



Proline Hydroxylation in Cell Wall Proteins: Is It Yet Possible to Define Rules?

Harold Duruflé^{1†}, Vincent Hervé^{1,2†}, Thierry Balliau³, Michel Zivy³, Christophe Dunand¹ and Elisabeth Jamet^{1*}

¹ Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, Toulouse, France, ² INRS – Institut Armand Frappier, Laval, Canada, ³ PAPPISO, GQE Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France

OPEN ACCESS

Edited by:

Ján A. Miernyk,
Agricultural Research Service (USDA),
United States

Reviewed by:

Ian S. Wallace,
University of Nevada, Reno,
United States
Li Tan,
University of Georgia, United States

*Correspondence:

Elisabeth Jamet
jamet@lrsv.ups-tlse.fr

[†]Co-first authors

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 20 June 2017

Accepted: 04 October 2017

Published: 17 October 2017

Citation:

Duruflé H, Hervé V, Balliau T, Zivy M,
Dunand C and Jamet E (2017)
Proline Hydroxylation in Cell Wall
Proteins: Is It Yet Possible to Define
Rules? *Front. Plant Sci.* 8:1802.
doi: 10.3389/fpls.2017.01802

Cell wall proteins (CWPs) play critical and dynamic roles in plant cell walls by contributing to developmental processes and response to environmental cues. Since the CWPs go through the secretion pathway, most of them undergo post-translational modifications (PTMs) which can modify their biological activity. Glycosylation is one of the major PTMs of CWPs and refers to *N*-glycosylation, *O*-glycosylation and glypiation. Each of these PTMs occurs in different amino acid contexts which are not all well defined. This article deals with the hydroxylation of Pro residues which is a prerequisite for *O*-glycosylation of CWPs on hydroxyproline (Hyp) residues. The location of Hyp residues is well described in several structural CWPs, but yet rarely described in other CWPs. In this article, it is studied in detail in five *Arabidopsis thaliana* proteins using mass spectrometry data: one of them (At4g38770, AtPRP4) is a structural CWP containing 32.5% of Pro residues arranged in typical motifs, the others are either rich (27–28%, At1g31580 and At2g10940) or poor (6–8%, At1g09750 and At3g08030) in Pro residues. The known rules of Pro hydroxylation allowed a good prediction of Hyp location in AtPRP4. However, they could not be applied to the other proteins whatever their Pro content. In addition, variability of the Pro hydroxylation patterns was observed within some amino acid motifs in all the proteins and new patterns of Pro hydroxylation are described. Altogether, this work shows that Hyp residues are present in more protein families than initially described, and that Pro hydroxylation patterns could be different in each of them. This work paves the way for completing the existing Pro hydroxylation code.

Keywords: *Arabidopsis thaliana*, cell wall protein, hydroxyproline, mass spectrometry, proline hydroxylation, proline-rich protein, post-translational modification

INTRODUCTION

Cell wall proteins (CWPs) are important players in plant cell walls, otherwise mainly constituted of polysaccharides, and eventually of phenolic compounds around differentiated lignified cells (Carpita and Gibeau, 1993). They have been involved in the remodeling of cell wall polymers networks by hydrolysing covalent bounds, inserting newly synthesized polysaccharides, cross-linking together structural proteins, proteins and polysaccharides or polysaccharides and phenolic compounds (Franková and Fry, 2013; Cosgrove, 2015). Together with extracellular peptides, some CWPs have also been involved in signaling, thus allowing cell-to-cell communication

(Matsubayashi and Sakagami, 2006; Lamport and Várnai, 2013; Tavormina et al., 2015). Altogether, CWPs and peptides contribute to both developmental processes and response to environmental cues (Tenhaken, 2015; Borassi et al., 2016). Most of them undergo post-translational modifications (PTMs) during their transport through the secretion pathway (Faye et al., 2005; Kim and Brandizzi, 2016) which can modify their conformation, their biological activity and/or their ability to interact with cell wall components (Lannoo and Van Damme, 2015; Baer and Millar, 2016; Strasser, 2016). As an example, the site-directed mutagenesis of the *N*-glycosylation motifs of a class III peroxidase was shown to reduce its thermal stability, its catalytic activity and to modify its conformation (Lige et al., 2001).

During recent years, proteomics has facilitated a better knowledge of the plant cell wall proteome by increasing its coverage thanks to the design of specific strategies able to recover protein extracts enriched in extracellular proteins from organs of several model plants and crops (Lee et al., 2004; Albenne et al., 2013; Komatsu and Yanagawa, 2013; Rodríguez-Celma et al., 2016). Beyond the identification of proteins, technological advances have also permitted description of their PTMs. Various methods have been developed to address this particular question. In particular, immobilized affinity chromatography (IMAC) and lectin-affinity chromatography have allowed studying protein phosphorylation and glycosylation, respectively (Ytterberg and Jensen, 2010; Nakagami et al., 2012; Ruiz-May et al., 2014; Canut et al., 2016).

Glycosylation is one of the major PTMs of CWPs and refers to *N*-glycosylation, *O*-glycosylation and addition of glycosphosphatidylinositol (GPI)-anchors, also named glypiation (Faye et al., 2005). Each of these PTMs occurs on specific amino acid sequences. *N*-glycosylation is the best described. It occurs on Asn residues in Asn-X-Ser/Thr motifs, where X cannot be a Pro residue. In these motifs, the hydroxyl functional group of Ser/Thr residues was shown to be required in the transglycosylation reaction on the Asn residue (Bause and Legler, 1981), whereas the presence of a Pro residue modifies the local conformation of the protein, thus preventing its *N*-glycosylation (Bause, 1983). The different structures of *N*-glycans are well-known, thus allowing systematic search in mass spectrometry (MS) data obtained in conditions preserving glycan-peptide bonds (Ruiz-May et al., 2012). GPI-anchors are transferred by a transamidase to a carboxy-terminal GPI-attachment signal peptide which can be predicted by bioinformatics (Eisenhaber et al., 2003). Several targeted proteomic studies have contributed to the identification of GPI-anchored proteins and some of them could be released from plasma membrane fractions by a phospholipases C or D which cleave GPI-anchors (Borner et al., 2003; Elortza et al., 2003; Elortza et al., 2005). *O*-glycosylation is the most complex type of glycosylation. In plant CWPs, it can occur on Ser and hydroxyproline (Hyp) residues (Faye et al., 2005). Galactose can be linked to Ser and Hyp residues whereas arabinose can only be linked to Hyp residues (Canut et al., 2016). According to the so-called Hyp contiguity hypothesis initially proposed for hydroxyproline-rich proteins (HRGPs), contiguous Hyp residues are arabinosylated and clustered non-contiguous Hyp residues are galactosylated (Shpak et al., 2001). Then, glycosyltransferases

can extend the *O*-glycans in different ways depending on the initial pattern of Pro hydroxylation. The correct *O*-glycosylation of HRGPs was shown to be required for their conformation or their biological activity (Stafstrom and Staehelin, 1986; Velasquez et al., 2011).

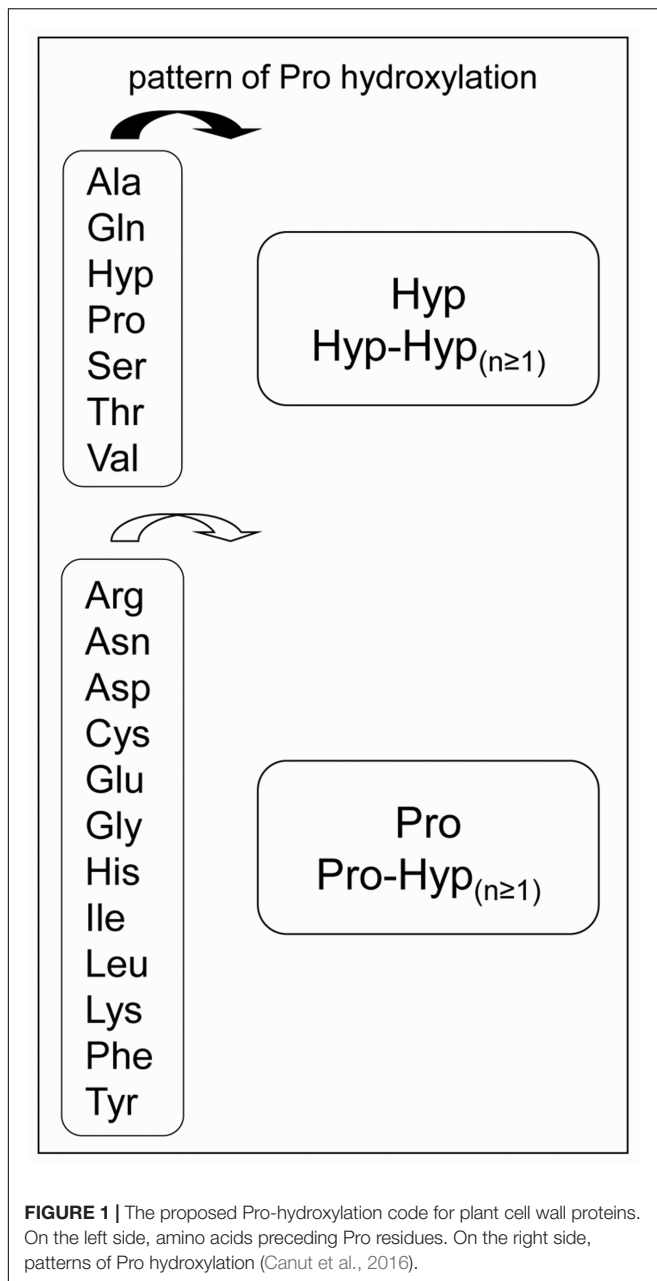
Pro hydroxylation is a major step for *O*-glycosylation, but it is still difficult to predict in which amino acid context it occurs. In a previous review, we have proposed an extended Pro hydroxylation code (Canut et al., 2016), based on both (i) the initial Pro hydroxylation code (Kieliszewski and Lamport, 1994) and (ii) additional experimental LC-MS/MS and Edman sequencing data. The extended code has taken into account more protein and peptide families than the former one, including structural proteins like HRGPs and Pro/Hyp-rich proteins, solanaceous lectins, allergens, systemins and CLE peptides. Briefly, Pro residues could be hydroxylated when they are located after Ala, Gln, Hyp, Pro, Ser, Thr, and Val residues, whereas the first Pro residue following the other amino acids could not be hydroxylated (**Figure 1**). Only little information is available regarding Trp and Met residues. In a large mutagenesis screen performed on the amino acids surrounding the only Pro residue of sporamin shown to be hydroxylated, Trp and Met were not shown to favor the hydroxylation of the following Pro residue (Shimizu et al., 2005).

In this article, our aim was to test the extended Pro hydroxylation code on a new set of CWPs including non-structural CWPs. We have thus selected five CWPs with various contents in Pro residues. Three of them were rich in Pro residues among which AtPRP4 which is a structural CWP (Fowler et al., 1999) and two of them were poor in Pro residues. We have performed a deep data mining on two recent cell wall proteomic studies performed on rosettes and stems of *Arabidopsis thaliana* (Hervé et al., 2016; Duruflé et al., 2017). From the fine analysis of MS data, we have compared the observed patterns of Pro/Hyp location to the predicted ones according to the Pro hydroxylation extended code. The limits of the existing extended Pro hydroxylation code are discussed and new motifs are described.

RESULTS

Mapping of Hyp Residues

For this analysis, we have taken advantage of two cell wall proteomics studies which have lead to the identification of numerous CWPs, 361 in rosettes and 302 in stems, i.e., 397 different CWPs (Hervé et al., 2016; Duruflé et al., 2017). This body of data corresponded to three independent experiments (two for rosettes and one for stems), each of them including three biological replicates. The parameters used for peptide identification included a possible mass delta of 15.99 Da for each Pro residue, corresponding to its hydroxylation. As an example, in one of the rosette experiment, 79% of the identified CWPs were predicted to be *N*-glycosylated (presence of the PS00001 PROSITE motif), whereas 17.5% had at least one peptide carrying a Hyp residue. Among the latter and in addition to the



proteins described below, there were lectins, Asp proteases, lipases acylhydrolases of the GDSL family, and class III peroxidases.

Five CWPs were selected on the basis of the following criteria: (i) their abundance in these aerial organs, as shown by the high number of sequenced peptides for each of them (from 89 to 533, depending on the protein); and (ii) a high sequence coverage (from 26 to 72% of the mature protein) (Table 1). The MS data corresponding to these five proteins are given in Supplementary Table S1. In addition, none of them has already been shown to contain Hyp residues. At4g38770 (AtPRP4) is a Pro-rich protein and its gene was shown to be expressed in aerial organs (Fowler et al., 1999). At1g09750 is a

predicted Asp protease. At1g31580 (ECS1/CXc750) was assumed to be involved in resistance mechanisms (Aufsatz et al., 1998). At3g08030 (AthA2-1) has a predicted DUF642 domain (Vázquez-Lobo et al., 2012). Finally, At2g10940 is a protein showing homology to non-specific plant lipid transfer proteins. Three out of these CWPs have amino acid sequences rich in Pro residues as calculated from their mature sequence: AtPRP4 (32.5% Pro), At1g31580 (27.9%), and At2g10940 (27.3%). The two others, At1g09750 and At3g08030, are poor in Pro residues (7.3 and 6.4%, respectively). Contrarily to the three former proteins which exhibit Pro-rich motifs, the latter ones have dispersed Pro residues.

The extended Pro hydroxylation code was applied to predict the location of Hyp residues in the five amino acid sequences and to compare them to the observed ones. All details are given in Supplementary Figure S1, and simplified views of AtPRP4 and At2g10940 are shown in Figures 2 and 3. The sequences including the predicted Hyp residues are shown on the left and the observed ones are framed on the right of each figure. No obvious difference could be found between the three datasets regarding the frequency of occurrence of the different peptide variants according to the location of Pro and Hyp residues (Pro/Hyp peptide variants). In particular, no difference could be found between the rosette and the stem samples. The MS/MS data were manually checked as shown in Supplementary Figure S2 for two peptides of AtPRP4. All the Pro/Hyp locations were confirmed with the exception of two motifs in the amino acid sequence of At1g09750 (GPM and LPM). We could not discriminate between the hydroxylation of a Pro residue and the oxidation of a Met residue (Supplementary Table S2B and Figure S1). Thus, we did not retain the hypothesis of a Pro hydroxylation in these motifs in the following.

Several observations could be done. (i) A very high proportion of Pro residues was hydroxylated in the three Pro-rich proteins (69/92 in AtPRP4, 8/10 in At1g31580, 36/49 in At2g10940). (ii) Only a few Hyp residues could be found in proteins poor in Pro residues, and rarely at predicted positions (none in At1g09750; only at two out of nine predicted possibilities, and two at unexpected positions, but at a very low frequency, in At3g08030). (iii) For a given peptide, several variants could be observed. For example, two variants of GFDHPFPLPPPPLPPFLK and three variants of YSPPVEVPPVPVYEPPOKK were found in AtPRP4: GFDHPFPLPOOLELPOFLK as predicted, and GFDHPFOLPOOLELPOFLK; YSOOVEVOOOVOVYEPPOKK as predicted, YSOOVEVOOOVOVYEPPOKK, and YSOOVEVOOOVOVYEOOQKK (Supplementary Figure S2). (iv) The observed discrepancies between the predicted and the observed Pro/Hyp locations could be either a Pro instead of a Hyp residue or vice-versa. (v) Some discrepancies could be systematically observed, as the FQOR motif in At1g31580 instead of the predicted FPOR motif.

This survey has allowed the fine mapping of Pro/Hyp residues in the five selected CWPs. The next issue was to know how efficiently the extended Pro hydroxylation code could predict their location.

TABLE 1 | Sequence coverage and number of peptides analyzed by LC-MS/MS for each protein.

	Size of the protein (number of amino acids)	Average percentage of sequence coverage	Total number of sequenced peptides
At4g38770 (AtPRP4)	419	51	131
At1g09750	428	26	450
At1g31580	68	60	89
At3g08030	353	72	533
At2g10940	264	60	250

The list of the sequenced peptides is provided in Supplementary Table S1. The percentage of sequence coverage is calculated for the mature protein sequence and takes the results of the three experiments into account.



Efficiency of the Prediction of the Location of the Pro/Hyp Residues

For each protein sequence, the total number of Pro and Hyp positions was recorded and compared to the number of correct predictions (Table 2 and Supplementary Table S2). The percentage of mis-predictions was found to range from 5.1 to 46.7%. The best prediction was obtained for AtPRP4 which is a HRGP, i.e., a canonical protein with regard to the

proposed rule. Except one motif in peptide 3 (Figure 2 and Supplementary Figure S1A), VPOOV instead of the predicted VOOOV, all the other predicted motifs were found at least once. Some variability was observed within 12 motifs located in seven peptides (numbered 1, 2, and 4–8 on Figure 2). The KPPPK motif was the most variable one, with the following variants: KPOOK as predicted (peptides 3–5 and 7), KPPOK (peptides 4, 5, and 7), KOPPK (peptides 5 and 7), and KPPPK (peptide 7). Other motifs including three Pro/Hyp residues were also variable,

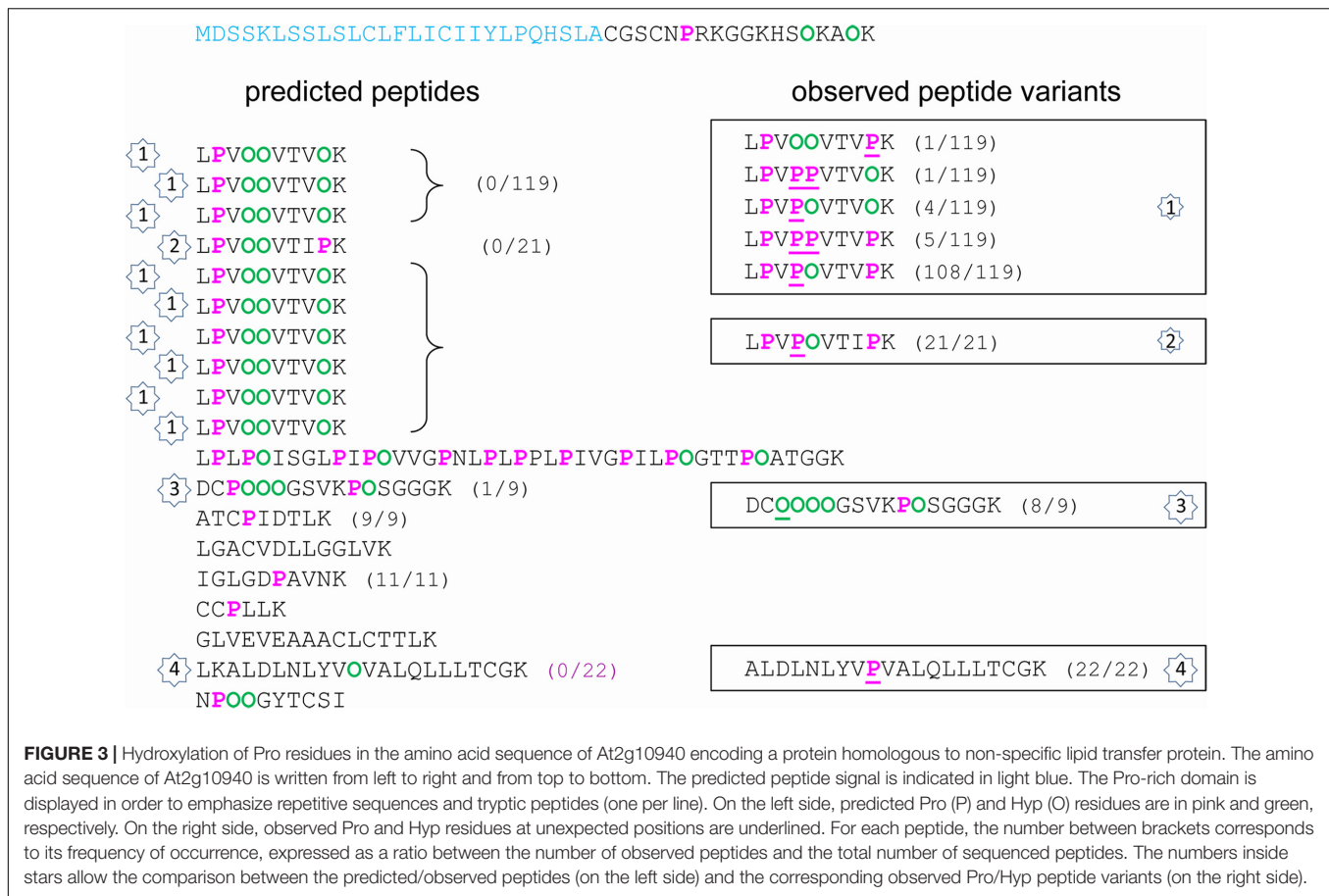


TABLE 2 | Efficiency of the prediction of Pro/Hyp location in CWP amino acid sequences according to the proposed rules.

Accession number	Number of analyzed Pro/Hyp positions	Number of unexpected location of Pro/Hyp residues	Percentage of mis-predictions
At4g38770	868	44	5.1
At1g09750	343	104	30.3
At1g31580	439	154	35.1
At3g08030	502	225	44.8
At2g10940	602	281	46.7

Five CWP sequences were analyzed by LC-MS/MS (Supplementary Table S1) and the percentage of mis-predictions was calculated (Supplementary Table S2).

such as EPPPK in peptide 2 (EPOOK as predicted, EPOPK and EPOOK), HPPPV in peptide 4 (HPOOV as predicted and HPOPO), CPPPV in peptide 8 (CPOOV as predicted and CPPPV). The other cases of variability concerned shorter motifs such as FPL (FOL in peptide 1), VPV/I (VPV in peptide 5 and VPI in peptide 8), KPPT/V (KOPT in peptide 6, KPPV in peptide 8). However, all these Pro/Hyp peptide variants were not the prevailing forms of the motifs (see Supplementary Table S2).

The prediction of Pro and Hyp location was much less efficient for the four other proteins, irrespectively, to their Pro content. Regarding the two proteins with a low Pro content, the percentage of mis-prediction of Pro/Hyp location was very high (30.3% for At1g09750 and 44.8% for At3g08030).

Hyp was only found in At3g08030, but in solely three motifs (VOF, GOH, and LOL) and at a low frequency (Table 3). Regarding the two proteins with a high percentage of Pro residues (At1g31580 and At2g10940), the situation was very different. Although they both exhibited a high percentage of Pro residues, the proposed rules did not allow reaching a high level of correct prediction of Pro/Hyp location. For At1g31580, peptide variants could mostly be observed for short motifs like RPI/R/T in peptides 1 and 3 (RPI/R/T as predicted and ROI/R/T as observed) and VPI/G in peptides 1 and 2 (VOI/G as predicted and VPI/G) (Supplementary Figure S2C). Two larger motifs were variable: FPPR in peptide 2 (FPOR as predicted and FOOR); LPPY in peptide 3 (LPOY as predicted

TABLE 3 | Exceptions to the proposed rules of Pro/Hyp location.

Accession number	Predicted location of Hyp residues	Observed motifs	Frequency of occurrence	
At4g38770	KPOOK	KPOOK	19/29	
		K P POK	5/29	
		K OP PK	4/29	
		K PP PK	1/29	
At1g09750	QOV	QOV	0/73	
		Q PV	73/73	
	VON	VON	0/21	
		V PN	21/21	
At1g31580	RPI	RPI	34/49	
		R PI	15/49	
	RPT	R OTRPT	23/3411/34	
		RPR	RPR	11/34
	VOG	R OR	23/34	
		VOG	33/55	
	VOI	V PG	22/55	
		VOI	52/55	
	FPOR	V PI	3/55	
		FPOR	55/55	
	At3g08030	LPOY	F OOR	55/55
			LPOY	9/34
		VOS	L PPY	23/34
			VOS	0/21
VOF		V PS	21/21	
		VOF	3/63	
VOH		V PF	60/63	
		VOH	0/20	
LPL		V PH	20/20	
		LPL	18/21	
SOL		L OL	3/21	
		SOL	0/48	
SOG		S PL	48/48	
		SOG	0/49	
VOOV	S PG	49/49		
	VOOV	1/140		
At2g10940	VOOV	V POV	133/140	
		V PPV	6/140	
		VOK	5/119	
	VOK	VOK	5/119	
		V PK	114/119	
	VOV	VOV V PV	0/22 22/22	

Only motifs observed at most 20 times are considered in this table.

and **LPPY**). With the exception of the **FPPR** motif in which Pro residues were always both hydroxylated and the **VPI** motif which was found in the **VOI** form in all but three cases out of 55, all the other variants were found in the one third/two third proportion between predicted and mis-predicted variants or vice versa. For At2g10940, variability was observed in four motifs (**Figure 3** and Supplementary Figure S2E): **VPPV** in peptides 1 and 2 (**VOOV** as predicted and **VPOV**); **VPK** in peptides 1 and 2 (**VOK** as predicted and **VPK**); **VPV** in peptide 4 (**VOV** as predicted and **VPV**); and **CPPPPG** in peptide 3 (**CPOOOG** as predicted and **COOOOG**).

A very high proportion of the observed motifs did not follow the extended Pro hydroxylation code for their first Pro residue.

DISCUSSION

This work has allowed mapping the Pro/Hyp residues in five CWPs and comparing the prediction of Pro/Hyp location according to a previously proposed extended Pro hydroxylation code (Kieliszewski and Lamport, 1994; Canut et al., 2016).

Among these five proteins, only AtPRP4 was assumed to contain Hyp residues because it is known as a structural CWP with a high content of Pro residues and canonical amino acid motifs such as KKPCPP (7 occurrences) and PPV (14 occurrences) (Showalter et al., 2010). However, to our knowledge, its pattern of Pro hydroxylation has not yet been described. Regarding the other four proteins, our results have allowed enlarging the number of protein families possibly modified at the post-translational level by the hydroxylation of Pro residues. They have also shown that the Pro hydroxylation patterns can be variable at a given amino acid position.

The prediction of Hyp residue location could be done with a high level of confidence in one of the so-called HRGPs using the extended Pro hydroxylation code probably because this code was designed from the analysis of such protein sequences (Kieliszewski and Lamport, 1994; Canut et al., 2016). They include proteins rich in Pro, Ala, Ser and Thr residues such as (i) extensins with repetitive S(P)_n ≥ 2 and YXY motifs, (ii) arabinogalactan proteins (AGPs) with AP/PA/SP/TP repeats and Pro-rich proteins (PRPs) with PPVX[KT], KKPCPP and PPV motifs (Showalter et al., 2010) or chimeric proteins containing a Pro-rich domain with XP_nY motifs (Hijazi et al., 2012; Canut et al., 2016). However, although prediction of the location of Pro/Hyp residues in the AtPRP4 sequence was very efficient (94.9% of successful predictions), some variability could be observed at a low frequency, particularly in the KPPPK motif, with 19 canonical Pro hydroxylation patterns (KPOOK) out of the 29 observed patterns and 10 variants (KPPOK, KOPPK, and KPPPK).

For the other proteins rich in Pro residues, many exceptions to the extended Pro hydroxylation code could be observed. In particular, in At1g31580, only about one third of the three amino acid-motifs had Hyp at the predicted location in RPI/T/R and VPG/I motifs and the FPPR motifs was systematically found with two Hyp residues (55/55). Besides, only one predicted VOOV motif could be recorded in At2g10940 out of the 140 observed peptides. The major variant was VPOV (133/140). A similar situation was found for the predicted VOK and VOV motifs (5/119 and 0/22 observations, respectively). For proteins having a low content in Pro residues, the prediction of Pro/Hyp location was also inefficient. Only a few Hyp residues could be observed and at a low frequency. Unexpected Hyp residues were also found in a previous study focused on class III peroxidases (Nguyen-Kim et al., 2016). For these proteins, a few Hyp residues were observed in CPN/Q/R, DPA, GPS/N, HPD, IPD, and LPA/Q/S motifs whereas some Pro residues were observed in APF/A, VPT, SPT/D, and TPG/L motifs. These results suggest that the extended Pro hydroxylation code cannot be used for such proteins. They also show that the rule established for the hydroxylation of the Pro residue within the EPA motif of sporamin cannot be applied (Shimizu et al., 2005). Based on mutagenesis of the surrounding amino acids, it was shown that the hydroxylation of the Pro residue required the following environment in tobacco BY-2 cells: [AVSTG]-P-[AVSTGA]-[GAVPSTC]-[APSD E].

Our results raise the question of the specificity of prolyl-4 hydroxylases (P4Hs) which have to recognize some features on the target protein at the level of its primary amino acid sequence or the secondary/tertiary structure. The specificity of three out of the 13 P4Hs of *A. thaliana* was characterized. P4H1 was shown to preferentially hydroxylate the second Pro residue in PPG motifs (Hieta and Myllyharju, 2002). All the peptides hydroxylated by P4H2 have at least three consecutive Pro residues and the third of them is preferentially hydroxylated (Tiainen et al., 2005). Finally, P4H5 was shown to hydroxylate Pro residues in SP₄ motifs in a sequential way, but never on the fourth Pro residue (Velasquez et al., 2015). Besides, P4H2 and P4H13 were assumed to complement the Pro hydroxylation pattern of SP₄ motifs in extensins (Velasquez et al., 2015). The characterization of additional P4Hs will give clues to understand this process which is probably tightly regulated because of its importance for biological activity. Indeed, this PTM is the first step prior to O-glycosylation: (i) poly-arabinoxylan in extensins; or (ii) complex O-glycans like type II arabinogalactans (AGs), type III AGs or peanut agglutinin (PNA) AGs in AGPs, allergens or AtAGP31, respectively (Hijazi et al., 2014). None of the five CWPs analyzed in this work is known to be O-glycosylated. However, a previous proteomic study based on affinity chromatography with PNA, a lectin specific for galactose residues, has allowed identifying a protein of the same family as At3g08030 (Zhang et al., 2011). The next step will consist in correlating the presence of Hyp residues to O-glycosylation. Finally, the presence of Pro/Hyp peptide variants raises the question of the role of Hyp residues as previously discussed for class III peroxidases (Nguyen-Kim et al., 2016). This variability could be incidental or contribute to the regulation of the biological activity of CWPs.

Altogether, Pro hydroxylation events are probably more abundant in CWPs than initially thought, but the precise rules of this PTM need additional experiments to be fully described. Some clues can be proposed from our results. For example, a Hyp residue is found in the VPX motifs of AtPRP4 (77 Pro hydroxylations among the 82 observed VPX motifs) and At1g31580 (85 out of 110), whereas it is mainly a Pro residue in the three other proteins (only 15 Pro hydroxylations out of the 406 observed VPX motifs). The Pro residues in the 97 observed SPX motifs of At3g08030 were never hydroxylated whereas only a few lack of Pro hydroxylation have been described in cell wall Pro-rich proteins (Canut et al., 2016). A systematic mining of MS data is now required to permit the identification of other CWPs or secreted peptides containing Hyp residues and to map them. This task is challenging because, as mentioned above, 17.5% of the CWPs identified in one of our rosette experiments had at least one peptide carrying a Hyp residue. However, the amount of MS data corresponding to all these proteins was not sufficient to perform a relevant statistical analysis and to propose yet a further expanded Pro hydroxylation code. Finally, such a code should probably take into account tissue-specific patterns as for O-glycosylation (Estevez et al., 2006) and protein families. This work paves the way for a better description of Pro hydroxylation patterns in CWPs.

MATERIALS AND METHODS

Extraction of Proteins from Cell Walls

Arabidopsis thaliana plants were cultivated in growth chambers at 22°C with a photoperiod of 16 h light/8 h dark. Rosettes and mature stems were collected after 4 and 6 weeks, respectively. The detailed description of the experiments is given in our previous articles (Hervé et al., 2016; Duruflé et al., 2017). Three biological replicates were performed for each experiment. Briefly, cell walls were purified as described (Feiz et al., 2006). Proteins were extracted from lyophilized cell walls in four steps using a 5 mM acetate buffer pH 4.6 complemented with 0.2 M CaCl₂ (two successive extractions) or 2 M LiCl (*ditto*) (Irshad et al., 2008). The four protein extracts were combined prior to further analysis.

Analysis of Proteins by LC/MS-MS and Bioinformatics

The same amount of each protein extract (40 µg) was analyzed by LC-MS/MS. In the case of rosettes, two types of analysis were performed: (i) the first one after separation of proteins by a short 1D-electrophoresis in three fractions prior to *in gely* tryptic digestion, (ii) the second one by shotgun analysis of the extracted proteins after tryptic digestion (Hervé et al., 2016). In the case of stems, only the second method was used (Duruflé et al., 2017). LC-MS/MS analyses were performed with a Q-exactive instrument (Thermo Fisher Scientific, Villebon-sur-Yvette, France) as described (Feiz et al., 2006; Hervé et al., 2016). All the MS/MS data were made publicly available in the PROTEOMICSdb¹ and WallProtDB databases². The following modifications were taken into account for peptide identification: Met oxidation, Pro hydroxylation, N-ter acetylation, N-ter deamidation of Glu, N-ter deamidation of Cys and loss of H₂O on N-ter Glu. The lists of peptides allowing the identification of the five CWP studied in detail in this article are given in

¹ <http://proteus.moulon.inra.fr/w2dpage/proticdb/angular/>

² <http://polebio.lrsv.ups-tlse.fr/WallProtDB/>

REFERENCES

- Albenne, C., Canut, H., and Jamet, E. (2013). Plant cell wall proteomics: the leadership of *Arabidopsis thaliana*. *Front. Plant Sci.* 4:111. doi: 10.3389/fpls.2013.00111
- Aufsatz, W., Amry, D., and Grimm, C. (1998). The ECS1 gene of *Arabidopsis* encodes a plant cell wall-associated protein and is potentially linked to a locus influencing resistance to *Xanthomonas campestris*. *Plant Mol. Biol.* 38, 965–976. doi: 10.1023/A:1006028605413
- Baer, B., and Millar, A. (2016). Proteomics in evolutionary ecology. *J. Proteomics* 135, 4–11. doi: 10.1016/j.jprot.2015.09.031
- Bause, E. (1983). Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* 209, 331–336. doi: 10.1042/bj2090331
- Bause, E., and Legler, G. (1981). The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the N-glycosylation step during glycoprotein biosynthesis. *Biochem. J.* 195, 639–644. doi: 10.1042/bj1950639
- Borassi, C., Sede, A., Mecchia, M., Salgado Salter, J., Marzol, E., Muschietti, J., et al. (2016). An update on cell surface proteins containing extensin-motifs. *J. Exp. Bot.* 67, 477–487. doi: 10.1093/jxb/erv455
- Borner, G. H., Lilley, K. S., Stevens, T. J., and Dupree, P. (2003). Identification of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A proteomic and genomic analysis. *Plant Physiol.* 132, 568–577. doi: 10.1104/pp.103.021170
- Canut, H., Albenne, C., and Jamet, E. (2016). Post-translational modifications of plant cell wall proteins and peptides: a survey from a proteomics point of view. *Biochim. Biophys. Acta* 1864, 983–990. doi: 10.1016/j.bbapap.2016.02.022
- Carpita, N., and Gibeaut, D. (1993). Structural models of primary cell walls in flowering plants, consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* 3, 1–30. doi: 10.1111/j.1365-3113.1993.tb00007.x
- Cosgrove, D. (2015). Plant cell wall extensibility: connecting plant cell growth with cell wall structure, mechanics, and the action of wall-modifying enzymes. *J. Exp. Bot.* 67, 463–476. doi: 10.1093/jxb/erv51
- Duruflé, H., San Clemente, H., Balliau, T., Zivvy, M., Dunand, C., and Jamet, E. (2017). Cell wall proteome analysis of *Arabidopsis thaliana* mature stems. *Proteomics* doi: 10.1002/pmic.201600449 [Epub ahead of print].
- Eisenhaber, B., Wildpaner, M., Schultz, C., Borner, G., Dupree, P., and Eisenhaber, F. (2003). Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence- and genome-wide studies for

Supplementary Table S1. The search for N-glycosylation motifs was performed with PROSITE³.

AUTHOR CONTRIBUTIONS

HD and VH performed the protein extractions from purified cell walls and contributed to the analyses of results. TB and MZ did the MS/MS analyses. CD and EJ initiated the research, designed the study and discussed the results. EJ coordinated the analysis of the results and the writing of the manuscript. All authors read and approved the final manuscript.

FUNDING

The authors are thankful to Université Paul Sabatier (Toulouse, France) and CNRS for supporting their research work. HD is granted by the Toulouse University and the Occitanie region. This work was also supported by the French Laboratory of Excellence project entitled “TULIP” (ANR-10-LABX-41; ANR-11-IDEX-0002-02).

ACKNOWLEDGMENTS

LC-MS/MS analyses were performed at the PAPPISO proteomics facility (pappiso.inra.fr). The authors wish to thank Dr. Hervé Canut for stimulating discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.01802/full#supplementary-material>

³ <http://prosite.expasy.org/>

- Arabidopsis and rice. *Plant Physiol.* 133, 1691–1701. doi: 10.1104/pp.103.023580
- Elortza, F., Nühse, T., Foster, L., Stensballe, A., Peck, S., and Jensen, O. (2003). Proteomic analysis of glycosylphosphatidylinositol-anchored membrane proteins. *Mol. Cell. Proteomics* 2, 1261–1270. doi: 10.1074/mcp.M300079-MCP200
- Elortza, F., Shabaz, M., Bunkenborg, J., Foster, L., Nühse, T., Brodbeck, U., et al. (2005). Modification-specific proteomics of plasma membrane proteins: identification and characterization of glycosylphosphatidylinositol-anchored proteins released upon phospholipase D treatment. *J. Proteome Res.* 5, 935–943. doi: 10.1021/pr050419u
- Estevez, J. M., Kieliszewski, M. J., Khitrov, N., and Somerville, C. (2006). Characterization of synthetic hydroxyproline-rich proteoglycans with arabinogalactan protein and extensin motifs in Arabidopsis. *Plant Physiol.* 142, 458–470. doi: 10.1104/pp.106.084244
- Faye, L., Boulaflous, A., Benchabane, M., Gomord, V., and Michaud, D. (2005). Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming. *Vaccine* 23, 1770–1778. doi: 10.1016/j.vaccine.2004.11.003
- Feiz, L., Irshad, M., Pont-Lezica, R. F., Canut, H., and Jamet, E. (2006). Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from Arabidopsis hypocotyls. *Plant Methods* 2:10. doi: 10.1186/1746-4811-2-10
- Fowler, T. J., Bernhardt, C., and Tierney, M. L. (1999). Characterization and expression of four proline-rich cell wall protein genes in Arabidopsis encoding two distinct subsets of multiple domain proteins. *Plant Physiol.* 121, 1081–1092. doi: 10.1104/pp.121.4.1081
- Franková, L., and Fry, J. (2013). Biochemistry and physiological roles of enzymes that 'cut and paste' plant cell-wall polysaccharides. *J. Exp. Bot.* 64, 3519–3550. doi: 10.1093/jxb/ert201
- Hervé, V., Duruflé, H., San Clemente, H., Albenne, C., Balliau, T., Zivy, M., et al. (2016). An enlarged cell wall proteome of *Arabidopsis thaliana* rosettes. *Proteomics* 16, 3183–3187. doi: 10.1002/pmic.201600290
- Hieta, R., and Myllyharju, J. (2002). Cloning and characterization of a low molecular weight prolyl 4-hydroxylase from *Arabidopsis thaliana*. Effective hydroxylation of proline-rich, collagen-like, and hypoxia-inducible transcription factor alpha-like peptides. *J. Biol. Chem.* 277, 23965–23971. doi: 10.1074/jbc.M201865200
- Hijazi, M., Durand, J., Pichereaux, C., Pont, F., Jamet, E., and Albenne, C. (2012). Characterization of the arabinogalactan protein 31 (AGP31) of *Arabidopsis thaliana*: new advances on the Hyp-O-glycosylation of the Pro-rich domain. *J. Biol. Chem.* 287, 9623–9632. doi: 10.1074/jbc.M111.247874
- Hijazi, M., Velasquez, S., Jamet, E., Estevez, J., and Albenne, C. (2014). An update on post-translational modifications of hydroxyproline-rich glycoproteins: toward a model highlighting their contribution to plant cell wall architecture. *Front. Plant Sci.* 5:395. doi: 10.3389/fpls.2014.00395
- Irshad, M., Canut, H., Borderies, G., Pont-Lezica, R., and Jamet, E. (2008). A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis thaliana*: confirmed actors and newcomers. *BMC Plant Biol.* 8:94. doi: 10.1186/1471-2229-8-94
- Kieliszewski, M. J., and Lamport, D. T. A. (1994). Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J.* 5, 157–172. doi: 10.1046/j.1365-313X.1994.05020157.x
- Kim, S.-J., and Brandizzi, F. (2016). The plant secretory pathway for the trafficking of cell wall polysaccharides and glycoproteins. *Glycobiology* 26, 940–949. doi: 10.1093/glycob/cww044
- Komatsu, S., and Yanagawa, Y. (2013). Cell wall proteomics of crops. *Front. Plant Sci.* 4:17. doi: 10.3389/fpls.2013.00017
- Lamport, D. T., and Várnai, P. (2013). Periplasmic arabinogalactan glycoproteins act as a calcium capacitor that regulates plant growth and development. *New Phytol.* 197, 58–64. doi: 10.1093/aob/mcu161
- Lannoo, N., and Van Damme, E. (2015). Review/N-glycans: the making of a varied toolbox. *Plant Sci.* 239, 67–83. doi: 10.1016/j.plantsci.2015.06.023
- Lee, S. J., Saravanan, R. S., Damasceno, C. M., Yamane, H., Kim, B. D., and Rose, J. K. (2004). Digging deeper into the plant cell wall proteome. *Plant Physiol. Biochem.* 42, 979–988. doi: 10.1016/j.plaphy.2004.10.014
- Lige, B., Shengwu, M., and Van Huystee, R. (2001). The effects of the site-directed removal of N-glycosylation from cationic peanut peroxidase on its function. *Arch. Biochem. Biophys.* 386, 17–24. doi: 10.1006/abbi.2000.2187
- Matsubayashi, Y., and Sakagami, Y. (2006). Peptide hormones in plants. *Annu. Rev. Plant Biol.* 57, 649–674. doi: 10.1146/annurev.arplant.56.032604.144204
- Nakagami, H., Sugiyama, N., Ishihama, Y., and Shirasu, K. (2012). Shotguns in the front line: phosphoproteomics in plants. *Plant Cell Physiol.* 53, 118–124. doi: 10.1093/pcp/pcr148
- Nguyen-Kim, H., San Clemente, H., Balliau, T., Zivy, M., Dunand, C., Albenne, C., et al. (2016). *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, 491–503. doi: 10.1002/pmic.2015010129
- Rodríguez-Celma, J., Ceballos-Laita, L., Grusak, M., Abadía, J., and López-Millán, A. (2016). Plant fluid proteomics: delving into the xylem sap, phloem sap and apoplastic fluid proteomes. *Biochim. Biophys. Acta* 1864, 991–1002. doi: 10.1016/j.bbapap.2016.03.014
- Ruiz-May, E., Hucko, S., Howe, K., Zhang, S., Sherwood, R., Thannhauser, T., et al. (2014). A comparative study of lectin affinity based plant N-glycoproteome profiling using tomato fruit as a model. *Mol. Cell. Proteomics* 13, 566–579. doi: 10.1074/mcp.M113.028969
- Ruiz-May, E., Thannhauser, T. W., Zhang, S., and Rose, J. (2012). Analytical technologies for identification and characterization of the plant N-glycoproteome. *Front. Plant Sci.* 3:150. doi: 10.3389/fpls.2012.00150
- Shimizu, M., Igasaki, T., Yamada, M., Yuasa, K., Hasegawa, J., Kato, T., et al. (2005). Experimental determination of proline hydroxylation and hydroxyproline arabinogalactosylation motifs in secretory proteins. *Plant J.* 42, 877–889. doi: 10.1111/j.1365-313X.2005.02419.x
- Showalter, A. M., Keppler, B., Lichtenberg, J., Gu, D., and Welch, L. R. (2010). A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. *Plant Physiol.* 153, 485–513. doi: 10.1104/pp.110.156554
- Shpak, E., Barbar, E., Leykam, J. F., and Kieliszewski, M. J. (2001). Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum*. *J. Biol. Chem.* 276, 11272–11278. doi: 10.1074/jbc.M011323200
- Stafstrom, J. P., and Staehelin, L. A. (1986). The role of carbohydrate in maintaining extensin in an extended conformation. *Plant Physiol.* 81, 242–246. doi: 10.1104/pp.81.1.242
- Strasser, R. (2016). Plant protein glycosylation. *Glycobiology* 26, 926–939. doi: 10.1093/glycob/cww023
- Tavormina, P., De Coninck, B., Nikonorova, N., De Smet, I., and Cammue, B. (2015). The plant peptidome: an expanding repertoire of structural features and biological functions. *Plant Cell* 27, 2095–2118. doi: 10.1105/tpc.15.00440
- Tenhaken, R. (2015). Cell wall remodeling under biotic stress. *Front. Plant Sci.* 5:771. doi: 10.3389/fpls.2014.00771
- Tiainen, P., Myllyharju, J., and Koivunen, P. (2005). Characterization of a second *Arabidopsis thaliana* prolyl 4-hydroxylase with distinct substrate specificity. *J. Biol. Chem.* 280, 1142–1148. doi: 10.1074/jbc.M411109200
- Vázquez-Lobo, A., Roujol, D., Zuñiga-Sánchez, E., Albenne, C., Piñero, D., Gamboa de Buen, A., et al. (2012). The highly conserved spermatophyte cell wall DUF642 protein family: phylogeny and first evidence of interaction with cell wall polysaccharides *in vitro*. *Mol. Phylogenet. Evol.* 63, 510–520. doi: 10.1016/j.ympev.2012.02.001
- Velasquez, S., Ricardi, M., Poulsen, C., Oikawa, A., Dilokpimol, A., Halim, A., et al. (2015). Complex regulation of prolyl-4-hydroxylases impacts root hair expansion. *Mol. Plant* 8, 734–746. doi: 10.1016/j.molp.2014.11.017

- Velasquez, S. M., Ricardi, M. M., Dorosz, J. G., Fernandez, P. V., Nadra, A. D., Pol-Fachin, L., et al. (2011). O-glycosylated cell wall proteins are essential in root hair growth. *Science* 332, 1401–1403. doi: 10.1126/science.1206657
- Ytterberg, A., and Jensen, O. (2010). Modification-specific proteomics in plant biology. *J. Proteomics* 73, 2249–2266. doi: 10.1016/j.jprot.2010.06.002
- Zhang, Y., Giboulot, A., Zivy, M., Valot, B., Jamet, E., and Albenne, C. (2011). Combining various strategies to increase the coverage of the plant cell wall glycoproteome. *Phytochemistry* 72, 1109–1123. doi: 10.1016/j.phytochem.2010.10.019

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

Copyright © 2017 Duruflé, Hervé, Balliau, Zivy, Dunand and Jamet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.