

# Prediction of Glycosylphosphatidylinositol-Anchored Proteins in Arabidopsis. A Genomic Analysis<sup>1</sup>

Georg H.H. Borner, D. Janine Sherrier<sup>2</sup>, Timothy J. Stevens, Isaiah T. Arkin<sup>3</sup>, and Paul Dupree\*

Department of Biochemistry, University of Cambridge, Cambridge, CB2 1QW, United Kingdom

Glycosylphosphatidylinositol (GPI) anchoring of proteins provides a potential mechanism for targeting to the plant plasma membrane and cell wall. However, relatively few such proteins have been identified. Here, we develop a procedure for database analysis to identify GPI-anchored proteins (GAP) based on their possession of common features. In a comprehensive search of the annotated Arabidopsis genome, we identified 167 novel putative GAP in addition to the 43 previously described candidates. Many of these 210 proteins show similarity to characterized cell surface proteins. The predicted GAP include homologs of  $\beta$ -1,3-glucanases (16), metallo- and aspartyl proteases (13), glycerophosphodiesterases (6), phytocyanins (25), multi-copper oxidases (2), extensins (6), plasma membrane receptors (19), and lipid-transfer-proteins (18). Classical arabinogalactan (AG) proteins (13), AG peptides (9), fasciclin-like proteins (20), COBRA and 10 homologs, and novel potential signaling peptides that we name GAPEPs (8) were also identified. A further 34 proteins of unknown function were predicted to be GPI anchored. A surprising finding was that over 40% of the proteins identified here have probable AG glycosylation modules, suggesting that AG glycosylation of cell surface proteins is widespread. This analysis shows that GPI anchoring is likely to be a major modification in plants that is used to target a specific subset of proteins to the cell surface for extracellular matrix remodeling and signaling.

Conventional membrane proteins possess one or more transmembrane domains that traverse the hydrophobic lipid bilayer. A glycosylphosphatidylinositol (GPI) anchor is an alternative means of attaching a protein to the membrane (for review, see Udenfriend and Kodukula, 1995), and it is found in all eukaryotic organisms. The C terminus of a GPI-anchored protein (GAP) is covalently attached via phosphoethanolamine and a conserved glycan to phosphatidylinositol or a ceramide (Kinoshita and Inoue, 2000). GAPs have many features that distinguish them from proteins with transmembrane domains. The anchor can be removed by the action of specific phospholipases (Griffith and Ryan, 1999), converting the protein into a water-soluble form. In many organisms, they are found specifically at the outer leaflet of the plasma membrane (PM). It is important that GPI anchoring can target proteins to the PM in a polarized or localized manner, for example to the apical membrane of polarized mammalian epithelial cells (Le Gall et al., 1995) or the axon of

neurons (Brown et al., 2000). As part of this targeting mechanism, GAPs are thought to associate with lipid rafts (Muniz and Riezman, 2000; Ikonen, 2001), and there is evidence for rafting of GAPs in plants (Peskkan et al., 2000). Therefore, this class of proteins forms a potentially important group of molecules involved in plant cell surface generation and remodeling (Sherrier et al., 1999).

GAPs have only relatively recently been discovered in plants. Arabinogalactan (AG) proteins (AGPs), cell surface proteoglycans, are the best characterized class of such proteins (Youl et al., 1998; Oxley and Bacic, 1999; Sherrier et al., 1999; Svetek et al., 1999). Already, 12 classical AGPs, five AG peptides, and 17 fasciclin-like AGPs have been predicted to be GPI anchored in Arabidopsis (Gaspar et al., 2001). There is evidence for anchoring of a nitrate reductase in sugar beet (*Beta vulgaris*) and barley (*Hordeum vulgare*; Kunze et al., 1997), and a purple acid phosphatase in *Spirodela oligorrhiza* (Nakazato et al., 1998; Nishikoori et al., 2001). However, in general, the types of plant proteins that use this mode of membrane attachment are unknown. We have previously shown that there are abundant GAPs that are not AGPs on the surface of Arabidopsis callus cells (Sherrier et al., 1999), and tobacco (*Nicotiana tabacum*) protoplasts also display a range of GAPs at their surface (Tacos et al., 1997). The few Arabidopsis proteins other than AGPs that have been proposed to be GPI anchored are COBRA (Schindelman et al., 2001), four blue-copper binding proteins (Nersissian et al., 1998), and four matrix metalloproteinases (MMPs; Maidment et al., 1999), yielding 43 candidates (including the AGPs, AG peptides, and fasciclin-like AGPs). All of these proteins are also

<sup>1</sup> This work was supported by the Biotechnology and Biological Sciences Research Council and by the European Commission. G.H.H.B. also received a scholarship from the Studienstiftung des Deutschen Volkes.

<sup>2</sup> Present address: Department of Plant and Soil Sciences and Delaware Biotechnology Institute, University of Delaware, Newark, DE 19717.

<sup>3</sup> Present address: Hebrew University Jerusalem, Department of Biological Chemistry, Alexander Silberman Institute of Life Sciences, Givat Ram, IL-91904 Jerusalem, Israel.

\* Corresponding author; e-mail p.dupree@bioc.cam.ac.uk; fax 44-1223-333345.

Article, publication date, and citation information can be found at [www.plantphysiol.org/cgi/doi/10.1104/pp.010884](http://www.plantphysiol.org/cgi/doi/10.1104/pp.010884).

predicted to be localized to the PM or cell wall. Moreover, polarized localization of certain AGPs (McCabe et al., 1997; Majewska-Sawka and Nothnagel, 2000) and the COBRA protein in root cells (Schindelman et al., 2001) has been reported.

GPI anchoring of proteins could serve several different functions. First, GAPs may serve as links between the PM and the cell wall. The GPI anchor would keep the protein attached to the membrane, while the carbohydrate moiety of proteins such as AGPs could interact with other cell wall components (Kohorn, 2000). Second, GPI anchoring may facilitate a fast switch between the two spatially close but distinct extracellular locations of the PM and the cell wall through cleavage by specific phospholipases. This cleavage could activate or deactivate a protein by separating it from substrates, ligands, or other interaction partners or by changing its conformation (Butikofer et al., 2001). Third, the released protein could become a structural component of the cell wall (Kohorn, 2000). The remnant of the anchor on the protein may serve to link it covalently to the cell wall, as is the case in yeast (*Saccharomyces cerevisiae*; Kollar et al., 1997). Fourth, the anchor may target the protein to a particular area of the PM underlying the cell wall where its function is needed. Fifth, GPI anchoring may cause cotargeting of otherwise unrelated proteins. Thus, proteins with complementary functions might become juxtaposed, increasing their working efficiency.

All known proteins that become GPI anchored have a cleavable N-terminal secretion signal for translocation into the endoplasmic reticulum. They also have a hydrophobic C terminus, which most likely forms a transient transmembrane domain. It is thought to function as a recognition signal for a transamidase, which cleaves the C-terminal hydrophobic propeptide and transfers the protein to a pre-fabricated GPI anchor (Kinoshita and Inoue, 2000). Certain sequence constraints exist for the cleavage site (the " $\omega$ -site") and the residues surrounding it (for review, see Udenfriend and Kodukula, 1995; Eisenhaber et al., 1998; Coussen et al., 2001). Therefore, it is possible to identify probable GAPs using these amino acid sequence features.

A computer-based algorithm for the identification of GAPs in metazoans and protozoans ("big-Pi") based on the analysis of the C-terminal sequence has recently been described (Eisenhaber et al., 1999; Coussen et al., 2001). Screens based on the N-terminal and C-terminal signals have been effective for analyzing the yeast genome (Caro et al., 1997; Hamada et al., 1998, 1999). However, no method optimized for large-scale screening of plant sequences is available to date. Here, we develop a procedure that effectively identifies GAPs from Arabidopsis. The 210 proteins identified are likely to function at the PM-cell wall interface.

## RESULTS AND DISCUSSION

There are no absolutely conserved sequence motifs that can be used to predict unambiguously the addition of a GPI anchor to a protein. Nevertheless, we found that the application of several simple sequence analysis tests was a highly selective procedure. All known GAPs from all organisms have an N-terminal signal sequence for targeting to the endoplasmic reticulum, a hydrophobic C-terminal sequence, and no internal transmembrane helices (TMHs; Udenfriend and Kodukula, 1995). Therefore, a search program termed "GPT" (GPI-anchoring Prediction Tool) was devised that performed three tests for the presence of these hydrophobic segments in primary amino acid sequences. The permissible location, length, and hydrophobicity of each of these segments could be independently varied. The possession of all three characteristics identified a potential GAP (see "Materials and Methods"). The list produced by GPT was refined using independent algorithms to confirm that each protein conformed to these expected characteristics. Only sequences that contained a potential  $\omega$  cleavage site (Udenfriend and Kodukula, 1995) were included in the list of predicted GAP. Because there are few experimentally demonstrated GAP in plants, this procedure was first optimized and tested using known yeast GAPs (see "Materials and Methods"). Yeast was chosen because a large number of GAPs have already been identified (Hamada et al., 1998, 1999), and a recent study demonstrated some conservation of the machinery involved in synthesis of GAP between plants and yeast (Takos et al., 2000). The procedure correctly identified 93% of known yeast GAP with less than 2% obvious false positives (see "Materials and Methods").

### Analysis of Proteins Encoded in the Arabidopsis Genome

Analysis of the Arabidopsis protein database (March 29, 2001) of 29,337 partially redundant sequences using GPT identified 443 potential GAP. Refinement of the analysis by SignalP V2.0 (Nielsen et al., 1997) and TMHMM (Sonnhammer et al., 1998), confirming the presence of a signal sequence and absence of TMHs, reduced the number of candidate sequences to 263. After elimination of sequences without a potential C-terminal  $\omega$  cleavage site, six of the remaining 202 candidates were removed because of redundancy. All 196 sequences obtained were sorted into classes based on homologies to other proteins. Just four sequences could be identified as false positives because they had known subcellular localizations other than the PM. A further four sequences were removed, as their prediction from genomic sequence was found to be incorrect. A renewed screen of the Arabidopsis protein database (dated August 8, 2001) resulted in the addition of four sequences. The list comprised 192 putative

GAPs (Table I). Based on high sequence similarity to these proteins, 18 further sequences, which did not pass through all the filters of the screen, were judged likely to be GAP. They were added to the list and are shown in parentheses in Table I, increasing the total number of candidate GAP to 210. These sequences represent approximately 0.8% of the estimated 25,498 genes (The Arabidopsis Genome Initiative, 2000). A putative function could be ascribed to 176 of the proteins based on sequence homology or on previous reports.

It is likely that a few candidate GAPs will not have been identified because the annotation of the Arabidopsis genome, particularly of small genes, is incomplete. It is clearly crucial for this analysis that proteins are correctly annotated, particularly at the N and C termini. Furthermore, SignalP and TMHMM V1.0 have prediction accuracies of 85% (Nielsen et al., 1997) and 97% (Krogh et al., 2001), respectively, suggesting that a few proteins may have been wrongly excluded. Nevertheless, we believe the proteins presented in Table I represent all the major types of GAP in Arabidopsis.

### Predicted Families of GAPs

#### *AGPs and AG Peptides (22)*

AGPs and AG peptides are members of the superfamily of cell wall Hyp-rich glycoproteins (HRGPs; Knox, 1995; Kieliszewski, 2001), and are characteristically highly glycosylated. They are thought to have multiple functions in plant growth and development (Showalter, 2001). "Classical" AGPs consist largely of multiple Hyp-rich glycosylation modules (glycomodules; Kieliszewski and Lamport, 1994), whereas the recently described AG peptides possess a single such glycomodule, being just 10 to 15 amino acids in their predicted mature form (Schultz et al., 2000; Gaspar et al., 2001). Classical AGPs and AG peptides have been predicted to become GPI anchored, and anchoring of AtAGP10 from Arabidopsis and a variety of AGPs from other plants has already been verified experimentally (Youl et al., 1998; Oxley and Bacic, 1999; Sherrier et al., 1999; Svetek et al., 1999; Schultz et al., 2000).

The analysis identified all 12 classical AGPs and four of the five AG peptides previously reported (Gaspar et al., 2001). It is interesting that one further putative classical AGP (AAC63647.1) and four novel AG peptides (BAB09787.1, AAG21543.1, CAB41186.1, and AAF01556.1) were also identified. The identification of so many members of this family of proteins confirmed that the screen was very thorough and effective in identifying candidate GAPs. The single AG peptide AtAGP16 that was not recognized by GPT has an unusually long 9-residue hydrophilic region C-terminal to the potential hydrophobic anchor addition signal, but was included in Table I on the basis of close similarity to the other AG peptides

and the previous prediction (Schultz et al., 2000). The potential significance of GPI anchoring of these proteins has been widely reviewed (Majewska-Sawka and Nothnagel, 2000; Schultz et al., 2000; Gaspar et al., 2001).

#### *Extensin-Related Proteins (6)*

Extensins, like AGPs, are members of the HRGP superfamily (Knox, 1995; Kieliszewski, 2001). They are characterized by long stretches of S(Hyp)<sub>4</sub> motifs that become arabinosylated glycomodules. This analysis predicted six extensin-related GAPs with several such glycomodules.

Extensins are thought to have structural roles in the cell wall. Their physical properties are highly variable due to specific glycosylation, hydroxylation, and crosslinking with other cell wall components (Kieliszewski and Lamport, 1994; Knox, 1995). One of these six extensin-related GAPs (CAB87672.1) also has probable AG glycomodules. Chimeric proteins with AG glycomodules and extensin glycomodules have been reported previously (Lind et al., 1994; Schultz et al., 1997; Bosch et al., 2001). There are no previous reports of GAPs related to extensins.

#### *Phytocyanin-Like Proteins (25)*

Phytocyanins are a superfamily of blue mono-copper binding proteins (Nersissian et al., 1998). This study identified 13 phytocyanins and 12 phytocyanin-related early nodulin-like proteins. Together, they probably form the largest group of GAPs in Arabidopsis.

All known phytocyanins have secretion signals and are thought to be extracellular proteins. There are four subfamilies: stellacyanins, uclacyanins, mavi-cyanins, and plantacyanins. The studied stellacyanins and uclacyanins have an HRGP domain and a hydrophobic C terminus (Nersissian et al., 1998), and it has previously been speculated that they are GPI anchored (van Gysel et al., 1993; Nersissian et al., 1998). The processing of N and C termini of cucumber (*Cucumis sativus*) stellacyanin supports this prediction (Nersissian et al., 1996). The HRGP domain contains probable AG glycomodules (see discussion of AG glycomodules), facilitating cell wall interactions (Nersissian et al., 1998). Eight of the sequences identified in this study can be classified as uclacyanins, and five as stellacyanins based on the composition of their putative copper-binding sites (Nersissian et al., 1998). These include all four previously reported Arabidopsis phytocyanins (van Gysel et al., 1993; Nersissian et al., 1998). It is interesting that not all of the Arabidopsis uclacyanins and stellacyanins identified here contain AG glycomodules (Table I).

Mono-copper binding proteins are believed to function as mobile electron carriers in electron-transport systems, but the substrates of the phyto-cya-

**Table 1.** Putative GPI-anchored proteins in Arabidopsis

Predictions have been submitted to Swissprot and Munich Information Centre for Protein Sequences (MIPS). Bold indicates probable AG glycomodules present; parentheses indicate sequence included based on high sequence identity; square brackets indicate sequence included after correction of annotation.

Protein Family	Accession Nos.	Total
Classical AGPs	<b>BAB09862.1;AAC77824.1;AAC77825.1;CAB89400.1;AAC77827.1;CAB42531.1;BAB11561.1;AAG40020.2;CAB78027.1;AAF01553.1;AAB87117.1;AAC63647.1;AAG41964.1</b>	13
AG peptides	<b>BAB01759.1;CAB38961.1;BAA97180.1;CAB87692.1;BAB09787.1;AAG21543.1;CAB41186.1;AAF01556.1;(AAG24284.1)</b>	9
Extensin related	<b>CAB87672.1</b> ;BAB10346.1;AAF63823.1;CAB78656.1;AAD55511.1;AAC00615.1	6
Phytocyanin-like (25)		
Stellacyanin-like	<b>AAC26242.1;AAC63847.1;AAB95306.1</b> ;CAA78771.1;CAA77089.1	5
Uclacyanin-like	<b>AAC32038.1;AAC27480.1;AAB07009.1;AAB47973.1;CAB87864.1</b> ;AAF18529.1;AAG21490.1;BAB02115.1	8
Early nodulin-like	<b>AAD23007.1;BAB10717.1;CAB79777.1;CAB38264.1;CAB40754.1;AAC63667.1;BAB01165.1;CAA22576.1;AAF26167.1;AAC32929.1;AAF68112.1;BAB02217.1</b>	12
COBRA family (11)		
COBRA	AAG12670.1;AAF02128.1;AAK56072.1;(AAB60732.1);(BAB10644.1);(CAC01762.1)	6
COBRA-like	<b>BAB10345.1;BAB00585.1</b> ;BAB01166.1;(CAA74765.1);(CAB38841.1)	5
Glycerophosphodiesterase-like	<b>BAB08565.1;CAB36515.1;BAA96908.1;BAB10996.1;AAF98209.1;(AAF98210.1)</b>	6
Hedgehog interacting protein (HIP)-like	BAA97210.1;BAB10221.1;[AAD55298.1]	3
Fasciclin-like	<b>BAB09240.1;AAD10681.1;AAB82617.1;CAB62325.1;AAD18116.1;CAA16540.1;CAB82694.1;AAK20858.1;BAB00834.1;BAB08377.1;AAD22328.1;BAB08232.1;BAB10980.1;AAK20859.1;AAD32933.1;CAB45494.1;(AAD21471.1);(CAC07928.1);(BAB08961.1);(AAF02137.1)</b>	20
β-1,3-Glucanase-like	BAB08443.1;AAB97119.1;BAB11001.1;BAB10375.1;AAF08558.1;AAB64040.1;BAB10263.1;AAD10143.1;AAF98573.1;BAB09480.1;CAB79832.1;BAA97291.1;AAC17632.1;CAB62612.1;AAK43968.1; (AAF98409.1)	16
Polygalacturonase	BAB00758.1	1
Proteases (13)		
Aspartyl-	<b>CAB62655.1;CAB62657.2</b> ;AAF26986.1;AAG42922.1;BAB09366.1;CAB92049.1;CAA21474.1	7
Metallo-	AAB61099.1;AAC00572.1;AAD14473.1;(CAB78706.1);(AAC31167.1)	5
Cystein-	CAB88124.1	1
Lipid transfer protein-like (LTPL)	<b>CAB77992.1;AAF78376.1;CAC05463.1;AAD48513.1;CAB45983.1;AAC77871.1;BAB01476.1;AAC16080.1;BAB10276.1;AAD15432.1;CAB83144.1;BAB11118.1;BAB01475.1;AAF36746.1;CAB68193.1;CAA16548.1;AAC16079.1;CAA16549.1</b>	18
Bp10-like	BAB08664.1;CAB41712.1	2
Receptor-like (19)		
S/T kinase type	<b>AAF19714.1;BAB11104.1;BAB11105.1;BAB11106.1</b> ;AAD12705.1;AAF26777.1;CAA18495.1;(AAF79910.1); <b>AAF79572.1;AAF07793.1</b> ;AAB81674.1	11
LRR type	<b>CAB37461.1</b> ;CAB77890.1;BAB02092.1;AAD50026.1;AAD55468.1;AAD21728.1;CAB81444.1;BAB09219.1	8
GAPEP	BAB08290.1;BAB08815.1;AAF14841.1;BAB09700.1;CAB62355.1; (BAA96979.1);(AAF14840.1);(BAB09698.1)	8
Other	<b>CAB87275.1;BAA97185.1;AAF02148.1</b> ;BAB10290.1;AAF97962.1;CAB78588.1	6
Unknown	<b>CAB36779.1;AAC28508.1</b> ;AAF99747.1;AAF82207.1;AAC78534.1;BAB08342.1;AAG12864.1;AAD20685.1;BAA95716.1;CAB79630.1;BAB09299.1;AAC24095.1;AAC26689.1;AAC64895.1;AAF19686.1;AAF79851.1;BAB09930.1;AAD20925.1;AAF04894.1;BAB08298.1;BAB02017.1;CAB71905.1;CAB93722.1;CAB89342.1;CAC01831.1	25
Hypothetical	BAB08533.1;BAB08294.1;AAF79800.1;AAC24384.1;BAB02002.1;CAB88347.1;CAB81604.1;CAB88306.1;AAG50914.1	9
Total		210

nins are unknown. Their redox-reactive center is much more accessible than that of plastocyanins, and this makes the phytocyanins more likely to interact with low  $M_r$  compounds than with protein electron mediators (Hart et al., 1996; Nersissian et al., 1998). The current view on phytocyanins holds that they are redox-active proteins most likely involved in the primary defense response, oxidative stress, or in lignin formation at the cell surface (Drew and Gatehouse, 1994; Nersissian et al., 1998; Miller et al., 1999; Ezaki et al., 2000; Zhang et al., 2000). Copper ions have recently been implicated in hydroxyl radical-mediated cell wall remodeling (Fry, 1998; Fry et al., 2001). It is an intriguing possibility that the phytocyanin-like proteins identified here might be involved in this process.

Two members of a family of phytocyanin-related early nodulins, MtENOD16 and MtENOD20, have been reported in the model legume *Medicago truncatula* and were predicted to become GPI anchored (Greene et al., 1998). These phytocyanin-related early nodulins lack the metal-binding-site Cys, suggesting that these proteins do not bind copper (Greene et al., 1998). Of the 25 proteins identified in this study, 12 lack this Cys, and therefore belong to the early nodulin-like subgroup of phytocyanin-related proteins. Eight have probable AG glycomodules.

#### COBRA Family Proteins (11)

COBRA is important for root formation in Arabidopsis (Benfey et al., 1993; Schindelman et al., 2001). In this screen, COBRA and two additional sequences that show 64% and 71% amino acid identity to COBRA were identified. Three further proteins (probably incorrectly annotated) were included in Table I based on high sequence identity. A second family of five proteins with lower levels of similarity to COBRA were also identified. Three of these have probable AG glycomodules at three conserved positions.

Based on COBRA's primary sequence, the protein was predicted to become GPI anchored (Schindelman et al., 2001). Although anchoring has yet to be directly demonstrated, the targeting to the PM supports the prediction (Schindelman et al., 2001). In mutant COBRA plants, root cells expand radially rather than longitudinally, suggesting that COBRA is necessary for oriented cell expansion. The mutant roots also have decreased levels of cellulose, implying a role for COBRA in cell wall maintenance. It is possible that COBRA directs cellulose synthesis to specific areas of the cell surface (Schindelman et al., 2001).

#### Glycerophosphodiesterase-Like Proteins (6)

Five closely related proteins showed low levels of similarity to bacterial glycerophosphodiesterases (Porcella et al., 2000). A sixth probably incorrectly

annotated protein was included in Table I based on high sequence identity. Proteomic analysis of callus GAP demonstrated that AtGPIP2 is one of these proteins (Sherrier et al., 1999; P. Dupree, M. Mann, and B. Kuester, unpublished data). This supports the prediction that these proteins are GPI anchored. A related PM glycerophosphodiesterase homolog from mammals has recently been implicated in lipid metabolism and signal transduction, but this protein is not GPI anchored (Zheng et al., 2000). Like many of the proteins found in the screen (see below), all of these glycerophosphodiesterase-like proteins contain probable AG glycomodules. It has already been shown that AtGPIP2 has AG glycosylation (Sherrier et al., 1999).

#### HIP-Like Proteins (3)

The screen identified two novel proteins showing distant homology to Hedgehog interacting protein (HIP; Chuang and McMahon, 1999). A third incorrectly annotated similar protein was included in Table I. Proteomic analysis indicated that one of these HIP-like proteins is AtGPIP1, the major GAP expressed in Arabidopsis callus (Sherrier et al., 1999; P. Dupree, M. Mann, and B. Kuester, unpublished data), confirming the prediction.

In various animals, Hedgehog is a diffusible protein morphogen that affects cell patterning and development (Zeng et al., 2001). HIP, a mouse PM protein, is a negative regulator of Hedgehog function, perhaps by sequestering free Hedgehog (Chuang and McMahon, 1999). There is no known homolog of Hedgehog in Arabidopsis. However, given the extensive homology, it seems plausible that AtGPIP1 may have developmental importance by interacting with an extracellular protein, and we are currently investigating this possibility.

#### Fasciclin-Like Proteins (20)

Two distinct families of proteins with one or two fasciclin-like (fas1) domains were identified. The larger family, whose members contain stretches of AG glycomodules alternating with fasciclin-like domains, has been called fasciclin-like AGPs (Schultz et al., 2000). In addition to the 14 proteins initially identified in the screen, a further four highly related sequences that do not have hydrophobic C termini were included in Table I because they are probably incorrectly annotated. One fasciclin-like protein has already been shown to possess AG glycosylation (AtAGP8; FLA8) (Schultz et al., 2000), and the existence of 16 homologs in the Arabidopsis genome has been reported (Gaspar et al., 2001). In addition to these sequences, we identified one further fasciclin-like AGP (AAF02137.1).

The two members of the second family of fasciclin-like proteins (AAD32933.1 and CAB45494.1) contain

a single fas1 domain. They are only very distantly related to the fasciclin-like AGPs, with different residues conserved in the fas1 domain. Also, they do not contain AG glycomodules.

Fasciclins are developmentally important cell surface proteins in various organisms ranging from algae (Huber and Sumper, 1994) to humans (Kawamoto et al., 1998; Kim et al., 2000). In fruit fly (*Drosophila melanogaster*), they appear to play a major role in growth cone guidance (Grenningloh et al., 1991; Prokop, 1999). Some fruit fly fasciclins such as Fas1 are GPI anchored and have been shown to function in signaling via Tyr kinases (Elkins et al., 1990). Thus, the plant fasciclin-like proteins are good candidates for receptors or cell adhesion molecules (Schultz et al., 2000).

#### Cell Wall Hydrolytic Enzymes (17)

Fifteen of the identified sequences have significant sequence identity to known  $\beta$ -1,3-glucanases. A sixteenth protein with a weak  $\omega$  cleavage site was included in Table I based on high sequence similarity. In addition, a polygalacturonase-like protein was identified. There are no previous reports of GPI-anchored cell wall hydrolytic enzymes from plants, but there are GPI-anchored cell wall-modifying glucanases in yeast (Caro et al., 1997; Hamada et al., 1998).

$\beta$ -1,3-Glucanases form at least two large families in plants. One comprises extracellular and acidic enzymes, whereas members of the other family are vacuolar and basic (Van den Bulcke et al., 1989). All 16 predicted GPI-anchored glucanases have an acidic pI, consistent with an extracellular localization. The  $\beta$ -1,3-glucanases are "pathogenesis-related" proteins that are induced in response to pathogen attack (Vogeli-Lange et al., 1988; Stintzi et al., 1993; Beffa and Meins, 1996). They may break down  $\beta$ -1,3-glucan of fungal cell walls, which in turn can produce elicitors to trigger a complex defense response and increase resistance to infection (Klarzynski et al., 2000).  $\beta$ -1,3-glucanases have also been implicated in certain developmental processes. Plant  $\beta$ -1,3-glucan (callose) breakdown is considered crucial for pollen development (Bucciaglia and Smith, 1994), and callose is also synthesized and broken down at the forming cell plate (Samuels et al., 1995; Otegui and Staehelin, 2000). PM or cell wall-localized GPI-anchored  $\beta$ -1,3-glucanases would be ideally localized for site-specific remodeling of cell wall components. The GPI anchor may cause localized targeting of the enzymes to those parts of the cell where the restructuring takes place.

#### Proteases (13)

A total of 13 proteases were identified as putatively GPI anchored. They fall into two main subfamilies of

aspartyl or metalloproteases, plus one Cys protease-like protein.

One family of seven members shows homology to a chloroplast nucleoid-binding aspartyl protease from tobacco (Nakano et al., 1997; Murakami et al., 2000). The homology is restricted to the protease domain of the chloroplast protein (Nakano et al., 1997), and the putative DNA-binding domain is absent. Two of the seven proteins have probable AG glycomodules, further supporting the predicted cell surface localization.

A second family consists of five MMPs (Maidment et al., 1999). At1-MMP, which has no consensus  $\omega$  cleavage site, and At4-MMP, which failed the SignalP test, were included based on high sequence identity. A soybean (*Glycine max*) MMP appears to be cell wall associated (Pak et al., 1997), and it was previously suggested that At1-, At2-, At3-, and At5-MMPs might be GPI anchored (Maidment et al., 1999).

Extracellular proteases could be involved in the remodeling and degradation of plant extracellular matrix (ECM) proteins. MMPs have been studied in detail in animals (Massova et al., 1998; Nagase and Woessner, 1999), where they degrade proteins of the ECM and are thus involved in a multitude of physiological processes. Some mammalian MMPs favor Hyp-rich proteins such as collagen. The plant cell wall contains an abundance of Hyp-rich proteins such as HRGPs (Knox, 1995; Kieliszewski, 2001), and many of the proteins found in this screen contain HRGP domains.

#### LTPL Proteins (18)

Lipid transfer proteins (LTPs) are a group of extracellular proteins of many proposed functions (Kader, 1996, 1997). In this screen, 18 sequences with distant homology to Arabidopsis LTPs were identified. They belong to the family of LTP-like (LTPL) proteins (Kader, 1997). Nine of these proteins have probable AG glycomodules.

Like the LTPLs identified here, plant LTPs are thought to be secreted, as they have signal peptides. There are many reports confirming their extracellular or cell wall localization (Bernhard et al., 1991; Sterk et al., 1991; Segura et al., 1993; Thoma et al., 1993, 1994). The 140- to 204-amino acid LTPLs identified in this screen are somewhat larger than LTPs, which are less than 120 amino acids. Even though plant LTPs are divergent, they all have four conserved disulfide bonds that may increase extracellular protein stability. This arrangement of invariable Cys residues is conserved in the 18 LTPLs, suggesting that they have a similar fold as the LTPs and that they are also extracellular.

The function of LTPs is still somewhat unclear. Several studies have established their ability to bind lipids, and they are capable of nonspecific lipid transfer in vitro. They may function in the deposition of

cutin monomers for the formation of extracellular waxes. Better established are the antipathogenic properties of certain LTPs. Furthermore, expression of LTPs can be induced by biotic and abiotic stress: fungal, bacterial, and viral infection, temperature, light, and osmotic stress (for review, see Kader, 1996, 1997). An LTP-like protein, stylar Cys-rich adhesin, has recently been implicated in pollen tube cell adhesion (Mollet et al., 2000; Park et al., 2000).

#### *Bp10-Like (2)*

Two of the putative GPI-anchored Arabidopsis proteins are highly similar to Bp10 from oilseed rape (*Brassica napus*; Albani et al., 1992). Bp10 is a pollen-specific gene with homology to ascorbate oxidases, which belong to the family of multi-copper oxidases. Ascorbate has been shown to be very important for the crosslinking of cell wall components and redox reactions on the plant cell surface during cell elongation (Cordoba and Gonzalez-Reyes, 1994). High ascorbate oxidase activity in the cell wall can be correlated with areas of rapid cell growth (Smirnov and Wheeler, 2000). Ascorbate has also been suggested to function in scission of plant cell wall polysaccharides (Fry, 1998; Fry et al., 2001). Thus, the Bp10 homologs reported here might be involved in cell wall remodeling.

#### *Receptor-Like Proteins (19)*

A group of 19 proteins with homology to receptor-like kinases (RLKs) were identified. Plant Ser/Thr RLKs are characterized by an extracellular receptor domain, a single TMH, and a cytoplasmic kinase domain. They can be further classified according to the nature of their extracellular domain (for review, see McCarty and Chory, 2000). All sequences reported here are homologous only to the extracellular receptor-domains of plant RLKs. The putative GPI-anchoring signal is C-terminal to this receptor domain, and the proteins lack a cytosolic kinase domain. We propose that they are GPI-anchored receptors.

Eleven sequences with homology to plant Ser/Thr kinases were identified. They fall into four subfamilies. Six proteins (BAB11105.1, BAB11104.1, BAB11106.1, AAF19714.1, AAD12705.1, and AAF26777.1) have two DUF26 domains (pfam01657), which are common in plant Ser/Thr kinases. The first four of these are related to the receptor domain of RLK3 from Arabidopsis, which is induced in response to pathogens and oxidative stress (Czernic et al., 1999) and have a single probable AG glycomodule near the C terminus. The second subfamily comprises two sequences (CAA18495.1 and AAF79910.1) showing very high similarity to PRK5, an Arabidopsis receptor protein kinase whose receptor domain is related to the pathogenesis-related-5 family (Wang et al., 1996).

One of these proteins (AAF79910.1) has a charge in the C-terminal hydrophobic domain. One sequence (AAF79572.1) has a lectin-related receptor domain; lectin-type Ser/Thr RLKs have been reported, but their function is unknown (Herve et al., 1999). The remaining two sequences (AAF07793.1 and AAB81674.1) show low levels of similarity to uncharacterized putative receptor kinases from Arabidopsis. Both have two LysM domains (smart00257), which are thought to bind peptidoglycan in bacteria (Bateman and Bycroft, 2000), but their function in eukaryotes is unknown.

The remaining eight sequences from this screen are homologous to the Cf-2/Cf-5 family of tomato (*Lycopersicon esculentum*) disease resistance proteins (Dickinson et al., 1993; Dixon et al., 1998) and to the meristem and organ development protein Clavata2 from Arabidopsis (Jeong et al., 1999). Both families are related to Leu-rich repeat RLKs, but have very short cytoplasmic tails and lack the kinase activity. Clavata2 has been suggested to function in heterodimeric receptors in conjunction with RLKs (Jeong et al., 1999). The same could be true for the putative GAP identified in this screen. It is interesting to note the variable number of Leu-rich repeats in the sequences reported here, ranging from nine to 26. A similar variability has been described within the Cf-5 family (Dixon et al., 1998), and may provide a means of altering the binding specificity of the receptor domain.

#### *GAPEPs (8)*

A family of five sequences with no detectable homology to known proteins was identified in the screen. In their unprocessed form, they are only 69 to 70 amino acids long, leaving a mere 17 to 18 amino acids after cleavage of the predicted signal peptides and hydrophobic C termini. Therefore, we propose the name GAPEP (GPI-anchored peptide) (AtGAPEP1–5, BAB08290.1, BAB08815.1, AAF14841.1, BAB09700.1, and CAB62355.1). This novel family has at least three more members that lack a consensus  $\omega$  cleavage site (AtGAPEP6–8, BAA96979.1, AAF14840.1, and BAB09698.1). The GAPEPs have 65% sequence similarity in their predicted mature forms and are charged, with basic and acidic residues in conserved positions. A similarly small GAP, CD52, has been reported in mammals (Domagala and Kurpisz, 2001). We speculate that these peptides could be signaling molecules.

#### *Other Miscellaneous GAPs (6)*

The remaining functionally annotated proteins were the embryo-specific protein 3 (CAB87275.1) and a homolog (BAA97185.1); the auxin-induced protein AIR12 (AAF02148.1; Neuteboom et al., 1999); a protein related to the oilseed rape pollen-specific protein BNM1 (BAB10290.1; Treacy et al., 1997); a Bcp1-like

protein, probably involved in plant male fertility (AAF97962.1; Xu et al., 1993, 1995); and a Gly-rich protein (CAB78588.1).

#### *Unknown and Hypothetical Proteins (34)*

Of the 210 sequences in Table I, 34 could not be assigned a function on the basis of homology to known proteins. There were 25 unknown proteins whose sequence has been predicted from genomic DNA, and whose expression has been confirmed by existence of at least one expressed sequence tag, and nine hypothetical proteins whose sequence has been predicted from genomic DNA, but without further evidence for their expression.

#### **Anticipated and Unanticipated Families of GAP**

All three Arabidopsis proteins that we could use as a positive control were correctly identified: AtGPIP1 and AtGPIP2 (Sherrier et al., 1999) and AtAGP10 (Schultz et al., 2000). In addition, all their correctly annotated homologs were identified. Other protein families that had previously been suggested to have GPI-anchored members were also identified: AG peptides (Schultz et al., 2000), fasciclin-like proteins (Gaspar et al., 2001), phytocyanins (Nersissian et al., 1998), MMPs (Maidment et al., 1999), and the COBRA protein (Schindelman et al., 2001). In addition, the number of putative GAPs from these families was enlarged by four AG peptides, one putative classical AGP, three fasciclin-like proteins, 21 phytocyanin-like proteins, and 10 COBRA-related proteins.

Hitherto, several families have had no reported GPI-anchored members in plants. However, the  $\beta$ -1,3-glucanases, aspartyl proteases, extensin-related proteins, LTPL proteins, and receptor-like proteins have homologs with known PM or extracellular localizations, compatible with the notion of GPI anchoring. Furthermore, known GAPs from yeast include glucanases and proteases (Caro et al., 1997; Hamada et al., 1998, 1999), supporting the prediction of GPI-anchored analogs in Arabidopsis.

It is interesting to note that not all secreted proteins have GPI-anchored relatives. Of all the cell wall hydrolases, only  $\beta$ -1,3-glucanases and one potential polygalacturonase were found. There were a few weak cellulase and pectin methylesterase candidates which we cannot exclude are GPI anchored (not shown). It is notable that there were no candidate xyloglucan endotransglycosylases (Fry et al., 1992; Campbell and Braam, 1999) or expansins (Cosgrove, 2000). Although some receptor-kinase like proteins appear to have GPI anchors, we found no putative wall-associated kinases (Kohorn, 2001). It is surprising that there were no homologs of nitrate reductase (Kunze et al., 1997) or purple acid phosphatase (Nakazato et al., 1998; Nishikoori et al., 2001). Homologs of a variety of GAPs of mammalian cells such as the

folate receptor or prion proteins were also not found, probably reflecting the large differences between the extracellular environments of plants and animals.

#### **AG Glycomodules Are Widespread in GAP**

One of the most surprising features of this analysis is that over 40% of the proteins contain probable AG glycomodules. In addition to the 40 classical AGPs, AG peptides, and fasciclin-like AGPs that we identified, a further 49 candidates for AG glycosylation were found (bold in Table I). Thus, it appears that a significant proportion of GPI-anchored cell-surface proteins may become AG glycosylated.

AG glycosylated proteins possess one or more small Hyp-rich glycomodules (Shpak et al., 1999, 2001; Schultz et al., 2000; Kieliszewski, 2001). There are two different types of Hyp-rich glycomodules defined by the contiguity hypothesis (Kieliszewski and Lampion, 1994). First, AG glycomodules contain clustered noncontiguous Hyps that direct attachment of complex AG-type chains. Second, arabinosylated glycomodules contain clustered contiguous Hyps. This hypothesis has been corroborated by studies using synthetic oligopeptides, allowing reliable prediction of glycosylation of glycomodules (Shpak et al., 1999, 2001). Consistent with this idea, AG glycomodules based on a "PAPAP" motif are frequently found in Arabidopsis classical AGPs (Schultz et al., 2000). Moreover, all the AG peptides have a single such AG glycomodule (Schultz et al., 2000). The six GPI-anchored glycerophosphodiesterase-like proteins have just a few potential sites for AG glycosylation that loosely satisfy the criteria, yet at least one becomes AG glycosylated (Sherrier et al., 1999). The remaining 83 proteins identified here have AG peptide modules as defined by the contiguity hypothesis, often with perfect or imperfect PAPAP motifs or runs of four or more XP motifs, and therefore we believe that most of these are likely to become AG glycosylated.

AG glycosylation appears to be of major importance for several specific classes of putative GAPs. In particular, two-thirds of the phytocyanin-like proteins have probable AG glycomodules. Other families with a substantial proportion of putatively AG glycosylated members include the LTPLs, the receptor-like proteins, the COBRA-like proteins, and the glycerophosphodiesterase-like family. It is interesting that no glucanases, no HIP-like proteins, and only two proteases seem likely to become thus modified.

What could be the possible function of the glycosylation? It is becoming clear that AGPs are very important for cell development, implying they might be involved in signaling and adhesion (for review, see Majewska-Sawka and Nothnagel, 2000). The carbohydrate chains could physically link the proteins to the ECM (Kohorn, 2000) or could bind substrates



and ligands. The latter seems plausible for the fasciclin-like and the receptor-like proteins. Alternatively, the carbohydrate moiety might conceal the C terminus of the protein, thus rendering it inaccessible to other proteins such as proteases. AG glycosylation would be particularly suitable as a protective shield because only a few Hyps suffice for the attachment of large carbohydrate side chains (Kieliszewski, 2001). In the case of the phytocyanins, the AGP domain might prevent contact between the reactive N terminus and components of the PM. It is also possible that just as glycosylation is increasingly being found to be important in protein sorting in mammalian cells (Benting et al., 1999; Hauri et al., 2000), AG glycosylation may provide a targeting signal that directs the proteins through the secretory pathway to a specific part of the cell surface.

## PERSPECTIVE

Cell walls are highly dynamic structures. Cell elongation may require localized wall loosening and expansion, and other events like trichome development, root tip growth, pollen tube growth, and wounding require highly localized and temporally controlled remodeling events. Under these conditions, GPI anchoring may provide the mechanism necessary for specific wall component targeting. Thus, it is particularly apposite that our analysis revealed families of GAPs that are likely to function in ECM modification such as the proteases, glucanases, and COBRA proteins. Wall remodeling may also be the role of phytocyanins and the putative multi-copper oxidases homologous to Bp10.

The growing pollen tube is a paradigm for extreme cell polarization requiring specific protein targeting to the growing tip, and GPI anchoring may be especially important in pollen. Several of the sequences found in this screen are related to pollen-specific proteins: the Bp10-like proteins and the BNM1 protein. The major family of  $\beta$ -1,3-glucanases described here are related to pollen-specific  $\beta$ -1,3-glucanases (Huecas et al., 2001), and callose, a  $\beta$ -1,3-glucan, is an important polysaccharide in pollen walls (Rhee and Somerville, 1998). Furthermore, AGPs have also been shown to function in pollen growth (Roy et al., 1998; Bosch et al., 2001).

Our analysis supports a role of GPI anchoring in signal generation and reception. The peptides that appear to have a GPI anchor (GAPEPs) are too small to have an enzymatic activity, and we speculate that they could function in intercellular signaling. AG peptides are also potential signaling molecules (Schultz et al., 2000), and AGPs have already been implicated in signaling and development (Nothnagel, 1997; Majewska-Sawka and Nothnagel, 2000; Schultz et al., 2000). It is perhaps surprising to find numerous PM receptor-like proteins that have no cytosolic domains. However, there are many reports

of GAPs participating in signal transduction across the PM in animals (Marmor and Julius, 2000; Saarna, 2000; Wang, 2001).

The present study is the most comprehensive analysis of Arabidopsis GAPs to date. Many of the identified proteins can be broadly categorized as being involved in signaling, adhesion, stress response, and cell wall remodeling, providing some insight into the purposes of GPI anchoring. The number and variety of proteins reported here establish the GPI anchor as a major protein modification and means of protein targeting in Arabidopsis, and more generally in plants.

## MATERIALS AND METHODS

The yeast (*Saccharomyces cerevisiae*) protein database was retrieved from the ftp server for the Genome Databases Group (Department of Genetics, Stanford University School of Medicine; <ftp://genome-ftp.stanford.edu/pub/yeast>). The annotated Arabidopsis protein database was retrieved from The Arabidopsis Information Resource ([ftp://tairpub:tairpub@ftp.Arabidopsis.org/home/tair/Sequences/blast\\_datasets](ftp://tairpub:tairpub@ftp.Arabidopsis.org/home/tair/Sequences/blast_datasets)).

In the first phase of the procedure, a Perl-based program, GPT, was used to screen the database for sequences of potential GAPs. The program performed three pass-or-fail checks for hydrophobic stretches in amino acid sequences. Only proteins with a putative N-terminal signal peptide, a hydrophobic C terminus, and no internal TMHs were identified as potentially GPI anchored. For all calculations of summed hydrophobicities, GPT used the thermodynamic Goldman-Engelman-Steitz hydrophathy scale (Engelman et al., 1986).

Because there are few experimentally demonstrated GAP in plants, GPT was trained and benchmarked using a set of 35 known GAP from yeast (Hamada et al., 1998) to find the most stringent parameter settings for the hydrophobic segments that would allow the identification of all 35 proteins. A further 11 known yeast GAPs (Hamada et al., 1999) were used as jack-knife control. The most stringent parameters, which still allowed correct identification of all sequences in the training set, were determined as follows: signal peptide length of 10 amino acids, hydrophobicity of  $-14 \text{ kJ mol}^{-1}$ ; hydrophobic signal length at C terminus of 12 amino acids, hydrophobicity of  $-18 \text{ kJ mol}^{-1}$ ; TMH length of 17 amino acids, hydrophobicity of  $-34 \text{ kJ mol}^{-1}$ . The hydrophobic stretch of the putative signal peptide was restricted to the first 30 residues. The program was allowed to ignore up to five charged or hydrophilic residues at the extreme C terminus. With these optimized parameters, GPT generated a list of 159 potential GAP from the yeast protein database of 8,996 (partly redundant) sequences.

In the second phase of the procedure, the list produced by GPT was refined to remove false positives. Proteins were first analyzed with SignalP V2.0 (Nielsen et al., 1997), setting the cut-off to  $P \geq 0.9$  certainty (<http://www.cbs.dtu.dk/services/SignalP-2.0/#submission>). The sequences that passed through this screen were checked for TMHs

using TMHMM (Sonnhammer et al., 1998; <http://www.cbs.dtu.dk/services/TMHMM-1.0>). The final test was a screen for the presence of a suitable cleavage site near the C terminus, employing the rules established by Udenfriend and Kodukula (1995). Proteins narrowly failing these tests but highly similar to proteins that passed the tests were not discarded. Duplicated sequences were removed.

In the yeast optimization, 34 of the 35 sequences initially used to define the GPT parameters and nine of the 11 jack-knife control sequences were correctly identified. In addition, 23 further candidates were identified. Fourteen of these have been reported in previous in silico studies of yeast GAPs (Caro et al., 1997; Hamada et al., 1998). One was a false positive. The remaining eight sequences are candidates for novel yeast GAPs.

All sequence alignments and calculations of sequence identities were performed with Clustal W (Thompson et al., 1994) on the Network Protein Sequence Analysis server of the Pôle Bio-Informatique Lyonnaise ([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_clustalw.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html)). BLAST searches (Altschul et al., 1990, 1997) for homologous sequences from all organisms were performed on the Cambridge Biological Sciences server (BLAST 2.1; <http://www.bio.cam.ac.uk/cgi-bin/blast2/blastallrs.pl>). BLAST searches of the Arabidopsis genome were performed using The Arabidopsis Information Resource BLAST 2.0 (<http://www.Arabidopsis.org/Blast>). Searches for conserved domains were performed using Reverse Position-Specific Blast (Altschul et al., 1997) on the National Center for Biotechnology Information server (<http://www.ncbi.nlm.nih.gov/BLAST>).

## ACKNOWLEDGMENT

We thank William Matthews for his help in the initial stages of this project.

Received September 27, 2001; returned for revision November 8, 2001; accepted January 7, 2002.

## LITERATURE CITED

- Albani D, Sardana R, Robert LS, Altosaar I, Arnison PG, Fabijanski SF (1992) A *Brassica napus* gene family which shows sequence similarity to ascorbate oxidase is expressed in developing pollen: molecular characterization and analysis of promoter activity in transgenic tobacco plants. *Plant J* 2: 331–342
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402
- Bateman A, Bycroft M (2000) The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *J Mol Biol* 299: 1113–1119
- Beffa R, Meins F Jr (1996) Pathogenesis-related functions of plant  $\beta$ -1,3-glucanases investigated by antisense transformation: a review. *Gene* 179: 97–103
- Benfey PN, Linstead PJ, Roberts K, Schiefelbein JW, Hauser MT, Aeschbacher RA (1993) Root development in *Arabidopsis*: four mutants with dramatically altered root morphogenesis. *Development* 119: 57–70
- Benting JH, Rietveld AG, Simons K (1999) N-Glycans mediate the apical sorting of a GPI-anchored, raft-associated protein in Madin-Darby canine kidney cells. *J Cell Biol* 146: 313–320
- Bernhard WR, Thoma S, Botella J, Somerville CR (1991) Isolation of a cDNA clone for spinach lipid transfer protein and evidence that the protein is synthesized by the secretory pathway. *Plant Physiol* 95: 164–170
- Bosch M, Knudsen JS, Derksen J, Mariani C (2001) Class iii pistil-specific extensin-like proteins from tobacco have characteristics of arabinogalactan proteins. *Plant Physiol* 125: 2180–2188
- Brown O, Cowen RL, Preston CM, Castro MG, Lowenstein PR (2000) Subcellular post-transcriptional targeting: delivery of an intracellular protein to the extracellular leaflet of the plasma membrane using a glycosyl-phosphatidylinositol (GPI) membrane anchor in neurons and polarised epithelial cells. *Gene Ther* 7: 1947–1953
- Bucciaglia PA, Smith AG (1994) Cloning and characterization of Tag 1, a tobacco anther  $\beta$ -1,3-glucanase expressed during tetrad dissolution. *Plant Mol Biol* 24: 903–914
- Butikofer P, Malherbe T, Boschung M, Roditi I (2001) GPI-anchored proteins: now you see 'em, now you don't. *FASEB J* 15: 545–548
- Campbell P, Braam J (1999) Xyloglucan endotransglycosylases: diversity of genes, enzymes and potential wall-modifying functions. *Trends Plant Sci* 4: 361–366
- Caro LH, Tettelin H, Vossen JH, Ram AF, van den Ende H, Klis FM (1997) In silico identification of glycosyl-phosphatidylinositol-anchored plasma-membrane and cell wall proteins of *Saccharomyces cerevisiae*. *Yeast* 13: 1477–1489
- Chuang PT, McMahon AP (1999) Vertebrate hedgehog signaling modulated by induction of a hedgehog-binding protein. *Nature* 397: 617–621
- Cordoba F, Gonzalez-Reyes JA (1994) Ascorbate and plant cell growth. *J Bioenerg Biomembr* 26: 399–405
- Cosgrove DJ (2000) Loosening of plant cell walls by expansins. *Nature* 407: 321–326
- Coussen F, Ayon A, Le Goff A, Leroy J, Massoulie J, Bon S (2001) Addition of a glycosylphosphatidylinositol to acetylcholinesterase. processing, degradation, and secretion. *J Biol Chem* 276: 27881–27892
- Czernic P, Visser B, Sun W, Savoure A, Deslandes L, Marco Y, Van Montagu M, Verbruggen N (1999) Characterization of an Arabidopsis receptor-like protein kinase gene activated by oxidative stress and pathogen attack. *Plant J* 18: 321–327
- Dickinson MJ, Jones DA, Jones JD (1993) Close linkage between the Cf-2/Cf-5 and Mi resistance loci in tomato. *Mol Plant-Microbe Interact* 6: 341–347
- Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JD (1998) The tomato Cf-5 disease resistance gene and six

- homologs show pronounced allelic variation in Leu-rich repeat copy number. *Plant Cell* **10**: 1915–1925
- Domagala A, Kurpisz M** (2001) CD52 antigen: a review. *Med Sci Monit* **7**: 325–331
- Drew JE, Gatehouse JA** (1994) Isolation and characterization of a pea pod cDNA-encoding a putative blue copper protein correlated with lignin deposition. *J Exp Bot* **45**: 1873–1884
- Eisenhaber B, Bork P, Eisenhaber F** (1998) Sequence properties of GPI-anchored proteins near the  $\omega$ -site: constraints for the polypeptide binding site of the putative transamidase. *Protein Eng* **11**: 1155–1161
- Eisenhaber B, Bork P, Eisenhaber F** (1999) Prediction of potential GPI-modification sites in proprotein sequences. *J Mol Biol* **292**: 741–758
- Elkins T, Zinn K, McAllister L, Hoffmann FM, Goodman CS** (1990) Genetic analysis of a *Drosophila* neural cell adhesion molecule: interaction of fasciclin I and Abelson Tyr kinase mutations. *Cell* **60**: 565–575
- Engelman DM, Steitz TA, Goldman A** (1986) Identifying nonpolar transbilayer helices in amino acid-sequences of membrane-proteins. *Annu Rev Biophys Chem* **15**: 321–353
- Ezaki B, Gardner RC, Ezaki Y, Matsumoto H** (2000) Expression of aluminum-induced genes in transgenic Arabidopsis plants can ameliorate aluminum stress and/or oxidative stress. *Plant Physiol* **122**: 657–665
- Fry SC** (1998) Oxidative scission of plant cell wall polysaccharides by ascorbate-induced hydroxyl radicals. *Biochem J* **332**: 507–515
- Fry SC, Dumville JC, Miller JG** (2001) Fingerprinting of polysaccharides attacked by hydroxyl radicals in vitro and in the cell walls of ripening pear fruit. *Biochem J* **357**: 729–737
- Fry SC, Smith RC, Renwick KF, Martin DJ, Hodge SK, Matthews KJ** (1992) Xyloglucan endotransglycosylase, a new wall-loosening enzyme activity from plants. *Biochem J* **282**: 821–828
- Gaspar Y, Johnson KL, McKenna JA, Bacic A, Schultz CJ** (2001) The complex structures of arabinogalactan-proteins and the journey toward understanding function. *Plant Mol Biol* **47**: 161–176
- Greene EA, Erard M, Dedieu A, Barker DG** (1998) MtENOD16 and 20 are members of a family of phytoecyanin-related early nodulins. *Plant Mol Biol* **36**: 775–783
- Grenningloh G, Rehm EJ, Goodman CS** (1991) Genetic analysis of growth cone guidance in *Drosophila*: fasciclin II functions as a neuronal recognition molecule. *Cell* **67**: 45–57
- Griffith OH, Ryan M** (1999) Bacterial phosphatidylinositol-specific phospholipase C: structure, function, and interaction with lipids. *Biochim Biophys Acta* **1441**: 237–254
- Hamada K, Fukuchi S, Arisawa M, Baba M, Kitada K** (1998) Screening for glycosylphosphatidylinositol (GPI)-dependent cell wall proteins in *Saccharomyces cerevisiae*. *Mol Gen Genet* **258**: 53–59
- Hamada K, Terashima H, Arisawa M, Yabuki N, Kitada K** (1999) Amino acid residues in the  $\omega$ -minus region participate in cellular localization of yeast glycosylphosphatidylinositol-attached proteins. *J Bacteriol* **181**: 3886–3889
- Hart PJ, Nersissian AM, Herrmann RG, Nalbandyan RM, Valentine JS, Eisenberg D** (1996) A missing link in cupredoxins: crystal structure of cucumber stellacyanin at 1.6 Å resolution. *Protein Sci* **5**: 2175–2183
- Hauri H, Appenzeller C, Kuhn F, Nufer O** (2000) Lectins and traffic in the secretory pathway. *FEBS Lett* **476**: 32–37
- Herve C, Serres J, Dabos P, Canut H, Barre A, Rouge P, Lescure B** (1999) Characterization of the *Arabidopsis* lecRK-a genes: members of a superfamily encoding putative receptors with an extracellular domain homologous to legume lectins. *Plant Mol Biol* **39**: 671–682
- Huber O, Sumper M** (1994) Algal-CAMs: isoforms of a cell adhesion molecule in embryos of the alga Volvox with homology to *Drosophila* fasciclin I. *EMBO J* **13**: 4212–4222
- Huecas S, Villalba M, Rodriguez R** (2001) Ole e 9, a major olive pollen allergen is a 1,3- $\beta$ -glucanase: isolation, characterization, amino acid sequence, and tissue specificity. *J Biol Chem* **276**: 27959–27966
- Ikonen E** (2001) Roles of lipid rafts in membrane transport. *Curr Opin Cell Biol* **13**: 470–477
- Jeong S, Trotochaud AE, Clark SE** (1999) The Arabidopsis CLAVATA2 gene encodes a receptor-like protein required for the stability of the CLAVATA1 receptor-like kinase. *Plant Cell* **11**: 1925–1934
- Kader JC** (1996) Lipid transfer proteins in plants. *Annu Rev Plant Physiol Plant Mol Biol* **47**: 627–654
- Kader JC** (1997) Lipid-transfer proteins: a puzzling family of plant proteins. *Trends Plant Sci* **2**: 66–70
- Kawamoto T, Noshiro M, Shen M, Nakamasu K, Hashimoto K, Kawashima-Ohya Y, Gotoh O, Kato Y** (1998) Structural and phylogenetic analyses of RGD-CAP/ $\beta$ ig-h3, a fasciclin-like adhesion protein expressed in chick chondrocytes. *Biochim Biophys Acta* **1395**: 288–292
- Kieliszewski MJ** (2001) The latest hype on Hyp-O-glycosylation codes. *Phytochemistry* **57**: 319–323
- Kieliszewski MJ, Lamport DT** (1994) Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J* **5**: 157–172
- Kim JE, Kim SJ, Lee BH, Park RW, Kim KS, Kim IS** (2000) Identification of motifs for cell adhesion within the repeated domains of transforming growth factor- $\beta$ -induced gene,  $\beta$ ig-h3. *J Biol Chem* **275**: 30907–30915
- Kinoshita T, Inoue N** (2000) Dissecting and manipulating the pathway for glycosylphosphatidylinositol-anchor biosynthesis. *Curr Opin Chem Biol* **4**: 632–638
- Klarzynski O, Plesse B, Joubert JM, Yvin JC, Kopp M, Kloareg B, Fritig B** (2000) Linear  $\beta$ -1,3 glucans are elicitors of defense responses in tobacco. *Plant Physiol* **124**: 1027–1038
- Knox JP** (1995) The extracellular matrix in higher plants: developmentally regulated proteoglycans and glycoproteins of the plant cell surface. *FASEB J* **9**: 1004–1012
- Kohorn BD** (2000) Plasma membrane-cell wall contacts. *Plant Physiol* **124**: 31–38

- Kohorn BD** (2001) WAKs: cell wall associated kinases. *Curr Opin Cell Biol* **13**: 529–533
- Kollar R, Reinhold BB, Petrakova E, Yeh HJC, Ashwell G, Drgonova J, Kapteyn JC, Klis FM, Cabib E** (1997) Architecture of the yeast cell wall:  $\beta(1\rightarrow6)$ -glucan interconnects mannoprotein,  $\beta(1\rightarrow3)$ -glucan, and chitin. *J Biol Chem* **272**: 17762–17775
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL** (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567–580
- Kunze M, Riedel J, Lange U, Hurwitz R, Tischner R** (1997) Evidence for the presence of GPI-anchored PM-NR in leaves of *Beta vulgaris* and for PM-NR in barley leaves. *Plant Physiol Biochem* **35**: 507–512
- Le Gall AH, Yeaman C, Muesch A, Rodriguez-Boulan E** (1995) Epithelial cell polarity: new perspectives. *Semin Nephrol* **15**: 272–284
- Lind JL, Bacic A, Clarke AE, Anderson MA** (1994) A style-specific Hyp-rich glycoprotein with properties of both extensins and arabinogalactan proteins. *Plant J* **6**: 491–502
- Maidment JM, Moore D, Murphy GP, Murphy G, Clark IM** (1999) Matrix metalloproteinase homologs from Arabidopsis: expression and activity. *J Biol Chem* **274**: 34706–34710
- Majewska-Sawka A, Nothnagel EA** (2000) The multiple roles of arabinogalactan proteins in plant development. *Plant Physiol* **122**: 3–10
- Marmor MD, Julius M** (2000) The function of GPI-anchored proteins in T cell development, activation and regulation of homeostasis. *J Biol Regul Homeost Agents* **14**: 99–115
- Massova I, Kotra LP, Fridman R, Mobashery S** (1998) Matrix metalloproteinases: structures, evolution, and diversification. *FASEB J* **12**: 1075–1095
- McCabe PF, Valentine TA, Forsberg LS, Pennell RI** (1997) Soluble signals from cells identified at the cell wall establish a developmental pathway in carrot. *Plant Cell* **9**: 2225–2241
- McCarty DR, Chory J** (2000) Conservation and innovation in plant signaling pathways. *Cell* **103**: 201–209
- Miller JD, Arteca RN, Pell EJ** (1999) Senescence-associated gene expression during ozone-induced leaf senescence in Arabidopsis. *Plant Physiol* **120**: 1015–1024
- Mollet JC, Park SY, Nothnagel EA, Lord EM** (2000) A lily stylar pectin is necessary for pollen tube adhesion to an in vitro stylar matrix. *Plant Cell* **12**: 1737–1750
- Muniz M, Riezman H** (2000) Intracellular transport of GPI-anchored proteins. *EMBO J* **19**: 10–15
- Murakami S, Kondo Y, Nakano T, Sato F** (2000) Protease activity of CND41, a chloroplast nucleoid DNA-binding protein, isolated from cultured tobacco cells. *FEBS Lett* **468**: 15–18
- Nagase H, Woessner JF Jr** (1999) Matrix metalloproteinases. *J Biol Chem* **274**: 21491–21494
- Nakano T, Murakami S, Shoji T, Yoshida S, Yamada Y, Sato F** (1997) A novel protein with DNA binding activity from tobacco chloroplast nucleoids. *Plant Cell* **9**: 1673–1682
- Nakazato H, Okamoto T, Nishikoori M, Washio K, Morita N, Haraguchi K, Thompson GA Jr, Okuyama H** (1998) The glycosylphosphatidylinositol-anchored phosphatase from *Spirodela oligorrhiza* is a purple acid phosphatase. *Plant Physiol* **118**: 1015–1020
- Nersissian AM, Immoos C, Hill MG, Hart PJ, Williams G, Herrmann RG, Valentine JS** (1998) Uclacyanins, stellacyanins, and plantacyanins are distinct subfamilies of phytocyanins: plant-specific mononuclear blue copper proteins. *Protein Sci* **7**: 1915–1929
- Nersissian AM, Mehrabian ZB, Nalbandyan RM, Hart PJ, Fraczekiewicz G, Czernuszewicz RS, Bender CJ, Peisach J, Herrmann RG, Valentine JS** (1996) Cloning, expression, and spectroscopic characterization of *Cucumis sativus* stellacyanin in its nonglycosylated form. *Protein Sci* **5**: 2184–2192
- Neuteboom LW, Ng JM, Kuyper M, Clijdesdale OR, Hooykaas PJ, van der Zaal BJ** (1999) Isolation and characterization of cDNA clones corresponding with mRNAs that accumulate during auxin-induced lateral root formation. *Plant Mol Biol* **39**: 273–287
- Nielsen JE, Brunak S, von Heijne G** (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**: 1–6
- Nishikoori M, Washio K, Hase A, Morita N, Okuyama H** (2001) Cloning and characterization of cDNA of the GPI-anchored purple acid phosphatase and its root tissue distribution in *Spirodela oligorrhiza*. *Physiol Plant* **113**: 241–248
- Nothnagel EA** (1997) Proteoglycans and related components in plant cells. *Int Rev Cytol* **174**: 195–291
- Otegui M, Staehelin LA** (2000) Syncytial-type cell plates: a novel kind of cell plate involved in endosperm cellularization of Arabidopsis. *Plant Cell* **12**: 933–947
- Oxley D, Bacic A** (1999) Structure of the glycosylphosphatidylinositol anchor of an arabinogalactan protein from *Pyrus communis* suspension-cultured cells. *Proc Natl Acad Sci USA* **96**: 14246–14251
- Pak JH, Liu CY, Huangpu J, Graham JS** (1997) Construction and characterization of the soybean leaf metalloproteinase cDNA. *FEBS Lett* **404**: 283–288
- Park SY, Jauh GY, Mollet JC, Eckard KJ, Nothnagel EA, Walling LL, Lord EM** (2000) A lipid transfer-like protein is necessary for lily pollen tube adhesion to an in vitro stylar matrix. *Plant Cell* **12**: 151–164
- Peskan T, Westermann M, Oelmuller R** (2000) Identification of low-density Triton X-100-insoluble plasma membrane microdomains in higher plants. *Eur J Biochem* **267**: 6989–6995
- Porcella SF, Raffel SJ, Schrupf ME, Schriefer ME, Dennis DT, Schwan TG** (2000) Serodiagnosis of louse-borne relapsing fever with glycerophosphodiester phosphodiesterase (GlpQ) from *Borrelia recurrentis*. *J Clin Microbiol* **38**: 3561–3571
- Prokop A** (1999) Integrating bits and pieces: synapse structure and formation in *Drosophila* embryos. *Cell Tissue Res* **297**: 169–186
- Rhee SY, Somerville CR** (1998) Tetrad pollen formation in quartet mutants of *Arabidopsis* is associated with persis-

- tence of pectic polysaccharides of the pollen mother cell wall. *Plant J* **15**: 79–88
- Roy S, Jauh GY, Hepler PK, Lord EM** (1998) Effects of Yariv phenylglycoside on cell wall assembly in the lily pollen tube. *Planta* **204**: 450–458
- Sarma M** (2000) GDNF: a stranger in the TGF- $\beta$  superfamily? *Eur J Biochem* **267**: 6968–6971
- Samuels AL, Giddings TH Jr, Staehelin LA** (1995) Cytokinesis in tobacco BY-2 and root tip cells: a new model of cell plate formation in higher plants. *J Cell Biol* **130**: 1345–1357
- Schindelman G, Morikami A, Jung J, Baskin TI, Carpita NC, Derbyshire P, McCann MC, Benfey PN** (2001) COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in *Arabidopsis*. *Genes Dev* **15**: 1115–1127
- Schultz CJ, Hauser K, Lind JL, Atkinson AH, Pu ZY, Anderson MA, Clarke AE** (1997) Molecular characterization of a cDNA sequence encoding the backbone of a style-specific 120-kD glycoprotein which has features of both extensins and arabinogalactan proteins. *Plant Mol Biol* **35**: 833–845
- Schultz CJ, Johnson KL, Currie G, Bacic A** (2000) The classical arabinogalactan protein gene family of *Arabidopsis*. *Plant Cell* **12**: 1751–1768
- Segura A, Moreno M, Garcia-Olmedo F** (1993) Purification and antipathogenic activity of lipid transfer proteins (LTPs) from the leaves of *Arabidopsis* and spinach. *FEBS Lett* **332**: 243–246
- Sherrier DJ, Prime TA, Dupree P** (1999) Glycosylphosphatidylinositol-anchored cell-surface proteins from *Arabidopsis*. *Electrophoresis* **20**: 2027–2035
- Showalter AE** (2001) Arabinogalactan-proteins: structure, expression and function. *Cell Mol Life Sci* **58**: 1399–1417
- Shpak E, Barbar E, Leykam JF, Kieliszewski MJ** (2001) Contiguous Hyp residues direct Hyp arabinosylation in *Nicotiana tabacum*. *J Biol Chem* **276**: 11272–11278
- Shpak E, Leykam JF, Kieliszewski MJ** (1999) Synthetic genes for glycoprotein design and the elucidation of Hyp-O-glycosylation codes. *Proc Natl Acad Sci USA* **96**: 14736–14741
- Smirnoff N, Wheeler GL** (2000) Ascorbic acid in plants: biosynthesis and function. *Crit Rev Biochem Mol Biol* **35**: 291–314
- Sonnhammer E, von Heijne G, Krogh A** (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In J Glasgow, T Littlejohn, F Major, R Lathrop, D Sankoff, C Sensen, eds, *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology*. AAAI Press, Menlo Park, CA, pp 175–182
- Sterk P, Booij H, Schellekens GA, Van Kammen A, De Vries SC** (1991) Cell-specific expression of the carrot EP2 lipid transfer protein gene. *Plant Cell* **3**: 907–921
- Stintzi A, Heitz T, Prasad V, Wiedemann-Merdinoglu S, Kauffmann S, Geoffroy P, Legrand M, Fritig B** (1993) Plant “pathogenesis-related” proteins and their role in defense against pathogens. *Biochimie* **75**: 687–706
- Svetek J, Yadav MP, Nothnagel EA** (1999) Presence of a glycosylphosphatidylinositol lipid anchor on rose arabinogalactan proteins. *J Biol Chem* **274**: 14724–14733
- Takos AM, Dry IB, Soole KL** (1997) Detection of glycosylphosphatidylinositol-anchored proteins on the surface of *Nicotiana tabacum* protoplasts. *FEBS Lett* **405**: 1–4
- Takos AM, Dry IB, Soole KL** (2000) Glycosylphosphatidylinositol-anchor addition signals are processed in *Nicotiana tabacum*. *Plant J* **21**: 43–52
- The Arabidopsis Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis*. *Nature* **408**: 796–815
- Thoma S, Hecht U, Kippers A, Botella J, De Vries S, Somerville C** (1994) Tissue-specific expression of a gene encoding a cell wall-localized lipid transfer protein from *Arabidopsis*. *Plant Physiol* **105**: 35–45
- Thoma S, Kaneko Y, Somerville C** (1993) A non-specific lipid transfer protein from *Arabidopsis* is a cell wall protein. *Plant J* **3**: 427–436
- Thompson JD, Higgins DG, Gibson TJ** (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680
- Treacy BK, Hattori J, Prud’homme I, Barbour E, Boutilier K, Baszczyński CL, Huang B, Johnson DA, Miki BL** (1997) Bnm1, a Brassica pollen-specific gene. *Plant Mol Biol* **34**: 603–611
- Udenfriend S, Kodukula K** (1995) How glycosylphosphatidylinositol-anchored membrane proteins are made. *Annu Rev Biochem* **64**: 563–591
- Van den Bulcke M, Bauw G, Castresana C, Van Montagu M** (1989) Characterization of vacuolar and extracellular  $\beta$ -(1, 3)-glucanases of tobacco: evidence for a strictly compartmentalized plant defense system. *Proc Natl Acad Sci USA* **86**: 2673–2677
- van Gysel A, Van Montagu M, Inzé D** (1993) A negatively light-regulated gene from *Arabidopsis* encodes a protein showing high similarity to blue copper-binding proteins. *Gene* **136**: 79–85
- Vogeli-Lange R, Hansen-Gehri A, Boller T, Meins F** (1988) Induction of the defense-related glucanohydrolases,  $\beta$ -1,3-glucanase and chitinase, by tobacco mosaic virus infection of tobacco leaves. *Plant Sci* **4**: 171–176
- Wang X, Zafian P, Choudhary M, Lawton M** (1996) The PR5K receptor protein kinase from *Arabidopsis* is structurally related to a family of plant defense proteins. *Proc Natl Acad Sci USA* **93**: 2598–2602
- Wang Y** (2001) The role and regulation of urokinase-type plasminogen activator receptor gene expression in cancer invasion and metastasis. *Med Res Rev* **21**: 146–170
- Xu H, Davies SP, Kwan BY, O’Brien AP, Singh M, Knox RB** (1993) Haploid and diploid expression of a *Brassica campestris* anther-specific gene promoter in *Arabidopsis* and tobacco. *Mol Gen Genet* **239**: 58–65
- Xu H, Knox RB, Taylor PE, Singh MB** (1995) Bcp1, a gene required for male fertility in *Arabidopsis*. *Proc Natl Acad Sci USA* **92**: 2106–2110

- Youl JJ, Bacic A, Oxley D** (1998) Arabinogalactan-proteins from *Nicotiana glauca* and *Pyrus communis* contain glycosylphosphatidylinositol membrane anchors. *Proc Natl Acad Sci USA* **95**: 7921–7926
- Zeng X, Goetz JA, Suber LM, Scott WJ Jr, Schreiner CM, Robbins DJ** (2001) A freely diffusible form of Sonic hedgehog mediates long-range signaling. *Nature* **411**: 716–720
- Zhang Y, Sederoff RR, Allona I** (2000) Differential expression of genes encoding cell wall proteins in vascular tissues from vertical and bent loblolly pine trees. *Tree Physiol* **20**: 457–466
- Zheng B, Chen D, Farquhar MG** (2000) MIR16, a putative membrane glycerophosphodiester phosphodiesterase, interacts with RGS16. *Proc Natl Acad Sci USA* **97**: 3999–4004