



HAL
open science

Comparison of mass spectrometry data and bioinformatics predictions to assess the bona fide localization of proteins identified in cell wall proteomics studies

Artur Pinski, David Roujol, Cécile Pouzet, Luc Bordes, Hélène San Clemente, Laurent Hoffmann, Elisabeth Jamet

► To cite this version:

Artur Pinski, David Roujol, Cécile Pouzet, Luc Bordes, Hélène San Clemente, et al.. Comparison of mass spectrometry data and bioinformatics predictions to assess the bona fide localization of proteins identified in cell wall proteomics studies. *Plant Science*, 2021, 310, pp.110979. 10.1016/j.plantsci.2021.110979 . hal-03464644

HAL Id: hal-03464644

<https://ut3-toulouseinp.hal.science/hal-03464644>

Submitted on 2 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Comparison of mass spectrometry data and bioinformatics predictions to assess the *bona fide* localization of proteins identified in cell wall proteomics studies

Artur Pinski^{a,b}, David Roujol^a, Cécile Pouzet^c, Luc Bordes^c, Hélène San Clemente^a, Laurent Hoffmann^a, Elisabeth Jamet^{a,*}

Affiliations

^a *Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, Auzeville Tolosane, France*

^b *Institute of Biology, Biotechnology and Environmental Protection, Faculty of Natural Sciences, University of Silesia in Katowice, 40-032 Katowice, Poland*

^c *FR AIB-TRI Imaging Platform facilities, Université de Toulouse, CNRS, Auzeville Tolosane, France*

* Corresponding author:

Email address: jamet@lrsv.ups-tlse.fr (E. Jamet)

Keywords: bioinformatics prediction, cell wall, confocal microscopy, proteome, mass spectrometry, sub-cellular localization, TagRFP

Abbreviations: CWP, cell wall protein; GPI, glycosylphosphatidylinositol; PI-AP, GPI-anchored protein; SP, signal peptide

ABSTRACT

Plant cell walls have complex architectures made of polysaccharides among which cellulose, hemicelluloses, pectins and cell wall proteins (CWPs). Some CWPs are anchored in the plasma membrane through a glycosylphosphatidylinositol (GPI)-anchor. The secretion pathway is the classical route to reach the extracellular space. Based on experimental data, a canonical signal peptide (SP) has been defined, and bioinformatics tools allowing the prediction of the sub-cellular localization of proteins have been designed. In the same way, the presence of GPI-anchor attachment sites can be predicted using bioinformatics programs. This article aims at comparing the bioinformatics predictions of the sub-cellular localization of proteins assumed to be CWPs to mass spectrometry (MS) data. The sub-cellular localization of a few CWPs exhibiting particular features has been checked by cell biology approaches. Although the prediction of SP length is confirmed in most cases, it is less conclusive for GPI-anchors. Three main observations were done: (i) the variability observed at the N-terminus of a few mature CWPs could play a role in the regulation of their biological activity; (ii) one protein was shown to have a double sub-cellular localization in the cell wall and the chloroplasts; and (iii) peptides were found to be located at the C-terminus of several CWPs previously identified in GPI-anchored proteomes, [thus raising the issue of their actual anchoring to the plasma membrane.](#)

1. Introduction

Plant cell walls have complex architectures made of polymers of different kinds: polysaccharides among which cellulose, hemicelluloses and pectins [1]; cell wall proteins (CWPs) [2] and aromatic or lipidic compounds such as lignin [3] or suberin [4] in secondary walls synthesized after the achievement of the growing phase. Although present in minor amounts, less than 10% of the mass, the CWPs play critical roles in the modifications and remodeling of these polymers, as well as in signaling and in the overall scaffolding structuring the cell wall. Indeed, the composition and structure of plant cell walls can vary during plant development and upon biotic and abiotic stresses [5, 6]. Many proteomics studies especially focused on the extracellular compartment, either the cell wall, the apoplast or the plasma membrane outer surface have been developed over recent years (for reviews, see [2, 7, 8]). The plant cell wall protein content is now well-described with about half of the predicted CWPs of *Arabidopsis thaliana* identified at least once (see *WallProtDB*, <http://www.polebio.lrsv.ups-tlse.fr/WallProtDB/>, [9]).

The secretion pathway is the classical route to reach the extracellular space. A canonical signal peptide (SP) has been defined, based on experimental data [10, 11]. It comprises three domains: a short N-region (1-5 amino acids) which is positively charged, a H-region (7-15 amino acids) which is hydrophobic and has an α -helix structure, and an uncharged C-region (3-7 amino acids) which has a β -sheet structure. There is a short, negatively-charged region (1-6 amino acids) downstream the SP cleavage site. However, there might be a competition between different possible cleavage sites [12]. Some proteins can then be anchored in the plasma membrane through a glycosylphosphatidylinositol (GPI)-anchor and are called GPI-anchored proteins GPI-APs [13]. This GPI-anchor is grafted in the endoplasmic reticulum at a so-called ω -site located inside a C-terminal SP (C-SP). This C-SP contains a linker region of about 11 residues upstream $\omega-1$ which has no predicted secondary structure, a region from $\omega-1$ to $\omega+2$ containing small side-chain residues, a spacer region between $\omega+3$ and $\omega+9$, and a hydrophobic tail from $\omega+10$ to the C-terminus [14]. The GPI-anchor can be cleaved by phospholipases C or D, thus releasing the protein to the [apoplast](#) [15, 16]. This property has been used to characterize GPI-anchored proteomes (for a review, see [7]).

Most studies [have pointed to](#) the presence of unexpected proteins in cell wall proteomes obtained using (i) either destructive protocols leading to the purification of enriched cell walls fractions, (ii) or non-destructive protocols aiming at eluting the CWPs from intercellular spaces without breaking the plasma membrane [2]. The identification of proteins lacking canonical SP in cell walls has raised the question of the existence of alternative secretion

[pathways](#) [17-19]. Indeed in mammals, several proteins devoid of classical signal peptides could be found in extracellular spaces. Several alternative mechanisms of protein secretion have been proposed and partly demonstrated [20]. These findings have led to the design of the SecretomeP software to address this issue [21]. In plants, a single case has been documented. A sunflower jacalin-related protein devoid of *bona fide* SP has been identified in extracellular fluids and it has been immunolocalized in the cell wall [22]. Besides a few moonlighting proteins have been described, like the rice α -amylase [23] or an *A. thaliana* non-specific lipid transfer protein (AtLTP2/AtLTP1.4) [24]. These two proteins have been shown to be present in both cell walls and plastids.

Altogether, the sub-cellular localization of proteins can be predicted, but it is critical to check it experimentally. In this respect, biochemical as well as cell biology approaches can be used, *i.e.* purification of cell wall compartments or *in vivo* localization of proteins fusions using fluorescent tags or immunolocalization. However, the possible contribution of mass spectrometry (MS) data is still underestimated. This study aims at comparing the bioinformatics predictions of sub-cellular localization, either in the cell wall or at the cell surface, and MS data. The availability of large MS datasets has allowed conducting a statistical analysis to define the N- and the C-terminus of proteins identified in several cell wall proteomics studies. These studies were performed on proteins extracted with salt solutions from purified cell walls of two plants: *A. thaliana* roots, leaves and stems [25-29], and *Brachypodium distachyon* leaves and stem internodes [30, 31]. Most of the predictions were confirmed by the experimental mapping of peptides. However, we observed some variability around the N-terminus length of the mature proteins around the predicted SP processing site and the presence of the C-terminal peptides predicted to be cleaved for GPI-anchor grafting at the ω -site. The sub-cellular localization of [five](#) proteins showing such characteristics was checked by transient expression in *Nicotiana benthamiana* epidermal cells of fusion proteins carrying the TagRFP (Red Fluorescent Protein) or the T-Sapphire as fluorescent proteins stable at the acidic pH of the [apoplast](#) [32].

2. Material and Methods:

2.1. Bioinformatics

The MS data used for this work have been previously published in the following articles: *A. thaliana* roots [25], leaves [26-28], and floral stems [27, 29]; *B. distachyon* young and mature leaves and stems [30, 31]. They are all available in public repositories either at *ProticDB* (all MS data, <http://proteus.moulon.inra.fr/w2dpage/proticdb/angular/#/>, keyword: cell wall) or at *WallProtDB* (cell wall proteomics data, <http://www.polebio.lrsv.ups-tlse.fr/WallProtDB/>). The treatment of these datasets was adjusted to take into account both

tryptic and semi-tryptic peptides, *i.e.* peptides which were cleaved at the C-terminal side of arginine or lysine by trypsin at one end but not at the other. For the prediction of GPI anchoring sites of *A. thaliana* proteins, we have also used the data from Takahashi *et al.* [33].

The *ProtAnnDB* tool has been used to predict the sub-cellular localization of the identified proteins [34]. It collects the information provided by a series of bioinformatics programs such as: TargetP [35], SignalP [36], LocTree3 [37], Phobius [38], Predotar [39], TMHMM [40], GPIsom [41], PredGPI [14] and Aramemnon [42]. In this work, a protein is considered a possible CWP if it has a signal peptide predicted by at least two different bioinformatics programs, and if it is lacking an intracellular retention signal or a predicted trans-membrane domain.

2.1. Transient expression of fusion proteins in N. benthamiana leaf epidermal cells

Genomic DNA was prepared using the Wizard® Genomic DNA Purification kit according to the manufacturer's recommendations (Promega France, Charbonnières les Bains, France). Total RNAs were extracted using the SV Total RNA Isolation System (Promega) as recommended and cDNAs were synthesized using the ProtoScript® II First Strand cDNA Synthesis Kit (New England Biolabs France, Evry-Courcouronnes, France). The coding regions of the five genes of interest were amplified by PCR from *A. thaliana* genomic DNA when they had no intron (*At1g78820*, *At3g62820* and *At5g60490*) or from cDNAs prepared from leaf total RNA (*At5g59310* and *At4g37800*). The primers used for PCR amplification prior to cloning and for sequencing are listed in Table S1.

The full-length coding sequences of *At1g78820*, *At5g59310* and *At4g37800* were successively cloned by recombination in the pDONR207 (Invitrogen, Carlsberg, CA, USA) and in the pEAQ-gwTR vector [43] using the Gateway® technology to generate constructs allowing the production of “protein of interest::TagRFP” fusion proteins. The coding sequences of the predicted mature proteins encoded by *At3g62820* and *At5g60490* were successively cloned in the pDONR207 (Invitrogen) and in the pEAQ-SPTRgw [43] vectors to generate constructs allowing the production of “chitinase SP::TagRFP::mature protein of interest” fusion proteins. All the constructs have been sequenced at both cloning steps of cloning to ensure the absence of mutation (Eurofins Genomics, Les Ulis, France).

Additional constructs were generated for *At5g59310*. Site-directed mutagenesis was performed using the Q5® Site-Directed Mutagenesis kit (New England Biolabs, Ipswich, MA, USA). The first construct was obtained in the pEAQ-SPTRgw vector and allowed producing a protein deleted from its signal peptide “ Δ SP-*At5g59310*::TagRFP”. The second one was

obtained in a new version of the pEAQ vector called pEAQ-SPDgw-TSaph allowing the production of the following fusion protein: “SP-At5g11420::ΔSP-At5g59310::T-Sapphire” (See Fig. S1 for the description of the pEAQ-SPDgw-TSaph binary vector) [44].

The final constructs were then transferred to the *Agrobacterium tumefaciens* GV3101::pMP90 strain [45] as described [43]. The pm-yb CD3-1006 marker [46] was co-infiltrated with the newly designed constructs in order to clearly distinguish the plasma membrane from the cell wall after plasmolysis with glycerol 15%. Cell walls were stained with a calcofluor 0.1% (m/v) for 10 min.

Microscopy observations were performed 2.5 to 4 days after leaf infiltration [43]. A Leica TCS SP8 AOBS confocal laser-scanning microscope controlled by the Leica Application Confocal Software (LAS) (Leica, Nanterre, France) was used. Calcofluor was excited at 405 nm and detected in the 439–472 nm range. Chlorophyll was excited at 488 nm and detected in the 680–720 nm range. YFP was excited at 514 nm and detected in the 525–550 nm range. TagRFP was excited at 561 nm and detected in the 590–630 nm range. T-Sapphire was excited at 405 nm and detected in the 500–550 nm range. No image enhancement was performed. The images have been analyzed with Leica Application Suite X software (Leica Microsystems).

The presence of the fusion proteins in the *N. benthamiana* leaves was assessed by Western blot analysis. Briefly, total proteins were extracted with the following two-step procedure: (i) after addition of 2 μL/mg leaf material of sodium acetate buffer 4 mM pH 4.6, CaCl₂ 200 mM, β-mercaptoethanol 0.1% (v/v), protease inhibitor 0.15% (v/v) (Sigma-Aldrich Chemie SARL, Saint-Quentin Fallavier, France), the mixture was vortexed for 30 min; the same volume of denaturing solution (urea 6 M, glycerol 10% (m/v), β-mercaptoethanol 10% (v/v), SDS 5% (m/v)) was added and it was vortexed again for 15 min prior to centrifugation at 20,000 *g*. The supernatant was collected and it was centrifuged again at the same speed for 10 min. This new supernatant was analyzed by SDS-PAGE [47] and western blotting. The detection of the proteins of interest was done using TagRFP antibodies (Evrogen, Moscow, Russia) at a dilution of 1:5,000. Goat anti-rabbit secondary antibodies (Sigma-Aldrich Chemie SARL) coupled to alkaline phosphatase were used at a dilution of 1:10,000. A 10 mL solution containing NBT (33 μL of a 50 mg/mL solution in DMF 70% (v/v))/BCIP (33 μL of a 50 mg/mL solution in DMF 100% (v/v)) in TBS buffer pH 9.5 was used to detect the alkaline phosphatase activity.

3. Results and Discussion:

3.1. Bioinformatics analysis of MS data to experimentally define the N-terminus of proteins predicted to be secreted

Our analysis included MS data from different proteomics studies on roots, leaves or stems of two model plants, *A. thaliana* and *B. distachyon* according to the workflow described in Fig. S2. Altogether, they allowed the identification of 2,866 and 1,187 unique proteins, respectively, corresponding to 616,741 and 180,658 peptides, 26,246 and 11,046 unique peptides, respectively (Table 1). Out of these, 754 and 480 proteins, respectively, were predicted to be CWPs (Table 1). To evaluate the consistency between the prediction by bioinformatics programs of the SP cleavage site and the experimental data, we mainly focused on semi-tryptic peptides for which the N-terminal amino acid should result from the signal peptidase activity. For many proteins (about 75% of the CWPs), we were unable to find any peptides in the proximity of the predicted SP cleavage site marking these proteins as “undetermined” (Table 2). It should be noted that the full coverage of a given protein by MS analysis in a complex protein mixture is not often achieved because it depends on many parameters such as the abundance of the protein, the distribution of the tryptic cut sites, the physicochemical characteristics of the peptides and the mass spectrometer configuration. However, for 177 proteins of *A. thaliana* and 101 proteins of *B. distachyon*, we could identify peptides in the proximity of the predicted SP cleavage site. In most cases (on average 87%), the sequence of the identified peptides matched the prediction of at least one program (SignalP, TargetP or Phobius) and we classified these proteins as “SP as predicted” (Table 2). The analysis revealed differences in the accuracy of predictions made by the three bioinformatic tools showing that SignalP was the most accurate for *A. thaliana* - 136 out of 153 proteins (86%) classified as “SP as predicted” – closely followed by the predictions made by TargetP (132 correct predictions) (Table S2, sheet 3). For the *B. distachyon* proteins, SignalP predictions aligned with MS data for 79 out of 88 proteins (90%) and TargetP produced 78 correct predictions (Table S3, sheet 3). In both species, Phobius did not perform as well as TargetP and SignalP, but the differences were minute (125 and 77 correct predictions for *A. thaliana* and *B. distachyon*, respectively).

Further analysis of the MS data suggested shorter SP than predicted by bioinformatics for nine proteins in *A. thaliana* and two in *B. distachyon* (Table 2). The difference between predicted and observed peptides was usually less than three amino acids (Table S2, sheet 1 and Table S3, sheet 1). In four cases for *A. thaliana* and five for *B. distachyon*, MS data suggested SP longer than predicted (Table 2; Table S2, sheet 1; and Table S3, sheet 1). Interestingly, in 11 proteins in *A. thaliana* and six in *B. distachyon*, we observed peptides overlapping with the predicted SP cleavage site, which suggest that these proteins do not have SP, and thus, could not be CWPs (Table 2). To test this hypothesis, we selected three

among the latter *A. thaliana* proteins and studied their sub-cellular localization using TagRFP fusion proteins produced in *N. benthamiana* epidermal cells and observed by confocal microscopy (see section 3.2). In 40 and 29 proteins of *A. thaliana* and *B. distachyon*, respectively, we have observed variations in SP lengths that usually differ by a few amino acids at the C-terminus (Table 2). There were observed to vary from 2 to 10 amino acids (Tables S2 and S3). The trypsin activity cannot explain their existence as they do not have any arginine or lysine amino acid at their N-terminus. Interestingly, we observed this variation in SP cleavage site in both species in a few cases of lipid transfer proteins and GDSL lipase/esterase-like proteins (Table S2, sheet 2, and Table S3, sheet 2). In the case of three proteins in each species (At1g29670, At1g48750, At4g16260, Bradi3g30300, Bradi4g25750, and Bradi4g44410), we observed the presence of numerous semi-tryptic peptides that differ from one to 15 amino acids at the N-terminus of the mature proteins (Table 3, Tables S2 and S3). Considering the frequency of peptides presence in the MS data, we found that the peptide being the most frequent agrees with the bioinformatic prediction by SignalP, TargetP or Phobius.

For example, in At1g29670 (Table 3), the QAQAQVPCFFVFGDSLVDNGN>NNGLISIAR peptide indicating a cleavage at the site predicted by TargetP and Phobius (position 25) was identified 210-times in the MS data, while the other peptides sum up to 76-times, with the shortest peptide (underlined) being present 65-times. In the MS data, we also observed twice the peptide predicted by SignalP (position 27). The presence of the shortest peptide is particularly interesting. Indeed, At1g29670 is predicted to be a GDSL-motif lipase/esterase and this peptide (SLVDNGN>NNGLISIAR) lacks the first half of the GDSL conserved motif, where **S** constitutes the active site [48]. Another interesting case was that of At5g59310 which is annotated as a non-specific lipid transfer protein (LTP4/AtLTP1.11) [49]. Altogether 223 semi-tryptic peptides could be mapped to the N-terminus of the protein: 16 (7%) overlap the predicted SP, 101 (45%) match with the predicted SP, and 106 (48%) correspond to a shorter mature protein. Among the latter, 88 peptides (39%) do not contain one (Cys1) out of the eight conserved cysteine residues (Cys1-Cys8) which are involved in four disulfide bridges stabilizing the tertiary structure of non-specific lipid transfer proteins [50]. The lack of the Cys1-Cys5 disulfide bridge could lead to a non-functional protein.

The variability observed at the N-terminus of mature proteins could be explained by signal peptidase cutting in more than one site, as postulated by von Heijne (1984) [12]. Digestion at more than one site would lead to the N-terminal heterogeneity among the mature chains, as observed in our MS data. However, it should not be excluded that some of the identified semi-tryptic and non-tryptic peptides are the result of a peptidase activity either during the preparation of proteins extracts or *in muro*. Indeed, the proportion of semi-tryptic peptides

was evaluated to 9% in the analyzed MS data and proteases represent about 12% of cell wall proteomes [2]. The degradation of the N-terminus of the mature proteins could contribute to the regulation of their biological activity either by the removal of part of an active enzymatic site as for a lipase/acylesterase (At1g29670) or by the destabilization of a secondary structure as for a non-specific lipid transfer protein (At5g59310).

3.2. Bioinformatics analysis of MS data to experimentally define the C-terminus of proteins predicted to be GPI-APs

We complemented the N-terminal analysis with a comparison between GPI-APs ω -site prediction and MS data (see Fig. S2 for a description of the overall workflow). Among the putative CWPs, we identified 109 proteins predicted to be GPI-APs in *A. thaliana*, based on the review by Zhou [7]. According to our bioinformatic analysis, we also identified 126 putative GPI-APs in *B. distachyon* (Table 1). The absence of peptides overlapping the region downstream of the predicted ω -site indicates that the protein could be a GPI-AP, but does not prove it (see one example in Table 4). However, the presence of peptides overlapping the region downstream of the predicted ω -site, *i.e.* the C-SP, indicates an intact C-SP; thus, suggesting that the protein lacks the GPI-anchor. In our MS data combined with those from Takahashi *et al.* [33], we observed such peptides for *A. thaliana* in 24 proteins (see one example in Table 4; Table S2, sheets 4-6). Additionally, we have found peptides overlapping with the predicted C-SP cleavage site for three proteins, suggesting that they could not be GPI-APs (Table 2). Considering the current number of experimentally identified GPI-APs in *A. thaliana* summarized by Zhou (2019) [7] (327 proteins) reduced by the number of proteins for which we identified peptides overlapping with predicted ω -sites (24 proteins) and peptides overlapping with SP cleavage site (3 proteins), the final number of possible GPI-APs amounts to 300 (Table 2). Basing the prediction of protein-coding genes on Araport11 genome release from the TAIR database (27,655 protein-coding genes), the percentage of the possible GPI-APs is 1.08%. This percentage value falls within the values predicted for eukaryotes, that is 0.21% to 2.01% [7]. Finally, in *B. distachyon* data, we have found 29 peptides overlapping with the C-SP (Table 2). Altogether, 97 proteins remain as possible GPI-APs. To evaluate whether the predictions for GPI-APs were consistent with MS data, we selected two *A. thaliana* proteins with predicted C-SP to check their sub-cellular localization using fluorescent fusion proteins (see section 3.2). A similar approach in identifying GPI-APs by analyzing of the tryptic peptides overlapping with predicted ω -site was undertaken by Elorza *et al.* (2003) [16]. The authors did not observe such peptides for any of 44 putative GPI-APs in *A. thaliana*. Our results agree with these findings as we also did not found any peptides that could suggest that these proteins are not GPI-APs (34 out of 44 proteins described by Elorzta *et al.* (2003) [16] were identified in this study) (Table S2, sheet 4).

Our research shows how vital and still relevant are basic studies related to the proper identification of the GPI-APs, combining experimental and bioinformatic approaches, especially in plants. Among the proteins classified in our study as **probably** “not GPI-APs”, At4g35220, was previously shown to be localized in the cytosol by looking at the sub-cellular localization of a At4g35220::GFP fusion protein in stably transformed *A. thaliana* plants [51]. Interestingly, our analysis of the At4g35220 N-terminus suggested a SP shorter than predicted. However, this finding does not exclude the possibility of the lack of a SP that could direct At4g35220 to the **apoplast**. Besides, three proteins mentioned in Zhou (2019) [7] (At2g32240, At1g16240, and At5g46860) were disqualified from being GPI-APs due to the presence of peptides overlapping with the SP cleavage site. Regarding At2g32240, a semi-tryptic peptide identified in our analysis starts with the N-terminal methionine, thus excluding it from being a CWP, and subsequently a GPI-AP. Indeed, At2g32240 was shown to be anchored to the cytoplasmic surface of the endoplasmic reticulum membrane through a C-terminal transmembrane domain and a short tail domain, via a tail-anchoring mechanism. At1g16240 (SYNTAXIN-51, SYP51) has been shown to be an interaction partner of VESICLE-ASSOCIATED MEMBRANE PROTEIN 711 (VAMP711) possibly having a function as an “interfering” SNARE, indicating that it is not a CWP [52]. At5g46860 (SYNTAXIN-22, SYP22) was shown to be located on a subpopulation of prevacuolar compartments/endosomes closely associated with the vacuolar membrane [53].

While for *A. thaliana*, there are a few experimental analyses dedicated to identifying GPI-APs, none have been focused on *B. distachyon*. GPI-APs are essential for cell-to-cell signaling, cell differentiation and plant reproduction [54, 55]. As an example, Desnoyer and Palanivelu (2020) [55] highlighted the involvement of GPI-APs in *A. thaliana* gametophyte generation through an analysis of the expression profiles of the corresponding genes. Thus unravelling the roles of GPI-APs in grasses is of paramount importance. Such studies can be augmented through a specific computational approach to identify and characterize GPI-anchored peptides in the proteomic experiments [56]. The grass cell walls differ from dicot cell walls regarding the major structural polysaccharides, their linkages, and the abundance of phenolic compounds, proteins, and pectins [57]. As shown for the BdXFUC1 xyloglucan α -fucosidase, *A. thaliana* should not be considered a good model for *B. distachyon* and other species in which orthologs lack SP [58]. The authors also showed that *Bd2g51280* (*BdBGLC2*) encodes a membrane-bound xyloglucan β -glucosidase.

Altogether, the prediction of SP using classical bioinformatics programs usually fitted with the experimental N-terminus and Target-P was the program giving the best results. In some cases, the N-terminus peptides of mature proteins could be shorter or longer than predicted or the position of the N-terminus of mature proteins could be variable. The situation

was more contrasted with the prediction of the C-SP. Indeed, a significant proportion of proteins show peptides overlapping with the predicted C-SP. Also, there were important discrepancies between previously described phospholipase-released proteomes and the actual occurrence of C-terminal peptides.

3.3. Experimental sub-cellular localization of five proteins of interest

As mentioned above, we have selected three proteins showing peptides mapped in the predicted SP, and one protein with peptides downstream the predicted ω -site. We have also considered one protein with no peptide downstream the predicted ω -site as a positive control. Sub-cellular localization was performed using TagRFP or T-Sapphire fusion proteins produced in *N. benthamiana* epidermal cells and observation by confocal microscopy. The presence of the full-length fusion proteins in the observed tissues was checked by western blotting of total proteins using an antibody specific for the TagRFP for all the fusion proteins (Fig. S3).

The first set of three proteins predicted to be secreted, but identified with peptides located in their predicted SP was as follows: At1g78820 which is predicted to be a D-mannose binding lectin; At4g37800 which is annotated as AtXTH7, a xyloglucan endotransglucosylase/hydrolase [59]; and At5g59310 annotated as LTP4/AtLTP1.11 [49]. Their coding sequence was cloned in translational fusion with that encoding the TagRFP (see Fig S1 for the description of the constructs). In two cases, the fusion proteins were found to be located in cell walls, after plasmolysis of the cells: At4g37800::TagRFP (Fig. 1 A-F) and At1g78820::TagRFP (Fig. S4). The TagRFP fluorescence co-localized with that of calcofluor staining cellulose in cell walls (see Fig.1 B, C, and E for At4g37800::TagRFP; Fig S4 B, C, and E for At1g78820::TagRFP), but not to that of the pm-yb CD3-1006 marker specific for the plasma membrane (see Fig. 1 D and F for At4g37800::TagRFP; Fig. S4 D and F for At1g78820::TagRFP). In conclusion, both fusion proteins were targeted to the cell wall. The presence of a peptide overlapping the predicted SP could thus be due to incidental variability. Unfortunately, in both cases the presence of an arginine residue in the close proximity to the predicted SP cleavage site did not allow precisely mapping the N-terminus of the proteins.

Regarding At5g59310::TagRFP, the situation was more complicated (Fig. 1 G-L). Indeed the fusion protein was located in both the cell wall and the chloroplasts (Fig. 1 H, I, and K). No co-localization could be observed with the pm-yb CD3-1006 marker (Fig. 1 I, J and L). This double localization was already observed for another protein of the same family (At2g38530, AtLTP2/ AtLTP1.4) [24]. It was assumed that the protein was first targeted to the

secretory pathway, prior being sorted to the cell wall and the chloroplasts. Consistently with this finding, At2g38530 was also identified in a chloroplast proteome [60]. To dig deeper into this issue, we have made two additional constructs to produce (i) a fusion protein devoid of SP (Δ SP-At5g59310::TagRFP) and (ii) a fusion protein targeted to the cell wall by another SP, *i.e.* that of At5g11420 which was previously shown to target At5g11420 to the cell wall [61] (SP-At5g11420:: Δ SP-At5g59310::T-Sapphire). Δ SP-At5g59310::TagRFP could not be detected in the cell wall (Fig. 2 C, D and F), but could only be found in chloroplasts (Fig. 2 B, C, E and F). On the contrary, SP-At5g11420:: Δ SP-At5g59310::T-Sapphire was not found in chloroplasts (Fig. 2 H, I and K). It was only found in the cell wall as shown after the plasmolysis of the epidermal cells (Fig. 2 G and I-L). We were thus able to distinguish the two sub-cellular localizations. These new experiments show that the SP of At5g59310 was able to target the protein to the cell wall, but they also suggest that the combination of this SP and some signal inside the mature protein was required to target it at the same time to chloroplasts. Indeed, previous findings have underlined that the chloroplast transit peptide and N-terminal unfolded region in mature protein act together, allowing efficient targeting to the chloroplasts with an optimal total length of both elements estimated to be 76 amino acids [62]. Changing the *bona fide* SP of At5g59310 to that of At5g11420 prevented the double sub-cellular localization of the mature protein. Without its native SP, At5g59310 was only located in chloroplasts. It should be mentioned that a chloroplast transit peptide can be predicted at the N-terminus of the mature protein using TargetP with a probability of 0.831. This feature could explain the sole chloroplastic localization of Δ SP-At5g59310::TagRFP.

The second set of proteins included CWPs predicted to be GPI-APs: At3g62820 which is a pectin methylesterase inhibitor (PMEI) with a C-SP predicted by PredGPI (specificity: 99.7%; probable); and At5g60490 which is a fasciclin-like arabinogalactan protein (AtFLA12) predicted as a GPI-AP by both GPIsom and PredGPI (specificity: 100%; highly probable). Peptides overlapping the predicted C-SP were found in the MS data for At3g62820, but not for At5g60490 (see Table S2, sheets 1 and 5). For this particular study, the SP of the proteins was replaced with that of the At3g12500 chitinase [43, 63] thereafter named SPC, to focus on the C-SP functionality. SPC::TagRFP::At3g62820 was found to be located in the cell wall only as shown by observation after epidermal cell plasmolysis (see Fig. S5). On the contrary, SPC::TagRFP::At5g60490 was found to be located both at the plasma membrane and in the cell wall (Fig. 3 A-F). In addition, a weak fluorescence was observed in the **apoplast**, suggesting that the GPI-anchor of the fusion protein could be cleaved by a phospholipase. Phospholipases C and D have been identified in cell wall proteomes (see *WallProtDB*, <http://www.polebio.lrsv.ups-tlse.fr/WallProtDB/>). As for the other proteins, the presence of the full length fusion protein was checked by western blotting (see Fig. S2). In

the case of At5g60490, the molecular mass of the fusion protein was larger than expected most probably due to *N*- and *O*-glycosylations as assumed for AtFLA16 [64]. These results show that the prediction of C-SP was correct for At5g60490, but not for At3g62820 as inferred from the analysis of MS data. Another *A. thaliana* FLA predicted to be a GPI-AP (AtFLA4, also named SALT-OVERLY SENSITIVE 5 or SOS5) has been previously localized at the plasma membrane and in the mucilage pocket of mucilage mother cells after stable expression of a mCitrine::SOS5 construct [65]. [Such a double localization at the plasma membrane and in the cell wall was also observed for the *A. thaliana* LTPG GPI-AP \(At1g27950\) expressed as a YFP N-terminal fusion protein \[66\].](#) In conclusion, as for cell wall proteomes which contain many intracellular proteins [2], the identification of a protein in a GPI-anchored proteome does not seem to ensure that this protein is a *bona fide* GPI-AP and the bioinformatics predictions do not seem to be completely reliable. MS data and cell biology approaches can thus bring relevant information to determine if a protein can be considered a GPI-AP.

4. Conclusion

The starting point of this study was the search for an additional criterion to decide whether a protein identified in a cell wall proteome could be a *bona fide* CWP. The mining of MS data proved to be efficient to confirm SP predictions [for most of the analyzed](#) proteins (93%) using the usual bioinformatics programs like TargetP, SignalP and Phobius, SignalP being slightly more efficient. Some variability was observed at the N-terminus of some proteins which could be incidental or contribute to biological activity as for a GDSL lipase/acyltransferase or a non-specific transfer protein. The C-SP case was more complex with [24% of the analyzed](#) proteins predicted to be GPI-APs using PredGPI, GPIsom and Big-PI exhibiting peptides covering the C-SP. Then, some proteins identified in dedicated GPI-anchored proteomes were not confirmed to be GPI-APs. Altogether, in addition to identification and quantification, MS data could be more widely used to experimentally determine the N- and the C-terminus of proteins identified in the cell wall or GPI-anchored proteomes. In particular, they could allow checking the functionality of predicted GPI-anchoring ω -sites and highlight specific mechanisms of regulation of biological activity. [However, the issue of proteins possibly secreted through alternative pathways remains to be solved using experimental approaches. Dedicated plant-specific bioinformatic programs also need to be developed.](#)

Acknowledgements

The authors wish to thank the Paul Sabatier-Toulouse 3 University and the CNRS for supporting their research. AP has been the recipient of a four month-Erasmus fellowship for his stay in France. The authors also thank Marie Picard who has participated in the molecular biology work during her internship as a technician, Dr Eugénie Robe and Dr Philippe Ranocha for the gift of cDNAs, Thierry Balliau and Dr Michel Zivy for the MS analyses, as well as Dr Alexander Betekhtin and Pr Christophe Dunand for stimulating exchanges. The microscopy observations have been performed on the FR AIB Imaging platform (https://www.fraib.fr/fraib_eng/Technological-platforms/Cell-Imaging).

Contributions

AP and HSC have performed the bioinformatics work. AP and DR have done the molecular biology work. LH, DR, CP and LB have performed the microscopy observations. EJ has designed the experiments and written the first draft of the manuscript. All the authors have contributed to the writing of the final version.

The authors have all approved the manuscript and declared no conflict of interest.

References

- [1] N.C. Carpita, D.M. Gibeaut, Structural models of primary cell walls in flowering plants, consistency of molecular structure with the physical properties of the walls during growth, *Plant J.*, 3 (1993) 1-30, <https://doi.org/doi:10.1111/j.1365-313x.1993.tb00007.x>.
- [2] C. Albenne, H. Canut, E. Jamet, Plant cell wall proteomics: the leadership of *Arabidopsis thaliana*, *Front. Plant Sci.*, 4 (2013) 111, <https://doi.org/10.3389/fpls.2013.00111>.
- [3] C. Rathgeber, H. Cuny, P. Fonti, Biological basis of tree-ring formation: A crash course, *Front. Plant Sci.*, 7 (2016) 734, <https://doi.org/10.3389/fpls.2016.00734>.
- [4] T. Andersen, M. Barberon, N. Geldner, Suberization - The second life of an endodermal cell, *Curr. Opin. Plant Biol.*, 28 (2015) 9-15, <https://doi.org/10.1016/j.pbi.2015.08.004>.
- [5] L. Novaković, L. Guo, A. Bacic, A. Sampathkumar, K. Johnson, Hitting the wall—Sensing and signaling pathways involved in plant cell wall remodeling in response to abiotic stress, *Plants*, 7 (2018) 89, <https://doi.org/10.3390/plants7040089>.
- [6] H. Le Gall, F. Philippe, J.-M. Domon, F. Gillet, J. Pelloux, C. Rayon, Cell wall metabolism in response to abiotic stress, *Plants*, 4 (2015) 112-166, <https://doi.org/10.3390/plants4010112>.
- [7] K. Zhou, Glycosylphosphatidylinositol-anchored proteins in *Arabidopsis* and one of their common roles in signaling transduction, *Front. Plant Sci.*, 10 (2019) 1022, <https://doi.org/10.3389/fpls.2019.01022>.
- [8] M. Calderan-Rodrigues, J. Guimarães Fonseca, F. De Moraes, L. Vaz Setem, A. Carmanhanis Begossi, C. Labate, Plant cell wall proteomics: A focus on monocot species, *Brachypodium distachyon*, *Saccharum* spp, and *Oryza sativa*, *Int J Mol Sci*, 20 (2019) 1975, <https://doi.org/10.3390/ijms20081975>.

- [9] H. San Clemente, E. Jamet, *WallProtDB*, a database resource for plant cell wall proteomics, *Plant Methods*, 11 (2015) 2, <https://doi.org/doi:10.1186/s13007-015-0045-y>.
- [10] H. Owjia, N. Nezafata, M. Negahdaripoura, A. Hajiebrahimia, Y. Ghasemia, A comprehensive review of signal peptides: Structure, roles, and applications, *Eur J Cell Biol*, 97 (2018) 422-441, <https://doi.org/10.1016/j.ejcb.2018.06.003>.
- [11] G. von Heijne, Patterns of amino acids near signal-sequence cleavage sites, *FEBS J.*, 133 (1983) 17-21, <https://doi.org/10.1111/j.1432-1033.1983.tb07424.x>.
- [12] G. Von Heijne, How signal sequences maintain cleavage specificity, *J. Mol. Biol.*, 173 (1984) 243-251, [https://doi.org/10.1016/0022-2836\(84\)90192-x](https://doi.org/10.1016/0022-2836(84)90192-x).
- [13] A. Kinoshita, N. Inoue, Dissecting and manipulating the pathway for glycosylphosphatidylinositol-anchor biosynthesis, *Curr. Opin. Chem. Biol.*, 4 (2000) 632-638, [https://doi.org/10.1016/s1367-5931\(00\)00151-4](https://doi.org/10.1016/s1367-5931(00)00151-4).
- [14] A. Pierleoni, P.L. Martelli, R. Casadio, PredGPI: a GPI-anchor predictor, *BMC Bioinformatics*, 9 (2008) 392, <https://doi.org/10.1186/1471-2105-9-392>.
- [15] F. Elortza, M. Shabaz, J. Bunkenborg, L. Foster, T. Nühse, U. Brodbeck, S. Peck, O. Jensen, Modification-specific proteomics of plasma membrane proteins: Identification and characterization of glycosylphosphatidylinositol-anchored proteins released upon phospholipase D treatment, *J. Proteome Res.*, 5 (2006) 935-943, <https://doi.org/doi:10.1021/pr050419u>.
- [16] F. Elortza, T. Nühse, L. Foster, A. Stensballe, S. Peck, O. Jensen, Proteomic analysis of glycosylphosphatidylinositol-anchored membrane proteins, *Mol. Cell. Proteomics*, 2 (2003) 1261-1270, <https://doi.org/10.1074/mcp.M300079-MCP200>.
- [17] J.K.C. Rose, S.-J. Lee, Straying off the highway: Trafficking of secreted plant proteins and complexity in the plant cell wall proteome, *Plant Physiol.*, 153 (2010) 433-436, <https://doi.org/doi:10.1104/pp.110.154872>.
- [18] E. Jamet, C. Albenne, G. Boudart, M. Irshad, H. Canut, R. Pont-Lezica, Recent advances in plant cell wall proteomics, *Proteomics*, 8 (2008) 893-908, <https://doi.org/10.1002/pmic.200700938>.
- [19] D. Robinson, Y. Ding, L. Jiang, Unconventional protein secretion in plants: a critical assessment, *Protoplasma*, 253 (2016) 31-43, <https://doi.org/10.1007/s00709-015-0887-1>.
- [20] W. Nickel, M. Seedorf, Unconventional mechanisms of protein transport to the cell surface of eukaryotic cells, *Ann. Rev. Cell Dev. Biol.*, 24 (2008) 287-308, <https://doi.org/10.1146/annurev.cellbio.24.110707.175320>.
- [21] J.D. Bendtsen, L.J. Jensen, N. Blom, G. von Heijne, S. Brunak, Feature based prediction of non-classical and leaderless protein secretion, *Protein Eng. Des. Sel.*, 17 (2004) 349-356, <https://doi.org/10.1093/protein/gzh037>.
- [22] M. Pinedo, M. Regente, M. Elizalde, I. Quiroga, L.A. Pagnussat, J. Jorin-Novo, A. Maldonado, L. de la Canal, Extracellular sunflower proteins: evidence on non-classical secretion of a jacalin-related lectin, *Protein Pept. Lett.*, 19 (2012) 270-276, <https://doi.org/10.2174/092986612799363163>.
- [23] M.H. Chen, L.F. Huang, H.M. Li, Y.R. Chen, S.M. Yu, Signal peptide-dependent targeting of a rice alpha-amylase and cargo proteins to plastids and extracellular compartments of plant cells, *Plant Physiol.*, 135 (2004) 1367-1377, <https://doi.org/10.1104/pp.104.042184>.
- [24] A. Jacq, C. Pernot, Y. Martinez, F. Domergue, B. Payré, E. Jamet, V. Burlat, V. Pacquit, The Arabidopsis Lipid Transfer Protein 2 (AtLTP2) is involved in cuticle-cell wall interface integrity and in etiolated hypocotyl permeability, *Front. Plant Sci.*, 8 (2017) 263, <https://doi.org/10.3389/fpls.2017.00263>.
- [25] H. Nguyen-Kim, H. San Clemente, T. Balliau, M. Zivy, C. Dunand, C. Albenne, E. Jamet, *Arabidopsis thaliana* root cell wall proteomics: Increasing the proteome coverage using a combinatorial peptide ligand library and description of unexpected Hyp in peroxidase amino acid sequences, *Proteomics*, 16 (2016) 491-503, <https://doi.org/10.1002/pmic.201500129>.
- [26] V. Hervé, H. Duruflé, H. San Clemente, C. Albenne, T. Balliau, M. Zivy, C. Dunand, E. Jamet, An enlarged cell wall proteome of *Arabidopsis thaliana* rosettes, *Proteomics*, 16 (2016) 3183-3187, <https://doi.org/10.1002/pmic.201600290>.

- [27] H. Duruflé, P. Ranocha, T. Balliau, C. Dunand, E. Jamet, Transcriptomic and cell wall proteomic datasets of rosettes and floral stems from five *Arabidopsis thaliana* ecotypes grown at optimal or sub-optimal temperature, *Data Brief*, 27 (2019) 104581, <https://doi.org/10.1016/j.dib.2019.104581>.
- [28] H. Duruflé, V. Hervé, P. Ranocha, T. Balliau, M. Zivy, J. Chourré, H. San Clemente, V. Burlat, C. Albenne, S. Déjean, E. Jamet, C. Dunand, Cell wall adaptation of two contrasted ecotypes of *Arabidopsis thaliana*, Col and Sha, to sub-optimal growth conditions: an integrative study, *Plant Sci.*, 263 (2017) 183-193, <https://doi.org/10.1016/j.plantsci.2017.07.015>.
- [29] H. Duruflé, H. San Clemente, T. Balliau, M. Zivy, C. Dunand, E. Jamet, Cell wall proteome analysis of *Arabidopsis thaliana* mature stems, *Proteomics*, (2017) <https://doi.org/10.1002/pmic.201600449>.
- [30] T. Douché, H. San Clemente, V. Burlat, D. Roujol, B. Valot, M. Zivy, R. Pont-Lezica, E. Jamet, *Brachypodium distachyon* as a model plant toward improved biofuel crops: Search for secreted proteins involved in biogenesis and disassembly of cell wall polymers, *Proteomics*, 13 (2013) 2438-2454, <https://doi.org/10.1002/pmic.201200507>.
- [31] T. Douché, B. Valot, T. Balliau, H. San Clemente, M. Zivy, E. Jamet, Cell wall proteomic datasets of stems and leaves of *Brachypodium distachyon*, *Data Brief*, (2021) <https://doi.org/10.1016/j.dib.2021.106818>.
- [32] A. Stoddard, V. Rolland, I see the light! Fluorescent proteins suitable for cell wall/apoplast targeting in *Nicotiana benthamiana* leaves, *Plant Direct*, 3 (2019) e00112, <https://doi.org/10.1002/pld3.112>.
- [33] D. Takahashi, Y. Kawamura, M. Uemura, Cold acclimation is accompanied by complex responses of glycosylphosphatidylinositol (GPI)-anchored proteins in *Arabidopsis*, *J. Exp. Bot.*, 67 (2016) 5203-5215, <https://doi.org/10.1093/jxb/erw279>.
- [34] H. San Clemente, R. Pont-Lezica, E. Jamet, Bioinformatics as a tool for assessing the quality of sub-cellular proteomic strategies and inferring functions of proteins: plant cell wall proteomics as a test case, *Bioinform. Biol. Insights*, 3 (2009) 15-28, <https://doi.org/10.4137/bbi.s2065>.
- [35] O. Emanuelsson, S. Brunak, G. Von Heijne, H. Nielsen, Locating proteins in the cell using TargetP, SignalP and related tools, *Nat. Protoc.*, 2 (2007) 953-971, <https://doi.org/10.1038/nprot.2007.131>.
- [36] J. Almagro Armenteros, K. Tsirigos, C. Sønderby, T. Petersen, O. Winther, S. Brunak, G. von Heijne, H. Nielsen, SignalP 5.0 improves signal peptide predictions using deep neural networks, *Nature Biotechnol.*, 37 (2019) 420-423, <https://doi.org/10.1038/s41587-019-0036-z>.
- [37] T. Goldberg, M. Hecht, T. Hamp, T. Karl, G. Yachdav, N. Ahmed, U. Altermann, P. Angerer, S. Ansorge, K. Balasz, M. Bernhofer, A. Betz, L. Cizmadija, K. Do, J. Gerke, R. Greil, V. Joerdens, M. Hastreiter, K. Hembach, M. Herzog, M. Kalemanov, M. Kluge, A. Meier, H. Nasir, U. Neumaier, V. Prade, J. Reeb, A. Sorokoumov, I. Troshani, S. Vorberg, S. Waldraff, J. Zierer, H. Nielsen, B. Rost, LocTree3 prediction of localization, *Nucleic Acids Res.*, 42(Web Server issue) (2014) W350–W355, <https://doi.org/10.1093/nar/gku396>.
- [38] L. Käll, A. Krogh, E. Sonnhammer, A combined transmembrane topology and signal peptide prediction method, *J. Mol. Biol.*, 338 (2004) 1027-1036, <https://doi.org/10.1016/j.jmb.2004.03.016>.
- [39] I. Small, N. Peeters, F. Legeai, C. Lurin, Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences, *Proteomics*, 4 (2004) 1581-1590, <https://doi.org/10.1002/pmic.200300776>.
- [40] E.L. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences, *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 6 (1998) 175-182.
- [41] N. Fankhauser, P. Mäser, Identification of GPI anchor attachment signals by a Kohonen self-organizing map, *Bioinformatics*, 21 (2005) 1846-1852, <https://doi.org/10.1093/bioinformatics/bti299>.

- [42] R. Schwacke, A. Schneider, E. van der Graaff, K. Fischer, E. Catoni, M. Desimone, W.B. Frommer, U.I. Flugge, R. Kunze, ARAMEMNON, a novel database for Arabidopsis integral membrane proteins, *Plant Physiol.*, 131 (2003) 16-26, <https://doi.org/10.1104/pp.011577>.
- [43] F. Berthold, D. Roujol, C. Hemmer, E. Jamet, C. Ritzenthaler, L. Hoffmann, C. Schmitt-Keichinger, Inside or outside? A new collection of Gateway vectors allowing plant protein subcellular localization or over-expression, *Plasmid*, 105 (2019) 102436, <https://doi.org/10.1016/j.plasmid.2019.102436>.
- [44] O. Zapata-Hommer, O. Griesbeck, Efficiently folding and circularly permuted variants of the Sapphire mutant of GFP, *BMC Biotechnol.*, 3 (2003) 5, <https://doi.org/10.1186/1472-6750-3-5>.
- [45] M. Shamloul, J. Trusa, V. Mett, V. Yusibov, Optimization and utilization of *Agrobacterium*-mediated transient protein production in *Nicotiana*, *J. Vis. Exp.*, 86 (2014) e51204, <https://doi.org/10.3791/51204>.
- [46] B. Nelson, X. Cai, A. Nebenführ, A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants, *Plant J.*, 51 (2007) 1126-1136, <https://doi.org/10.1111/j.1365-313X.2007.03212.x>.
- [47] U. Laemmli, Cleavage of the structural proteins during the assembly of the head of bacteriophage T4, *Nature*, 227 (1970) 680-685, <https://doi.org/10.1038/227680a0>.
- [48] L. Ding, M. Li, W. Wang, J. Cao, Z. Wang, K. Zhui, Y. Yang, Y. Li, X. Tan, Advances in plant GDSL lipases: from sequences to functional mechanisms, *Acta Physiol Plant*, 41 (2019) 151, <https://doi.org/10.1007/s11738-019-2944-4>.
- [49] K. Chae, B. Gonong, S. Kim, C. Kieslich, D. Morikis, S. Balasubramanian, E. Lord, A multifaceted study of stigma/style cysteine-rich adhesin (SCA)-like Arabidopsis lipid transfer proteins (LTPs) suggests diversified roles for these LTPs in plant growth and reproduction, *J. Exp. Bot.*, 61 (2010) 4277-4290, <https://doi.org/10.1093/jxb/erq228>.
- [50] N. Wang, C. Lee, C. Cheng, W. Lo, Y. Yang, M. Chen, Construction and analysis of a plant non-specific lipid transfer protein database (nsLTPDB), *Bmc Genomics*, 13 (Suppl 1) (2012) S9, <https://doi.org/10.1186/1471-2164-13-S1-S9>.
- [51] E. Davis, D. Chen, D. DeWald, J. Shope, J. MacAdama, Y. Wu, Searching for new cell wall protein genes in Arabidopsis, *Plant Physiol Biochem*, 47 (2009) 81-85, <https://doi.org/10.1016/j.plaphy.2008.10.003>.
- [52] W. Heard, J. Sklena, D. Tome, S. Robatzek, A. Jones, Identification of regulatory and cargo proteins of endosomal and secretory pathways in *Arabidopsis thaliana* by proteomic dissection, *Mol. Cell. Proteomics*, 14 (2015) 1796-1813, <https://doi.org/10.1074/mcp.M115.050286>.
- [53] K. Ebine, Y. Okatani, T. Uemura, T. Goh, K. Shoda, M. Niihama, M. Morita, C. Spitzer, M. Otegui, A. Nakano, T. Ueda, A SNARE complex unique to seed plants is required for protein storage vacuole biogenesis and seed development of *Arabidopsis thaliana*, *Plant Cell*, 20 (2008) 3006-3021, <https://doi.org/10.1105/tpc.107.057711>.
- [54] G.H. Borner, K.S. Lilley, T.J. Stevens, P. Dupree, Identification of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A proteomic and genomic analysis, *Plant Physiol.*, 132 (2003) 568-577, <https://doi.org/doi:10.1104/pp.103.021170>.
- [55] N. Desnoyer, R. Palanivelu, Bridging the GAPs in plant reproduction: a comparison of plant and animal GPI-anchored proteins, *Plant Reprod*, 33 (2020) 129-142, <https://doi.org/10.1007/s00497-020-00395-9>.
- [56] M. Omaetxebarria, F. Elortza, E. Rodríguez-Suárez, K. Aloria, J. Arizmendi, O. Jensen, R. Matthiesen, Computational approach for identification and characterization of GPI-anchored peptides in proteomics experiments, *Proteomics*, 7 (2007) 1951-1960, <https://doi.org/10.1002/pmic.200700068>.
- [57] J. Vogel, Unique aspects of the grass cell wall, *Curr. Opin. Plant Biol.*, 11 (2008) 301-307, <https://doi.org/10.1016/j.pbi.2008.03.002>.
- [58] D. Rubianes, E. Valdivia, G. Revilla, I. Zarra, J. Sampedro, Xyloglucan exoglycosidases in the monocot model *Brachypodium distachyon* and the conservation of xyloglucan disassembly in

- angiosperms, *Plant Mol. Biol.*, 100 (2019) 495-509, <https://doi.org/10.1007/s11103-019-00875-1>.
- [59] J.K. Rose, J. Braam, S.C. Fry, K. Nishitani, The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature, *Plant Cell. Physiol.*, 43 (2002) 1421-1435, <https://doi.org/10.1093/pcp/pcf171>.
- [60] M. Ferro, [AT_CHLORO](#), a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins, *Mol. Cell. Proteomics*, 9 (2010) 1063-1084, <https://doi.org/10.1074/mcp.M900325-MCP200>.
- [61] C. Albenne, H. Canut, L. Hoffmann, E. Jamet, Plant cell wall proteins: a large body of data, but what about runaways?, *Proteomes*, 2 (2014) 224-242, <https://doi.org/10.3390/proteomes2020224>.
- [62] B. Shen, C. Zhu, Z. Yao, L. Cui, J. Zhang, C. Yang, Z. He, X. Peng, An optimized transit peptide for effective targeting of diverse foreign proteins into chloroplasts in rice, *Sci. Rep.*, 7 (2017) <https://doi.org/10.1038/srep46231>.
- [63] J. Haseloff, K. Siemering, D. Prasher, S. Hodge, Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly, *Proc. Natl. Acad. Sci. USA*, 94 (1997) 2122-2127, <https://doi.org/10.1073/pnas.94.6.2122>.
- [64] E. Liu, C. MacMillan, T. Shafee, Y. Ma, J. Ratcliffe, A. van de Meene, A. Bacic, J. Humphries, K. Johnson, Fasciclin-Like Arabinogalactan-protein 16 (FLA16) is required for stem development in Arabidopsis, *Front. Plant Sci.*, 11 (2020) 615392, <https://doi.org/10.3389/fpls.2020.615392>.
- [65] J. Griffiths, M. Crepeau, M. Ralet, G. Seifert, H. North, Dissecting seed mucilage adherence mediated by FEI and SOS5, *Front. Plant Sci.*, 7 (2016) 1073, <https://doi.org/10.3389/fpls.2016.01073>.
- [66] A. DeBono, T. Yeats, J. Rose, D. Bird, R. Reinhard Jetter, L. Kunst, L. Samuels, Arabidopsis LTPG is a glycosylphosphatidylinositol-anchored lipid transfer protein required for export of lipids to the plant surface, *Plant Cell*, 21 (2009) 1230-1238, <https://doi.org/10.1105/tpc.108.064451>.

Table 1. Overview of the analyzed data resulting from MS and bioinformatics analyses.

	<i>A. thaliana</i> ^a	<i>B. distachyon</i> ^a
Number of peptides	616,741	180,658
Number of unique peptides	26,246	11,046
Number of unique proteins	2,866	1,187
Number of predicted CWPs	754	480
Number of predicted GPI-APs	109 (out of 324 ^b)	126 ^c

- a. See Material and Methods for the description of the experimental work.
- b. Total number of proteins assumed to be GPI-APs as in [7].
- c. Proteins predicted to be GPI-APs according to the bioinformatics programs listed in Material and Methods.

Table 2. Comparison of MS data to bioinformatic predictions of presence of peptide signal (SP) or of GPI-APs ω -site prediction.

	<i>A. thaliana</i> ^a	<i>B. distachyon</i> ^a
CWPs^b		
SP as predicted	153	88
SP shorter than predicted	9	2
Presence of peptides overlapping with the predicted SP cleavage site	11	6
SP longer than predicted	4	5
Variation in SP length	40	29
Undetermined	577	379
Possible CWPs	743	474
GPI-APs^c		
Presence of peptides overlapping with region downstream of the ω -site	24	29
Presence of peptides overlapping with SP cleavage site	3	0
Possible GPI-APs	82 (297 ^d)	97

a. See Material and Methods for the description of the experimental work.

b. The total number of analyzed CWPs were 754 and 480 for *A. thaliana* and *B. distachyon*, respectively (see Table 1, penultimate line).

c. The total number of analyzed GPI-APs were 109 and 126 for *A. thaliana* and *B. distachyon*, respectively (see Table 1, last line).

d. Number of possible GPI-APs according to Zhou [7] and our MS data (324-27, see explanation in the text).

Table 3. Prediction of SP vs MS data: Examples of the different cases encountered during the comparison.

The amino-acid sequences in the upper rows are those of the N-terminus of the proteins. The sequences highlighted in grey correspond to SP bioinformatic prediction with TargetP. The tryptic cut sites are underlined (K, R). The sequences below are those of the peptides identified by MS.

SP as predicted: At1g02335 (germin)

MMNSRISIIIALSCIMITSIRAYDPDALQDLCVADKSHGTKLNGFPCKET
YDPDALQDLCVADK

SP shorter than predicted: At1g71950 (protease inhibitor)

MQTFAPGTRVYHLIIIVFFFFFCLSSFFASIVMADEYTGEATGSSEAKVH
ADEYTGEATGSSEAKK

Presence of peptides overlapping the predicted SP cleavage site: At4g37800 (XTH7)

MVVSLFSSRNVFYTLSLCLFAALYQPVMSRPAKFEDDFRIAWSDTHITQI
LSLCLFAALYQPVMSRR

SP longer than predicted: Bradi5g13550 (GH5)

MASSLTPLLVLGLLLVLASARAISISLPALPLSTASRWVVGADGRRVKL
ISLPALPLSTASR

Variation in the N-terminus length:

At1g29670 (GDSL lipase/esterase-like)

MESYLTKWCVVLVLLCFGFSVVKAQAQAQVPCFFVFGDSLVDNGNNGLISAR
QAQAQVPCFFVFGDSLVDNGNNGLISAR
VPCFFVFGDSLVDNGNNGLISAR
FFVFGDSLVDNGNNGLISAR
VFGDSLVDNGNNGLISAR
GDSLVDNGNNGLISAR
SLVDNGNNGLISAR

At3g12145 (LRR domains)

MKLFVHLSIFFSILFITLPPSSYSCTENDKNALLQIKKALGNPPLLSSWNPR
SSYSCTENDKNALLQIKKALGNPPLLSSWNPR
CTENDKNALLQIKK
TENDKNALLQIKK

Undetermined: At1g11820 (GH17)

MAFTSMVSTVPVLFVFFFTLLLSANSSSLSHNIKVQEQDKDPFVGFNIGTDVSNLLSPTELVK
VQEQDKDPFVGFNIGTDVSNLLSPTELVK

Table 4. Prediction of GPI-APs ω -site prediction vs MS data: Examples of the different cases encountered during the comparison.

The amino-acid sequences in the upper rows are those of the C-terminus of the proteins. The sequences highlighted in grey correspond to C-SP bioinformatic prediction with GPIsom. The amino-acids in bold italics are the predicted ω -sites. The tryptic cut sites are underlined (K, R). The sequences below are those of the peptides identified by MS.

Possible GPI-AP: At3g06035 (unknown function)

GIGIGKEDDWIVVVLTTNTPEGSYSTATPTKQES**NG**FTFGIGLVSYLVIFMYSSFCFFLF [stop]
TTNTPEGSYSTATPTK

**Presence of peptides overlapping with region downstream of the ω -site:
 At1g71950 (SCPL46, serine carboxypeptidase)**

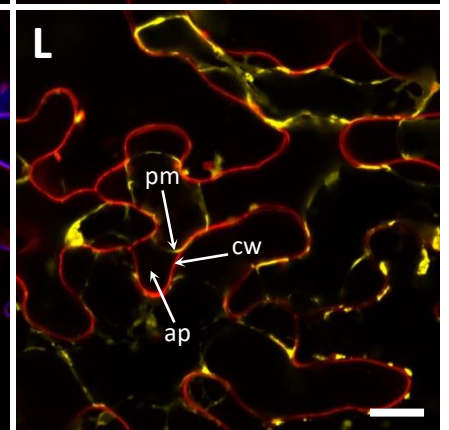
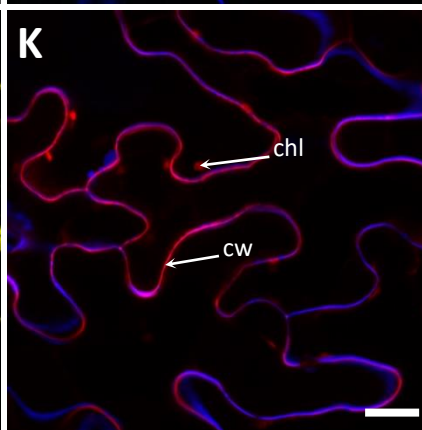
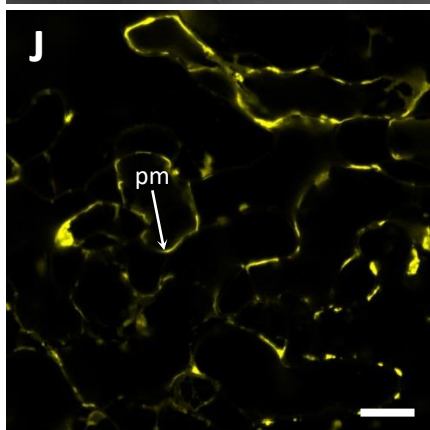
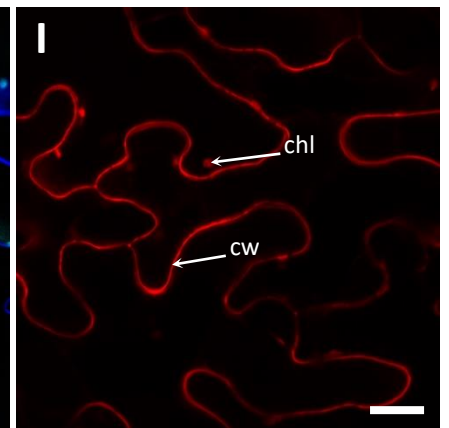
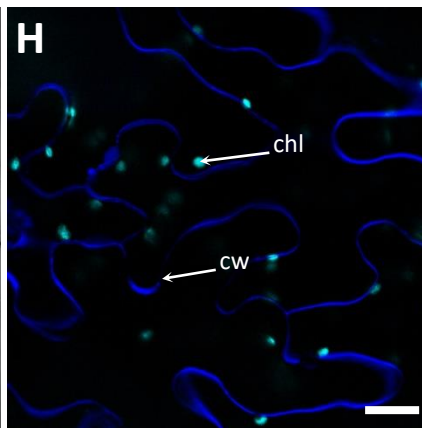
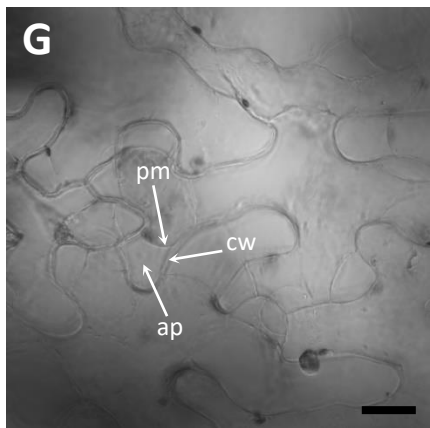
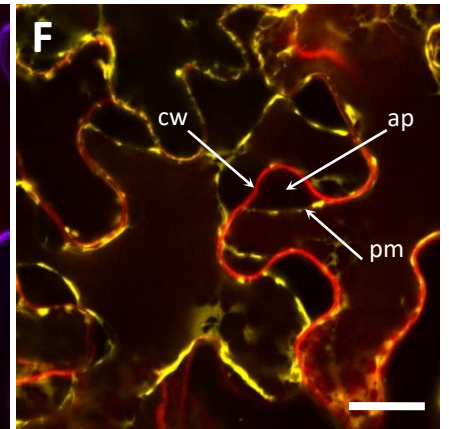
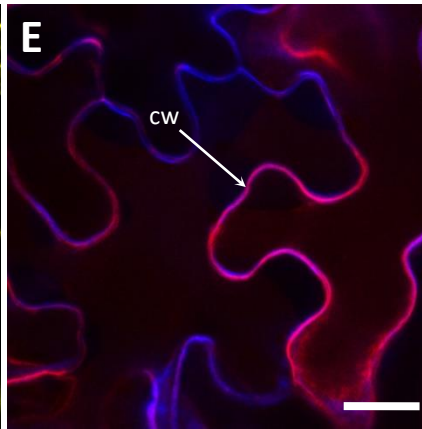
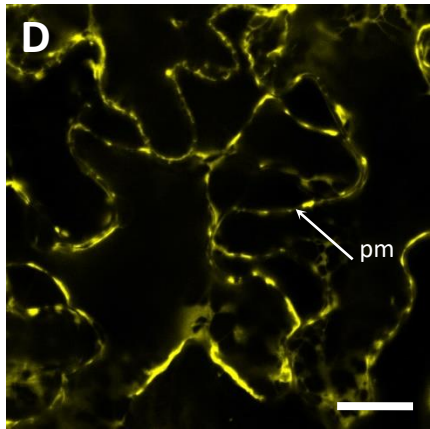
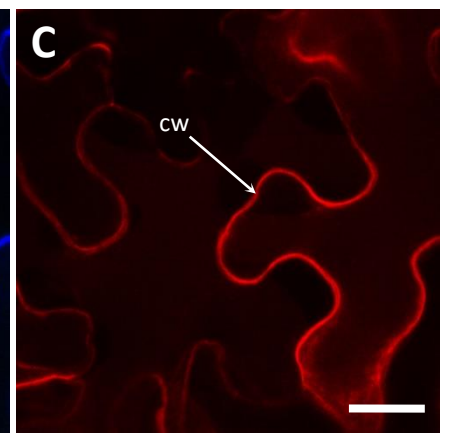
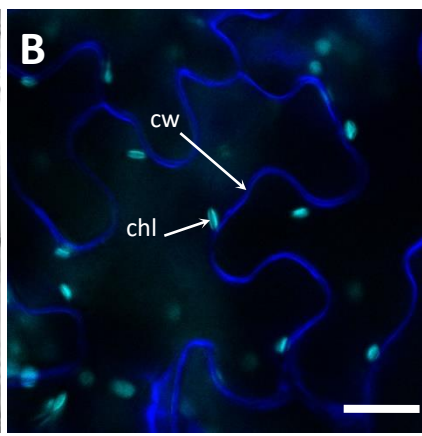
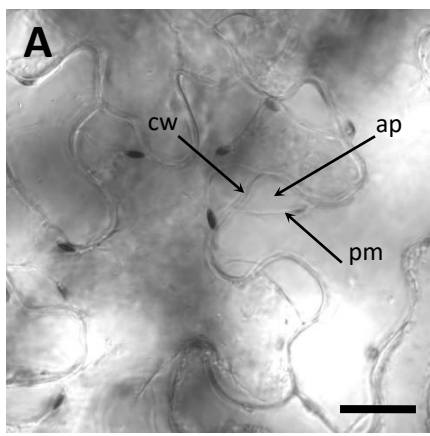
VPYRVWFAGQQVGGWTQVYGNTLAFATVR**GA**AHEVPFSQPARALVLFKAFLGGRPLPEEF [stop]
AFLGGRPLPEEF

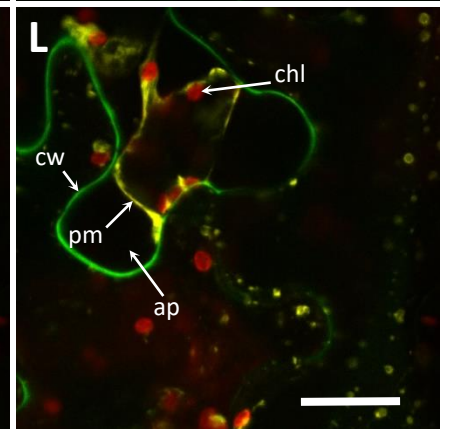
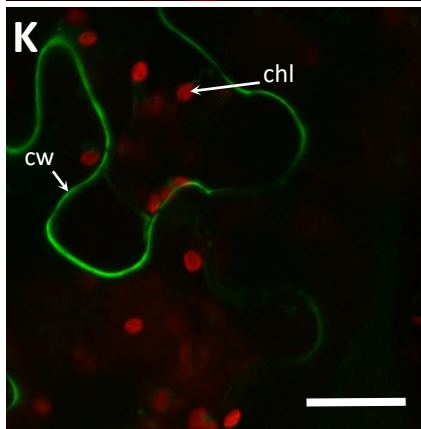
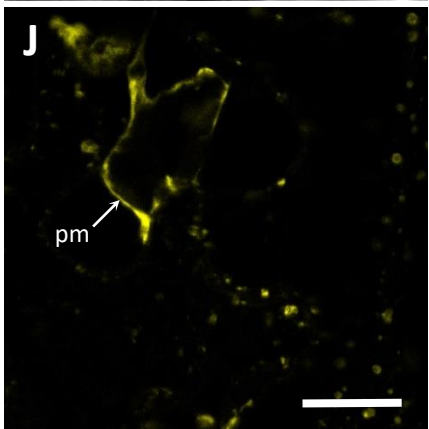
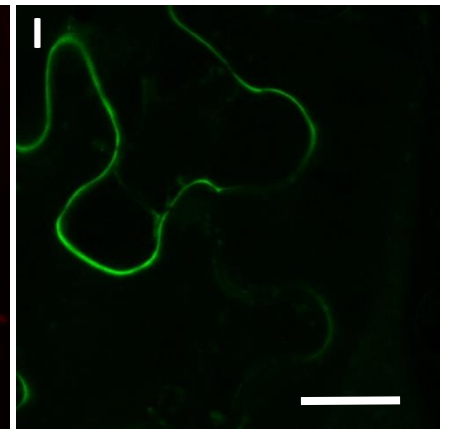
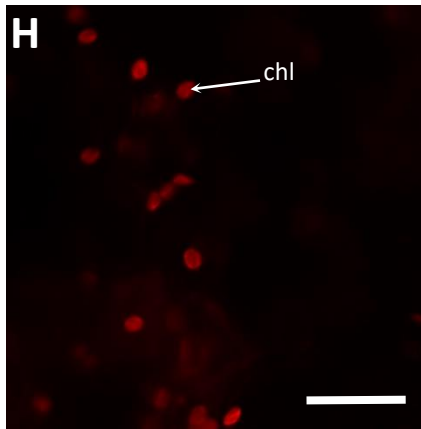
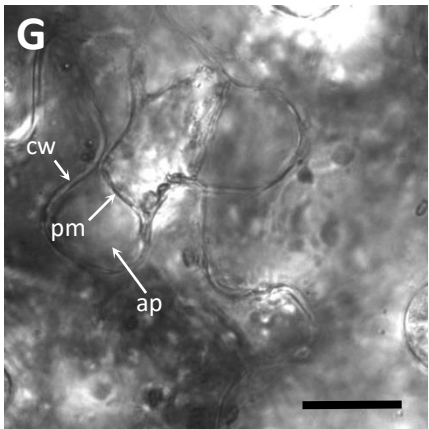
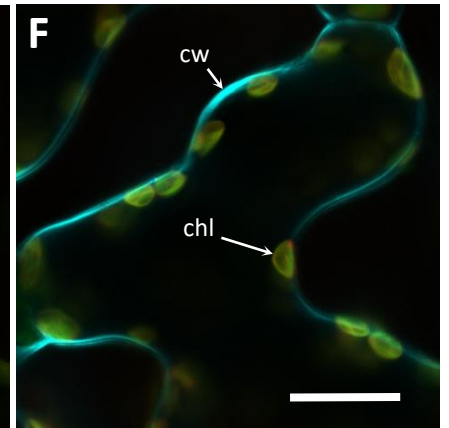
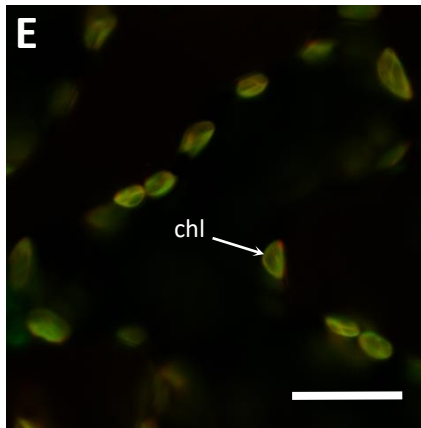
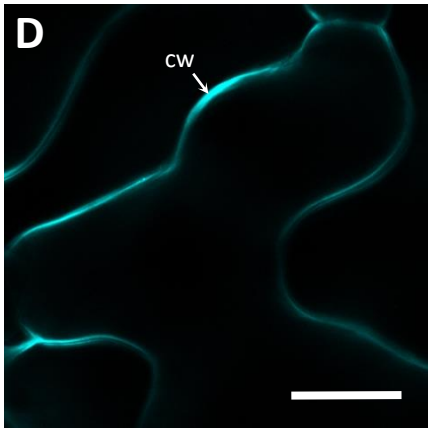
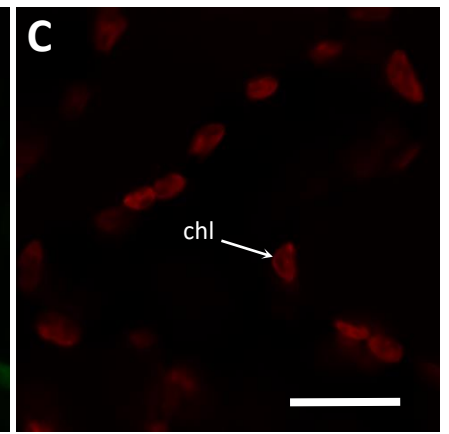
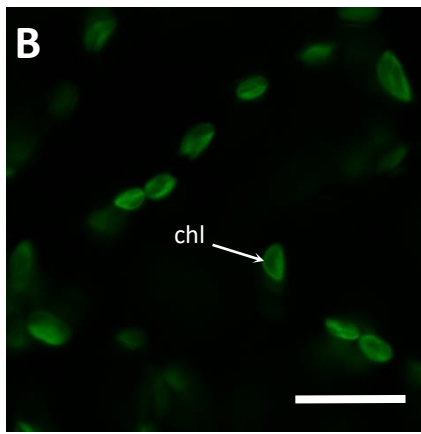
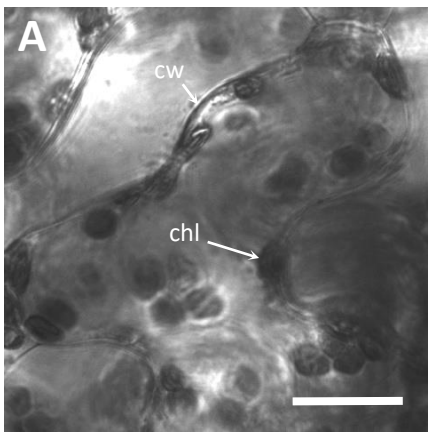
Legends to Figures

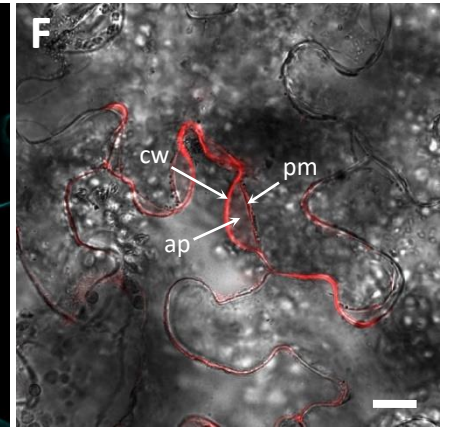
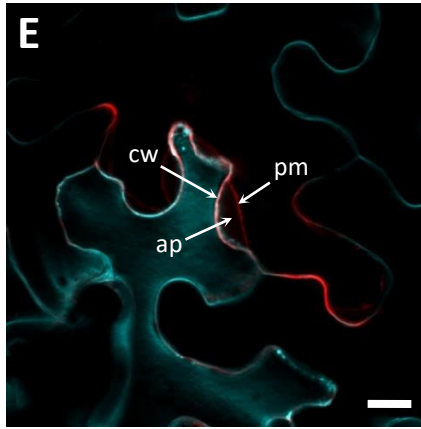
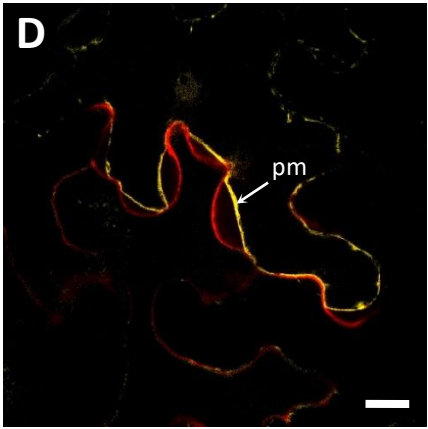
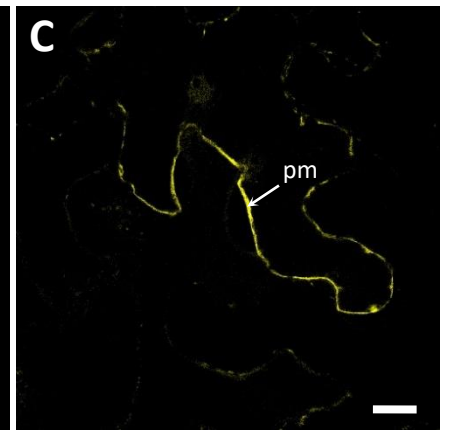
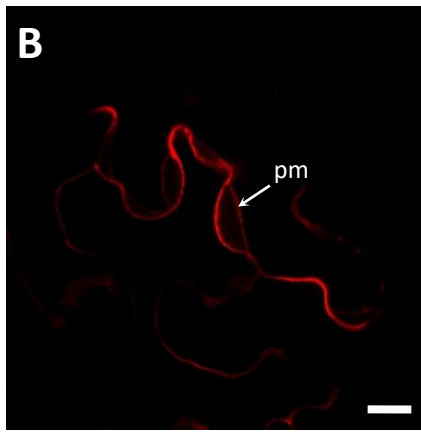
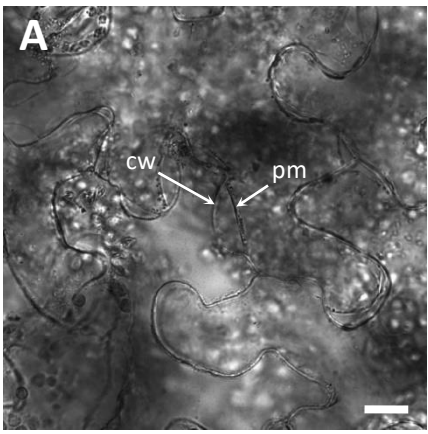
Fig. 1. Sub-cellular localization of At4g37800::TagRFP (top panel: **A-F**) and At5g59310::TagRFP (bottom panel: **G-L**). **A, G.** Bright field. **B, H.** Merge of calcofluor staining of cellulose in cell walls and chloroplasts auto-fluorescence. **C.** Detection of At4g37800::TagRFP. **D, J.** Detection of the plasma membrane marker pm-yb CD3-1006. **E.** Merge of calcofluor staining and C. **F.** Merge of C and D. **I.** Detection of At5g59310::TagRFP. **K.** Merge of calcofluor staining and I. **L.** Merge of I and J. The epidermal cells (top and bottom panels) have been observed after plasmolysis with 15% glycerol. ap: [apoplast](#); chl: chloroplast; cw: cell wall; pm: plasma membrane. Scale bars represent 20 μ m.

Fig. 2. Sub-cellular localizations of Δ SP-At5g59310::TagRFP (top panel: **A-F**) and SP-At5g11420:: Δ SP-At5g59310::T-Sapphire (bottom panel: **G-L**). **A, G.** Bright field. **B, H.** Chloroplasts auto-fluorescence. **C.** Detection of Δ SP-At5g59310::TagRFP. **D.** Calcofluor staining of cellulose in cell walls. **E.** Merge of B and C. **F.** Merge of B, C and D. **I.** Detection of SP-At5g11420:: Δ SP-At5g59310::T-Sapphire. **J.** Detection of the plasma membrane marker pm-yb CD3-1006. **K.** Merge of H and I. **L.** Merge of K and J. Note that no plasmolysis has been performed for the observations related to Δ SP-At5g59310::TagRFP (**A-F**). For the observations related to SP-At5g11420:: Δ SP-At5g59310::T-Sapphire (**G-L**), the epidermal cells have been observed after plasmolysis with 15% glycerol. ap: [apoplast](#); chl: chloroplast; cw: cell wall; pm: plasma membrane. Scale bars represent 20 μ m.

Fig. 3. Sub-cellular localization of SPC::TagRFP::At5g60490. **A.** Bright field. **B.** Detection of SPC::TagRFP::At5g60490. **C.** Detection of the plasma membrane marker pm-yb CD3-1006. **D.** Merge of B and C. **E.** Merge of calcofluor staining and B. **F.** Merge of A and B. The epidermal cells have been observed after plasmolysis with 15% glycerol. ap: [apoplast](#); chl: chloroplast; cw: cell wall; pm: plasma membrane. Scale bars represent 20 μ m.







Mass spectrometry data mining

Signal
peptide

Mature protein

GPI-anchor
signal peptide

Peptide

Peptide



Microscopic
observation of sub-
cellular localization

The image shows a microscopic view of a cell with sub-cellular localization. The cell is stained with a red dye, and the nucleus is stained with a blue dye. The text "Microscopic observation of sub-cellular localization" is overlaid on the image.

Bona fide cell wall
protein
identification