

The Vegetative Vacuole Proteome of *Arabidopsis thaliana* Reveals Predicted and Unexpected Proteins ^W

Clay Carter,¹ Songqin Pan,^{1,2} Jan Zouhar,¹ Emily L. Avila, Thomas Girke,³ and Natasha V. Raikhel⁴

Center for Plant Cell Biology, Botany and Plant Sciences, University of California, Riverside, California 92521

Vacuoles play central roles in plant growth, development, and stress responses. To better understand vacuole function and biogenesis we have characterized the vegetative vacuolar proteome from *Arabidopsis thaliana*. Vacuoles were isolated from protoplasts derived from rosette leaf tissue. Total purified vacuolar proteins were then subjected either to multidimensional liquid chromatography/tandem mass spectrometry or to one-dimensional SDS-PAGE coupled with nano-liquid chromatography/tandem mass spectrometry (nano-LC MS/MS). To ensure maximum coverage of the proteome, a tonoplast-enriched fraction was also analyzed separately by one-dimensional SDS-PAGE followed by nano-LC MS/MS. Cumulatively, 402 proteins were identified. The sensitivity of our analyses is indicated by the high coverage of membrane proteins. Eleven of the twelve known vacuolar-ATPase subunits were identified. Here, we present evidence of four tonoplast-localized soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs), representing each of the four groups of SNARE proteins necessary for membrane fusion. In addition, potential cargo of the *N*- and *C*-terminal propeptide sorting pathways, association of the vacuole with the cytoskeleton, and the vacuolar localization of 89 proteins of unknown function are identified. A detailed analysis of these proteins and their roles in vacuole function and biogenesis is presented.

INTRODUCTION

The generally recognized functions of the vacuole include storage (ions, metabolites, and proteins), digestion, pH and ion homeostasis, turgor pressure maintenance, biotic and abiotic defense responses, toxic compound sequestration, and pigmentation (De, 2000). More recently, it has been suggested that this organelle even plays important roles in tropic responses, plant development, and signal transduction (Kato et al., 2002a, 2002b; Morita et al., 2002; Surpin et al., 2003). Not surprisingly, evidence strongly suggests that the vacuole is an essential organelle for plant life (Rojo et al., 2001).

Complicating studies of the vacuole is the fact that there are at least two types of vacuoles found within plants (De, 2000). These have been termed the lytic vacuole (LV) and the protein storage vacuole (PSV). The LV is thought to be analogous to the mammalian lysosome and yeast vacuole, whereas the PSV serves as the main site of protein deposition within seeds and roots (Vitale and Raikhel, 1999). Even so, both types of vacuoles may serve, at least in part, in both storage and degradation

processes and may both be found within individual cells. In mature tissues, often the LV and PSV will fuse to form a large central vacuole that retains functional characteristics of both parent compartments.

Our understanding of the molecular actions of the plant vacuole has greatly increased in recent years. For example, it has been clearly demonstrated that multiple mechanisms exist for targeting proteins to the vacuole (Vitale and Raikhel, 1999), and we are also beginning to understand the details of vacuolar membrane trafficking (reviewed in Surpin and Raikhel, 2004). However, the functional contribution of various vacuolar proteins has primarily proceeded with targeted examinations of proteins previously known to exist within the vacuole. In particular, we are interested in the detailed understanding of how vesicles and their corresponding cargo are targeted to the vacuole. There are at least two distinct mechanisms for transporting soluble proteins to the plant vacuole. Proteins destined for the lumen of the vacuole generally contain a signal peptide (required for default entry into the endoplasmic reticulum [ER]) and also either have a cleavable sequence-specific propeptide (*N*-terminal propeptide [NTPP]; contains the core canonical sequence NPIR) or a non-sequence-specific cleavable *C*-terminal propeptide (CTPP) (Vitale and Raikhel, 1999). The specific characteristics defining a functional CTPP are not known, but in general are 10 to 20 amino acids in length and are highly hydrophobic. Heterologous expression of various CTPP-containing proteins demonstrates that *Arabidopsis thaliana* contains a fully functional CTPP pathway (Rojo et al., 2002). However, to date, no proteins endogenous to *Arabidopsis* have been shown to contain a functional CTPP. Furthermore, only a single *Arabidopsis* protein (aleurain, At5g60360) is known to contain a functional NTPP (Ahmed et al., 2000).

Besides the distinct NTPP and CTPP pathways to the vacuole, several vacuolar targeted proteins also contain undefined sorting

¹ These authors contributed equally to this work.

² Current address: W.M. Keck Proteomics Laboratory, Center for Plant Cell Biology, Botany and Plant Sciences, University of California, Riverside, CA 92521.

³ Current address: Bioinformatics Core Facility, Center for Plant Cell Biology, Botany and Plant Sciences, University of California, Riverside, CA 92521.

⁴ To whom correspondence should be addressed. E-mail nraikhel@ucr.edu; fax 951-827-2155.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Natasha V. Raikhel (nraikhel@ucr.edu).

^W Online version contains Web-only data.

Article, publication date, and citation information can be found at www.plantcell.org/cgi/doi/10.1105/tpc.104.027078.

signals. For example, it has recently been shown that some vacuolar proteins are also transported via ER bodies (precursor protease vesicles) that bud from the ER and directly fuse with the tonoplast (vacuole membrane) (Hara-Nishimura et al., 1998; Matsushima et al., 2003; Rojo et al., 2003a). Also, how integral membrane proteins are targeted to the tonoplast is essentially unknown (Jiang and Rogers, 1998). If we are to understand the complete array of vacuolar sorting determinants, a comprehensive picture of protein content of the plant vacuole is required. Analyzing the specific sequence motifs of vacuolar proteins may reveal unique determinants that are responsible for the observed localization of these proteins.

With the advent of the completely sequenced *Arabidopsis* genome, numerous studies have characterized large-scale protein expression under various conditions as well as in specific organelles (Peltier et al., 2000; Yamaguchi et al., 2000; Schubert et al., 2002; Balmer et al., 2003; reviewed in Cánovas et al., 2004; Friso et al., 2004; Heazlewood et al., 2004). Furthermore, many proteins encoded by the *Arabidopsis* genome (~35%) have no assigned or even putative functions (Tian et al., 2004). Proteomic methodologies can provide invaluable insights into these proteins' potential functions based upon subcellular localizations or changes in expression level in response to a stimulus.

Although much is known about the general functions of the plant vacuole, a detailed knowledge of proteins targeted to the vacuole and their underlying molecular functions is lacking. Yet, this is an essential step for understanding the biology of this organelle. The diverse functions of the vacuole suggest that a large array of proteins is required to conduct all of these processes. In this study, we have used complementary methodologies to identify high- and low-abundance proteins derived from the central vacuoles of rosette leaf tissue. These central vacuoles retain characteristics of both the LV and PSV and contain proteins that are delivered by the NTPP, CTPP, and other trafficking pathways. Here, we describe a comprehensive investigation of the vegetative vacuolar proteome of *Arabidopsis* and present a detailed analysis of the identified proteins and their possible roles in vacuole function. This is a critical step toward understanding vacuolar biogenesis and its involvement in plant biology.

RESULTS AND DISCUSSION

Purity of Samples

Of primary importance, the purity of the starting samples was carefully considered. Toward this end, central vacuoles were purified from mature leaves using a previously described method that was slightly modified (Ahmed et al., 2000). This method results in samples free of markers defining other endomembrane compartments, including the ER (SEC12/At2g01470), Golgi (VPS45/At1g70890 and SYP41/At5g26980), prevacuolar compartment (ELP/At3g52850 and SYP21/At5g16830), and cell plate (KNOLLE/At1g08560) (see Ahmed et al., 2000; Rojo et al., 2003b). Furthermore, vacuole preparations were stained with neutral red and visually checked for purity using bright-field microscopy. No obvious chloroplast contamination was ob-

served by this analysis (e.g., Figure 1A). In addition, we performed analyses to detect other potential contaminants using immunoblotting and fluorescence microscopy (see Discussion below and Supplemental Figure 1 online). Questions of purity and potential impacts of contaminants are discussed wherever appropriate below.

Methodology

To ensure a thorough coverage of the vacuole proteome, complementary proteomic methodologies were explored. A schematic showing the methods used is shown in Figure 1. The chosen methodologies ensure sensitive nonbiased data collection, especially for low abundance, extremely hydrophobic, acidic, or basic proteins (Whitelegge, 2002; Gu et al., 2003). Specifically, a combination of multidimensional liquid chromatography/tandem mass spectrometry (2-D LC MS/MS) (Figure 1A, 1) and one-dimensional (1-D) SDS-PAGE coupled with nano-liquid chromatography/tandem mass spectrometry (nano-LC MS/MS) (Figure 1A, 2) was used on total vacuole preparations. Also, a separate tonoplast-enriched fraction was analyzed by 1-D SDS-PAGE coupled with nano-LC MS/MS (Figure 1A, 3). After separation by liquid chromatography, a survey scan method was used to identify individual peptide precursor ions. Peptides were then sequentially selected for collision-induced dissociation analysis to generate fragment ion spectra (Figure 1B). The resultant fragment ion patterns were analyzed using the MASCOT MS/MS Ions Search tool (http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS).

Through these studies, 402 proteins were identified in the *Arabidopsis* vegetative vacuole. The number of proteins identified by each of the methods is shown in Figure 2 (these groups of proteins do not include 72 [15% of total] potential chloroplast or mitochondrial contaminants, see below). The fact that many proteins were found exclusively by a particular method reinforced the need for multiple approaches for generating a comprehensive proteome list (Whitelegge, 2002; Friso et al., 2004). The specific proteins identified by each method can be found in Table 1 and in Supplemental Table 1 online. Peptide scores for each of these proteins can also be found in Supplemental Table 2 online. A total of 46.5% of proteins were identified by a single peptide hit. However, all such proteins were verified by manual inspection of the data. Proteins were categorized into eight major groups, of which six are presented in Table 1. Indicative of the methods used, only 3 out of 25 glycosyl hydrolases (highly soluble proteins) were found in the thoroughly washed tonoplast fraction. These three hydrolases were also identified using the other methods. Furthermore, 38 of 45 identified transporter proteins (highly membrane associated) were found in the tonoplast fraction. Interestingly, a recently published investigation of the *Arabidopsis* tonoplast proteome identified a large number of glycosidases and proteases (mostly soluble proteins) within their samples (Shimaoka et al., 2004). However, this may be because of a lack of thorough washing of the vacuolar membranes. A list of overlapping proteins found by our group, Shimaoka et al. (2004), and another independent tonoplast examination (Szponarski et al., 2004) can be found in Supplemental Table 1 online.

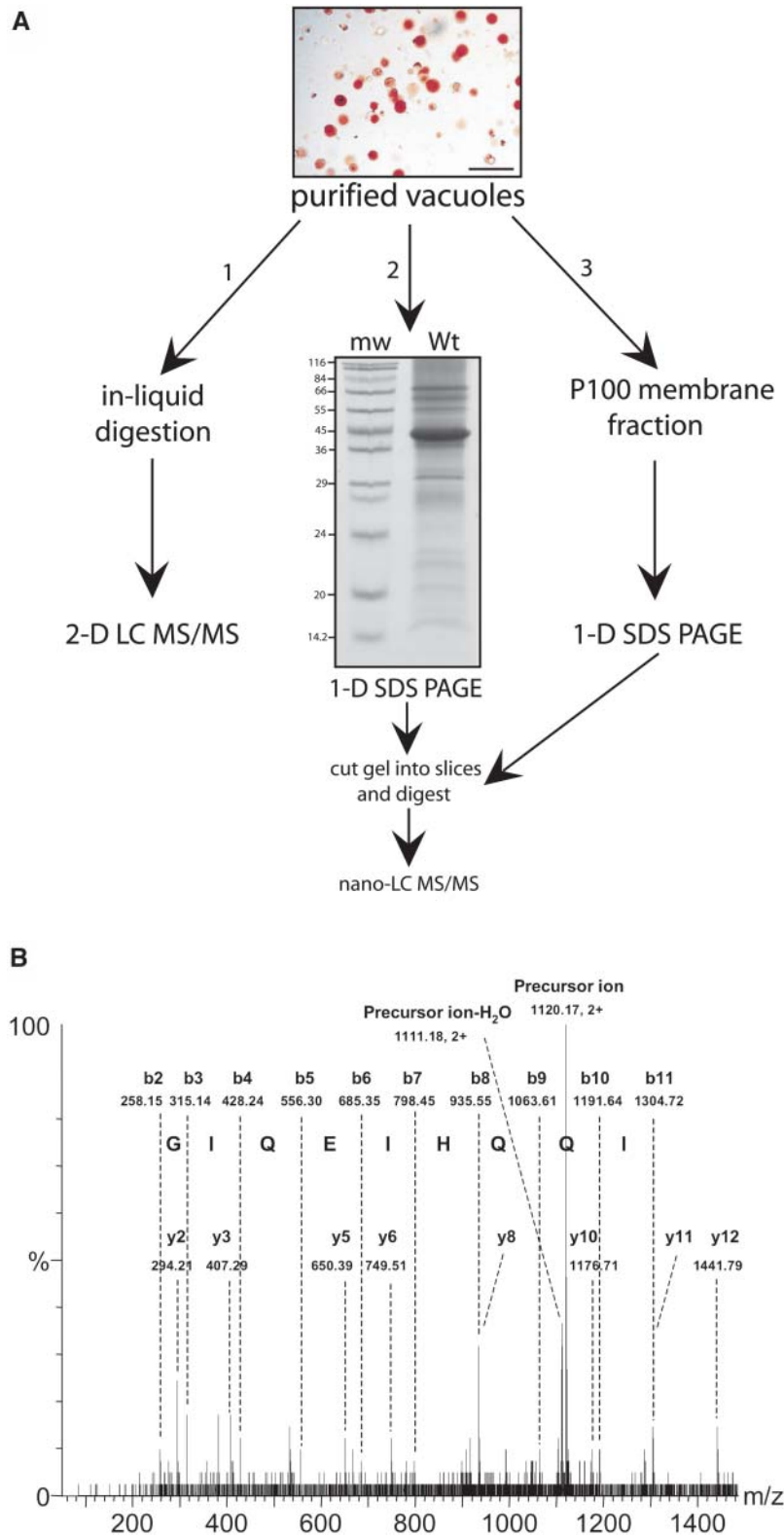


Figure 1. Schematic of Proteomic Studies Conducted on the Vegetative Vacuole.

(A) Purified vacuoles (bar = 50 μ m) were subjected to (1) in-liquid trypsin digestion followed by 2-D LC MS/MS or (2) 1-D SDS-PAGE followed by in-gel

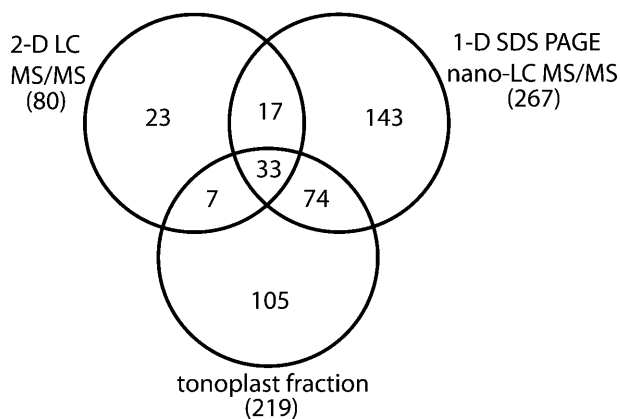


Figure 2. Distribution of Identified Proteins by Different Methods.

Overlap of the different protein sets is shown. Numbers in parentheses indicate the total number of proteins found by a particular method.

Surprisingly, SDS-PAGE analysis of total vacuolar proteins (Figure 1A, 2) revealed that a single protein at ~ 42 kD represented the bulk of vacuolar protein content. This protein was investigated separately using nano-LC MS/MS and matrix-assisted laser-desorption ionization time of flight analyses (data not shown) and was identified as a myrosinase-associated protein (MAP; At3g14210). Interestingly, we also identified five additional MAP gene products (At1g54000, At1g54010, At1g54020, At1g54030, and At1g14220) and the corresponding myrosinase gene products (At5g25980 and At5g2600). The reason that this protein accumulates to such high levels in the vacuoles of mature *Arabidopsis* leaves is unknown. However, proteins that accumulate to such high levels in other systems play roles in buffering changes in osmotic pressure, such as serum albumin in animals (Singh-Zocchi et al., 1999). Perhaps this MAP plays a role in controlling drastic osmotic fluctuations within the vacuole. It is also possible that the MAPs may be involved in modulating the enzyme activity of myrosinase.

A Question of Contamination

Our results indicate that there was little contamination from other components of the endomembrane system; however, several plastidial and mitochondrial proteins were identified. For the analyses described below (and above), the proteins previously known to exist in mitochondria and chloroplasts have been excluded. These proteins were determined to be mitochondrial or plastidial based upon functional annotations and by searching the *Arabidopsis* Mitochondrial Protein Database ([http://www.](http://www.mitoz.bcs.uwa.edu.au/apmdb/APMDB_Database.php)

[mitoz.bcs.uwa.edu.au/apmdb/APMDB_Database.php](http://www.mitoz.bcs.uwa.edu.au/apmdb/APMDB_Database.php)) and the Plastid Proteome Database (<http://cbsusrv01.tc.cornell.edu/users/ppdb/>). A list of these contaminant proteins can be found in Supplemental Table 1 online. The location of identification of these proteins is also indicated (whole vacuole versus tonoplast-specific) in Supplemental Table 1 online. Out of 72 potential contaminant proteins, 27 were found in whole vacuoles, 26 in tonoplast samples, and 19 were found in both. Whether or not these proteins represent true contamination remains to be evaluated. However, immunoblotting was later performed to help determine the level of mitochondrial and chloroplastic contamination. Supplemental Figure 1A online demonstrates that the mitochondrial (COX2) and chloroplast (HSP93) specific markers used were not detected in our vacuole preparations. In addition to immunoblotting assays, we analyzed purified vacuolar samples for potential chloroplast contamination using fluorescence microscopy (Supplemental Figure 1B online). In contrast with lysed protoplasts (the starting material for our organelle fractionation), no fluorescence related to chloroplasts or their fragments was detected in purified vacuolar samples.

Proteomic studies of mouse phagosomes (membrane-bound vesicles formed by an inward folding of the cell membrane to hold foreign matter taken into the cell by phagocytosis; occurs before fusion with the lysosome) have demonstrated the presence of plasma membrane, mitochondrial, cytoplasmic, and ER proteins in these preparations (Garin et al., 2001). The finding of ER proteins in these organelles was later found to have great biological significance based upon the observation that the ER serves as a source of membrane material for maturing phagosomes (Gagnon et al., 2002). Furthermore, proteins and even entire organelles are often targeted to the vacuole for degradation or autophagy processes (De, 2000). Indeed, via electron microscopy we routinely observe mitochondria and plastids engulfed within vacuoles of various tissues (V. Kovaleva and N.V. Raikhel, unpublished data). Proteomic studies on the plant vacuole may uncover proteins that at first appear to be contamination but in fact result from degradation processes.

Bioinformatic Analyses

All identified proteins were analyzed through multiple prediction programs to provide theoretical subcellular localizations, transmembrane domains, molecular masses, and isoelectric points. To ensure a thorough analysis, redundant methods were used for predictions of subcellular localization and transmembrane domains. The results of these prediction programs are shown in Table 1, Figure 3, and in Supplemental Table 1 online.

Figure 1. (continued).

digestion and LC MS/MS, or (3) tonoplast fractions were first enriched and then subjected to 1-D SDS-PAGE followed by in-gel digestion and LC MS/MS.

(B) Mass spectrometry output from LC MS/MS. The fragmentation spectrum (MS/MS) of a peptide derived from a low-abundance tonoplast SNARE protein, SYP22 (At5g46860), is shown. The precursor ion was doubly charged with a mass-to-charge ratio (m/z) 1120.17. The spectrum was matched by MASCOT database searching to a peptide, EQGIQEIHQIGEVNEIFK (amino acids 182 to 200), of SYP22. All matched y - and b -series ions are labeled. Amino acid residues are assigned based on the mass ladders generated by the b -series ions.

Table 1. Functional Categorization and Characterization of Identified Vacuolar Proteins

Locus	Description	MW	pI	Localization			TMDs			2-D LC MS/MS	1-D SDS- PAGE LC MS/MS	Tonoplast
Glycosidases												
At1g09010.1	Glycoside hydrolase family 2, low similarity to mannosidase	107,644	6.42	-	-	-	0	1	0		x	
At1g12240.1	Glycosyl hydrolase family 32, β -fructosidase	73,827	5.30	-	-	-	1	2	1	x	x	
At1g52400.1	Glycosyl hydrolase family 1, β -glucosidase (BG1)	60,442	7.23	S	S	S	1	0	1	x	x	
At1g55120.1	Glycosyl hydrolase, β -fructofuranosidase	67,050	5.52	S	S	S	1	1	0		x	
At1g62660.1	Glycosyl hydrolase family 32, identical to β -fructosidase	72,211	5.27	-	-	-	1	1	1	x		
At1g66270.1	Glycosyl hydrolase family 1, β -glucosidase	59,646	7.02	S	S	S	1	1	0		x	
At2g28100.1	Glycosyl hydrolase family 29 (α -L-fucosidase)	57,169	5.13	S	S	S	1	0	0		x	
At2g32810.1	Glycosyl hydrolase family 35 (β -galactosidase)	99,181	8.11	S	S	S	1	1	1		x	
At3g09260.1	Glycosyl hydrolase family 1, β -glucosidase	59,703	6.91	S	S	S	1	0	0		x	
At3g10740.1	Glycosyl hydrolase family 51, similar to arabinoxylan arabinofuranohydrolase isoenzyme	75,027	5.52	S	S	S	0	1	0		x	
At3g16850.1	Polygalacturonase, putative	49,107	5.65	S	S	S	1	0	0	x	x	
At3g26720.1	Glycosyl hydrolase family 38 (α -mannosidase)	115,203	6.70	S	S	S	1	0	1		x	x
At3g47050.1	Glycosyl hydrolase, family 3, β -D-glucan exohydrolase	67,234	5.08	-	-	-	0	0	0		x	
At3g55260.1	Glycosyl hydrolase family 20, similar to β -hexosaminidase	61,212	6.20	S	S	S	3	0	0	x	x	
At3g56310.1	Glycosyl hydrolase family 27 (α -galactosidase/melibiose)	48,345	4.52	S	S	S	1	1	1	x	x	
At3g57260.1	Glycosyl hydrolase family 17, similar to glucan endo-1,3- β -glucosidase	37,321	4.60	S	S	S	1	1	1		x	
At3g62110.1	Polygalacturonase, putative	51,979	6.44	S	S	S	1	0	0		x	
At3g62750.1	Glycosyl hydrolase family 1, β -glucosidase	55,155	6.72	S	S	S	1	0	1	x	x	
At4g27830.1	Glycosyl hydrolase family 1, β -glucosidase	57,061	5.32	S	S	S	0	0	0	x		
At5g10560.1	Glycosyl hydrolase family 3, β -xylosidase	87,168	6.33	C	S	S	0	0	0		x	
At5g11720.1	Glycosyl hydrolase family 31, similar to α -glucosidase	101,101	5.72	S	S	S	0	1	0	x	x	x
At5g13690.1	α -N-Acetylglucosaminidase	92,672	7.60	S	S	S	0	0	0		x	
At5g13980.2	Glycosyl hydrolase family 38 (α -mannosidase)	115,885	6.41	S	S	S	0	0	0		x	
At5g25980.1	Glycosyl hydrolase family 1, similar to myrosinase precursor	53,779	6.89	S	S	S	1	1	1	x	x	x
At5g26000.1	Glycosyl hydrolase family 1, myrosinase precursor	61,115	5.70	S	S	S	0	0	0	x	x	
Protein Degradation												
At1g02305.1	Expressed protein (cathepsin B-like Cys protease)	40,016	6.91	S	-	S	1	2	1		x	
At1g09850.1	Cys protease XBCP3	48,056	6.42	S	S	S	1	0	0		x	x
At1g11910.1	Asp proteinase-related	54,596	5.20	S	S	S	2	0	1	x	x	x
At1g15000.1	Ser carboxypeptidase-related	49,205	5.19	S	S	S	0	1	0		x	
At1g36340.1	Ubiquitin-conjugating enzyme	17,815	9.63	-	M	-	0	0	0		x	
At1g47128.1	Cys proteinase RD21A	50,949	5.05	S	S	S	1	1	0	x	x	x
At1g62290.1	Asp protease-related	55,731	6.27	S	S	S	1	0	1		x	
At1g78670.1	γ -Glutamyl hydrolase-related	39865	6.17	S	S	S	2	1	0		x	
At1g78680.1	γ -Glutamyl hydrolase (γ -Glu-X carboxypeptidase/conjugase)	38,625	7.16	S	S	S	2	1	0		x	
At2g35780.1	Ser carboxypeptidase-related	51,504	6.92	-	S	S	0	0	0		x	
At3g09790.1	Polyubiquitin (UBQ8)	71,752	9.52	-	-	-	0	0	0			x
At3g10410.1	Ser carboxypeptidase III, putative	57,284	5.03	S	S	S	0	2	1		x	
At3g14067.1	Subtilisin-like Ser protease, putative	81,800	6.76	S	S	S	1	1	0	x	x	x

(Continued)

Table 1. (continued).

Locus	Description	MW	pI	Localization			TMDs				2-D LC MS/MS	1-D SDS- PAGE LC MS/MS	Tonoplast
At3g28220.1	Expressed protein (meprin and TRAF homology domain-containing protein)	42,869	8.80	–	–	–	1	1	1	x	x	x	
At3g51730.1	Expressed protein, saposin precursor	24,080	4.74	S	S	S	0	1	0		x		
At4g01610.1	Cathepsin B-like Cys protease, putative	39,400	6.09	S	S	S	1	2	0		x		
At4g02890.3	polyubiquitin (UBQ14)	34,175	7.82	–	–	–	0	0	0			x	
At4g12910.1	Ser carboxypeptidase-related	54,211	6.32	S	S	S	1	1	1		x		
At4g16190.1	Cys proteinase	41,246	6.97	S	S	S	1	1	1		x		
At4g20850.1	Expressed protein (Ser protease domain)	152,351	5.93	M	M	S	0	0	0		x	x	
At4g30810.1	Serine carboxypeptidase-related	53,038	6.27	S	S	S	1	1	0		x	x	
At4g36190.1	Expressed protein (Ser carboxypeptidase S28 family)	55,416	7.17	S	S	S	0	3	1	x	x		
At4g36195.1	Expressed protein (Ser carboxypeptidase S28 family)	54,756	6.50	S	S	S	1	0	1	x	x		
At5g03240.2	Polyubiquitin (UBQ3) identical to Gl:928809	34,262	7.82	–	–	–	0	0	0			x	
At5g23210.2	Ser carboxypeptidase-related	45,708	8.96	–	–	–	0	0	0			x	
At5g60360.1	Cys proteinase AALP	38,941	6.70	S	S	S	1	0	1		x		
At5g65760.1	Hydrolase, α/β fold family; similar to SPP42785 lysosomal pro-X carboxypeptidase precursor	58,555	5.11	S	S	S	1	1	1		x		
Stress Response													
At1g02920.1	Glutathione transferase, putative	23,580	6.61	–	M	–	0	0	0		x		
At1g02930.1	Glutathione transferase, putative	23,468	6.18	–	M	–	0	0	0		x		
At1g03860.1	Prohibitin 2-related B-cell receptor associated protein	31,793	9.87	S	–	S	1	1	0			x	
At1g21250.1	Wall-associated kinase 1	81,194	5.43	S	S	S	1	1	1			x	
At1g24450.1	Ribonuclease III-related	20,724	10.11	S	S	S	2	2	1		x		
At1g25570.1	Leu-rich repeat protein-related	68,955	5.10	S	S	S	1	1	0		x		
At1g30360.1	ERD4 protein	81,918	9.61	S	S	S	10	9	10			x	
At1g32400.1	Senescence-associated protein family	31,538	5.10	S	S	S	5	4	4		x	x	
At1g52040.1	Jacalin lectin	50,149	5.33	–	–	–	0	0	0			x	
At1g65820.1	Microsomal glutathione S-transferase, putative	16,568	9.32	S	S	S	3	3	3			x	
At1g67360.2	Stress-related protein	26,407	9.41	–	–	S	1	0	0		x	x	
At1g71695.1	Peroxidase, putative	39,542	8.39	S	S	S	2	1	1	x	x		
At1g74020.1	Strictosidine synthase	35,275	5.43	S	S	S	0	0	0	x	x	x	
At1g75040.1	Pathogenesis-related protein 5 (PR-5)	25,234	4.54	S	S	S	1	1	1		x		
At2g15970.1	Cold acclimation protein	21,587	9.23	–	M	–	5	4	5		x		
At2g30860.1	Glutathione transferase, putative	24,128	6.64	C	M	–	0	1	0		x		
At2g30870.1	Glutathione transferase, putative	24,212	5.39	M	C	C	1	1	0		x		
At2g33380.1	RD20 protein, induced by abscisic acid during dehydration	26,582	4.98	–	–	–	1	1	0		x	x	
At2g39330.1	Jacalin lectin	50,426	7.30	–	–	–	0	0	0			x	
At2g39780.1	S-like ribonuclease RNS2	29,135	6.38	S	S	S	1	0	0		x		
At3g05500.1	Stress-related protein-related	27,177	8.56	–	–	–	0	0	0			x	
At3g16470.1	Jacalin lectin	48,479	4.91	–	–	–	0	0	0		x	x	
At3g22060.1	Receptor protein kinase-related	27,846	8.90	S	S	S	1	1	1		x		
At3g27280.1	Prohibitin-related	30,620	7.82	S	–	S	1	0	0		x	x	
At3g32980.1	Peroxidase, putative	38,829	6.65	C	S	S	0	0	0		x	x	
At3g49110.1	Peroxidase	38,924	6.86	S	S	S	1	1	1	x			
At3g49120.1	Peroxidase, putative	38,814	7.70	S	S	S	1	1	0	x	x		
At3g57020.1	Strictosidine synthase-related	41,440	7.01	S	S	S	1	1	1			x	
At4g02520.1	Glutathione transferase, putative	24,111	6.32	–	M	–	0	0	0		x	x	
At4g08770.1	Peroxidase, putative	38,185	7.72	S	S	S	0	0	0	x			
At4g08780.1	Peroxidase, putative	38,068	7.72	S	S	S	0	0	0	x			
At4g16500.1	Cys proteinase inhibitor-like protein	12,538	9.72	S	S	S	1	1	1		x		
At4g23670.1	Major latex protein (MLP)-related	17,500	6.34	–	–	–	0	0	0		x		
At4g25090.1	Respiratory burst oxidase (NADPH oxidase), putative	96,846	9.41	–	–	–	5	5	4		x		
At5g01820.1	CBL-interacting protein kinase 14	50,280	8.21	–	–	–	1	1	0		x		
At5g18100.1	Copper/zinc superoxide dismutase (CSD3)	16,923	7.73	–	–	–	0	0	0		x		

(Continued)

Table 1. (continued).

Locus	Description	MW	pI	Localization			TMDs			2-D LC MS/MS	1-D SDS- PAGE LC MS/MS	Tonoplast
At5g24770.1	Vegetative storage protein, Vsp2	29,825	6.28	S	S	S	1	0	0	x		
At5g24780.1	Vegetative storage protein, Vsp1	30,244	5.39	S	S	S	1	0	0	x	x	
At5g40770.1	Prohibitin (gbAAC49691.1)	30,382	7.89	S	-	S	1	0	0		x	
At5g44020.1	Vegetative storage protein-related	31,039	8.03	S	S	S	1	1	0	x	x	
At5g49760.1	Receptor protein kinase-related	104,691	6.08	S	S	S	3	2	2	x	x	
At5g58150.1	Leu-rich repeat transmembrane protein kinase, putative	86,594	7.54	M	S	S	2	1	1	x		
At5g62740.1	Hypersensitive-induced response protein	31,413	5.07	-	-	-	0	0	0	x	x	
Membrane Fusion and Remodeling												
At1g02130.1	Ras-related GTP binding protein ARA-5	22,631	4.77	-	-	S	0	0	0		x	
At1g12360.1	KEULE	75,066	8.08	-	-	-	0	0	0	x		
At1g16240.1	Syntaxin of plants 51 (SYP51)	25,871	8.22	-	-	-	1	1	1	x	x	
At1g18070.1	EF-1- α -related GTP binding protein, putative	59,225	4.96	-	-	-	1	0	0	x		
At1g18210.2	Calcium binding protein, putative	18,332	3.95	-	-	-	0	0	0	x		
At1g22740.1	Ras-related GTP binding protein (Rab7)	22,927	5.09	M	M	M	0	1	0	x	x	
At1g27480.1	Expressed protein similar to lecithin:cholesterol acyltransferase precursor	48,290	8.78	S	S	S	1	2	1	x	x	
At1g28580.1	Lipase, putative	43,007	4.97	S	S	S	4	2	0	x		
At1g31550.1	Lipase, putative	43,079	5.54	S	-	S	3	0	0	x		
At1g32410.2	Expressed protein (vacuolar protein sorting 55 family protein)	15,007	4.19	S	-	S	4	2	4	x	x	
At1g35720.1	Ca ²⁺ -dependent membrane binding protein annexin	36,186	5.02	-	-	-	0	0	0	x		
At1g43620.1	Sterol glucosyltransferase-related	68,327	5.47	-	-	-	3	2	0		x	
At1g45201.1	Conserved hypothetical protein contains Pfam profile: PF01764 lipase	54,776	9.04	-	C	-	0	2	1		x	
At1g52280.1	GTP binding protein, putative	23,051	4.94	S	S	M	0	1	0	x		
At1g53590.1	C2 domain-containing protein	84,884	5.81	S	S	S	1	1	2		x	
At1g54000.1	Myrosinase-associated protein, putative; contains Pfam profile PF00657: GDSL-like lipase/acylhydrolase	43,176	7.12	S	S	S	1	1	1	x		
At1g54010.1	Myrosinase-associated protein, putative; contains Pfam profile PF00657: lipase/acylhydrolase with GDSL-like motif	43,126	8.19	S	S	S	1	1	1	x		
At1g54020.1	Myrosinase-associated protein, putative; contains InterPro entry IPR001087 lipolytic enzyme, G-D-S-L family	32,438	7.84	C	-	-	0	0	0	x	x	
At1g54030.1	Myrosinase-associated protein-related; contains Pfam profile PF00657: GDSL-like lipase/acylhydrolase	46,065	6.71	C	-	-	1	1	0	x	x	
At1g64460.1	Phosphatidylinositol 3- and 4-kinase	33,158	5.08	-	-	-	0	0	0	x		
At1g65320.1	CBS domain containing protein	46,538	7.27	-	-	-	2	2	0	x	x	
At1g70490.2	ADP-ribosylation factor 1 (ARF1), putative	20,575	6.96	C	-	M	1	0	0		x	
At1g74210.1	Glycerophosphodiester phosphodiesterase-related	45,004	7.11	S	S	M	0	0	0	x		
At1g79590.1	Syntaxin of plants 52 (SYP52)	26,091	8.70	-	-	-	1	1	1		x	
At2g01540.1	C2 domain-containing protein	20,067	6.91	-	-	-	0	0	0	x	x	
At2g07050.1	Cycloartenol synthase [(S)-2,3-epoxysqualene mutase] (CAS1)	86,016	6.01	-	-	-	0	2	0		x	
At2g14720.2	Spot 3 protein and vacuolar sorting receptor AtELP2a	69,795	5.11	S	S	S	1	1	1		x	
At2g20990.1	C2 domain-containing protein, similar to Ca ²⁺ -dependent lipid-binding protein (CLB1)	61,727	7.67	S	S	S	1	1	1	x		
At2g21160.1	Signal sequence receptor, α subunit	28,150	4.84	S	S	S	2	1	1	x		
At2g38760.1	Annexin-related	36,238	6.38	-	-	-	0	0	0		x	
At2g45200.1	cis-Golgi SNARE protein, putative	26,452	9.64	-	-	-	1	1	1	x		

(Continued)

Table 1. (continued).

Locus	Description	MW	pI	Localization			TMDs			2-D LC MS/MS	1-D SDS- PAGE LC MS/MS	Tonoplast
At3g03520.1	Phosphoesterase family	59,080	5.12	–	–	–	0	0	0			x
At3g14210.1	Myrosinase-associated protein, putative; contains Pfam profile:PF00657 lipase/acylhydrolase with GDSL-like motif	44,043	7.82	S	S	S	1	1	1	x	x	x
At3g14220.1	GDSL-motif lipase/hydrolase protein, similar to myrosinase-associated proteins	40,231	6.85	S	S	S	1	0	1		x	
At3g22845.1	Expressed protein (mp24/gp25L/p24 protein-related)	24,264	6.86	–	–	–	1	1	1		x	
At3g46060.1	Ras family GTP binding protein	23,817	8.02	–	S	–	2	0	0			x
At3g56190.1	α -Soluble NSF attachment protein	32,737	4.98	–	–	–	0	0	0		x	x
At3g60340.2	Palmitoyl protein thioesterase precursor, putative	37,066	5.00	S	S	S	1	2	1		x	x
At3g62290.1	ADP-ribosylation factor	20,590	6.96	M	M	–	1	0	0			x
At4g04910.1	Component of vesicle-mediated transport-related, similar to <i>N. tabacum</i> N-ethylmaleimide sensitive fusion protein	81,470	5.98	M	C	–	0	0	0		x	
At4g17170.1	GTP binding protein (At-RAB2)	23,147	7.50	S	–	–	0	0	0		x	x
At4g17530.1	Ras-related small GTP binding protein RAB1c	22,300	5.04	–	–	S	0	0	0		x	
At4g20260.2	Endomembrane-associated protein	24,566	4.65	–	–	–	0	0	0		x	x
At4g21160.2	Zinc finger and C2 domain protein (ZAC)	37,128	4.90	–	–	–	0	0	0		x	
At4g24840.1	Brefeldin A-sensitive Golgi protein-like	84,602	5.49	–	–	–	0	0	0		x	
At4g32150.1	Synaptobrevin-related protein	25,021	9.39	M	S	S	1	1	1			x
At4g35790.2	Phospholipase D-related	97,762	7.17	–	–	–	0	1	0			x
At5g13710.1	24-Sterol C-methyltransferase	38,251	6.12	–	–	–	0	0	0		x	x
At5g39510.1	v-SNARE AtVT1a (ZIG)	24,933	10.10	–	–	–	1	1	1		x	x
At5g46860.1	Syntaxin of plants SYP22 (VAM3p)	29,464	6.37	–	–	M	1	1	1			x
At5g47710.1	C2 domain-containing protein	18,317	7.30	–	–	–	0	0	0		x	
Transporters												
At1g11260.1	Glucose transporter	57,593	9.19	M	–	–	12	11	12			x
At1g12840.1	Vacuolar ATP synthase subunit C-related	42,602	5.20	M	–	M	0	0	0	x	x	x
At1g15690.1	Inorganic pyrophosphatase-related	80,803	4.89	S	S	S	16	14	14	x	x	x
At1g20260.1	Vacuolar H ⁺ -ATPase subunit B-related	36,341	4.56	–	–	–	0	0	0	x		
At1g20840.1	Transporter-related	79,468	5.06	M	–	S	11	10	10			x
At1g30400.1	Glutathione S-conjugate ABC transporter (AtMRP1)	181,911	6.08	–	–	–	15	14	14			x
At1g64200.1	H ⁺ -transporting ATPase protein-related	27,067	6.11	–	–	–	0	0	0		x	
At1g76030.1	Vacuolar ATP synthase subunit B	54,090	4.73	–	–	–	0	0	0	x	x	x
At1g78900.1	ATPase 70-kD subunit-related	68,795	4.86	–	–	–	0	0	0	x	x	x
At1g80310.1	Sulfate transporter-related	49,534	8.56	C	–	–	10	8	9			x
At2g02040.1	His transport protein (PTR2-B)	64,404	5.34	–	–	–	11	10	10			x
At2g18960.1	ATPase 1, plasma membrane-type (proton pump 1)	104,207	6.67	–	–	–	10	9	10			x
At2g21410.1	Vacuolar proton-ATPase subunit-related	93,089	5.24	–	–	–	6	7	6	x	x	x
At2g26975.1	Copper transport protein-related	15,788	8.50	C	–	–	2	2	3		x	
At2g28900.1	Membrane channel protein-related	15,464	9.59	C	–	–	4	0	0		x	x
At2g38170.1	Calcium exchanger (CAX1)	50,159	6.16	–	–	C	10	9	8	x	x	x
At2g41560.1	Potential calcium-transporting ATPase 4	112,733	5.94	–	–	–	10	9	8	x	x	x
At2g43330.1	Sugar transporter family	54,795	4.94	C	–	–	12	9	12			x
At2g47800.1	Glutathione-conjugate transporter AtMRP4	169,064	7.70	C	C	–	17	14	15	x		
At2g48020.2	Sugar transporter, putative	49,682	8.96	–	–	–	12	10	12		x	x
At3g01280.1	Porin-related	29,408	9.23	–	M	–	0	0	0			x
At3g01390.2	Vacuolar membrane ATPase subunit G (AVMA10)	12,379	5.82	–	–	–	0	0	0		x	x
At3g03720.1	Cationic amino acid transporter-related	63,621	6.09	–	–	–	14	13	14			x
At3g16240.1	δ Tonoplast integral protein (δ -TIP)	25,009	5.43	–	–	S	7	6	6	x	x	x
At3g26520.1	Tonoplast intrinsic protein, putative	25,831	4.72	–	–	–	6	6	7	x	x	x

(Continued)

Table 1. (continued).

Locus	Description	MW	pI	Localization	TMDs	2-D LC MS/MS	1-D SDS- PAGE LC MS/MS	Tonoplast
At3g26590.1	MATE efflux protein	54,304	5.73	- - -	12 11 11			x
At3g28710.1	Adenosine triphosphatase-related, similar to vacuolar adenosine triphosphatase subunit D	40,774	4.82	- - -	0 1 0	x	x	x
At3g28715.1	Expressed protein, similar to vacuolar adenosine triphosphatase subunit D	40,770	4.75	- - -	0 1 0	x	x	x
At3g30390.1	Amino acid transporter	49,500	6.50	- - -	11 10 11			x
At3g42050.1	Vacuolar ATP synthase subunit H-related	50,267	7.04	- - -	0 1 0	x	x	x
At3g57330.1	Potential calcium-transporting ATPase 11	111,928	6.27	- - C	10 9 10	x	x	x
At3g58730.1	V-ATPase subunit D (vATPD)	29,041	10.17	S - M	0 0 0		x	x
At3g62700.1	Glutathione-conjugate transporter, putative	172,121	8.72	M - -	17 14 16	x		x
At4g02620.1	V-ATPase-related	14,242	6.53	- - -	2 1 0		x	x
At4g03560.1	Two-pore calcium channel (TPC1)	84,856	4.62	- - -	12 14 11	x	x	x
At4g11150.1	H ⁺ -transporting ATPase chain E, vacuolar	26,042	6.31	- - -	0 0 0	x	x	x
At4g23710.1	V-ATPase subunit G (vag2 gene)	11,724	5.28	- - -	0 0 0		x	
At4g35300.1	Transporter-related, low similarity to hexose transporter	79,709	5.35	S - S	11 10 11			x
At4g38510.1	Probable H ⁺ -transporting ATPase	54,288	4.78	- - -	0 0 0	x	x	x
At4g38920.1	H ⁺ -pumping ATPase 16- kD proteolipid, vacuolar (ava-p1)	16,554	8.68	S - S	4 4 4		x	x
At4g39080.1	Proton pump-related vacuolar proton ATPase 100-kD subunit	92,817	5.77	- - -	6 7 6	x	x	x
At5g14120.1	Nodulin-related protein	63,132	7.54	M M S	13 11 12	x		
At5g15090.1	Voltage-dependent anion-selective channel protein hsr2	29,193	8.68	- M -	0 0 0			x
At5g39040.1	ABC transporter family protein	69,086	8.28	M - -	6 5 4	x		x
At5g64870.1	Nodulin-related	52,816	7.29	M - -	0 0 0			x
At5g67330.1	NRAMP metal ion transporter 4 (NRAMP4)	56,368	4.72	- - -	12 11 12	x		
Cytoskeleton								
At1g49240.1	Actin 8	41,845	5.29	- - -	0 0 0			x
At1g52410.1	Myosin-related protein	84,001	4.29	S S S	1 1 0		x	x
At1g59610.1	Dynamin-related protein	100,211	9.59	- - -	0 0 0			x
At1g64330.1	Expressed protein, myosin heavy chain-related	64,477	4.82	- - -	0 0 0		x	x
At3g58160.1	Myosin heavy chain MYA3	141,397	5.95	- - -	1 0 0		x	
At3g60190.1	Dynamin-related protein 4 (ADL4)	69,787	7.59	C - -	0 0 0		x	x
At3g60840.1	Microtubule associated protein (MAP65/ASE1)	73,430	7.60	- - -	0 0 0		x	
At5g41790.1	Myosin heavy chain-related protein	149,945	4.55	- - -	0 0 0		x	x
At5g42080.1	Dynamin-related protein	68,155	8.51	- - -	0 0 0			x
At5g55400.1	Fimbrin	79,771	4.88	M C C	0 0 0		x	

Locus numbers and descriptions are based on TAIR annotations. Subcellular localization predictions are listed in order by iPSORT, PREDOTAR, and TargetP (-, no signal; S, signal peptide; C, chloroplast transit peptide; M, mitochondrial transit peptide). Transmembrane domain (TMD) predictions are listed in order by HMMTOP, THUMBUP, and TMHMM. The fraction/method in which each protein was identified is also indicated. MW, predicted molecular weight.

The size and pI distribution of these proteins, based upon predicted gene models, is shown in Figures 3A and 3B, respectively. These analyses demonstrate that proteins with a wide range of sizes and isoelectric points were identified and that there are no observable gaps, also indicating that there is little bias or information loss with the methods used. The predicted mean and median pI of all identified vacuolar proteins are 6.69 and 6.39, respectively. These are more acidic than the mean (7.40) and median (7.32) pI for all predicted proteins of the Arabidopsis genome (calculated from predicted protein masses available at The Arabidopsis

Information Resource [TAIR] Web site: <http://arabidopsis.org/tools/bulk/protein/index.jsp>). Also, the average and median predicted molecular masses of all Arabidopsis proteins are 46.6 and 39.8 kD, respectively, versus the 51.5 (mean) and 43.1 kD (median) of the identified vacuolar proteins. Thus, the average predicted mass of the identified vacuolar proteins is slightly larger than that of the average Arabidopsis protein. Of course, the mature masses and isoelectric points of many of these proteins may be different than the predicted ones as a result of processing and degradation within the vacuole.

Targeting algorithms predicted signal peptides (cleavable N-terminal signal for entry into ER) in 149 (37% of total, PREDOTAR), 150 (37%, iPSORT), and 169 (42%, TargetP) of the identified proteins, with 131 (33%) of these predicted by all three search algorithms. Overall, 179 proteins (45%) were predicted to have a signal peptide by at least one of these prediction programs. Transmembrane domains were also predicted using HMMTOP, TMHMM, and THUMBUP (see Methods), and results from this analysis can be found in Supplemental Table 1 online. The distribution of proteins with transmembrane domains predicted by HMMTOP is shown in Figure 3C. After removal of all proteins with signal peptides (often the signal peptide is predicted as a transmembrane domain), another 112 proteins were predicted to have at least one transmembrane domain by one or more prediction programs. Although the prediction programs used are not perfect, overall, these results indicate that most proteins targeted to the vacuole are either delivered through the ER/Golgi or are membrane associated. Of course, there is the possibility that certain proteins containing signal peptides or transmembrane domains may be targeted to the vacuole via degradation or recycling processes. In addition, many identified proteins not predicted to have signal peptides or transmembrane domains are known to be associated with integral membrane proteins or other aspects of vacuole function, including nine peripheral vacuolar ATPase (V-ATPase) subunits (Figure 4) and three polyubiquitin proteins.

Vacuolar Sorting Signals

Some proteins previously shown to contain a functional CTPP include tobacco (*Nicotiana tabacum*) PR-5 (osmotin), tobacco chitinase, barley (*Hordeum vulgare*) lectin, and horseradish (*Armoracia lapathifolia*) peroxidase C1a (HRP C1a) (Bednarek and Raikhel, 1991; Neuhaus et al., 1991; Melchers et al., 1993; Matsui et al., 2003). To date, no endogenous Arabidopsis proteins have been demonstrated to contain a functional CTPP. However, the CTPP pathway was proven to be fully functional in Arabidopsis plants, as recombinant CTPP proteins from other plants species were properly targeted to the vacuole via dense vesicles (Ahmed et al., 2000; Rojo et al., 2002). This leads us to the conclusion that there are endogenous Arabidopsis proteins carrying functional CTPP sorting signals that have not yet been identified. In our study, we identified Arabidopsis PR-5 (At1g75040), three jacalin-like lectins (At1g52040, At2g39330, and At3g16470), and six putative peroxidases (At1g71695, At3g32980, At3g49110, At3g49120, At4g08770, and At4g08780) in our samples. When the vacuolar peroxidase sequences are aligned with all other Arabidopsis peroxidases from the Arabidopsis genome (84 proteins), it becomes clear that the vacuolar peroxidases contain a highly hydrophobic C-terminal extension not found in other peroxidases. The aligned C-terminal regions of the identified vacuolar peroxidases with closely related Arabidopsis genes and HRP C1a are shown in Figure 5A. The functional CTPP of HRP C1a is nearly identical to the C termini of the identified Arabidopsis vacuolar peroxidases, and the cleavage site (S338) is also conserved (Figure 5). Only three other peroxidases containing a similar C-terminal extension are found in the Arabidopsis genome (At2g38380,

At2g38390, and At5g06730). Thus, we have identified six out of nine peroxidases that likely contain CTPPs. These peroxidases as well as the identified lectins and PR-5 all contain hydrophobic C termini similar to known CTPP proteins (data not shown) and most likely represent native targets to study CTPP transport mechanisms in Arabidopsis.

Our analyses of vacuolar proteins only identified aleurain (At5g60360) and a putative Cys protease (At4g16190) that contain the core canonical NTPP sequence NPIR. Furthermore, in Arabidopsis, only aleurain has been previously shown to be transported via the NTPP pathway (Ahmed et al., 2000). To identify other cargo potentially transported by the NTPP pathway, proteins predicted to contain a signal peptide and two or fewer transmembrane domains (142 total proteins) were searched for divergent versions of the NTPP signal. This sequence was based upon other known sequence-specific vacuolar sorting signals (sweet potato [*Ipomoea batatas*] sporamin, barley aleurain, potato [*Solanum tuberosum*] PT20, castor bean [*Ricinus communis*] ricin, and pumpkin [*Cucurbita* sp] 2S albumin) and determined to be [N/L]-[P/I/L]-[I/P]-[R/N/S] (Matsuoka and Nakamura, 1991; Holwerda et al., 1992; Koide et al., 1999; Matsuoka and Neuhaus, 1999; Frigerio et al., 2001). The results from this analysis can be found in the supplemental data online. This analysis identified 42 proteins in this subset with potential NTPP signals. An alignment of nine of these putative NTPP-containing proteins is shown in Figure 5B. These proteins all contain an NTPP variant immediately following the predicted signal peptide cleavage site (data not shown). Some vacuolar proteins contain functional variants of the NTPP sequence in the middle or at the C terminus of the protein (Frigerio et al., 2001). Thus, the other 31 identified proteins that contain an NTPP variant far downstream from the signal peptide cleavage site (see supplemental data online) may also be localized to the vacuole by this sorting determinant. These proteins represent new native targets to study the NTPP pathway in Arabidopsis.

Based upon the large number of predicted soluble proteins found in our samples and the small number of NTPP- and CTPP-containing proteins, it is clear that alternative mechanisms to the NTPP and CTPP pathways likely exist for transporting proteins to the vacuole. As discussed previously, it is now known that ER bodies represent a newly identified means of protein transport to the vacuole. Some of these proteins contain putative ER retention signals (e.g., β -glucosidase) (Matsushima et al., 2003), whereas others do not (e.g., invertase and vacuolar processing enzymes) (Rojo et al., 2003a). Indeed, our samples identified nine proteins that contain potential ER retention signals (discussed below). Yet, the means by which a number of the other identified proteins are transported to the vacuole are unknown. To address this finding, several bioinformatics analyses were performed to identify other potential vacuolar sorting determinants.

The programs Gibbs Motif Sampler, MEME, and Block Maker were used for identifying novel targeting motifs (pattern discovery approach) in the same protein set used for NTPP signal prediction. The results from these analyses can be found at <http://bioinfo.ucr.edu/projects/VacuoleProteomics/Overview.html>. Unfortunately, under a wide range of parameter settings no obvious new patterns could be identified using these tools. As

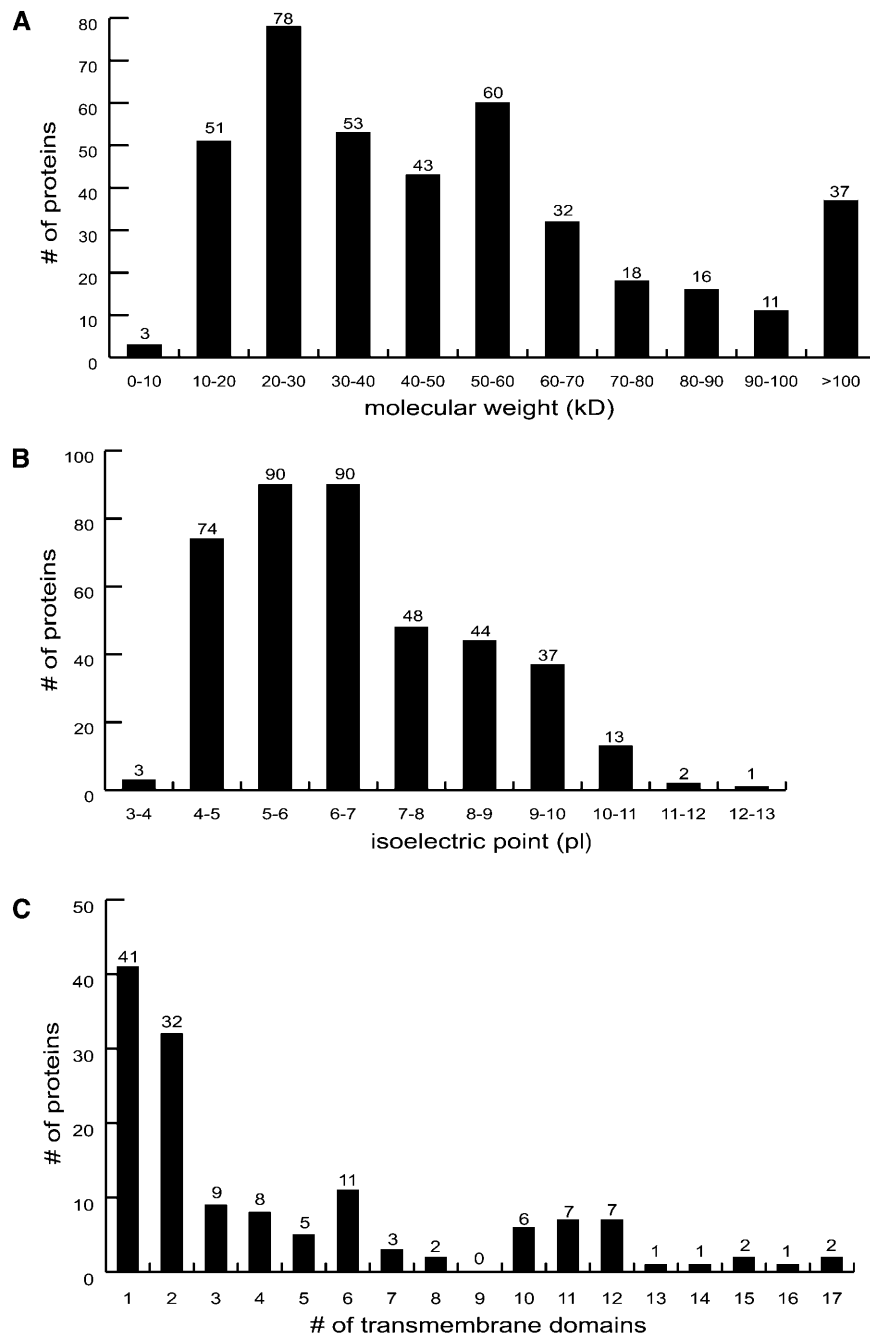


Figure 3. Bioinformatic Analyses of the Identified Proteins.

(A) Size distribution of the identified proteins in 10-kD increments.

(B) Distribution of proteins versus isoelectric point (pI).

(C) Number of predicted transmembrane domains identified in vacuolar proteins as determined by the HMMTOP method. Note that all proteins with a predicted signal peptide and a single predicted transmembrane domain were removed before this analysis because the transmembrane domain prediction programs used usually identify signal peptides as a transmembrane domain.

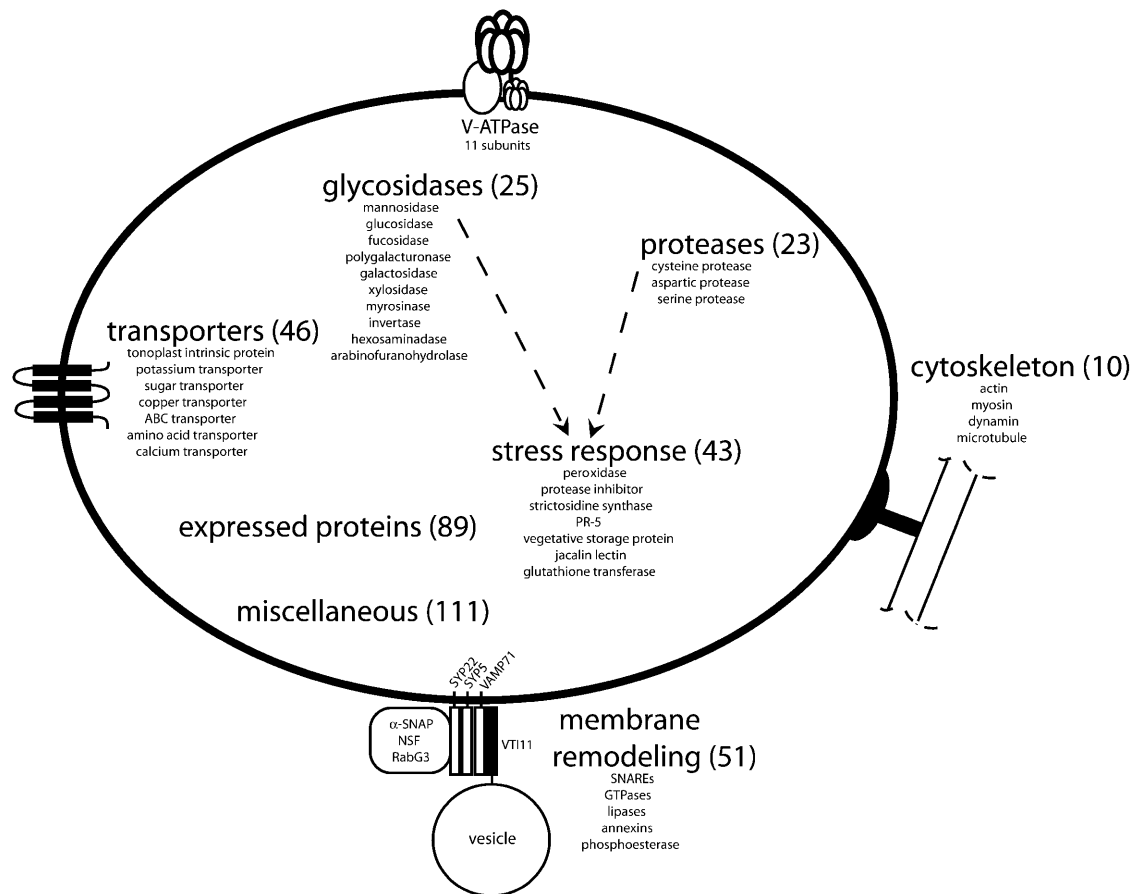


Figure 4. Schematic of the Plant Vacuole and Functional Categorization of the Identified Vacuolar Proteins.

Identified proteins were grouped into eight major categories, and the number of proteins belonging to each group are shown in parentheses. Some of the specific activities of each group and a model of the tonoplast SNARE-pin complex are also shown.

a parameter test, the known ER retention signal KDEL could be easily identified with all three programs in a test protein set that contained different variants of this motif. To identify potential non-sequence-specific signals, the physicochemical properties and hydrophobicity profiles of all proteins in this set were analyzed with several proteomics tools of the EMBOSS package (<http://www.hgmp.mrc.ac.uk/Software/EMBOSS/Apps/>). The obtained histograms of these analyses can be accessed at the same Web site. Visual inspections of these plots did not reveal noticeable patterns of compositional properties in these proteins. These negative results indicate that many vacuolar proteins may be targeted to the vacuole via more complex mechanisms than obvious sequence or amino acid distribution patterns.

Identification of a Putative Tonoplast SNARE-Pin Complex and Its Implications to Protein Trafficking

The vacuole is a major component of the plant endomembrane system, and the correct targeting of vesicles and their cargo is the primary basis for the compartmentalization and identity of the

vacuole. The primary control of membrane fusion in eukaryotes is mediated by soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) proteins (reviewed in Bock et al., 2001). Data on SNARE localization to the tonoplast has been accompanied with some level of controversy. In previous observations, SYP22 (At5g46860) was found on the tonoplast of meristematic cells (Sato et al., 1997); however, other studies have suggested that SYP22 might be localized exclusively to the prevacuolar compartment (Sanderfoot et al., 1999). Recently, we identified a protein complex directly interacting with SYP22 that is localized to the tonoplast (Rojo et al., 2003b), providing indirect evidence of SYP22-tonoplast localization. Here, we present additional evidence that the Qa-SNARE, SYP22, is localized to the tonoplast (Figure 1B) along with four other SNARE proteins. Tonoplast localization of SYP22 was also detected by Szponarski et al. (2004); however, the protein (NP_568671/At5g46860) was misannotated as an expressed mitochondrial protein. Two proteins, Qc-SNAREs SYP51 (At1g16240) and SYP52 (At1g79590), may have redundant functions (hereafter referred to as SYP5s) (Sanderfoot et al., 2001). A third identified protein is VTI11 (At5g39510), a Qb-SNARE that likely acts as

A

P00433 (HRP)	NCRVVNSNS [▼] LLHDMVE-VVDFVSSM-----	353
At1g71695*	NCSARNTQSFMSVLEEGIEEAIISMI-----	358
At3g32980*	NCRVVNSNSLLHDVVE-IVDFVSSM-----	352
At3g49110*	NCRVVNSNSLLHDVVD-IVDFVSSM-----	354
At3g49120*	NCRVVNSNSLLHDVVD-IVDFVSSM-----	353
At4g08770*	NCRVVNSKSKIMDVVEDALEFASSM-----	346
At4g08780*	NCRVVNSKSKIMDVVEDALEFASFM-----	346
At2g38380	NCRVVNPRIRVVENDDGVVSSI-----	349
At5g06730	DCRVVNGQSSATEAGDIQLQSDGFPVSADM	358
At5g06720	DCKKVNGS-----	335
At4g36430	NCRKINS-----	331
At3g50990	ICRRVNHVD-----	336
At5g66390	ICRRVNHAY-----	336
At2g18140	KCRKINNS-----	337
At1g14550	ICSFVN-----	310
At5g19880	DCRRVN-----	329
At5g58400	SCRPN-----	325

B

At1g67240	31	LGFD <u>NLI</u> SLCS	44
At1g78680	23	KAATI <u>LLPS</u> QTGFD	36
At2g20990	26	VLFFV <u>LLPN</u> DKDP	39
At3g10410	32	TFERS <u>NLPS</u> TRAEK	45
At4g16190	29	TDGFV <u>NPIR</u> QVVPE	42
At4g23200	41	RVILS <u>LLPS</u> NVTSH	54
At4g29260	55	VTWD <u>LIPS</u> ICVDS	68
At5g60360	25	GFDES <u>NPIR</u> MVSDG	38
At5g65760	25	LSSSK <u>LLPR</u> FPRYT	38

Figure 5. Proteins with Putative Vacuolar Sorting Determinants.

(A) Alignment of the C-terminal region of vacuolar peroxidases with closely related peroxidases from Arabidopsis. All identified vacuolar peroxidases (labeled with an asterisk) have a hydrophobic C-terminal extension that may serve as a CTPP sorting signal. Only three peroxidases in the Arabidopsis genome, At2g38380, At2g38390 (data not shown), and At5g06730, have a similar C-terminal extension but were not found in the vegetative vacuole. Horseradish peroxidase C1a is also included in the alignment (GenBank accession no. P00433), and the native site of peptide cleavage is indicated with an arrowhead.

(B) Aligned regions of putative NTPP-containing proteins. Proteins containing the degenerate signal [N/L]-[P/I/L]-[I/P]-[R/N/S] immediately following the predicted signal peptide cleavage site are shown, and the putative NTPP signals are in bold and underlined.

a v-SNARE for the SYP22 t-SNARE complex (Sanderfoot et al., 2001). VTI11 and SYP5s were previously identified as interacting partners of SYP22 based on immunoprecipitation and colocalization experiments. A fourth identified SNARE protein is VAMP711 (At4g32150), an R-SNARE from the longin family. The closely related VAMP712 (At2g25340) and VAMP714 (At5g22360) were previously identified in a tonoplast-derived sample (Szponarski et al., 2004), and based on protein similarity, VAMP711/712/714 may also have redundant functions. Interestingly, these four SNARE proteins were the only SNAREs identified in our tonoplast sample (Table 1, Figure 4). From our data, all four groups of SNARE proteins necessary for membrane fusion (Bock et al., 2001) are represented, and these findings are unprecedented among recently published data. Importantly, the SNARE proteins were entirely

absent in proteomics data presented by Shimaoka et al. (2004), and only two groups of SNARE proteins were identified by Szponarski et al. (2004). The identification of all four groups of SNARE proteins within our highly enriched tonoplast fraction suggests that we have identified a SNARE-pin complex responsible for vacuolar membrane fusion at the tonoplast (Figure 4). We also identified several proteins that likely interact with the proposed SNARE-pin complex, including GTP binding proteins AtRabG3b (AtRab75/At1g22740) and AtRabG3d (AtRab72/At1g52280), which are homologous to yeast GTPase Ypt7p, and two proteins (At4g04910 and At3g56190), which are homologous to yeast Sec18p/NSF and Sec17p/ α -SNAP. These yeast homologs are known to be involved in membrane fusion at the tonoplast (reviewed in Wickner, 2002). Here, we present experimental evidence that these proteins, currently annotated based on in silico analyses, are likely to be orthologous to their yeast counterparts.

Localization of VAMP71 together with the SYP22/VTI11/SYP5 complex to the tonoplast suggests its predicted role as an R-SNARE in a SNARE-pin formation and represents a unique application of a proteomics approach. Given the presence of redundancy in the longin gene family (Sanderfoot et al., 2000), identification of the fourth SNARE protein in the SNARE-pin complex using traditional biochemical or genetic methods would likely require analysis of plants carrying multiple mutations. Considering the embryo or gametophyte lethality of mutants in genes encoding SNAREs or SNARE-associated proteins (Sanderfoot et al., 2000; Rojo et al., 2001; Surpin et al., 2003), genetic approaches may not always elucidate SNARE protein function. Therefore, a proteomics approach proves to be an appropriate method to test these hypotheses.

Expressed/Hypothetical Proteins

Of the ~28,000 predicted protein-coding genes in Arabidopsis, ~65% have been classified into functional categories based on homology (Tian et al., 2004). Approximately 35% of all Arabidopsis proteins, including many plant-specific proteins, have no predicted function. Thus, the functions of a substantial proportion of the Arabidopsis genome remain elusive.

Of the 402 proteins we identified, 89 (22%) were annotated as either expressed or hypothetical proteins. Of these proteins, 46 (52%) were predicted to have a signal peptide and another 20 (22%) have one or more predicted transmembrane domains (see Supplemental Table 1 online). Thus, a large percentage of these proteins are likely directly associated with the endomembrane system. These proteins were also examined for conserved domains using BLAST searches (<http://www.ncbi.nlm.nih.gov:80/BLAST/>; Altschul et al., 1990) to provide clues to their activities. These analyses did not identify conserved domains associated with known activities; however, the identification of these proteins in the plant vacuole may aid future research in understanding their respective functions.

Functional Categorization

Much can be inferred into vacuole function and biogenesis based upon the proteins identified. All proteins were classified into

various categories based upon the annotations and Gene Ontology categorizations from TAIR (www.arabidopsis.org). We classified all proteins into the eight major categories shown in Figure 4. The identities of the proteins in six of these categories are also shown in Table 1 (others are shown in Supplemental Table 1 online). These categories include membrane remodeling and fusion/protein trafficking, transporters, biotic and abiotic stress response, protein degradation, glycosidases, cytoskeleton, miscellaneous, and expressed/hypothetical proteins. Many proteins identified in the individual groups represent new findings.

Membrane Remodeling and Fusion/Protein Trafficking

Aside from the proteins directly involved in membrane fusion discussed above (e.g., SNAREs and associated factors), a large number of enzymes with the capability to affect tonoplast physicochemical properties were discovered. The largest group of these proteins includes nine putative lipases. The products of phospholipase activity (e.g., phosphatidic acid and phosphoinositols) play key roles in signal transduction, cytoskeletal structure, the respiratory burst, and endomembrane trafficking in animals (Wang, 2004). Similar findings are occurring for plant species. Phosphatidic acid derived from phospholipase D activity is required for tobacco pollen tube growth and is suggested to play a role in the regulation of Golgi trafficking in the secretory pathway (Potocky et al., 2003). Furthermore, a putative phospholipase (SGR2) is required for shoot gravitropism in Arabidopsis (Kato et al., 2002a), a process that is highly membrane trafficking dependent (Surpin and Raikhel, 2004). It is hypothesized that SGR2 activity may alter membrane structure or produce signaling lipids (Kato et al., 2002a). Thus, besides serving degradative purposes in the vacuole, the lipases found within our samples may be involved in similar processes as those discussed above.

Protein Degradation

A major function of the plant vacuole is protein processing and degradation. The large central vacuoles of plants are often thought of as being analogous to the yeast vacuole or animal lysosome. In yeast, vacuolar proteases are synthesized as precursors, undergoing many posttranslational modifications before a final active protein is formed. This process is regulated by a tightly controlled proteolytic cascade (for review, see van Den Hazel et al., 1996). Many studies have indicated that a similar system occurs in plants. For example, T-DNA knockout mutants of the vacuolar processing enzyme family do not degrade and improperly process various proteins (Rojo et al., 2003a; Shimada et al., 2003; Gruis et al., 2004).

In this study, 23 potential proteases were identified (Table 1, Figure 4), of which only three were previously known to reside in the Arabidopsis vacuole (RD21 [At1g47128], aleurain [At5g60360], and carboxypeptidase Y [At3g10410]) (Ahmed et al., 2000; Yamada et al., 2001; Rojo et al., 2003a). These proteins represent Asp, Cys and Ser-type proteases. Only eight proteases from our data set were also reported by Shimaoka et al. (2004), including four Ser proteases (At2g35780; At3g10410, AtCPY; At4g20850; At4g36195), two Cys proteases (At1g02305;

At1g47128, RD21), and two Asp proteases (At1g11910 and At1g62290). The functions of these proteases in protein degradation in the vacuole can now be efficiently studied by reverse genetics approaches. Furthermore, the identification of these various proteases (and their potential T-DNA knockouts) now allows us to decipher the hierarchy of the putative proteolytic cascade in plant vacuoles.

Transporters

Another major function of the plant vacuole is maintenance of turgor pressure and storage of metabolites and ions. Forty-six different gene products involved in transport were identified (Table 1, Figure 4). The major transporter group consisted of subunits of the V-ATPase.

The central vacuole is maintained at a pH of ~ 5.5 (Sze et al., 2002). This acidic environment is produced through the activity of a V-ATPase and is required for providing energy for the transport of ions and organic metabolites. It has also been suggested that the proton gradient is required for protein sorting and membrane fusion events (Matsuoka et al., 1997; Wickner and Haas, 2000; Peters et al., 2001). The Arabidopsis V-ATPase contains 12 distinct functional subunits, encoded by 26 genes (Sze et al., 2002). Because of the extensive gene duplication and corresponding protein sequence identities (Sze et al., 2002), we analyzed our data in respect to the protein function as a distinct subunit rather than to the number of particular gene products. Recently, a proteomic examination of the vacuolar tonoplast identified seven functional subunits of the V-ATPase (Szponarski et al., 2004), and an independent study by Shimaoka et al. (2004) reported 11 functional subunits.

In this study, we identified all but a single subunit, VHA-e, of the functional plant V-ATPase (17 total gene products) (Sze et al., 2002). The nonidentified protein is predicted to be 8 kD in size, which may explain why it was not identified (e.g., may be lost at the concentrating step or run off the gel). This subunit was not identified in either this study or by Shimaoka et al. (2004). Overall, the identification of 11 of the 12 V-ATPase subunits indicates the depth of coverage of the vacuole proteome reported here.

Many other proteins involved in ion and metabolite transport across the tonoplast were identified. In total, we report 29 transporters or transport-related proteins, other than the subunits of a V-ATPase. Notable among these proteins are several calcium, sugar, and ABC transporters (Table 1). Remarkably, three of these proteins were also identified in both independent studies by Szponarski et al. (2004) and Shimaoka et al. (2004), including AVP1 subunit of V-type H⁺-PPase (At1g15690) and ABC transporters AtMRP1 and AtMRP4. Many identified proteins are members of large gene families, of which we have identified a subset that are presumably localized to the tonoplast.

Glycosidases

Glycosyl hydrolases are involved in many processing events in plants ranging from general metabolism to defense (Melchers et al., 1993; Roitsch et al., 2003). A total of 25 potential glycosyl hydrolase gene products were identified in the vegetative vacuole (Table 1, Figure 4). These glycosidases represent a wide

array of predicted activities, including β -fructofuranosidase, β -glucosidase, α -L-fucosidase, β -galactosidase, α -mannosidase, β -D-glucan exohydrolase, β -hexosaminidase, α -galactosidase/melibiose, glucan endo-1,3- β -glucosidase, polygalacturonase, β -xylosidase, and α -glucosidase. Only six of these proteins were reported by Shimaoka et al. (2004) and none by Szponarski et al. (2004). Thus, the protein data presented here form a successful attempt to comprehensively describe a subset of luminal vacuolar proteins.

Interestingly, three identified β -glucosidases contain ER retention signals at their respective C termini (At1g52400, REEL; At1g66270, RDEL; At3g09260, KDEL). An ER retention signal was recently suggested to be responsible for the vacuolar localization of one of these glucosidases (At3g09260) (Matsushima et al., 2003). Thus, we have identified two more examples of ER retention signals that may function in vacuolar targeting of proteins under certain conditions.

Biotic and Abiotic Stress Response

Another major function of plant vacuoles includes stress and defense responses. Defense-related proteins previously shown to localize to the vacuole from various species include proteinase inhibitors, chitinases, peroxidases, osmotin (*PR-5*), glycosidases, berberine bridge enzyme, strictosidine synthase, and lectins (Ryan, 1980; McKnight et al., 1990; Bednarek and Raikhel, 1991; Neuhaus et al., 1991; Melchers et al., 1993; Kelly et al., 1998; Bird and Facchini, 2001; Matsui et al., 2003).

We also found a large number of the above defense-related proteins to be localized to the central vacuole (Table 1, Figure 4). We identified six putative peroxidases (At1g71695, At3g32980, At3g49110, At3g49120, At4g08770, and At4g08780), three jacalin-like lectins (At1g52040, At2g39330, and At3g16470), two strictosidine synthase-like proteins (At1g74020 and At3g57020), *PR-5* (At1g75040), and a putative Cys protease inhibitor (At4g16500). Interestingly, three proteins from a prohibitin family (At1g03860, At3g27280, and At5g40770) were also identified by Shimaoka et al. (2004). Moreover, many of the glycosidases and proteases discussed above are also likely involved in direct defense or stress responses.

Cytoskeleton

All organelles share intimate contacts with the cytoskeleton (Buchanan et al., 2000). The direct involvement of the cytoskeleton and the vacuole in the gravitropic response has also been recently documented (Blancaflor, 2002; Kato et al., 2002b). A small number of proteins potentially involved in anchoring the vacuole to the cytoskeleton were identified. These include members of the actin (At1g49240), myosin (At1g52410, At1g65330, At3g58160, and At5g41790), microtubule-associated protein (At3g60840), and fimbrin (At5g55400) families. We also identified three members of the dynamin protein family (At1g59610, At3g60190, and At5g42080). Mammalian dynamins are known to be associated with actin (Schafer, 2004), suggesting their plant counterparts may be involved in cytoskeleton-related processes at the tonoplast. Interestingly, one of two possible dynamins, At1g59610 (also identified in Shimaoka

et al., 2004) or At1g10290 (the latter not identified here), has been suggested to play a role in the transport of cargo proteins from the *trans*-Golgi network to the vacuole (Jin et al., 2001).

Miscellaneous

At total of 111 (28%) proteins do not fall into the large, easily categorized groups discussed above (see Supplemental Table 1 online). A large number of these proteins are primarily associated with the ER, including 40S and 60S ribosomal proteins, luminal binding protein, calcineurin, calnexin, calreticulin, and calmodulin (Boyce et al., 1994). At first look, these proteins may appear to be contamination. However, the presence of these proteins in the vacuole can easily be explained. For example, dense vesicles coated with ribosomes (ER bodies/protease precursor vesicles) have been shown to bud directly from the rough ER and fuse directly with vacuole tonoplast (Hara-Nishimura et al., 1998; Matsushima et al., 2003; Rojo et al., 2003a). The finding of ribosomal proteins exclusively in the tonoplast fraction exemplifies this fact (see Supplemental Table 1 online). In contrast with Szponarski et al. (2004), where the tonoplast sample was derived from *Arabidopsis* suspension cells, we isolated vacuoles from mature *Arabidopsis* leaves. It is not surprising that the data set from the suspension cell-derived tonoplast sample did not contain many ribosomal proteins, as ER bodies are mainly associated with vegetative tissues (Hayashi et al., 2001). Furthermore, as discussed previously, ER retention signals can act as efficient vacuolar sorting signals and promote vacuolar targeting of proteins that escape the ER (Gomord et al., 1997). Consistent with this observation, the stress resulting from the generation of protoplasts may induce the overproduction of the ER-resident proteins (such as those listed above), potentially resulting in their vacuolar delivery and identification in our samples.

Conclusions

In this study, we purified vacuoles away from other cellular compartments and provide a thorough coverage of the *Arabidopsis* vegetative vacuolar proteome. This is an essential step for systems biology approaches in general (Girke et al., 2003) and in particular toward understanding the function of this dynamic organelle (Hicks et al., 2004). Although the vacuole has been extensively studied, the molecular bases for its many physiological roles remain largely unknown. These results present testable hypotheses for determining the molecular components of many processes, including tonoplast fusion, vacuole biogenesis, vacuolar sorting determinants, cytoskeletal attachments, and protein degradation. This study also sets a baseline that may be used to compare the protein profiles of the vacuoles from wild-type and mutant or specifically treated plants.

METHODS

Plant Material

Wild-type *Arabidopsis thaliana* ecotype Columbia was grown in soil at 19°C with a cycle of 12 h light/12 h dark.

Reagents and Chemicals

Reagents and chemicals were obtained from Fisher Scientific (Pittsburgh, PA) or from Sigma (St. Louis, MO) unless otherwise noted.

Vacuole Isolation

Vacuoles were purified from healthy green rosette leaves from 35-d-old plants. Briefly, 2 g of leaf tissue were cut into thin slices with a razor blade and incubated for 4 h at 21°C in protoplasting solution containing 0.3 g of macerozyme R-10 (0.55 units/mg) (Serva, Heidelberg, Germany), 0.3 g of cellulase Onozuka R-10 (1.0 units/mg) (Serva), and 0.12 g of CaCl₂ dissolved in 30 mL of wash buffer (0.4 M mannitol/10 mM Mes, pH 5.7). Protoplasts were recovered by filtering the protoplasts away from undissolved tissue followed by centrifugation for 20 min at 500 rpm (57g) in a Beckman GS-6R centrifuge (GH-3.8 rotor; Fullerton, CA). Protoplasts were washed two times in wash buffer and then lysed in a solution containing 0.2 M mannitol, 10% ficoll, 10 mM EDTA, and 5 mM NaPO₄, pH 8.0. Density ultracentrifugation was then performed as previously described (Ahmed et al., 2000) except that BSA was omitted from all solutions. This method results in the layering of enriched vacuoles and allows for the efficient separation of central vacuoles from other endomembrane compartments (Ahmed et al., 2000; Rojo et al., 2003a). Yields were approximately 50 μg of protein per preparation.

Proteomics of Purified Samples

Total Vacuole Proteome Analyses

To ensure thorough protein coverage, two separate methods were used for total vacuole proteome analyses. One methodology involved trypsin digestion of protein in liquid followed by multidimensional LC MS/MS. These samples were processed as previously described (Rojo et al., 2003a). For the second method, vacuolar preparations were concentrated with a Millipore 10,000 molecular weight cutoff centrifugal concentrator (Bedford, MA), and total protein concentration was determined (Bradford, 1976). Two hundred and fifty micrograms of total protein was subjected to standard 18-cm 14% SDS-PAGE (Laemmli, 1970). The resulting gel was briefly stained with Coomassie Brilliant Blue R 250 and cut into thin horizontal slices (60 total, ~2.5 mm width). Gel slices were then digested with trypsin and processed essentially as described previously (Wang et al., 2000).

The in-gel digest of individual gel bands was analyzed by a μLC-nano electron-spray ionization MS/MS using a Micromass Q-TOF API US instrument (Waters, Milford, MA). The Atlantis dC₁₈ column (3 μm, 100 Å, 100-μm internal diameter, 15 cm long) was used for LC peptide separation (Waters). A 300-μm internal diameter × 5-mm-long C₁₈ guard column (Dionex, Sunnyvale, CA) was placed at the front of the analytical column to preconcentrate the tryptic peptides through a 10-port switch valve connection. The μL-pickup sample loading method was configured with a 20-μL sample loop in the autosampler for maximal loading (6.4 μL) with zero sample loss. Samples were loaded and desalted for 3 min using 0.2% formic acid with a flow rate of 20 μL/min. The solvent components for peptide elution were as follows: mobile phase A was 0.1% formic acid and 5% acetonitrile, and mobile phase B was 0.1% formic acid and 95% acetonitrile. The duration of the complete LC method was for 140 min as follows: 0 to 3 min, 5% B; at 5 min, 12% B; at 95 min, 45% B; at 107 min, 70% B; at 110 to 115 min, 90% B; at 120 to 140 min, 5% B. A 1-h blank washing method was introduced between each sample. A manually configured flow splitter was used to apply 0.4-μL/min flow rate to the guard and analytical column, which was connected with a 10-μm internal diameter PicoTip nanospray emitter (New Objective, Woburn, MA). The MS instrument parameters were tuned to achieve the highest sensitivity and resolution as possible.

Tonoplast Proteome Analyses

For tonoplast isolation, vacuolar preparations were centrifuged at 100,000g for 1 h, and the supernatant was removed. Membrane fractions were then washed in 1 × PBS and recentrifuged three times. Protein from the resultant membrane preparation was quantified by the method of Lowry et al. (1951). Two hundred micrograms of protein was boiled in 2 × (final concentration) SDS-PAGE buffer and subjected to a standard 18-cm 14% SDS-PAGE (Laemmli, 1970). The gel was stained briefly in Coomassie Brilliant Blue R 250 and processed using the second method described above.

Bioinformatic Analyses

Protein Identification

For protein identification, the Internet-based MASCOT MS/MS Ions Search tool (http://www.matrixscience.com/cgi/search_form.pl?FORM-VER=2&SEARCH=MIS) was used to manually search against the National Center for Biotechnology Information (NCBI) nonredundant database. Search parameters were set as follows: taxonomy, Arabidopsis; enzyme, trypsin; variable modifications, acetyl (N-term), oxidation (M), phospho (ST), phospho (Y), pyro-glu (N-term E), pyro-glu (N-term Q); mass values, monoisotopic; protein mass, unrestricted peptide; mass tolerance, ± 150 ppm; fragment mass tolerance, ± 0.2 D; max missed cleavages, 1. The parameters used for positive protein identification were essentially identical to those previously described (Mootha et al., 2003). Proteins with a minimum of one peptide match (with a significant ion score of ≥25 according to MASCOT) and containing a sequence tag with at least three amino acids in a row in either Y- or B-ion series, but not matching to any other proteins, were considered as positive and listed in the final summary. All proteins with only a single peptide match were verified by manual inspection of Y- and B-ion series to identify the minimal sequence tag of three consecutive ions (see Supplemental Table 2 online). Also, eight proteins with a single peptide match but without a three-amino acid sequence tag were also included after individual investigation. Each of these proteins had multiple numbers of two amino acid sequence tags, significant MASCOT scores (≥25), and no secondary hits (see Supplemental Table 2 online). For matches representing protein families, only one member is presented. All proteins identified through the NCBI database (and corresponding annotations) were verified by BLAST searches against the TAIR database (<http://arabidopsis.org/Blast/>).

Protein Characterization

All identified proteins were run through a series of prediction programs to provide theoretical subcellular localizations, molecular masses, transmembrane domains, and isoelectric points. Subcellular localizations were predicted by three programs, including iPSORT (<http://hypothesiscreator.net/iPSORT/>), PREDOTAR 1.03 (<http://genoplante-info.infobiogen.fr/predotar/predotar.html>; includes signal peptide prediction), and TargetP (<http://www.cbs.dtu.dk/services/TargetP/>) (Nielsen et al., 1997; Emanuelsson et al., 2000; Bannai et al., 2002). Transmembrane domains were predicted by the three programs HMMTOP (<http://www.enzim.hu/hmmtop/>), TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>), and THUMBUP (http://theory.med.buffalo.edu/Softwares-Services_files/thumbup.htm) (Tusnady and Simon, 1998; Krogh et al., 2001; Zhou and Zhou, 2003). Protein masses and isoelectric points were determined with the TAIR bulk protein search tool (<http://www.arabidopsis.org/tools/bulk/protein/index.jsp>).

Pattern Searching

Physicochemical properties, hydrophobicity profiles, protein masses, and isoelectric points were determined with the bulk protein search tool from TAIR (<http://www.arabidopsis.org/tools/bulk/protein/index.jsp>) and

the EMBOSS package (<http://www.hgmp.mrc.ac.uk/Software/EMBOSS/Apps/>). Motif finding of known patterns was performed with regular expressions in Perl (<http://www.perl.org/>). For pattern discovery of unknown motifs, we used Gibbs Motif Sampler (http://bayesweb.wadsworth.org/cgi-bin/gibbs.8.pl?data_type=protein), MEME (<http://meme.sdsc.edu/meme/website/meme.html>), and Block Maker (http://blocks.fhcrc.org/blocks/make_blocks.html).

The data set of 142 proteins (see supplemental data online) used for NTPP prediction was derived from proteins predicted to contain a signal peptide and two or fewer transmembrane domains. Note that transmembrane domain (TMD) prediction programs usually identify signal peptides as TMDs, thus accounting for one of the allowed TMDs in this protein set. Furthermore, to address the possibility of erroneous TMD predictions, we allowed a second TMD in the proteins included in this list. This protein set was scanned with regular expression searches for the presence of two related motifs. Searches were performed with pattern 1, [N/L]-[P/I/L]-[I/P], and pattern 2, [N/L]-[P/I/L]-[I/P]-[R/N/S]. The pattern discovery programs Gibbs Motif Sampler, MEME, and Block Maker were used for identifying potentially new targeting motifs in the same protein set. Overall sequence similarity between sequences and biased amino compositions can generate false positive patterns using these programs. Therefore, related sequences with >50% sequence identity were removed with the BLASTCLUST software, and low-complexity regions were masked with the CAST program. The removal of related sequences resulted in a subset of 110 protein sequences that were used for motif discovery. More detailed results can be viewed at <http://bioinfo.ucr.edu/projects/VacuoleProteomics/Overview.html>.

ACKNOWLEDGMENTS

The authors would like to thank Glenn Hicks for critical readings of the manuscript. This work was supported by the Department of Energy (Grant DE-FG03-02ER15295/A000 to N.V.R.), by the National Science Foundation (Graduate Research Fellowship to E.L.A.), and by the National Institutes of Health (National Research Service Award Postdoctoral Fellowship to C.C.).

Received August 19, 2004; accepted September 3, 2004.

REFERENCES

- Ahmed, S.U., Rojo, E., Kovaleva, V., Venkataraman, S., Dombrowski, J.E., Matsuoka, K., and Raikhel, N.V. (2000). The plant vacuolar sorting receptor AtELP is involved in transport of NH₂-terminal propeptide-containing vacuolar proteins in *Arabidopsis thaliana*. *J. Cell Biol.* **149**, 1335–1344.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
- Balmer, Y., Koller, A., del Val, G., Manieri, W., Schurmann, P., and Buchanan, B.B. (2003). Proteomics gives insight into the regulatory function of chloroplast thioredoxins. *Proc. Natl. Acad. Sci. USA* **100**, 370–375.
- Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., and Miyano, S. (2002). Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* **18**, 298–305.
- Bednarek, S.Y., and Raikhel, N.V. (1991). The barley lectin carboxyl-terminal propeptide is a vacuolar protein sorting determinant in plants. *Plant Cell* **3**, 1195–1206.
- Bird, D.A., and Facchini, P.J. (2001). Berberine bridge enzyme, a key branch-point enzyme in benzyloquinoline alkaloid biosynthesis, contains a vacuolar sorting determinant. *Planta* **213**, 888–897.
- Blancaflor, E.B. (2002). The cytoskeleton and gravitropism in higher plants. *J. Plant Growth Regul.* **21**, 120–136.
- Bock, J.B., Matern, H.T., Peden, A.A., and Scheller, R.H. (2001). A genomic perspective on membrane compartment organization. *Nature* **409**, 839–841.
- Boyce, J.M., Coates, D., Fricker, M.D., and Evans, D.E. (1994). Genomic sequence of a calnexin homolog from *Arabidopsis thaliana*. *Plant Physiol.* **106**, 1691.
- Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254.
- Buchanan, B.B., Gruissem, W., and Jones, R.L. (2000). *Biochemistry and Molecular Biology of Plants*. (Rockville, MD: American Society of Plant Physiologists).
- Cánovas, F.M., Dumas-Gaudot, E., Recorbet, G., Jorin, J., Mock, H.P., and Rossignol, M. (2004). Plant proteome analysis. *Proteomics* **4**, 285–298.
- De, D.N. (2000). *Plant Cell Vacuoles*. (Collingwood, Australia: CSIRO Publishing).
- Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016.
- Frigerio, L., Jolliffe, N.A., Di Cola, A., Felipe, D.H., Paris, N., Neuhaus, J.M., Lord, J.M., Ceriotti, A., and Roberts, L.M. (2001). The internal propeptide of the ricin precursor carries a sequence-specific determinant for vacuolar sorting. *Plant Physiol.* **126**, 167–175.
- Friso, G., Giacomelli, L., Ytterberg, A.J., Peltier, J.B., Rudella, A., Sun, Q., and Wijk, K.J. (2004). In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: New proteins, new functions, and a plastid proteome database. *Plant Cell* **16**, 478–499.
- Gagnon, E., Duclos, S., Rondeau, C., Chevet, E., Cameron, P.H., Steele-Mortimer, O., Paiement, J., Bergeron, J.J., and Desjardins, M. (2002). Endoplasmic reticulum-mediated phagocytosis is a mechanism of entry into macrophages. *Cell* **110**, 119–131.
- Garin, J., Diez, R., Kieffer, S., Dermine, J.F., Duclos, S., Gagnon, E., Sadoul, R., Rondeau, C., and Desjardins, M. (2001). The phagosome proteome: Insight into phagosome functions. *J. Cell Biol.* **152**, 165–180.
- Girke, T., Ozkan, M., Carter, D., and Raikhel, N.V. (2003). Towards a modeling infrastructure for studying plant cells. *Plant Physiol.* **132**, 410–414.
- Gomord, V., Denmat, L.A., Fitchette-Laine, A.C., Satiat-Jeunemaitre, B., Hawes, C., and Faye, L. (1997). The C-terminal HDEL sequence is sufficient for retention of secretory proteins in the endoplasmic reticulum (ER) but promotes vacuolar targeting of proteins that escape the ER. *Plant J.* **11**, 313–325.
- Gruis, D., Schulze, J., and Jung, R. (2004). Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases. *Plant Cell* **16**, 270–290.
- Gu, S., Chen, J., Dobos, K.M., Bradbury, E.M., Belisle, J.T., and Chen, X. (2003). Comprehensive proteomic profiling of the membrane constituents of a *Mycobacterium tuberculosis* strain. *Mol. Cell Proteomics* **2**, 1284–1296.
- Hara-Nishimura, I.I., Shimada, T., Hatano, K., Takeuchi, Y., and Nishimura, M. (1998). Transport of storage proteins to protein storage vacuoles is mediated by large precursor-accumulating vesicles. *Plant Cell* **10**, 825–836.
- Hayashi, Y., Yamada, K., Shimada, T., Matsushima, R., Nishizawa, N.K., Nishimura, M., and Hara-Nishimura, I. (2001). A proteinase-storing body that prepares for cell death or stresses in the epidermal cells of *Arabidopsis*. *Plant Cell Physiol.* **42**, 894–899.

- Heazlewood, J.L., Tonti-Filippini, J.S., Gout, A.M., Day, D.A., Whelan, J., and Millar, A.H. (2004). Experimental analysis of the Arabidopsis mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell* **16**, 241–256.
- Hicks, G.R., Rojo, E., Hong, S., Carter, D.G., and Raikhel, N.V. (2004). Germinating pollen has tubular vacuoles, displays highly dynamic vacuole biogenesis, and requires VACUOLESS1 for proper function. *Plant Physiol.* **134**, 1227–1239.
- Holwerda, B.C., Padgett, H.S., and Rogers, J.C. (1992). Proaleurain vacuolar targeting is mediated by short contiguous peptide interactions. *Plant Cell* **4**, 307–318.
- Jiang, L., and Rogers, J.C. (1998). Integral membrane protein sorting to vacuoles in plant cells: Evidence for two pathways. *J. Cell Biol.* **143**, 1183–1199.
- Jin, J.B., Kim, Y.A., Kim, S.J., Lee, S.H., Kim, D.H., Cheong, G.W., and Hwang, I. (2001). A new dynamin-like protein, ADL6, is involved in trafficking from the trans-Golgi network to the central vacuole in Arabidopsis. *Plant Cell* **13**, 1511–1526.
- Kato, T., Morita, M.T., Fukaki, H., Yamauchi, Y., Uehara, M., Niihama, M., and Tasaka, M. (2002a). SGR2, a phospholipase-like protein, and ZIG/SGR4, a SNARE, are involved in the shoot gravitropism of Arabidopsis. *Plant Cell* **14**, 33–46.
- Kato, T., Morita, M.T., and Tasaka, M. (2002b). Role of endodermal cell vacuoles in shoot gravitropism. *J. Plant Growth Regul.* **21**, 113–119.
- Kelly, P.J., Bones, A., and Rossiter, J.T. (1998). Sub-cellular immunolocalization of the glucosinolate sinigrin in seedlings of *Brassica juncea*. *Planta* **206**, 370–377.
- Koide, Y., Matsuoka, K., Ohto, M., and Nakamura, K. (1999). The N-terminal propeptide and the C terminus of the precursor to 20-kilo-dalton potato tuber protein can function as different types of vacuolar sorting signals. *Plant Cell Physiol.* **40**, 1152–1159.
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L.L. (2001). Predicting transmembrane protein topology with a Hidden Markov Model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
- Lowry, O.H., Rosebrough, N.J., Farr, A.L., and Randall, R.J. (1951). Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**, 265–275.
- Matsui, T., Nakayama, H., Yoshida, K., and Shinmyo, A. (2003). Vesicular transport route of horseradish C1a peroxidase is regulated by N- and C-terminal propeptides in tobacco cells. *Appl. Microbiol. Biotechnol.* **262**, 517–522.
- Matsuoka, K., Higuchi, T., Maeshima, M., and Nakamura, K. (1997). A vacuolar-type H⁺-ATPase in a nonvacuolar organelle is required for sorting of soluble vacuolar protein precursors in tobacco cells. *Plant Cell* **9**, 533–546.
- Matsuoka, K., and Nakamura, K. (1991). Propeptide of a precursor to a plant vacuolar protein required for vacuolar targeting. *Proc. Natl. Acad. Sci. USA* **88**, 834–838.
- Matsuoka, K., and Neuhaus, J.-M. (1999). *Cis*-elements of protein transport to the plant vacuoles. *J. Exp. Bot.* **50**, 165–174.
- Matsushima, R., Kondo, M., Nishimura, M., and Hara-Nishimura, I. (2003). A novel ER-derived compartment, the ER body, selectively accumulates a beta-glucosidase with an ER-retention signal in Arabidopsis. *Plant J.* **33**, 493–502.
- McKnight, T.D., Roessner, C.A., Devagupta, R., Scott, A.I., and Nessler, C.L. (1990). Nucleotide sequence of a cDNA encoding the vacuolar protein strictosidine synthase from *Catharanthus roseus*. *Nucleic Acids Res.* **18**, 4939.
- Melchers, L.S., Sela-Buurlage, M.B., Vloemans, S.A., Woloshuk, C.P., Van Roekel, J.S., Pen, J., van den Elzen, P.J., and Cornelissen, B.J. (1993). Extracellular targeting of the vacuolar tobacco proteins AP24, chitinase and beta 1,3-glucanase in transgenic plants. *Plant Mol. Biol.* **21**, 583–593.
- Mootha, V.K., et al. (2003). Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* **115**, 629–640.
- Morita, M.T., Kato, T., Nagafusa, K., Saito, C., Ueda, T., Nakano, A., and Tasaka, M. (2002). Involvement of the vacuoles of the endodermis in the early process of shoot gravitropism in Arabidopsis. *Plant Cell* **14**, 47–56.
- Neuhaus, J.M., Sticher, L., Meins, F., Jr., and Boller, T. (1991). A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole. *Proc. Natl. Acad. Sci. USA* **88**, 10362–10366.
- Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6.
- Peltier, J.B., Friso, G., Kalume, D.E., Roepstorff, P., Nilsson, F., Adamska, I., and van Wijk, K.J. (2000). Proteomics of the chloroplast: Systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. *Plant Cell* **12**, 319–341.
- Peters, C., Bayer, M.J., Bühler, S., Andersen, J.S., Mann, M., and Mayer, A. (2001). Trans-complex formation by proteolipid channels in the terminal phase of membrane fusion. *Nature* **409**, 581–588.
- Potocky, M., Elias, M., Profotova, B., Novotna, Z., Valentova, O., and Zarsky, V. (2003). Phosphatidic acid produced by phospholipase D is required for tobacco pollen tube growth. *Planta* **217**, 122–130.
- Roitsch, T., Balibrea, M.E., Hofmann, M., Proels, R., and Sinha, A.K. (2003). Extracellular invertase: Key metabolic enzyme and PR protein. *J. Exp. Bot.* **54**, 513–524.
- Rojo, E., Gillmor, C.S., Kovaleva, V., Somerville, C.R., and Raikhel, N.V. (2001). VACUOLELESS1 is an essential gene required for vacuole formation and morphogenesis in Arabidopsis. *Dev. Cell* **1**, 303–310.
- Rojo, E., Sharma, V.K., Kovaleva, V., Raikhel, N.V., and Fletcher, J.C. (2002). CLV3 is localized to the extracellular space, where it activates the Arabidopsis CLAVATA stem cell signaling pathway. *Plant Cell* **14**, 969–977.
- Rojo, E., Zouhar, J., Carter, C., Kovaleva, V., and Raikhel, N.V. (2003a). A unique mechanism for protein processing and degradation in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **100**, 7389–7394.
- Rojo, E., Zouhar, J., Kovaleva, V., Hong, S., and Raikhel, N.V. (2003b). The AtC-VPS protein complex is localized to the tonoplast and the prevacuolar compartment in Arabidopsis. *Mol. Biol. Cell.* **14**, 361–369.
- Ryan, C.A. (1980). Wound-regulated synthesis and vacuolar compartmentation of proteinase inhibitors in plant leaves. *Curr. Top. Cell. Regul.* **17**, 1–23.
- Sanderfoot, A.A., Assaad, F.F., and Raikhel, N.V. (2000). The Arabidopsis genome. An abundance of soluble N-ethylmaleimide-sensitive factor adaptor protein receptors. *Plant Physiol.* **124**, 1558–1569.
- Sanderfoot, A.A., Kovaleva, V., Bassham, D.C., and Raikhel, N.V. (2001). Interactions between syntaxins identify at least five SNARE complexes within the Golgi/prevacuolar system of the Arabidopsis cell. *Mol. Biol. Cell* **12**, 3733–3743.
- Sanderfoot, A.A., Kovaleva, V., Zheng, H., and Raikhel, N.V. (1999). The t-SNARE AtVAM3p resides on the prevacuolar compartment in Arabidopsis root cells. *Plant Physiol.* **121**, 929–938.
- Sato, M.H., Nakamura, N., Ohsumi, Y., Kouchi, H., Kondo, M., Hara-Nishimura, I., Nishimura, M., and Wada, Y. (1997). The AtVAM3 encodes a syntaxin-related molecule implicated in the vacuolar assembly in *Arabidopsis thaliana*. *J. Biol. Chem.* **272**, 24530–24535.

- Schafer, D.A.** (2004). Regulating actin dynamics at membranes: A focus on dynamin. *Traffic* **5**, 463–469.
- Schubert, M., Petersson, U.A., Haas, B.J., Funk, C., Schroder, W.P., and Kieselbach, T.** (2002). Proteome map of the chloroplast lumen of *Arabidopsis thaliana*. *J. Biol. Chem.* **277**, 8354–8365.
- Shimada, T., et al.** (2003). Vacuolar processing enzymes are essential for proper processing of seed storage proteins in *Arabidopsis thaliana*. *J. Biol. Chem.* **278**, 32292–32299.
- Shimaoka, T., Ohnishi, M., Sazuka, T., Mitsuhashi, N., Hara-Nishimura, I., Shimazaki, K.I., Maeshima, M., Yokota, A., Tomizawa, K.I., and Mimura, T.** (2004). Isolation of intact vacuoles and proteomic analysis of tonoplast from suspension-cultured cells of *Arabidopsis thaliana*. *Plant Cell Physiol.* **45**, 672–683.
- Singh-Zocchi, M., Andreasen, A., and Zocchi, G.** (1999). Osmotic pressure contribution of albumin to colloidal interactions. *Proc. Natl. Acad. Sci. USA* **96**, 6711–6715.
- Surpin, M., and Raikhel, N.** (2004). Traffic jams affect plant development and signal transduction. *Nat. Rev. Mol. Cell Biol.* **5**, 100–109.
- Surpin, M., Zheng, H., Morita, M.T., Saito, C., Avila, E., Blakeslee, J.J., Bandyopadhyay, A., Kovaleva, V., Carter, D., Murphy, A., Tasaka, M., and Raikhel, N.** (2003). The VTI family of SNARE proteins is necessary for plant viability and mediates different protein transport pathways. *Plant Cell* **15**, 2885–2899.
- Sze, H., Schumacher, K., Muller, M.L., Padmanaban, S., and Taiz, L.** (2002). A simple nomenclature for a complex proton pump: VHA genes encode the vacuolar H(+)-ATPase. *Trends Plant Sci.* **7**, 157–161.
- Szponarski, W., Sommerer, N., Boyer, J.C., Rossignol, M., and Gibrat, R.** (2004). Large-scale characterization of integral proteins from *Arabidopsis* vacuolar membrane by two-dimensional liquid chromatography. *Proteomics* **4**, 397–406.
- Tian, G.W., et al.** (2004). High-throughput fluorescent tagging of full-length *Arabidopsis* gene products in planta. *Plant Physiol.* **35**, 25–38.
- Tusnády, G.E., and Simon, I.** (1998). Principles governing amino acid composition of integral membrane proteins: Applications to topology prediction. *J. Mol. Biol.* **283**, 489–506.
- van Den Hazel, H.B., Kielland-Brandt, M.C., and Winther, J.R.** (1996). Review: Biosynthesis and function of yeast vacuolar proteases. *Yeast* **12**, 1–16.
- Vitale, A., and Raikhel, N.V.** (1999). What do proteins need to reach different vacuoles? *Trends Plant Sci.* **4**, 149–155.
- Wang, X.** (2004). Lipid signaling. *Curr. Opin. Plant Biol.* **7**, 329–336.
- Wang, Y., Sun, J., and Chitnis, P.** (2000). Proteomic study of the peripheral proteins from thylakoid membranes of the cyanobacterium *Synechocystis* sp. PCC6803. *Electrophoresis* **21**, 1746–1754.
- Whitelegge, J.P.** (2002). Plant proteomics: BLASTing out of a MudPIT. *Proc. Natl. Acad. Sci. USA* **99**, 11564–11566.
- Wickner, W.** (2002). Yeast vacuoles and membrane fusion pathways. *EMBO J* **21**, 1241–1247.
- Wickner, W., and Haas, A.** (2000). Yeast homotypic vacuole fusion: A window on organelle trafficking mechanisms. *Annu. Rev. Biochem.* **69**, 247–275.
- Yamada, K., Matsushima, R., Nishimura, M., and Hara-Nishimura, I.** (2001). A slow maturation of a cysteine protease with a granulin domain in the vacuoles of senescing *Arabidopsis* leaves. *Plant Physiol.* **127**, 1626–1634.
- Yamaguchi, K., von Knoblauch, K., and Subramanian, A.R.** (2000). The plastid ribosomal proteins. Identification of all the proteins in the 30 S subunit of an organelle ribosome (chloroplast). *J. Biol. Chem.* **275**, 28455–28465.
- Zhou, H., and Zhou, Y.** (2003). Predicting the topology of transmembrane helical proteins using mean burial propensity and a hidden-Markov-model based method. *Protein Sci.* **12**, 1547–1555.

The Vegetative Vacuole Proteome of *Arabidopsis thaliana* Reveals Predicted and Unexpected Proteins

Clay Carter, Songqin Pan, Jan Zouhar, Emily L. Avila, Thomas Girke and Natasha V. Raikhel
Plant Cell 2004;16;3285-3303; originally published online November 11, 2004;
DOI 10.1105/tpc.104.027078

This information is current as of April 24, 2019

Supplemental Data	/content/suppl/2004/11/04/tpc.104.027078.DC1.html
References	This article cites 79 articles, 37 of which can be accessed free at: /content/16/12/3285.full.html#ref-list-1
Permissions	https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&cissn=1532298X&WT.mc_id=pd_hw1532298X
eTOCs	Sign up for eTOCs at: http://www.plantcell.org/cgi/alerts/ctmain
CiteTrack Alerts	Sign up for CiteTrack Alerts at: http://www.plantcell.org/cgi/alerts/ctmain
Subscription Information	Subscription Information for <i>The Plant Cell</i> and <i>Plant Physiology</i> is available at: http://www.aspb.org/publications/subscriptions.cfm