Abstract

Cell walls are complex structures surrounding plant cells. They provide not only mechanical support and protection against environmental changes, but also a mean for cell-to-cell communication. They are mainly constituted of polysaccharides (about 90% of their mass) and proteins. Cell wall proteins (CWPs) play critical roles because they contribute to the plasticity of the cell wall architecture during development and in response to biotic and abiotic environmental changes. Their systematic identification has started in the 2000’s with the sequencing of the genome of the Arabidopsis thaliana model plant and the development of adapted mass spectrometry (MS) technologies. Since then, many other plants have been studied among which plants of agronomical interest. The description of cell wall proteomes has fully benefited not only from the improvement of MS technologies, but also from better sample preparation and peptide separation prior to MS analysis. Bioinformatics has also played critical roles by designing software allowing protein identification, annotation and quantification, as well creating MS data repositories.

Keywords: Cell wall proteins; Mass Spectrometry; Polysaccharides; Arabidopsis thaliana

Introduction

Plant cell walls constitute an extracellular compartment playing many roles during development and adaptation to environmental biotic and abiotic changes [1-4]. They are named primary cell walls as long as cells are growing, and they become secondary cell walls when cells differentiate to specific functions. The major components of primary cell walls are polysaccharides which fall into three categories: cellulose, hemicelluloses and pectins [5]. Secondary cell walls may contain additional polymers such as lignins [6]. All these compounds are present in different proportions depending on plant tissues and on plant species and their structure can be modified thanks to cell wall proteins (CWPs). Although present in minor amounts in cell walls, CWPs play major roles in cell wall structure, plasticity and signaling [7-9]. CWPs can degrade, ligate or even modify polysaccharides, e.g. polygalacturonases which are able to degrade pectic homogalacturonans [10], xylanol glucan endo-transglycosyl hydrolases (XTHs) which cut and religate hemicellulosic xylglucans [11], and pectin methylesterases (PMEs) which demethylate homogalacturonans [12]. These modifications have consequences on the properties of cell walls. For a long time, cell walls models only included the so-called
Bioinformatics & Proteomics Open Access Journal

structural proteins which were assumed to form covalently linked networks giving rigidity to cell walls and protecting cells from pathogen invasion [5,13]. Besides, a few CWP families were characterized from biochemical studies. However, a clear picture of the cell wall proteome was not available, thus leading to an under-estimation of the physiological roles of cell walls.

By the end of the 1990’s, proteomics studies started to develop thanks to impressive progresses in mass spectrometry (MS) technologies and to the availability of newly sequenced genomes. The first plant genome to be sequenced was that of the model plant Arabidopsis thaliana [14]. Then, the genomes of several plants of economical interest became available, such as those of rice [15], tomato [16], Medicago truncatula [17], and linum [18]. Numerous proteomics studies have then published among which those restricted to organelles, such as chloroplasts [19], mitochondria [20] and nuclei [21]. The case of cell walls was more puzzling because cell walls are an open compartment difficult to purify [22]. It contains minor amount of proteins and can be easily contaminated by intracellular proteins including the very abundant photosynthesis proteins. This mini-review aims at summarizing the main steps of the successful story of plant cell wall proteomics which started about twenty years ago with the first description of a small cell wall proteome by Robertson et al. (1997) [23].

In this founding article, cell suspension cultures were washed with salt solutions in order to elute CWPs from their walls without breaking their plasma membranes, thus using a so-called non-destructive method [23]. Twenty proteins were identified after separation by 1D-electrophoresis (1D-E) and Edman N-terminal sequencing. Later on, many other strategies were designed to increase the size of cell wall proteomes [24]. They include: (i) the improvement of cell wall purification procedures to limit the contamination by intracellular proteins in the so-called destructive methods [25]; (ii) the diversification of the salt solutions used to elute CWPs [26,27]; (iii) the introduction of affinity chromatography to separate proteins according to their charge or to their N-glycosylation status [28-32]; (iv) the use of 1D-E for protein separation prior to tryptic digestion because CWPs are mostly basic glycoproteins poorly separated by 2D-E [33]; (v) the use of combinatorial peptide ligand library (CPLL) chromatography to get access to minor CWPs [34,35]. More recently, shotgun analyses omitting the protein separation step have been successfully performed [36] and the first systematic quantitative analysis of CWPs has been published [37]. Altogether, many different types of organs have been analyzed including roots, leaves, inflorescences, fruits, seeds and cell suspension cultures (for an overview, see WallProtDB, www.polebio.lrsv.ups-tlse.fr/WallProtDB/). Plant cell wall proteomics studies have also greatly benefited from the improvement of peptide separation prior to MS analysis and the increase performance of mass spectrometers, from MALDI-TOF MS to LC-MS/MS. Finally, the careful bioinformatics annotation of the identified proteins has allowed (i) better distinguishing proteins predicted to be secreted from intracellular proteins, and (ii) assigning predicted functions to more than 85% of the identified proteins. In this purpose, two bioinformatics tools have been designed in our team: (i) ProtAnnDB (www.polebio.lrsv.ups-tlse.fr/ProtAnnDB/) is a pipeline of prediction of sub-cellular localization and functional domain using bioinformatics programs publicly available [38]; (ii) WallProtDB is a plant cell wall proteomics database collecting published cell wall proteomes after curated annotation of the identified proteins, based on the presence of predicted functional domains and on experimental work [39]. To facilitate the comparisons between cell wall proteomes, CWPs are grouped in nine functional classes in WallProtDB [33]: (i) proteins acting on polysaccharides like glycoside hydrolases [7,40]; (ii) oxi-do-reductases like class III peroxidases [41]; (iii) proteases [42]; (iv) proteins possibly related to lipid metabolism like non-specific lipid transfer proteins (ns-LTPs) [43]; (v) proteins possibly involved in signaling like arabinogalactan proteins (AGPs) [44]; (vi) proteins having interacting domains with proteins or polysaccharides like lectins [45]; (vii) structural proteins like extensins [46]; (viii) miscellaneous proteins; and (ix) proteins of yet unknown function. It should be mentioned that the functional class comprising the structural proteins is the smallest because of the difficulty to extract such proteins which are covalently linked to the other cell wall components [47,48]. This grouping helps getting an overview of newly described cell wall proteomes, but as all classifications, it has limitations and it has to evolve to take into account newly characterized proteins.

Nowadays, the size of newly described organ cell wall proteomes has been increased to 250 to 400 CWPs [27,34,36], i.e. at the most twenty times as large as the first described cell wall proteome. Presently, the larger plant cell wall proteome is that of A. thaliana which comprises more than 900 CWPs and covers about half of the predicted one (see WallProtDB). The second larger cell wall proteome is that of Brachypodium distachyon, which a monocot model plant, with nearly 600 CWPs.

CWP s route through the secretion pathway where they undergo post-translational modifications (PTMs) [49]. The best described are N-glycosylation, Pro hydroxylation and O-glycosylation. All these PTMs can be critical for protein structure and/or biological activity. Proteomics has also brought new information regarding this PTMs. Affinity chromatography on the concanavalin A (ConA) lectin has allowed the separation of N-glycoproteins and their identification. In some cases, the location and structure of N-glycans could also be described [29,30,50]. More recently, some studies have addressed the question of the localization of hydroxyproline (Hyp) residues resulting from the hydroxylation of Pro residues by prolyl-hydroxylation. Hyp residues were initially described in structural proteins such as extensins and in AGPs [51-53]. It seems that they are present in more protein families and that there is some variability in their distribution, so that it is yet difficult to propose a universal Pro hydroxylation code [34,54]. Finally, O-glycosylations are very difficult to describe because of the complexity of the structure of O-glycans. They require dedicated studies and thus cannot yet been included in omics strategies [55,56].

The importance of bioinformatics in proteomics flowcharts needs to be stressed (Figure 1). First of all, the access to annotated genomes with well-predicted open reading frames is critical. Prediction of exon-intron junctions has to be accurate and, if possible, curated with sequenced RNAs as for RefSeq sequences [57]. Indeed, MS-based proteomics is based on the comparison between mass lists of tryptic peptides, which are fragmented or not depending on the MS instrument, and theoretical mass lists calculated from the protein sequences present in databases [58,59]. Then, as stated above, when proteins are identified, it can be important to predict both their sub-cellular localization, especially to check the quality of sub-cellular proteomics experiments, and their functional domains [60,61]. This information allows assessing the biological role of proteins of interest and designing relevant experiments to demonstrate it [38,62]. Finally, public data repositories have been set up to permit sharing MS data [63].

Altogether, cell wall proteomics studies have brought a large amount of information during the twenty last years and they have provided an overview of cell wall proteomes. In particular, many more proteins families playing roles in polysaccharide modifications are included in the present description of cell walls [7]. In the mean time, a lot of new functions have been described for CWP s thanks to genetics studies. For example, the role of CWP s in cell-to-cell communication has been illustrated by several examples including the release of signaling peptides by extracellular proteases to induce cell differentiation or plant defense mechanisms [64]. Next important issues in plant cell wall proteomics concern the identification of more structural CWP s, a better characterization of CWP PTMs, advances in the description of the extracellular peptidome and the
systematic quantification of proteins. Finally, the question of the existence of alternative non-conventional routes of protein secretion also needs to be solved [65]. Indeed, all the published cell wall proteomes mention the presence of proteins predicted to be intracellular [24]. However, to our knowledge, only a sunflower jacalin devoid of predicted signal peptide, has been shown to be secreted through the release of exosomes [66]. Some alternative mechanisms have been described in animal cells [67], but remain to be established for plant cells.

Acknowledgments

The author is thankful to the Paul Sabatier-Toulouse 3 University and to the Centre National de la Recherche Scientifique (CNRS) for their financial support. This work was also granted by the IdEx project (ANR-11-IDEX-0002-02). The author wishes to thank her present and former students and colleagues for fruitful collaborations. She also apologizes for the many articles of interest which could not be quoted in this short article.

References


