Arabinogalactan-proteins (AGPs) are complex glycoconjugates that are commonly found at the cell surface and in secretions of plants. Their location and diversity of structures have made them attractive targets as modulators of plant development but definitive proof of their direct role(s) in biological processes remains elusive. Here we overview the current state of knowledge on AGPs, identify key challenges impeding progress in the field and propose approaches using modern bioinformatic, (bio)chemical, cell biological, molecular and genetic techniques that could be applied to redress these gaps in our knowledge.

**INTRODUCTION**

Arabinogalactan-proteins (AGPs), ubiquitous cell surface proteoglycans in both terrestrial and aquatic plants (and algae), are proposed to play essential roles in a range of plant growth and development processes, including cell expansion, cell division, reproductive development, somatic embryogenesis, xylem differentiation, abiotic stress responses, and hormone signaling pathways (Seifert and Roberts, 2007; Ellis et al., 2010). These roles emerge from largely indirect evidence and from the "recognition" potential arising from the incredible diversity of their glycan and protein backbone moieties as well as their location; attached to the outer leaflet of the plasma membrane by a glycosylphosphatidylinositol anchor and in some instances cross-linked into the wall. Despite intense research to unravel AGP function, their molecular mechanism(s) of action remain elusive. AGPs exhibit complexity at many levels: First, AGP protein backbone genes are part of large gene families, and this makes the study of AGP function through characterization of single AGP mutants a major challenge due to gene redundancy (Ma and Zhang, 2010; Showalter et al., 2003). Second, AGP protein backbones are highly glycosylated, hindering production of antibodies specifically directed to the protein moiety that would allow for identification and isolation of single AGPs. Third, the variety of monosaccharides present in AGP carbohydrate moieties has considerably complicated how an AGP ability in Yariv binding has considerably complicated how an AGP key research challenges.

**STRUCTURES**

Arabinogalactan-proteins belong to the hydroxyproline-rich superfamily of glycopeptides (Schultz et al., 2002; Johnson et al., 2003) being composed largely of carbohydrate (90–98% w/w) with some protein typically rich in the amino acids, Hyp/Pro, Ala, Ser/Thr, that is usually covalently modified with a GPI anchor at the C-terminus (see Figure 1). Historically, AGPs were defined if they met three criteria: the presence of arabinogalactan chains, a Hyp-rich protein backbone, and the ability to bind to a class of synthetic phenyloxa dyes, the β-glucosyl Yariv reagent (see Fu et al., 1996). The significant advances in our knowledge of their carbohydrate structures, protein backbone sequences, and variability in Yariv binding has considerably complicated how an AGP is defined. For instance, the diversity of protein backbones has led to a classification of the AGPs into different sub-classes based on the presence/absence of particular motifs/domains (Johnson et al., 2003). The carbohydrate moiety is typically in the form of type II arabinogalactans (AGs) although some AGPs also contain short arabinio-oligosaccharide chains (Figure 1; Fincher et al., 1983; Tan et al., 2004, 2010; Ellis et al., 2010). Type II AGs have also
been reported either as free polysaccharides (Ponder and Richards, 1997) or as side chains of rhamnogalacturonan I (RG-I; Caffall and Mohnen, 2009). The existence of different forms of type II AGs raises a few questions. Are free type II AGs generated from AGPs by hydrolases in the wall or synthesized de novo? Are AG side chains of RG-I derived from either AGPs by transglycosylases or from covalently linked RG-I-AGP complexes? To understand how this diversity impacts biological function, we face the challenge of isolating "individual" AGPs and sequencing their glycans (and protein backbones).
Another aspect of AGP research is the intriguing possibility that they are one form of covalent cross linker for wall matrix polysaccharides. In the early 1970s, Keegstra et al. (1973) hypothesized that Rha residues on AG side chains of AGPs might be attachment sites for RG-I. Since then, AGPs/AGs have been reported to form complexes with pectins (Yamada et al., 1987; Saulnier et al., 1988; Kwon and Morrisan, 1991; Pellerin et al., 1995; Yamada, 2000; Duan et al., 2003, 2004) and xyloses (Kwon and Morvan, 1995). However, residues involved in the covalent cross-link between AGPs and wall polysaccharides have not been defined. Several major challenges must be addressed to determine if AGP polysaccharide complexes (APCs) exist and to determine the structure and function of any such complexes.

CHALLENGE 1: ISOLATION AND PURIFICATION OF AGPs

The incredible heterogeneity of AGP structures has hampered purification of individual AGPs. As a consequence, most studies on AGPs have been on a family of molecules and often in the presence of contaminating polymers. There are a few examples of AGPs purified by a combination of traditional chromatographic methods (for example, ion exchange/lectin affinity/gel permeation using chaotropic reagents) and/or Yariv precipitation that are “pure” AGPs; for example, the AGPs from tobacco floral tissues (Gane et al., 1995) and larix AG exudates (Ponder and Richards, 1997). The application of molecular biology techniques to both isolate heterologously expressed AGP protein backbones or synthetic peptides as green fluorescent protein tagged (GFP)-fusion is to isolate heterologously expressed AGP protein backbones or synthetic peptides as green fluorescent protein tagged (GFP)-fusion proteins by the Kieliszewski/Showalter, Matsuoka, and Somerville laboratories (Shpak et al., 1999; Zhao et al., 2002; Tan et al., 2004; Shimizu et al., 2005; Estévez et al., 2006) was an ingenious innovation that allowed the purification of AGPs with a single protein backbone and therefore the study of inherent glycan heterogeneity. However, the low DP of these glycans raises some questions on the fidelity of glycosylation in heterologous/high expression systems.

Thus, a combination of purification techniques is necessary to purify relatively homogeneous AGPs (and APC complexes extracted as described below). These techniques take advantage of the heterogeneity of AGPs and wall polysaccharides including size, charge, hydrophobicity (Serpe and Nothnagel, 1996; Lampert et al., 2011), the ability to co-precipitate with Yariv reagents, the availability of anti-AG antibodies (Pattathil et al., 2010), and the use of tagged heterologously expressed protein backbones.

CHALLENGE 2: EXTRACTION AND PURIFICATION OF PUTATIVE APCS FROM WALLS

Because pectins and non-cellulosic polysaccharides are embedded within the highly cross-linked wall, the first obstacle to studying putative APCs is to extract intact macromolecules from the wall, especially from secondary walls. Traditional methods to release polysaccharides from the wall include either the use of wall-specific degrading enzymes (York et al., 1986) or the extraction of walls with increasingly harsh solvents (Fry, 1988). Since the enzymatic and strong base treatments could also potentially break covalent linkages between AGPs and wall polysaccharides, the released polymers may only contain partial structural information of potential APCs and may still contain contaminating wall polysaccharides.

To avoid these extraction complications it may be possible to source APCs from potentially rich sources such as suspension culture media, especially of xylogenic calli, polysaccharide-rich seed mucilages, and exudates, such as gums (Defaye and Wong, 1986) and root mucilages (Dowdy et al., 1988) since these are released in a “solubilized” form.

CHALLENGE 3: SEQUENCING OF AGs AND APCs

Our current knowledge of AG carbohydrate sequences is based on experiments using tools that include monosaccharide composition, linkage analysis, chemical or enzymatic degradation of glycans, mass spectrometry (MS), and NMR analysis. Partial acid hydrolysis (Defaye and Wong, 1986), actomyosin, alkaline degradation, and Smith degradation (Churms et al., 1981; Rasic et al., 1987) have supported the basic structures summarized in Figure 1 and led to the suggestion that the AG glycans contain a backbone of β-(1,3)-galacto-oligosaccharides interrupted at regular intervals with a periodate-sensitive residue. However, few of the large AG chains have been de novo-sequenced due to the inherent biosynthetic heterogeneity and the current limitations of sequencing technologies.

The availability of linkage-specific enzymes has greatly assisted the sequencing of glycans although their lack of commercial accessibility has hampered progress. Thus in AG analysis the degradation of an AG-specific exo-β-(1,3)-galactanase that can bypass the β-(1,6)-galactosyl side chains (Tsumuraya et al., 1990; Kotake et al., 2005; Ichino et al., 2006). This enzyme, together with the recently characterized β-glucuronidases (Haque et al., 2005), α-arabinofuranosidase (Hata et al., 1992), and endo-β-(1,6)-galactanase form a enzyme tool specific for AG side-chain analysis which enabled Tryfona et al. (2010) to characterize some long β-(1,6)-galacto-oligosaccharide AG side chains with the aid of MS/MS fragmentation (see Oxley et al., 2004). A recent study of Arabidopsis AG±1 (Hijazi et al., 2012), a chimeric AGP, illustrates the power of a multipronged approach to purification and characterization of AGs.

Therefore, the best solution is to sequence small structural units of AGs, generated using a combination of chemical and enzymatic techniques, and then to re-construct models of the intact AGs. Discovery of new chemicals and enzymes that can selectively cleave AGs would facilitate future progress in the sequencing of the AG glycan chains.

BIOINFORMATICS

Genomics and its related technologies have revolutionized the study of biology, facilitated the development of other ‘omics platforms, and created a need for bioinformatics to handle the acquisition, storage, and analysis of the vast amount of data generated from ‘omics and ‘omics-related projects. The AGP field has greatly benefited from genomics and bioinformatics. Given that AGP protein backbone sequences often have low sequence similarity, BLAST-type searches typically identify only a few closely related AGP family members and, therefore, are not a particularly effective means to comprehensively identify members of the AGP family. In contrast, bioinformatics approaches have provided a
Tan et al. The research challenges of arabinogalactan-proteins

AGP protein backbones from the rice genome/proteome, including 13 classical AGPs, 15 AG-peptides, 3 non-classical AGPs, 3 early nodulin-like AGPs (eNod-like AGPs), 8 non-specific lipid transfer protein-like AGPs (nLTP-like AGPs), and 27 FLAs. A few other bioinformatic studies are reported for AGP protein backbones, but these studies were not focused exclusively on AGPs and/or concentrated only on one particular sub-class (e.g., GPI-anchored AGPs or FLAs). For example, Bioner et al. (2002, 2003) used bioinformatics to identify GPI-anchored proteins in Arabidopsis from genomic and proteomic data. In addition, Ishad et al. (2008) applied bioinformatic analysis to their cell wall proteomic data in Arabidopsis to identify several AGPs and Egelund et al. (2006) used bioinformatic analyses to identify 34 wheat and 24 rice AGPs that could be developed and used. This relies on our knowledge of AGP gene/protein families. Schultz et al. (2002) conducted the first comprehensive bioinformatics analysis to identify and characterize AGP genes/proteins from the Arabidopsis genome/proteome with respect to their protein backbones. This study was refined by Showalter et al. (2010), who conducted the only other comprehensive bioinformatics analysis for AGP protein backbones in rice. They identified 69 rice AGP protein backbones from the rice genome/proteome, including 13 classical AGPs, 15 AG-peptides, 3 non-classical AGPs, 3 early nodulin-like AGPs (eNod-like AGPs), 8 non-specific lipid transfer protein-like AGPs (nLTP-like AGPs), and 27 FLAs. A few other bioinformatic studies are reported for AGP protein backbones, but these studies were not focused exclusively on AGPs and/or concentrated only on one particular sub-class (e.g., GPI-anchored AGPs or FLAs). For example, Bioner et al. (2002, 2003) used bioinformatics to identify GPI-anchored proteins in Arabidopsis from genomic and proteomic data. In addition, Ishad et al. (2008) applied bioinformatic analysis to their cell wall proteomic data in Arabidopsis to identify several AGPs and Egelund et al. (2006) used bioinformatic analyses to identify 34 wheat and 24 rice AGPs that could be developed and used. This relies on our knowledge of AGP gene/protein families. Schultz et al. (2002) conducted the first comprehensive bioinformatics analysis to identify and characterize AGP genes/proteins from the Arabidopsis genome/proteome with respect to their protein backbones. This study was refined by Showalter et al. (2010), who conducted the only other comprehensive bioinformatics analysis for AGP protein backbones in rice. They identified 69 rice AGP protein backbones from the rice genome/proteome, including 13 classical AGPs, 15 AG-peptides, 3 non-classical AGPs, 3 early nodulin-like AGPs (eNod-like AGPs), 8 non-specific lipid transfer protein-like AGPs (nLTP-like AGPs), and 27 FLAs. A few other bioinformatic studies are reported for AGP protein backbones, but these studies were not focused exclusively on AGPs and/or concentrated only on one particular sub-class (e.g., GPI-anchored AGPs or FLAs).}

CHALLENGE 4: IDENTIFYING AND CLASSIFYING AGPs FROM OTHER SEQUENCED PLANT GENOMES

Over 30 plant and algal genomes/proteomes are now known (see http://en.wikipedia.org/wiki/List_of_sequenced_eukaryotic_genomes#Algae). It would be useful to apply either the current or improved bioinformatics programs to these various datasets. Suggested enhancements to the programs would include making the bioinformatic analysis more automated and integrating the programs for predicting signal peptides, GPI anchor addition sites, gene expression, co-expression analysis, etc. into a single program. In addition, based on existing protein sequence and carbohydrate data on AGPs, a bioinformatics program predicting sites of prolyl hydroxylation and corresponding sites and type of glycosylation (i.e., AGs and arabinose-oligosaccharides) could be developed and used. This relies on our knowledge that the types of O-glycosylation on the AGP protein backbone can be predicted from the Hyp-contiguity hypothesis that defines Hyp (arabinogalactosylation) as occurring on the clustered, non-contiguous Hyp residues separated by Ala or Ser residues in a protein backbone whereas blocks of contiguous Hyp residues, such as occur in extensins, are arabinosylated with short oligosaccharides (Kieliszewski and Lampert, 1994; Shpak et al., 1999; Goodrum et al., 2000; Zhao et al., 2002); N-glycosylation, is predicted by the universally conserved consensus amino acid sequence Asn-X-Ser/Thr, where X can be any amino acid except Pro. Similarly, the specificity of prolyl hydroxylation by prolyl-4-hydroxylase, although not as well defined in plants as in mammalian systems (Gorres and Raines, 2010), can be used together with the Hyp-contiguity hypothesis to inform design of bioinformatics programs.

CHALLENGE 5: APPLYING AND IMPROVING BIOINFORMATIC ANALYSES OF MICROARRAY DATA TO ELUCIDATE PATTERNS OF AGP (CO-)EXPRESSION

Unfortunately, not all of the sequenced plant genomes have extensive publicly available microarray data, unlike Arabidopsis and rice (e.g., see PLEXdb, http://www.plexdb.org/). Thus, in addition to generating new microarray data, it would be convenient to utilize and integrate expression analysis programs like Genevestigator and co-expression analyzer tools (see Table 1) to mine data and provide it in a more tailored manner. Analysis of such data can provide remarkable insight into the function (and functional redundancy) of AGP protein backbone genes as well as elucidate networks of AGPs and AGP-related genes involved in various metabolic pathways.

CHALLENGE 6: IMPROVING AND DEVELOPING NEW BIOINFORMATICS PROGRAMS TO ELUCIDATE MOLECULAR PHYLOGENIES OF AGP PROTEIN BACKBONE GENES

It would be interesting from an evolutionary standpoint to understand how AGPs are related within and between species, since such analysis may explain how the AGP family evolved and provide insight into AGP function. From a functional perspective, it would be useful to be able to identify AGP gene orthologs and paralogs. Software developers would use the gene families identified in Challenge 4 through packages summarized in Table 1 and the extensive web-based resources developed for studying gene ontology to focus on the AGP protein backbone genes. Suggested enhancements to the programs would include making the bioinformatic analysis more automated and integrating the programs for predicting signal peptides, GPI anchor addition sites, gene expression, co-expression analysis, etc. into a single program. In addition, based on existing protein sequence and carbohydrate data on AGPs, a bioinformatics program predicting sites of prolyl hydroxylation and corresponding sites and type of glycosylation (i.e., AGs and arabinose-oligosaccharides) could be developed and used. This relies on our knowledge that the types of O-glycosylation on the AGP protein backbone can be predicted from the Hyp-contiguity hypothesis that defines Hyp (arabinogalactosylation) as occurring on the clustered, non-contiguous Hyp residues separated by Ala or Ser residues in a protein backbone whereas blocks of contiguous Hyp residues, such as occur in extensins, are arabinosylated with short oligosaccharides (Kieliszewski and Lampert, 1994; Shpak et al., 1999; Goodrum et al., 2000; Zhao et al., 2002); N-glycosylation, is predicted by the universally conserved consensus amino acid sequence Asn-X-Ser/Thr, where X can be any amino acid except Pro. Similarly, the specificity of prolyl hydroxylation by prolyl-4-hydroxylase, although not as well defined in plants as in mammalian systems (Gorres and Raines, 2010), can be used together with the Hyp-contiguity hypothesis to inform design of bioinformatics programs.

Table 2
Table 1 | Bioinformatic programs used to identify and characterize AGPs.

<table>
<thead>
<tr>
<th>Program</th>
<th>Program use</th>
<th>Web address</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAST percentage calculator</td>
<td>Identification of AGP backbones</td>
<td><a href="http://www.adelaide.edu.au/directory/carolyn.schultz/underFiles">http://www.adelaide.edu.au/directory/carolyn.schultz/underFiles</a></td>
</tr>
<tr>
<td>BIO OHIO</td>
<td>Identification of AGP backbones and more</td>
<td><a href="http://code.google.com/p/prot-class/">http://code.google.com/p/prot-class/</a></td>
</tr>
<tr>
<td>Plant big-PI predictor</td>
<td>Identification of GPI anchor addition sites</td>
<td><a href="http://mendel.imp.ac.at/gpl/PlantServer.html">http://mendel.imp.ac.at/gpl/PlantServer.html</a></td>
</tr>
<tr>
<td>Genevestigator</td>
<td>Identification of gene expression</td>
<td><a href="https://www.genevestigator.ethz.ch/">https://www.genevestigator.ethz.ch/</a></td>
</tr>
</tbody>
</table>

Table 2 | The 21 putative GaTs from the Arabidopsis thaliana CAZy GT-family-31 and their proposed function.

<table>
<thead>
<tr>
<th>Sub-clade</th>
<th>Accession</th>
<th>Proposed function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>At1g33250, At5g12440, At4g10240, At4g23300, At1g05750, At4g11590, At3g77350, At3g14200, At1g52380, At4g23490, At1g07650</td>
<td>β-(1,3)-GalTs; substrate unknown!</td>
</tr>
<tr>
<td>7</td>
<td>At1g27120, At1g24690, At5g26820, At3g06440, At1g28810, At4g21060</td>
<td>β-(1,3)-GalTs involved in the synthesis of N- and O-glycans, and β-(1,3)-GlcNAcTs; substrate unknown!</td>
</tr>
<tr>
<td>10</td>
<td>At4g32120, At1g17340, At1g22015, At5g53340, At2g53000, At1g77810, At3g32430, At1g05170, At1g53290, At2g26100, At3g49600, At1g33430, At4g26940, At1g32930</td>
<td>β-GaTs involved in the synthesis of AGPs</td>
</tr>
<tr>
<td>11</td>
<td>At5g57500</td>
<td>β-(1,3)-GalTs; substrate unknown!</td>
</tr>
</tbody>
</table>

The accessions clustered into four plant-specific sub-groups according to Eglund et al. (2011). Substrate specificity were predicted based on secondary structure and conserved motifs shared with known β-(1,3)-GalTs (Qu et al., 2008).

Many mammalian, fungal, and bacterial GaTs have been identified, cloned, and biochemically characterized (Canarelli et al., 2009; Ellis et al., 2010). In contrast, only a few plant cell wall polysaccharide/proteoglycan-related GaTs have been characterized biochemically (Doblin et al., 2010). From studies of Arabidop- sius at the molecular and biochemical level (Strasser et al., 2007; Qu et al., 2008), and from assembly of mammalian proteoglycans, it is expected that AG glycan chains that decorate AGPs are synthesized by type II membrane-bound GaTs located in the Golgi apparatus. This includes members of CAZy GT-family-31 with putative β-(1,3)-GaT activity, that are suggested to be involved in synthesis of the β-(1,3)-Gal backbone in AG glycan chains (Qu et al., 2008, Eglund et al., 2011).

Early studies showed that the Golgi apparatus plays an important role in synthesis of β-(1,6)-Gal of the AG glycan chains of AGPs (Mascara and Fincher, 1982; Schibeci et al., 1994), whereas the initial enzyme in the AG biosynthetic pathway, adding the first Gal residue to a Hyp residue on the protein backbone (the Hyp-O-galactosyltransferase or HGT), is predominantly located in the ER (Oku et al., 2010). Outside of the development of in vitro assays to monitor GaT activity (Qu et al., 2008; Liang et al., 2010;
Tan et al. The research challenges of arabinogalactan-proteins

AG chains, several GalTs will be required, such as HGT(s), β-(1,3)-GalTs, and β-(1,6)-GalTs, and these enzymes will work co-ordinately to regulate the density, length, and sequence of the galactan chain. In addition, several GTs responsible for decorating termini of AG chains, i.e., arabinoxylosyltransferases (AxTIs), rhamnosyltransferases (RhTs), fucosyltransferases (FUTs), and glucuronoxylosyltransferases (GlcATs) are also involved. Recently, AIFUT4 and AIFUT6, two members of CAZy GT-family-37, were characterized as Golgi located ω(1,2)FUTs and are the first enzymes demonstrated to have a specific function in AG glycosylation (Wu et al., 2010). To ensure continued momentum in the field, we suggest a focused co-ordinated approach on three core challenges:

CHALLENGE 9: AN ALTERNATIVE APPROACH FOR THE IDENTIFICATION OF THE GLYCOSYLATION MACHINERY INVOLVED IN AG CHAIN SYNTHESIS

An alternative approach to the one described in Challenge 7, centered on the analysis of Gum Arabic, a tree exudate from the Acacia species, whose main fraction is an AG (Defaye and Wong, 1986; Randall et al., 1989; Al-Assaf et al., 2005). AG chains comprise as much as 90–98% of the gum exudate (Oman et al., 1993), thus making Gum Arabic-producing cells from the Acacia trees an obvious choice as starting material to identify enzymes involved in AG biosynthesis.

CHALLENGE 10: BIOSYNTHESIS OF PUTATIVE APCs

The challenge is to determine in which sub-cellular compartment putative APCs are assembled and by what mechanism? One possibility is that APCs are synthesized intracellularly in the ER/Golgi apparatus by multiple GTs (as proposed for AGPs and other non-cellulosic polysaccharides) by either en bloc transfer of pre-assembled oligosaccharides or stepwise sugar addition, followed by delivery into the wall. Another possibility is that APCs are assembled in the extracellular matrix, possibly by transglycosylases, a mechanism that has been well studied in xyloglucan remodeling within the wall (Rose et al., 2002) and is commonly utilized by yeast to modify their wall in response to abiotic/biotic stimuli (Kollar et al., 1997).

CHALLENGE 11: HETEROLOGOUS EXPRESSION SYSTEMS

Expression of non-cellulosic/cellulosic plant GTs in functional assay systems remains a key challenge. The past lack of success of this approach has been ascribed to the mismatch between biochemical assays and native activity, failure of the expressed protein to accumulate to sufficient levels, incorrect folding or improper post-translational modifications (Petersen et al., 2009). The most obvious choice would be to develop an “in planta” system, however, the endogenous GT activities can make it difficult to distinguish the specific activity of the expressed protein (Petersen et al., 2009). Prokaryotes, of which some have limited capacity for post-translational processing, pose other problems. We therefore suggest developing multiple heterologous expression systems to maximize the likelihood that at least one will allow for successful expression where the biochemical activity is retained. Additionally, testing new expression systems that may prove “universal” (e.g., Aspergillus), which has served as one of the preferred expression systems in the biotechnology industry, as well as cell-free expression systems may prove useful for heterologous expression of plant GTs.

CHALLENGE 12: A HIGH-THROUGHPUT ENZYME ACTIVITY SCREENING SYSTEM

The assignment of substrate specificity to GTs is often hindered by difficulties related to limited availability of relevant candidate acceptor molecules for biochemical assays. To overcome this challenge the next step should be to employ carbohydrate array technology (Moller et al., 2007) with AGPs/Gum Arabic-specific sugars and peptides, related acceptor substrates, i.e., natural acceptors from Gum Arabic and AGPs (e.g., β-(1,3)-galacto-oligosaccharides, generated by Smith degradation (see Challenge 3), de-arabinofuranosylated AGPs generated by mild acid, chemically synthesized β-(1,3)-Gal oligosaccharides and isolated AGP protein backbones) together with other “AGP-enriched” fractions from wild type, AGP GT mutants, and Gum Arabic exudates. Combining AGP-related arrays with established in vitro assays will facilitate a high-throughput screening system that can be used to test heterologously expressed candidate GTs in mixtures with either radio-labeled or fluorometrically tagged NDP-sugars as the donor to identify AGP-specific carbohydrate acceptor molecules on the array. Development of such a comprehensive screening system would be a significant step in identifying the many GTs responsible for AG biosynthesis.

FUNCTION

Arabinogalactan-protein glycan-specific antibodies and β-Glc AraViv reagent have been broadly used to investigate AGP activity in tissue culture and in planta (Seifert and Roberts, 2007; Ellis et al., 2010). The current use of these two indirect tools continues to provide information on AGP activity in new biological systems, e.g., European larch, Larix decidua (Rafinska and Bednarska, 2011), and little studied developmental processes, in this case, ovule development in gymnosperms, confirming the relevance and the conservation of function of these molecules within the plant kingdom. Unfortunately, the broad specificity of these techniques makes it impossible to assign function to a single AGP. This limitation has been partially overcome by genetic and molecular studies, including the characterization of AGP single or double mutants, RNAi and over-expressing lines, although these approaches also have complications.

The usefulness of reverse genetics approaches to investigate AGP backbone function is well demonstrated. The function of one cotton FLA, GhAGP24, in cotton fiber initiation and elongation (Li et al., 2010a) and four Arabidopsis members, FLA1, FLA3, FLA11, and FLA12 have recently been published (Li et al., 2010a,b; MacMillan et al., 2010; Johnson et al., 2011). Roles for FLA1 in lateral root and shoot development in tissue culture prior to cell-type specification (Johnson et al., 2011) and FLA3 in microspore development, possibly by participation in cellulose deposition within
the intine (Li et al., 2010b), have been described. FLAs 11 and 12 have also been implicated in the process of cellulose deposition, contributing to plant stem strength and elasticity by affecting cell wall integrity (MacMillan et al., 2010). Such a function is consistent with an earlier report by Shi et al. (2003) implicating FLA4/SOS5 in maintaining proper cell expansion under salt-stressed conditions. The apparent diversity of FLA function may be due to the ability of FLAs to mediate protein–protein interactions with cell wall or plasma membrane-associated ligands via their fasciclin-like domains, shown in other eukaryotic systems to facilitate cell adhesion.

In addition to FLA3, AGP6, and AGP11, two classical AGPs specifically expressed in pollen, have been demonstrated to be involved in the control of timing of pollen germination, as pollen of the agp11 double mutant germinates precociously inside the anthers (Coumbe et al., 2010). How the presence of AGP6 and AGP11 avoids precocious pollen germination is unknown, but it may occur by regulating water uptake.

The Lys-rich AGP sub-family has been the focus of several studies in tomato and Arabidopsis. Functional characterization of AtAGP18, one of the three Lys-rich AGPs, by over-expression of the genomic sequence in Arabidopsis indicates that AGP18 plays a role in vegetative growth and sexual reproduction (Acosta-Garcina and Vieille-Calzada, 2004; Zhang et al., 2011a). The bushy phenotype resembles that of tomato lines over-expressing LeAGP1 and is similar to tobacco plants over-producing cytokinins (Zhang et al., 2011b) leading to the suggestion that it may participate in a cytokinin signal transduction pathway as a co-receptor of cytokinins. A similar model has been proposed for FLA4/SOS5 in its interactions with two members of the leucine-rich repeat receptor-like kinase family, FEI1 and FEI2, shown by double mutant analyses to have non-additive genetic interactions with FLA4/SOS5 (Xu et al., 2008). SOS5 has been hypothesized to act as the ligand of a signal molecule that then either binds directly to FEI1/FEI2 or assists in presenting the signal molecule to FEI1/FEI2, initiating a signaling cascade that regulates the synthesis of cellulose and ultimately cell growth.

Several questions arise from this ligand model of AGP function. Given the effects on cellulose in the fla1 fla2 null mutant (MacMillan et al., 2010) and the abnormal cellulose deposition fla3 RNAi lines (Li et al., 2010b), may FLA1, FLA2, and FLA3, as well as other GPI-anchored or non-anchored AGPs, also be a part of this same network of components involved in wall sensing? Does this model explain the observation of AGPs as cell fate markers in tissues undergoing cell differentiation? Considering that the appearance of AGPs during specific developmental stages has been described using antibodies that recognize AGP-carbohydrate epitopes, is the heterogeneity of AGP glycosylation also involved in providing the necessary specificity to interact with different signal molecules and generate specific responses? What is the relevance of the presence and number of fasciclin domains of FLAs? Further investigation of the possible function of AGPs in wall sensing is of fundamental importance to uncover some of the components and mechanisms involved in the regulation of wall biosynthesis and ultimately plant cell growth. To address some of these challenges, we propose the use of the following experimental approaches, techniques, and resources:

**CHALLENGE 13: TARGETING FUNCTIONAL REDUNDANCY OF AGPs**

The application of multiple gene knock-down technologies such as double-stranded RNAs or artificial micro-RNAs could allow the silencing of putative redundant genes within the different AGP protein backbone subfamilies and therefore overcome the problems associated with functional redundancy. The detection of specific expression patterns and changes in transcript levels of AGP protein backbone genes have also assisted in directing the application of targeted experimental approaches to reveal their function, highlighting the importance of the availability and analysis of transcriptional data.

In addition, the use of co-expression gene network analyses to identify genes possibly related with AGP function, including those implicated in environmental sensing and signal transduction, would help to deepen our knowledge of the relationship, if any, between AGPs and the regulation of wall growth and integrity. The characterization of the promoters of AGP genes specifically expressed in pollen is generating detailed information of the tissue and spatiotemporal location of AGP transcripts that will allow the implementation of more targeted experimental approaches to test the function of pollen AGPs (Anand and Tyagi, 2010; Choi et al., 2010; Yang et al., 2011). However, when using transcriptional data as a guide to study gene function, we should be aware that in some cases mRNA levels have not been in agreement with protein levels (Yang et al., 2011).

**CHALLENGE 14: PRODUCTION OF SPECIFIC AGP PROTEIN BACKBONE ANTIBODIES**

The recent production of antibodies specifically recognizing the Lys-rich region of AaAGP17 and AtAGP19 protein backbones demonstrates both the veracity of this approach and also provides tools to study in more detail their tissue and cellular distribution and ultimately their function (Yang et al., 2011). Either these antibodies or alternatively antibodies to tagged versions of AGPs could be used in co-location and immunoprecipitation experiments to identify possible interacting partners.

**CHALLENGE 15: DETERMINING THE FUNCTIONAL SIGNIFICANCE OF AG GLYCAN CHAIN HETEROGENEITY**

One approach to address the functional importance of the glycan moiety of AGPs is to characterize AGP-specific GT mutants. Mutants implicated in AGP glycan moiety biosynthesis by transcript co-expression analysis could also be analyzed as single mutants and in combination with other GTs to potentially increase plant phenotypic severity. Limiting analyses to either single cell types (e.g., pollen/pollen tubes), or simple tissues with limited cell-types, would help in these analyses and provide a more restricted list of candidate GT genes. These genes could then be heterologously co-expressed and cellular fractions used in biochemical assays for functional assessment. While the initial aim of this work is to identify the GTs involved in AGP glycan synthesis, the underlying objective is to use these and other AGP mutants as functional assay systems to dissect the mechanism and pathway of AGP synthesis in greater detail. Such mutants are only useful as a means to manipulate AGPs if a visible and/or measurable or assayable AGP phenotype is observed.
CONCLUSION

In this brief overview we have attempted to summarize what we believe to be the major challenges facing the research community in attempting to unravel the structure, function, and biosynthesis of AGPs and to provide some indicators on how we might progress. In addition, we believe there is much to be learnt from advances in understanding the structural, functional, and biosynthetic properties of AGPs. The authors have made an effort to encourage colleagues and newcomers to this field and to provide a guide to advancing AGP research.

ACKNOWLEDGMENTS

Antony Bacic, Arianna Hernandez-Sanchez, and Monika S. Doblin acknowledge the financial support of the ARC Centre of Excellence for Plant Cell Walls. Arianna Hernandez-Sanchez is the recipient of scholarship support from the government of Mexico and the support of the University of Melbourne. Li Tan acknowledges the support of the ARC Centre of Excellence for Glycosciences and to provide some indicators on how we might progress.

REFERENCES


Kagey, K., Tamagawa, K., Kusumi, A., Kusumi, T., and Nakashima, F. (1973). The structure of plant cell walls. III. A model of the walls of suspension-cultured cyramce cells based on the
Tan et al. The research challenges of arabinogalactan-proteins


Conflict of Interest Statement: 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.

Received: 22 March 2012; accepted: 10 June 2012; published online: 27 June 2012.


This article was submitted to Frontiers in Plant Physiology, a specialty of Frontiers in Plant Science.

Copyright © 2012 Tan, Showalter, Egelund, Hernandez-Sanchez, Doblin and Bacic. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.