | -               | -                       | -          |
|                | GALT4   | At1g27120 | -               | -                       | -          |
|                | GALT5   | At1g74800 | -               | -                       | -          |
|                | GALT6   | At5g62620 | -               | -                       | -          |
|                | HPGT1   | At5g3340  | -               | -                       | -          |
|                | HPGT2   | At4g2120  | -               | -                       | -          |
|                | HPGT3   | At2g25300 | -               | -                       | -          |
|                | -       | At1g77810 | β-1,3-GALT      | Golgi                   | Qu et al., 2008 |
|                | KNS4    | At1g33430 | -               | -                       | Suzuki et al., 2017 |
|                | GALT31A | At1g32930 | β-1,6-GALT      | Golgi                   | Geshi et al., 2013 |
| GT37           | FUT4    | At2g15390 | α-1,2-FUT       | -                       | Wu et al., 2010; Liang et al., 2013 |
|                | FUT6    | At1g14080 | -               | -                       | Tryfona et al., 2014 |
| GT77           | RAY1    | At1g0630  | β-arabinofuranosyltransferase | Golgi         | Gille et al., 2013 |
|                | AQM1    | At1g27900 | GlcA O-methyltransferase | Golgi         | Temple et al., 2019 |
|                | AQM2    | At1g67330 | -               | -                       | -          |

*TABLE 1* | Information about the characterized transferases involved in the biosynthesis of AGP glycans in Arabidopsis thaliana.
organelles that mediate an unconventional protein secretory pathway in plants (Poulsen et al., 2014). Moreover, the enzyme complex containing AtGALT31A and AtGALT29A exhibited enhanced β-1,6-GALT activity when compared to AtGALT29 alone.

**Glucuronosyltransferases: Adding the Special Sugar**

AtGLCAT14A, AtGLCAT14B, AtGLCAT14C, AtGLCAT14D, and AtGLCAT14E are glucuronosyltransferases that belong to the GT14 family and add glucuronic acid to β-1,6- and β-1,3-galactose chains of AGPs. The activity of AtGLCAT14A, AtGLCAT14B, and AtGLCAT14C enzymes was confirmed by recombinant expression in P. pastoris and in vitro enzyme assays (Knoch et al., 2013; Dilokpimol and Geshi, 2014). AtGLCAT14A was transiently expressed in N. benthamiana and localized in the Golgi apparatus. AtGLCAT14A is co-localized with AtGALT31A in the Golgi and in

**TABLE 2 | Information about the mutants of the characterized transferases involved in the biosynthesis of AGP glycans in Arabidopsis thaliana.**

| Genes | Mutants | Mutation Localization | Mutant Phenotypes | References |
|-------|---------|------------------------|-------------------|------------|
| GLCAT14A | galt14a-1 (SALK_064313) | Exon | Reduced of GlcA substitution on AGPs (galt14a, galt14a-2, galt14b, galt14c, galt14d) | Knoch et al., 2013; Dilokpimol and Geshi, 2014; Lopez-Hernandez et al., 2020; Zhang et al., 2020 |
| GLCAT14B | galt14a-2 (SALK_043905) | Exon | Reduced of GlcA substitution on AGPs (galt14a-1, galt14a-2, galt14a-1, galt14aglc14b, galt14aglc14c, galt14aglc14d, galt14aglc14e) | Knoch et al., 2013; Dilokpimol and Geshi, 2014; Lopez-Hernandez et al., 2020; Zhang et al., 2020 |
| GLCAT14C | galt14a-4 (CRISPR-Cas9) | Exon | Reduced of GlcA substitution on AGPs (galt14a, galt14b, galt14c, galt14d, galt14e) | Knoch et al., 2013; Dilokpimol and Geshi, 2014; Lopez-Hernandez et al., 2020; Zhang et al., 2020 |
| GLCAT14D | galt14d-1 (GK363F05.01) | Exon | Reduced of GlcA substitution on AGPs (galt14a, galt14b, galt14c, galt14d, galt14e) | Knoch et al., 2013; Dilokpimol and Geshi, 2014; Lopez-Hernandez et al., 2020; Zhang et al., 2020 |

Dilokpimol et al., 2014

Geshi et al., 2013

Matsubayashi, 2015

Ogawa-Ohnishi and Matsubayashi, 2015

Tryfona et al., 2014

Qu et al., 2008

Suzuki et al., 2017

Geshi et al., 2013

Wu et al., 2010; Liang et al., 2013; Tryfona et al., 2014

Gille et al., 2013

Temple et al., 2019

Frontiers in Plant Science | www.frontiersin.org
December 2020 | Volume 11 | Article 610377
11
uncharacterized small compartments, but the two enzymes do not interact (Knoch et al., 2013).

AtGLCAT14 mutants have defective synthesis of AGs. gcat14a, gcat14b, gcat14d, gcat14e, gcat14agcat14b, gcat14bgcat14c, gcat14agcat14bgcat14c, gcat14agcat14bgcat14d, and gcat14agcat14bgcat14e mutants have reduced the content of GlcA on AGPs when compared to wild type (Knoch et al., 2013; Lopez-Hernandez et al., 2020; Zhang et al., 2020). Although this suggests some redundancy between these enzymes, the mutants show some preferential changes in GlcA on specific branch lengths of AG and, therefore, may have roles glucoronidating different parts of the AG or different AGPs (Lopez-Hernandez et al., 2020). An increase of Gal and Ara was detected in AG extracts of gcat14c, gcat14agcat14b, gcat14bgcat14c, and gcat14agcat14bgcat14c (Zhang et al., 2020). However, in gcat14a, an increase of Gal and a reduction of Ara were detected. The increase of galactosylation may result from the increase of O6 acceptor sites, which are shared by AtGLCAT14A and GALTs and, thus, the addition of GlcA may terminate Gal chain extension (Knoch et al., 2013). Indeed, the amount of glycosylated AGPs was increased in the mutants gcat14c, gcat14agcat14b, and gcat14agcat14bgcat14c, which is a result of the loss of GlcA residues, allowing for the elongation of the branched β-(1,3)- and β-(1,6)-galactans. On the other hand, Lopez-Hernandez et al. (2020) did not find evidence for an increase in galactan side chain length in the gcat14 mutants.

AtGLCAT14 mutants present several galactan phenotypes. gcat14a knockout mutants showed enhanced cell elongation during seedling growth (Knoch et al., 2013), gcat14b, gcat14c, gcat14agcat14b, gcat14bgcat14c, and gcat14agcat14bgcat14c presented reduced seed coat mucilage, gcat14agcat14b and gcat14agcat14bgcat14c exhibited delayed seed germination, reduced root hair length, gcat14agcat14b, gcat14agcat14bgcat14c, and gcat14agcat14bgcat14d reduced trichome branching, gcat14agcat14b and gcat14agcat14bgcat14d presented shorter inflorescences, gcat14bgcat14c and gcat14agcat14bgcat14c presented reduced silique length and seed set, and gcat14agcat14b and gcat14agcat14bgcat14c displayed a significant percentage of small and defective pollen that failed to germinate (Lopez-Hernandez et al., 2020; Zhang et al., 2020). gcat14agcat14bgcat14e triple mutant plants had severely limited seedling growth and were sterile (Lopez-Hernandez et al., 2020).

In addition, gcat14b, gcat14c, gcat14agcat14b, gcat14bgcat14c, gcat14agcat14bgcat14c, and gcat14agcat14bgcat14d showed a reduction in Ca\(^{2+}\) binding in AGP extracts compared to wild type (Lopez-Hernandez et al., 2020; Zhang et al., 2020), consistent with the model proposed by Lamport and Várnai (2013) in which AGPs can bind and store Ca\(^{2+}\) through GlcA in a reversible and pH-dependent way at the plasma membrane. It was demonstrated that in vitro AGPs could hold Ca\(^{2+}\) in a pH range of 4–5 and as the pH was lowered Ca\(^{2+}\) was released. AGPs fully released Ca\(^{2+}\) at pH 2.5 (Lamport and Várnai, 2013). This bound Ca\(^{2+}\) may be important for intracellular signaling. Indeed, many of the plant developmental phenotypes in gcat14 mutants can be suppressed by raising the Ca\(^{2+}\) concentration in the growth medium (Lopez-Hernandez et al., 2020).

Intracellular Ca\(^{2+}\) signals were disrupted in the gcat14agcat14bgcat14e mutant plants, which showed altered movement of the Ca\(^{2+}\) signal through the roots. Thus, GLCATs may play an important role in Ca\(^{2+}\) signaling as they determine the presence of GlcA on AGPs (Dilokpimol and Geshi, 2014).

**Gluconic Acid Methyltransferases: Changing GlcA**

In type II AGs from *Arabidopsis*, most of the GlcA substituted is 4-O-methylglucuronosyl (4-O-Me-GlcA; Tryfona et al., 2012). Recently, two GlcA-specific methyltransferases have been identified in *Arabidopsis* (Temple et al., 2019). Arabinogalactan methyltransferases I (AGM1) and 2 (AGM2), two family members of DUF579 family, have GlcA-O-methylation activity on AGPs. AGM1 fused to GFP localized in the Golgi apparatus when transiently expressed in tobacco leaves. An agm1 mutant showed reduced methylation of GlcA on root AGPs and in the double mutant *agm1agm2* there was no AG GlcA methylation of root AGPs. AGM1 and AGM2 are involved in methylation of GlcA of AG in root AGPs. An *agm1agm2* double mutant did not exhibit a growth or fertility phenotype, showing that GlcA methylation is not essential for viability. The biological role of GlcA methylation modification on AGPs is still unclear (Temple et al., 2019). However, it has been shown that the methyl group on GlcA is essential for the effectiveness of a signaling molecule in pollen tube guidance (Mizukami et al., 2016). In addition, 4-O-methylation of GlcA may change the Ca\(^{2+}\) binding affinity to GlcA thus modulating the calcium release response to pH (Lamport and Várnai, 2013). Moreover, the addition of the methyl group to GlcA prevents the addition of 4-linked sugars, such as Rha, and the extension of 4-linked side chains to the GlcA of AG as seen in APAPI (Tan et al., 2013; Temple et al., 2019).

**Fucosyltransferases: Continuing the Hard Work**

AtFUT4 and AtFUT6, two AGP-α-(1,2)-FUT of AGPs, belong to the plant GT37 CAZY family. Their enzymatic activities were demonstrated by heterologous expression in *N. tabacum* BY2 cells (Wu et al., 2010). Tobacco BY2 cells contain nonfucosylated AGPs but heterologous expression of these genes resulted in fucosylated AGPs (Wu et al., 2010). Knock-out mutants (*fut4*, *fut6*, and *fut4fut6*) grown under salt stress showed reduced root growth (Liang et al., 2013; Tryfona et al., 2014). This conditional phenotype reveals the importance of this sugar in root growth and salt sensitivity. Fuc was absent in *fut4* leaf AGPs, whereas the *fut4fut6* double mutant lacks fucose in both leaf and root AGPs (Liang et al., 2013; Tryfona et al., 2014). As *AtFUT4* is expressed in roots and leaves while *AtFUT6* is expressed mostly in roots, it is likely that *AtFUT4* and *AtFUT6* are both responsible for AG fucosylation in roots while *AtFUT4* is also responsible for fucosylation in leaves (Wu et al., 2010; Liang et al., 2013; Tryfona et al., 2014). The *fut4* and *fut6* single mutants had reduced Fuc content in root AGPs. The *fut6* mutant was not stained by eel lectin that binds specifically to terminal α-L-Fuc, indicating that *AtFUT6* likely adds terminal Fuc residues to AG polysaccharides (Liang et al., 2013). *AtFUT6-GFP* was
transiently expressed in tobacco leaves and localized to the Golgi apparatus (Wu et al., 2010).

Arabinosyltransferases: The Ghost Workers
Finally, ray1 mutants showed a reduced level of Ara in their AGPs in etiolated seedlings, roots, and rosette leaves of Arabidopsis, leading to its name REDUCED ARABINOSE YARIV1. Knockout mutants also exhibit reduced root growth, reduced rosette diameter, and delayed development of the inflorescence. AtRAY1 belongs to the GT77 family, and its heterologous expression in N. benthamiana demonstrated β-arabinofuranosyltransferase activity (Gille et al., 2013). However, only α-1,3-linked Ara, and not β-1,3-linked Ara, has been reported in AGPs so it remains unclear whether and how RAY1 functions in the biosynthesis of AGPs glycans (Showalter and Basu, 2016). There are no published candidates for the α-arabinofuranosyltransferases that transfer the main arabinose decoration on AGs.

Glycoside Hydrolases: Cutting All the Sugar
GHs (EC 3.2.1.x) hydrolyze the glycosidic bond between two sugars and are likely important for the metabolism of AGPs (Henrissat and Davies, 1997; Knoch et al., 2014). GHs are classified into 167 families in the CAZy database (Coutinho et al., 2003; Lombard et al., 2014). Hydrolysis of AGP glycans requires several GHs, such as β-galactosidases, β-galactanases, α-arabinofuranosidases, β-arabinopyranosidases, β-glucuronidases, α-fucosidases, α-rhamnosidases, and β-xyllosidases. These enzymes can be a useful tool to analyze AG sugars decorating AGPs (Knoch et al., 2014).

So far, five GHs of AGPs have been reported in plants (Figure 3). RsBGAL1, a β-galactosidase from Raphanus sativus, resides in the GH35 family and the protein when heterologously expressed in P. pastoris hydrolyzed β-(1,3)- and β-(1,6)-galactosyl residues (Kotake et al., 2005). RsAraf1 from R. sativus immature seeds encodes an α-arabinofuranosidase in the GH3 family. When expressed in Arabidopsis, RsAraf1 hydrolyzed α-arabinofuranosyl residues of AGPs (Kotake et al., 2006). AtGUS2 is a β-glucuronidase that belongs to the GH79 family. Knockout mutants showed increased levels of GCA while plants over-expressing AtGUS2 showed the opposite phenotype, no GCA in seedlings AGP fractions and increased cell elongation in seedlings, a similar phenotype of atgcat14a (Eudes et al., 2008). SlTBG1, a tomato (Solanum lycopersicum) β-galactosidase belonging to the GH35 family, has activity against β-(1,3)-linkages and β-(1,6)-linkages in a galactooligosaccharide, which may be involved in the hydrolysis of AGs from AGPs (Eda et al., 2014). AtAPSE (β-1-ARAPASE) is a β-l-arabinopyranosidase in the GH27 family. In an apse mutant, the amount of β-1-Arap residues of AGPs was higher compared to wild type, confirming its activity (Imaizumi et al., 2017).

AGPs in the Extracellular Space: Facing the Challenge
O-glycosylated GPI-anchored AGPs are transferred to the extracellular space, or more precisely to the cell surface apoplast, through the endomembrane system via post-Golgi vesicles, where they remain temporarily attached to the outer face of the plasma membrane (Schultz et al., 1998; Showalter, 2001). In response to intercellular or intracellular signals, GPI-anchored AGPs can be released from the plasma membrane by phosphatidylinositol-specific phospholipases that removes inositol from the diacylglycerol moiety (Schultz et al., 1998; Oxley and Bacic, 1999; Gillmor et al., 2005; Desnoyer and Palanivelu, 2020; Figure 2B). Up to the present, only one inositol phospholipid-specific phospholipase C (PI-PLC) has been identified, partially purified and characterized in plants (Butikofer and Brodbeck, 1993). Several GPI-anchored proteins such as FLAs, COBRA-like, and receptor-like proteins are released by phospholipases in Arabidopsis (Borner et al., 2003). This process may serve to regulate the amount of these glycoproteins in the plasma membrane and control the release of soluble AGPs into the extracellular matrix, where they may act as soluble signals for neighboring cells (Schultz et al., 1998). Nevertheless, some AGP members are not GPI-anchored. The sugars may be cleaved by GHs and released into the extracellular medium, where they may function as signaling molecules binding to specific receptors and triggering signaling pathways (Showalter, 2001; Figure 2B). Furthermore, AGPs may also be processed by proteolytic activities (Fincher et al., 1983; Faye et al., 2005) or transported by endocytotic multivesicular bodies to the vacuole, where they are degraded (Herman and Lamb, 1992; Samaj et al., 2000).

AGP RESEARCH OVERVIEW AND FUTURE DIRECTIONS
In this review, information on AGP biosynthesis accumulated in the last three decades was presented. Much has been accomplished, but there is even more to be done. With the sequencing of several plant genomes, bioinformatics opened doors to identify several candidate genes involved in AGP biosynthesis. However, a huge amount of data is generated and much of it is waiting for further analysis. A major challenge will be to conduct the more laborious biochemical analyses to test and support the findings from bioinformatic and genomic analyses, which moves at such a rapid pace.

This review has largely focused on AGP glycosylation given the cumulative supporting information of the importance of the 90% sugar moiety of AGPs. There is no doubt that AGPs play a range of important functions in plants. The new challenge is to define how AGPs, after all the processes involved with their biosynthesis, act in the cells. Specifically, it will be important to elucidate the mechanism of action by which AGPs, and specifically their sugar epitopes, function. This will mean identifying the molecules, which interact with these sugar residues.

Several other important and stimulating questions remain to be answered: do the enzymes involved in AGP synthesis act on a specific single AGP, a subset of AGPs or all AGPs? How many GTs are involved in AGP glycosylation and what
is the substrate specificity for each GT? Are the GTs involved in AGP biosynthesis associated with one another in enzyme complexes? The structure of AG is very variable, but is the pattern of glycosylation of AGPs cell/tissue-specific, stage-specific, or AGP-specific? What is the precise glycan structure of each AGP? Are the different patterns of glycosylation determined by the diverse amino acid sequences of AGPs? What implications do the different sugar compositions have in AGP function? Do AGPs follow a conventional or unconventional protein secretory pathway? Which AGPs localize in the plasma membrane? How does AGP turnover occur? These major challenges will be a stimulus for new future research developments on AGPs through the next 30 years.

REFERENCES

Albenne, C., Canut, H., and Jamet, E. (2013). Plant cell wall proteomics: the leadership of Arabidopsis thaliana. Front. Plant Sci. 4:111. doi: 10.3389/fpls.2013.00111

Anderson, C. T., and Kieber, J. I. (2020). Dynamic construction, perception, and remodeling of plant cell walls. Annu. Rev. Plant Biol. 71, 39–69. doi: 10.1146/annurev-arplant-081519-053846

Asif, M. H., Trivedi, P. K., Misra, P., and Nath, P. (2009). Prolyl-4-hydroxylase (AIP4HH) mediates and mimics low oxygen response in Arabidopsis thaliana. Funct. Integr. Genomics 9, 525–535. doi: 10.1007/s10142-009-0118-y

Aspinall, G. O., Molloy, J. A., and Craig, J. W. T. (1969). Extracellular polysaccharides of fungal cell walls. J. Gen. Microbiol. 57, 39–69. doi: 10.1128/JCM.47.1.1063–1070.2016

Atmodjo, M. A., Hao, Z., and Mohnen, D. (2013). Evolving views of pectin biochemistry. Annu. Rev. Plant Biol. 64, 747–779. doi: 10.1146/annurev-arplant-042811-105534

Bacete, L., Mélida, H., Miedes, E., and Molina, A. (2018). Plant cell wall-mediated immunity: cell wall changes trigger disease resistance responses. Plant J. 93, 614–636. doi: 10.1111/tpj.13807

Basu, D., Liang, Y., Liu, X., Himmelreich, K., Faik, A., Kieliszewski, M., et al. (2013). Functional identification of a hydroxyproline-O-galactosyltransferase specific for arabinogalactan protein biosynthesis in Arabidopsis. J. Biol. Chem. 288, 10132–10143. doi: 10.1074/jbc.M112.432609

Basu, D., Tian, L., DeBrosse, T., Poirier, E., Emch, K., Herock, H., et al. (2016). Glycosylation of a fasciclin-like arabinogalactan-protein (SOS9) mediates root growth and seed mucilage adherence via a cell wall receptor-like kinase (FEII1/FEI2) pathway in Arabidopsis. PLoS One 11:e0145092. doi: 10.1371/journal.pone.0145092

Basu, D., Tian, L., Wang, W., Bobbs, S., Herock, H., Travers, A., et al. (2015a). A small multigene hydroxyproline-O-galactosyltransferase family functions in arabinogalactan-protein glycosylation, growth and development in Arabidopsis. BMC Plant Biol. 15:295. doi: 10.1186/s12870-015-0670-7

Bundy, M. G. R., Kosentka, P. Z., Willet, A. H., Zhang, L., Miller, E., and Shpak, E. D. (2016). A mutation in the catalytic subunit of the glycosylphosphatidylinositol transamidase disrupts growth, fertility, and stomata formation. Plant Physiol. 171, 974–985. doi: 10.1104/pp.16.00339

Butikofer, P., and Brodbeck, U. (1993). Partial purification and characterization of a (glycosyl) inositol phospholipid-specific phospholipase C from peanut. J. Biol. Chem. 268, 17794–177802.

Canut, H., Albenne, C., and Jamet, E. (2016). Post-translational modifications of plant cell wall proteins and peptides: a survey from a proteomics point of view. Biochim. Biophys. Acta 1864, 983–990. doi: 10.1016/j.bbapap.2016.02.022

Cosgrove, D. J. (2005). Growth of the plant cell wall. Annu. Rev. Plant Biol. 56, 1063–1070. doi: 10.1146/annurev.arplant.56.041604.172317

Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003). An evolving view of the structure and function of plant cell wall G/c-1,3-GlcNAc were the glycosyltransferases that have been identified in Arabidopsis. Phytochemistry 64, 850–861. doi: 10.1016/S0031-9422(02)00487-2

Desnoyer, N., Howard, G., Jong, E., and Palanivelu, R. (2020). AtPIG-S, a predicted glycosylphosphatidylinositol transamidase subunit, is critical for pollen tube growth in Arabidopsis. BMC Plant Biol. 20:380. doi: 10.1186/s12870-020-02587-5

Desnoyer, N., and Palanivelu, R. (2020). Bridging the GAPs in plant reproduction: a comparison of plant and animal GPI-anchored proteins. Plant Reprod. 33, 129–142. doi: 10.1007/s00497-020-00395-9

AUTHOR CONTRIBUTIONS

JS organized and wrote the manuscript. RF assisted in writing the manuscript. SC proposed and corrected the manuscript. PD and AMS assisted in writing and revising the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The work described in this review was supported by the EU project 690946 – SexSeed – Sexual Plant Reproduction – Seed Formation, funded by H2020-MSCA-RISE-2015 and SeedWheels FCT Project – POCI-01-0145-FEDER-027839.
Silva et al. Biosynthesis of Arabinogalactan-Proteins
Yariv, J., Rapport, M. M., and Graf, L. (1962). The interaction of glycosides and saccharides with antibody to the corresponding phenylazo glycosides. *Biochem. J.* 85, 383–388. doi: 10.1042/bj0850383

Yeats, T. H., Bacic, A., and Johnson, K. L. (2018). Plant glycosylphosphatidylinositol anchored proteins at the plasma membrane-cell wall nexus. *J. Integr. Plant Biol.* 60, 649–669. doi: 10.1111/jipb.12659

Youl, J. J., Bacic, A., and Oxley, D. (1998). Arabinogalactan-proteins from *Nicotiana alata* and *Pyrus communis* contain glycosylphosphatidylinositol membrane anchors. *Proc. Natl. Acad. Sci. U. S. A.* 95, 7921–7926. doi: 10.1073/pnas.95.14.7921

Yuasa, K., Toyooka, K., Fukuda, H., and Matsuoka, K. (2005). Membrane-anchored prolyl hydroxylase with an export signal from the endoplasmic reticulum. *Plant J.* 41, 81–94. doi: 10.1111/j.1365-313X.2004.02279.x

Zhang, Y., Held, M. A., and Showalter, A. M. (2020). Elucidating the roles of three β-glucuronosyltransferases (GLCATs) acting on arabinogalactan-proteins using a CRISPR-Cas9 multiplexing approach in *Arabidopsis*. *BMC Plant Biol.* 20:221. doi: 10.1186/s12870-020-02420-5

**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.

Copyright © 2020 Silva, Ferraz, Dupree, Showalter and Coimbra. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.