

Mapping the *Arabidopsis* organelle proteome

Tom P. J. Dunkley*, Svenja Hester*, Ian P. Shadforth†, John Runions‡, Thilo Weimar*, Sally L. Hanton§, Julian L. Griffin*, Conrad Bessant†, Federica Brandizzi§, Chris Hawes‡, Rod B. Watson¶, Paul Dupree*, and Kathryn S. Lilley*||

*Department of Biochemistry, University of Cambridge, Downing Site, Cambridge CB2 1QW, United Kingdom; †Department of Analytical Science and Informatics, Cranfield University, Silsoe MK45 4DT, United Kingdom; ‡Research School of Biological and Molecular Sciences, Oxford Brookes University, Oxford OX3 0BP, United Kingdom; §Department of Biology, University of Saskatchewan, 112 Science Place, Saskatoon, SK, Canada S7N 5E2; and ¶Applied Biosystems, Lingley House, 120 Birchwood Boulevard, Warrington WA3 7QH, United Kingdom

Edited by Randy Schekman, University of California, Berkeley, CA, and approved February 27, 2006 (received for review August 11, 2005)

A challenging task in the study of the secretory pathway is the identification and localization of new proteins to increase our understanding of the functions of different organelles. Previous proteomic studies of the endomembrane system have been hindered by contaminating proteins, making it impossible to assign proteins to organelles. Here we have used the localization of organelle proteins by the isotope tagging technique in conjunction with isotope tags for relative and absolute quantitation and 2D liquid chromatography for the simultaneous assignment of proteins to multiple subcellular compartments. With this approach, the density gradient distributions of 689 proteins from *Arabidopsis thaliana* were determined, enabling confident and simultaneous localization of 527 proteins to the endoplasmic reticulum, Golgi apparatus, vacuolar membrane, plasma membrane, or mitochondria and plastids. This parallel analysis of endomembrane components has enabled protein steady-state distributions to be determined. Consequently, genuine organelle residents have been distinguished from contaminating proteins and proteins in transit through the secretory pathway.

endomembrane | localization of organelle proteins by isotope tagging | isotope tags for relative and absolute quantitation | organelle proteomics

Proteins are spatially organized according to their functions within the eukaryotic cell. Therefore, protein localization is an important step toward assigning functions to the thousands of uncharacterized proteins predicted by the genome-sequencing projects. Proteomics provides powerful tools for characterizing the protein contents of organelles. Confident protein localization, however, requires that either organelle preparations are free of contaminants or that techniques are used to discriminate between genuine organelle residents and contaminating proteins (1). Although reasonably pure preparations of some organelles, such as mitochondria, can be achieved, the components of the endomembrane system so far have proved recalcitrant to purification (2, 3). The constituent organelles of the endomembrane system have similar sizes and densities, making them difficult to separate. In addition, the proteins that reside within this system are in a constant state of flux. Endomembrane proteins traffic through the system en route to their final destination; for example, plasma membrane (PM) proteins travel although the endoplasmic reticulum (ER) and the Golgi apparatus before reaching the cell surface. Proteins within the endomembrane system also cycle between compartments; for example, ER residents continuously escape to the Golgi apparatus and are retrieved in COPI vesicles (4). Consequently, it is not sufficient merely to identify the proteins within a single organelle-enriched fraction. Instead, the steady-state distributions of proteins within the whole endomembrane system must be determined if a realistic insight into the subcellular localization of endomembrane proteins is to be achieved.

Localization of organelle proteins by isotope tagging (LOPIT) is a proteomic technique for protein localization that does not depend on the preparation of pure organelles (5). Organelles are partially separated first by using centrifugation through density gradients. Distributions of proteins within such gradients can be assessed then by measuring the relative abundance of proteins between fractions

along the length of the gradients. These distribution patterns can be visualized by using differential isotope tagging of proteins in different density gradient fractions in conjunction with mass spectrometry. The subcellular localization of proteins with hitherto unknown locations then can be determined by comparing their distributions to those of previously localized proteins, because proteins that belong to the same organelle will cofractionate in the density gradients. The applicability of LOPIT for discriminating between ER and Golgi apparatus-localized proteins already has been demonstrated (5). Here we have further developed the LOPIT technique by using isotope tags for relative and absolute quantitation (iTRAQ) in conjunction with 2D liquid chromatography. This study has resulted in the successful determination of 689 protein density gradient distributions, enabling the simultaneous assignment of 527 proteins to the ER, Golgi apparatus, vacuolar membrane, PM, or to the mitochondria and plastids in *Arabidopsis*. These results demonstrate that proteomic analysis of the major endomembrane components can be performed in parallel, enabling protein steady-state distributions between these organelles to be determined. Consequently, genuine residents of the ER, Golgi apparatus, vacuolar membrane, and PM have been distinguished from contaminants and proteins that are in transit through the secretory pathway by proteomics.

Results

Protein Identification and Quantitation. Protein localization by LOPIT is based on the principle that proteins belonging to the same organelle cofractionate after equilibrium density gradient centrifugation of subcellular membranes (5). Here, we have increased the resolution and protein coverage of the LOPIT technique by quantifying distributions of protein within a gradient by using iTRAQ labeling in conjunction with 2D liquid chromatography of peptides followed by tandem MS. iTRAQ enables relative protein levels in four samples to be determined in a single MS experiment. In this case, two overlapping four-plex iTRAQ comparisons were carried out across the region on iodixanol density gradients where the ER, Golgi apparatus, vacuolar membrane, and PM were enriched (determined by Western blotting; data not shown; also see Fig. 4, which is published as supporting information on the PNAS web site, for a scheme of the experimental design). The iTRAQ reagent has an *N*-hydroxysuccinimide ester group, which reacts with peptide *N* termini and the primary amines of the C-terminal lysines of tryptic peptides (6). The iTRAQ reagent also contains a balance (carbonyl) group and a pendant reporter group (*N*-methylpiperazine). The distribution of ¹³C, ¹⁵N, and ¹⁸O between the reporter group and balance group differs in the four versions of the iTRAQ reagent such that the reporter group masses differ successively by 1 Da.

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: ER, endoplasmic reticulum; GPI, glycosylphosphatidylinositol; iTRAQ, isotope tags for relative and absolute quantitation; LOPIT, localization of organelle proteins by isotope tagging; PCA, principle components analysis; PLS-DA, partial least squares discriminant analysis; PM, plasma membrane.

||To whom correspondence should be addressed. E-mail: k.s.lilley@bioc.cam.ac.uk.

© 2006 by The National Academy of Sciences of the USA

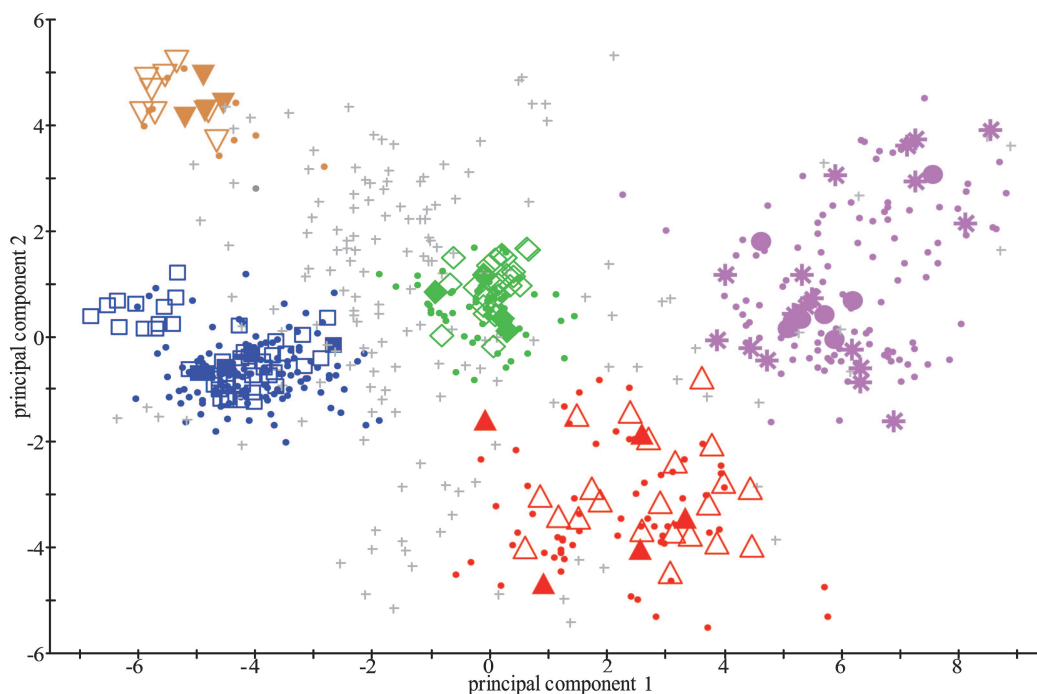


Fig. 1. PCA scores plot showing clustering of proteins according to their density gradient distributions and, therefore, localizations. iTRAQ reporter ion ratios were imported into SIMCA 10, logged, and preprocessed with unit-variance scaling before performing PCA analysis. Filled shapes indicate known organelle residents (marker proteins). Open shapes (or stars in the case of mitochondria/plastid) indicate proteins with predicted localizations that were confirmed based on their proximity to the corresponding marker proteins on the PCA scores plot. Small dots indicate proteins, without known or predicted localizations, that were assigned to an organelle by using PLS-DA, in conjunction with limited manual analysis. Small crosses indicate proteins that were not assigned to an organelle. Inverted triangles, vacuolar membrane; squares, ER; diamonds, PM; circles, known mitochondria/plastids; stars, predicted mitochondria/plastids; triangles, Golgi apparatus.

When an iTRAQ-labeled peptide is subjected to collision-induced dissociation, the iTRAQ tag fragments between the reporter group, balance group, and the peptide. The balance group is uncharged and, therefore, is not present in product ion spectra. The reporter group, however, retains a proton and the resulting reporter ions derived from the four versions of the iTRAQ reagent appear at 114.1 m/z , 115.1 m/z , 116.1 m/z , and 117.1 m/z in each product ion spectrum. Peptide quantitation is achieved by comparing the peak areas of these reporter ions. In this study, the reporter ion areas indicate relative protein levels in the four iTRAQ-labeled density gradient fractions (Fig. 5 *A–F*, which is published as supporting information on the PNAS web site).

To test the reproducibility of the LOPIT approach and to improve the accuracy of protein localization, two independent density gradient separations of *Arabidopsis* membranes were performed and two iTRAQ comparisons were performed on each. Two-dimensional liquid chromatography of peptides followed by tandem MS analysis of the iTRAQ-labeled samples resulted in the identification of 954 proteins in the first experiment (each with a minimum of two distinct, significant peptides) and 994 in the second experiment. The 689 proteins identified as being common to both experiments were used in subsequent analyses. Each protein was quantified in all four iTRAQ-labeling experiments. Quantitation of peptides common to two or more homologous proteins could lead to misleading localization results, particularly if the proteins reside in different organelles. Only peptides unique to a single identified protein therefore were quantified. For proteins where two or more peptides were quantified, the percentage standard deviation (PSD) of the iTRAQ ratios was calculated as a measure of technical error. The average PSD for the first experiment was 16.0% and for the second experiment 17.3%.

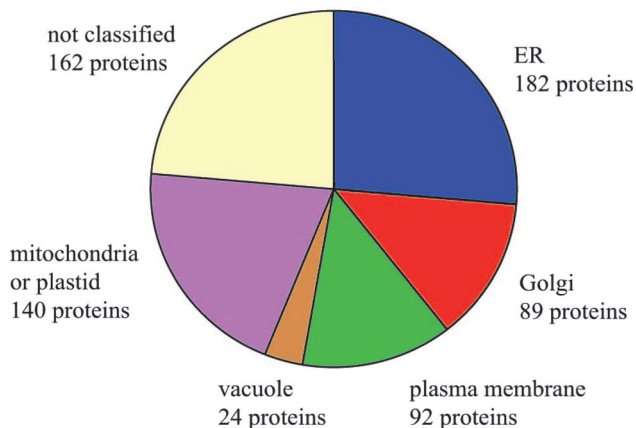
Protein Localization. The relative iTRAQ reporter ion intensities indicate distributions of individual proteins in the density gradient.

Reporter ion spectra for known vacuolar membrane, ER, PM, mitochondrial, and Golgi apparatus proteins are shown in Fig. 5 *A–E*. The profiles of these marker proteins are clearly distinct, demonstrating that residents of these organelles can be distinguished despite the fact that pure organelle preparations were not obtained. Localization of previously undescribed proteins is achieved by matching their profiles to the marker profiles. For example, the uncharacterized protein At1g31850 (GMT1; Fig. 5*F*) exhibits a very similar profile to the Golgi apparatus marker gtl6 (Fig. 5*E*) and therefore is most likely localized within the Golgi apparatus.

For each of the 689 identified proteins, six reporter ion ratios were derived from each of the two four-plex comparisons from each of the two experiments, yielding 24 data points for each protein (reporter ion ratios and normalized reporter ion areas in Table 2 and Table 3, respectively, which are published as supporting information on the PNAS web site). Clearly, manual analysis of such a data set is not possible; therefore, principle components analysis (PCA) was used as a data reduction tool and to identify patterns within it. Initially, PCA was performed on the two repeat experiments independently to confirm that protein clustering was reproducible (Fig. 6, which is published as supporting information on the PNAS web site). The data from the two experiments were then combined and PCA performed on the entire data set, resulting in the reduction of the combined data set to a two-variable scores plot (Fig. 1). The principal component scores for each protein describe the position of that protein in relation to these two new axes or principal components and the other proteins contributing to the model. The PCA scores plot represents a map of all 689 proteins, within which proteins with similar density gradient distributions and, therefore, localizations, are clustered. Annotation of 27 proteins with known localizations in *Arabidopsis* (organelle marker proteins) revealed the presence of five main clusters, corresponding

Table 1. Summary of localization results

Organelle	No. of proteins			LOPIT classification
	Known loc.	Predicted loc.	Predicted loc. confirmed by LOPIT	
ER	6	43	43	182
Golgi	5	23	22	89
PM	4	28	24	92
Vacuole	4	9	8	24
Mit./plastid	8	18	18	140
Unknown	662	541		162



"Known loc." refers to proteins with experimentally determined localizations in *Arabidopsis*. "Predicted loc." refers to proteins with predicted localizations based on sequence homology or the presence of motifs characteristic of a specific localization. "Predicted loc. confirmed by LOPIT" refers to proteins with predicted localizations that were confirmed based on their positions on the PCA scores plot relative to the organelle marker proteins. These proteins were used as the PLS-DA training set. "LOPIT classification" refers to proteins that were assigned to an organelle by using PLS-DA in conjunction with limited manual analysis.

to mitochondria/plastids (which were not resolved from each other within the density gradient used in this study), Golgi apparatus, ER, PM, and vacuolar membrane (Fig. 1 and Table 1). A literature survey was conducted to annotate a further 121 proteins with predicted localizations, based on sequence homology or the presence of motifs characteristic of a specific localization (Tables 1–3). The PCA positions of these proteins were examined. The predicted localizations of 115 of these proteins were confirmed based on their positions on the PCA scores plot relative to the organelle marker proteins (Fig. 1 and Table 2). The localizations of the remaining six proteins [CslE1 (At1g55850), CesA1 (At4g32410), CesA3 (At5g05170), COBL7 (At4g16120), vATPase V0a (At2g28520), and a glycosylphosphatidylinositol (GPI)-anchored protein (At1g29980)] are discussed below.

PCA alone is not suitable for assigning proteins to organelles because it is not possible to determine the positions and the shapes of the boundaries that define the organelle clusters. Consequently, partial least squares discriminant analysis (PLS-DA), which is a technique for classifying observations in multidimensional data sets based on training sets of observations with known class memberships, was used. The organelle markers and the confirmed organelle proteins (142 proteins; Table 1) were grouped into five classes, corresponding to the ER, Golgi apparatus, PM, vacuolar membrane, and mitochondria/plastids, and were used to build a PLS-DA model ($Q^2 = 0.602$). This model predicted the class memberships of the identified proteins. The classification was further cross referenced by checking the positions of the classified proteins on the PCA scores plot. Based on their high similarity to

the PCA scores and the reporter ion profiles of the organelle marker proteins, 29 proteins were manually reclassified at this point (marked with an asterisk in Tables 2 and 3). Using this approach, 527 proteins in total were assigned subcellular localizations to the ER (182), Golgi apparatus (89), PM (92), vacuolar membrane (24), or mitochondria/plastids (140) (Fig. 1 and Tables 1–3). Density gradient distributions of proteins assigned to the five classes are shown in Fig. 5. Proteins assigned to the same class have very similar profiles, thus demonstrating the validity of the multivariate classification approach. The 162 proteins that were not classified could belong to organelles that were not included in the PLS-DA training set, such as the prevacuolar compartment. The nonclassified proteins also could represent proteins localized to multiple compartments. The density gradient distributions and, therefore, PCA scores of these proteins are combinations of the density gradient profiles of the different organelles in which they reside. For example, if 95% of a protein was targeted to the ER and 5% to the Golgi apparatus, then the protein probably would cluster with the ER markers. However, a protein distributed equally between the ER and Golgi apparatus may exhibit an intermediate distribution and would most likely be unclassified after PLS-DA analysis. The possibility of the intermediate distribution matching another organelle profile cannot be excluded; however, there is no evidence to suggest that codistribution has occurred in this study.

Validation. The distinct iTRAQ profiles and PCA scores of the ER, Golgi apparatus, PM, vacuolar membrane, and mitochondria/plastid markers confirmed the validity of the LOPIT technique for distinguishing proteins resident in these organelles (Fig. 1). To further confirm the utility of the LOPIT approach, localizations of previously uncharacterized proteins representing targeting predictions to each of Golgi apparatus, ER, vacuolar membrane, and PM were verified by microscopy. In all, 22 proteins were fused to GFP and the fusion proteins expressed in *Nicotiana tabacum* leaves (ref. 7; see also Table 4 and *Supporting Methods*, which are published as supporting information on the PNAS web site). The subcellular distribution of each was then determined by confocal microscopy (Fig. 2 and Table 2; see also Fig. 7, which is published as supporting information on the PNAS web site). Sixteen of the GFP fusions were targeted as predicted, two were not targeted as predicted, and targeting of four fusions was inconclusive because although there was GFP expression, the subcellular structure marked was not assignable to a distinct organelle. GFP fusions to proteins predicted to target the ER and Golgi apparatus were shown to colocalize with known markers for these organelles as further demonstration of the LOPIT approach (Fig. 2). The false positive rate of localization of proteins to organelles by using the LOPIT approach from this relatively small data set thus could be estimated at $\approx 10\%$ (2 of 18).

Discussion

Development of the LOPIT technique by using iTRAQ in conjunction with 2D liquid chromatography of peptides followed by tandem MS has resulted in the determination of density gradient distributions of 689 proteins, of which 60% were predicted to contain at least one transmembrane domain by using HMMTOP 2.0 (www.enzim.hu/hmmtop). Based on these distributions, 527 proteins were assigned to the ER, Golgi apparatus, PM, vacuolar membrane, or mitochondria/plastid. Previous proteomic studies of the endomembrane system have been hindered by the presence of contaminating proteins, making it impossible to assign proteins to organelles. For example, proteomic analysis of a Golgi apparatus-enriched fraction derived from rat liver resulted in the identification of 421 proteins, of which $>55\%$ were contaminants (8). In contrast, LOPIT does not depend on the production of pure organelles. Consequently, this proteomic analysis shows that confident and parallel localization of proteins to the ER, PM, Golgi apparatus, or vacuolar membrane can be achieved. The remainder of this discussion will focus on the protein content revealed in these endo-

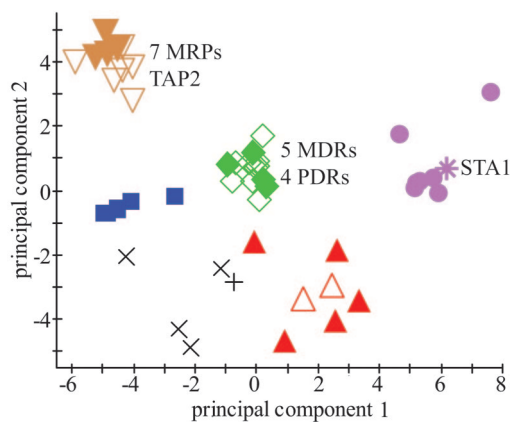


Fig. 3. PCA score plot showing positions of the ABC transporters, vacuolar sorting receptors, the unclassified V-ATPase V0a homolog, and GMT1 and GMT2 relative to the organelle marker proteins (filled shapes). Inverted triangles, vacuolar membrane and vacuolar ABC transporters; diamonds, PM and PM ABC transporters; circles, mitochondria/plastids; star, mitochondrial ABC transporter; diagonal crosses, vacuolar sorting receptors; horizontal cross, V-ATPase subunit V0a; triangles, Golgi apparatus, GMT1, and GMT2; squares, ER.

ABC transporters (MRP1, MRP2, MRP3, MRP4, MRP5, MRP6, MRP10, and TAP2) (21). In the study as a whole, 18 members of the ABC transporter superfamily have been assigned to specific organelles by LOPIT. The spatial segregation of ABC transporters belonging to the MDR, MRP, and PDR subfamilies is striking; every PDR and MDR family member is assigned to the PM, whereas all seven MRP are localized to the vacuolar membrane (Figs. 2 and 3). Components of the V1 (A, C, D, E, and H) and V0 (two V0a homologs and V0d) V-ATPase complexes are assigned to the vacuolar membrane (25). A third V0a homolog was identified but exhibits a density gradient distribution distinct from the vacuolar membrane markers. This protein was not classified and is positioned in a region of PCA scores plot that is close to the Golgi apparatus cluster. This region contains the four identified vacuolar-sorting receptors and, therefore, may correspond to the prevacuolar compartment (26). Differential localization of V0a homologs also has been observed in yeast; the V0a homolog Vph1 is localized to the vacuolar membrane, whereas Stv1 is localized to the late Golgi apparatus (25).

Golgi Apparatus. The 89 proteins assigned by LOPIT to the Golgi apparatus largely belong to three main classes: predicted glycosyltransferases, EMP70 proteins, and putative methyltransferases. Twenty-four predicted glycosyltransferases cited in CAZy (27) are localized to the Golgi apparatus, reflecting the fact that this organelle is specialized for glycosylation. These glycosyltransferases include enzymes involved in N-glycosylation, such as N-acetyl glucosaminyl transferase I, and cell wall biosynthetic enzymes, such as Quasimodo1 (28, 29). Interestingly, the cellulose synthase proteins Cesa1 and Cesa3 also are localized to the Golgi apparatus (30). The synthesis of cellulose occurs at the PM (31). However, localization of cellulose synthase proteins to compartments within the cell, including the Golgi apparatus, has been reported, and it has been suggested that cellulose synthase components cycle between the PM and an intracellular organelle to deliver cellulose synthase complexes to regions of cell wall synthesis (31, 32). The cellulose synthase-like (Csl) family are thought to be involved in Golgi apparatus-localized polysaccharide synthesis (30). In this study, CslC6, CslD2, and CslD3 are localized to the Golgi apparatus. However, CslE1 clearly fractionates with the ER, suggesting an alternative role for this protein.

Ten EMP70 family proteins are assigned to the Golgi apparatus. Nothing is known about the function of the EMP70 family, despite members being found in organisms as diverse as *Homo sapiens*, *Saccharomyces cerevisiae*, and *Arabidopsis*. The yeast EMP70 homolog Yer113c also has been localized to the Golgi apparatus (33). However, in COS-1 cells, a human EMP70 homolog has been localized to the endosome (34). In plants, the prevacuolar compartment, which contains the vacuolar sorting receptors (VSR), is thought to correspond to an endosomal compartment (26). The VSRs did not cluster with the Golgi apparatus-assigned proteins in this analysis (Fig. 3); therefore, it is unlikely that the EMP70 family is localized to the endosome in *Arabidopsis*.

Eighteen members of a previously uncharacterized plant-specific family are localized to the Golgi apparatus. Members of this family have putative methyltransferase activity, based on alignment with bacterial S-adenosylmethionine-dependent methyltransferases by using the sequence-structure alignment program FUGUE (35), and, therefore, were named the GMT (Golgi Methyl Transferase) family. The confirmation of the localization of GMT1 and GMT2 to the Golgi apparatus clearly is shown by confocal microscopy (Fig. 2d and data not shown).

Our previous ICAT LOPIT study focused on discriminating between proteins within the Golgi apparatus and the ER. Twenty-five of the Golgi apparatus proteins assigned here also were identified in that study, and the Golgi assignment is in agreement with the data in that study for all of these proteins.

Validation of LOPIT Predictions by GFP Localization. In nearly 90% of the cases studied at random, GFP fusions were targeted to the correct LOPIT-predicted organelle (16 of the 18 that gave a clear result). These data can be taken as confirmation of the efficacy of the LOPIT technique. In only two cases (At1g75140-GFP and At5g57800-GFP) was the predicted localization (ER and PM, respectively) different from the localization of the GFP-fusion protein (vacuolar membrane). This disagreement between the two techniques may reflect that fact that GFP-tagged *Arabidopsis* proteins were expressed highly in *Nicotiana tabacum* leaf tissue, and that the LOPIT localizations were carried out on root-derived material. It is possible that the locations of these proteins differ between tissue types. A few of the GFP fusion localizations (3 of 21) were inconclusive.

Concluding Remarks. The application of LOPIT to study the plant endomembrane system has resulted in the simultaneous localization of 527 proteins to the ER, PM, Golgi apparatus, vacuolar membrane, and mitochondria/plastid. The localization information provided in the supporting tables will prove an extremely useful resource for biologists investigating the endomembrane system. LOPIT enabled the protein contents of the PM, vacuolar membrane, Golgi apparatus, and ER to be analyzed in parallel, resulting in the determination of steady-state protein distributions between these organelles. This study represents a proteomic analysis of the ER, Golgi apparatus, PM, and vacuolar membrane that enables genuine organelle residents to be distinguished from contaminants and proteins in transit through the endomembrane system.

Experimental Procedures

Membrane Fractionation and iTRAQ Labeling. *Arabidopsis thaliana* callus cultures were established and maintained as described in ref. 36. Membranes were prepared and fractionated by using self-generating iodixanol density gradients (5). Two independent membrane preparations were performed. Density gradient fractions (500 μ l) were carbonate washed (800 μ l of 160 mM Na₂CO₃ at 4°C for 30 min). Membranes were pelleted by centrifugation (100,000 \times g for 25 min), washed (1 ml H₂O at 4°C) and repelleted (100,000 \times g for 10 min). Membranes were solubilized in 100 μ l of 25 mM triethylammonium bicarbonate (TEAB)/8 M urea/2% Triton X-100/0.1% SDS. Protein concentrations were determined by

using the BCA Protein Assay kit (Pierce). Two four-plex iTRAQ comparisons (Applied Biosystems) were performed across the first 12 of the 20 fractions collected from each membrane preparation. Comparison A: Fraction 1 (least dense) was labeled with iTRAQ reagent 114, fraction 4 with 115, fraction 7 with 116, and fractions 11 and 12 (which were pooled to give comparable protein concentrations to the other three fractions) with 117. Comparison B: Fraction 2 was labeled with iTRAQ reagent 114, fraction 5 with 115, fraction 8 with 116, and fractions 11 and 12 with 117. Samples (100 μ g of protein) were reduced [4 mM Tris(2-carboxyethyl)phosphine at 20°C for 1 h] and cysteines blocked [8 mM methyl methanethiosulfonate at 20°C for 10 min]. Samples were diluted (50 mM TEAB), such that the urea concentration was <1 M, digested with trypsin overnight at 37°C (Promega; 2.5 μ g was added at 0 and 1 h), and lyophilized. Lyophilized samples were resuspended in 100 μ l of 0.25 M TEAB/75% ethanol, added to one unit of the corresponding iTRAQ reagent and incubated for 1 h at 20°C. Residual reagents were hydrolyzed with 100 μ l of water (20°C for 15 min). The labeled peptides from each iTRAQ comparison were then pooled and lyophilized.

Cation Exchange Chromatography and LC-MS/MS Analysis. See *Supporting Methods* for more details.

MS Data Analysis and Protein Quantitation. MS data files were processed by using WIFF2DTA, which converts QSTAR data files (.wiff) into text files that contain m/z and intensity information (peak lists) for each product ion spectra (37). Each MS data file was processed twice to generate first, centroided peak lists that contain a single m/z value for each ion corresponding to the center of the peak, and second, uncentroided peak lists that contain m/z information for the entire peak. MASCOT 2.0.01 (Matrix Science, London) was used to search centroided peak lists against the MIPS *Arabidopsis* protein database (ftp://ftpmips.gsf.de/crest/arabiprot/arabi_all.proteins.v090704.gz; 26,719 entries). The modifications used were as follows: fixed, iTRAQ (K), iTRAQ (N-term), MMTS (C); variable, oxidation (M), iTRAQ (Y). The MS tolerance was 0.2 Da, and the MSMS tolerance was 0.5 Da. The MASCOT search results comprise a list of peptide identifications, each of which has a score that indicates the quality of the result. To determine the minimum peptide MASCOT score for accurate protein identification,

each peak list was searched against a version of the *Arabidopsis* database in which the protein sequences were reversed. The identifications that resulted from the reversed database search were used to calculate the false protein identification rate:

False protein identification rate = number of proteins identified from the reversed search/number of proteins identified from the normal search \times 100.

MASCOT peptide scores of at least 27 and 32 for the first and second experiments, respectively, resulted in false protein identification rates of <1% for proteins containing at least two peptides. Normalized iTRAQ reporter ion ratios were calculated from the uncentroided peak lists by using the recently developed I-TRACKER (38). Normalized reporter ion areas were calculated as follows: Normalized area A = area A/(area A + area B + area C + area D). The Genome Annotating Proteome Pipeline (GAPP) system (I.P.S. and C.B., unpublished work) was used to parse peptide identification and scoring information from the MASCOT output files and link these identifications to the quantitation data in a relational database (MYSQL 4.0; MYSQL AB, Uppsala). Peptides were quantified if at least three of the reporter ion peaks were above a threshold of 15 counts and if they had a MASCOT score of at least 20. In addition, only peptides that were unique to a single identified protein were quantified.

Multivariate Data Analysis. iTRAQ ratios were imported into SIMCA 10 (Umetrics, Umea, Sweden), logged, and preprocessed with unit-variance scaling. PCA and PLS-DA were performed as described previously, except that five organelle classes were used for the PLS-DA training set (5). For protein classification by PLS-DA, the lowest-scoring training set proteins were used as the threshold for class membership.

Construction of GFP Fusions and Confocal Microscopy. See *Supporting Methods* and Table 4 for more information.

We thank Zhinong Zhang for maintaining the *Arabidopsis* callus lines and Prof. R Jackson for his insightful comments about the manuscript. This work was supported by the Biotechnology and Biological Sciences Research Council, including a Council for Advancement and Support for Education award (to T.P.J.D.), the European Union Pharma Planta Consortium, and Applied Biosystems.

- Dunkley, T. P., Dupree, P., Watson, R. B. & Lilley, K. S. (2004) *Biochem. Soc. Trans.* **32**, 520–523.
- Gabalton, T. & Huynen, M. A. (2004) *Biochim. Biophys. Acta* **1659**, 212–220.
- Peck, S. C. (2005) *Plant Physiol.* **138**, 591–599.
- Hanton, S. L., Bortolotti, L. E., Renna, L., Stefano, G. & Brandizzi, F. (2005) *Traffic* **6**, 267–277.
- Dunkley, T. P. J., Watson, R., Griffin, J. L., Dupree, P. & Lilley, K. S. (2004) *Mol. Cell. Proteomics* **3**, 1128–1134.
- Ross, P. L., Huang, Y. L. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., et al. (2004) *Mol. Cell. Proteomics* **3**, 1154–1169.
- Brandizzi, F., Snapp, E. L., Roberts, A. G., Lippincott-Schwartz, J. & Hawes, C. (2002) *Plant Cell* **14**, 1293–1309.
- Wu, C. C., MacCoss, M. J., Mardones, G., Finnigan, C., Mogelsvang, S., Yates, J. R., 3rd, & Howell, K. E. (2004) *Mol. Biol. Cell.* **15**, 2907–2919.
- Galili, G., Sengupta-Gopalan, C. & Ceriotti, A. (1998) *Plant Mol. Biol.* **38**, 1–29.
- Young, B. P., Craven, R. A., Reid, P. J., Willer, M. & Stirling, C. J. (2001) *EMBO J.* **20**, 262–271.
- Kelleher, D. J., Karaoglu, D., Mandon, E. C. & Gilmore, R. (2003) *Mol. Cell* **12**, 101–111.
- Burda, P. & Aebi, M. (1999) *Biochim. Biophys. Acta* **1426**, 239–257.
- Werck-Reichhart, D., Bak, S. & Paquette, S. (2002) in *The Arabidopsis Book*, eds Somerville, C. & Meyerowitz, E. (Am. Soc. Plant Biol., Rockville, MD), 1–28.
- Beisson, F., Koo, A. J. K., Ruuska, S., Schwender, J., Pollard, M., Thelen, J. J., Paddock, T., Salas, J. J., Savage, L., Milcamps, A., et al. (2003) *Plant Physiol.* **132**, 681–697.
- Geisler, M., Axelsen, K. B., Harper, J. F. & Palmgren, M. G. (2000) *Biochim. Biophys. Acta* **1465**, 52–78.
- BarPeled, M. & Raikhel, N. V. (1997) *Plant Physiol.* **114**, 315–324.
- Pratelli, J., Sutter, J. U. & Blatt, M. R. (2004) *Trends Plant Sci.* **9**, 187–195.
- Zheng, H. Q., Kunst, L., Hawes, C. & Moore, I. (2004) *Plant J.* **37**, 398414.
- Kaiser, C. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 3783–3785.
- Cheng, S.-H., Willmann, M. R., Chen, H.-C. & Sheen, J. (2002) *Plant Physiol.* **129**, 469–485.
- Sanchez-Fernandez, R., Davies, T. G. E., Coleman, J. O. D. & Rea, P. A. (2001) *J. Biol. Chem.* **276**, 30231–30244.
- DeWitt, N. D., Hong, B., Sussman, M. R. & Harper, J. F. (1996) *Plant Physiol.* **112**, 833–844.
- Johanson, U., Karlsson, M., Johansson, I., Gustavsson, S., Sjovald, S., Fraysse, L., Weig, A. R. & Kjellbom, P. (2001) *Plant Physiol.* **126**, 1358–1369.
- Borner, G. H. H., Lilley, K. S., Stevens, T. J. & Dupree, P. (2003) *Plant Physiol.* **132**, 568–577.
- Nishi, T. & Forgac, M. (2002) *Nat. Rev. Mol. Cell Biol.* **3**, 94–103.
- Tse, Y. C., Mo, B. X., Hillmer, S., Zhao, M., Lo, S. W., Robinson, D. G. & Jiang, L. W. (2004) *Plant Cell* **16**, 672–693.
- Coutinho, P. M. & Henrissat, B. (1999) in *Recent Advances in Carbohydrate Bioengineering*, eds Gilbert, H. J., Davies, G., Henrissat, B. & Svensson, B. (R. Soc. Chem., Cambridge, U.K.), pp. 3–12.
- Bakker, H., Lommen, A., Jordi, W., Stiekema, W. & Bosch, D. (1999) *Biochem. Biophys. Res. Commun.* **261**, 829–832.
- Bouton, S., Leboeuf, E., Mouille, G., Leydecker, M. T., Talbotec, J., Granier, F., Lahaye, M., Hofte, H. & Truong, H. N. (2002) *Plant Cell* **14**, 2577–2590.
- Richmond, T. A. & Somerville, C. R. (2000) *Plant Physiol.* **124**, 495–498.
- Haigler, C. H. & Brown, R. M. (1986) *Protoplasma* **134**, 111–120.
- Gardiner, J. C., Taylor, N. G. & Turner, S. R. (2003) *Plant Cell* **15**, 17401748.
- Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S. & O’Shea, E. K. (2003) *Nature* **425**, 686–691.
- Schimmoller, F., Diaz, E., Muhlbauer, B. & Pfeffer, S. R. (1998) *Gene* **216**, 311–318.
- Shi, J. Y., Blundell, T. L. & Mizuguchi, K. (2001) *J. Mol. Biol.* **310**, 243–257.
- Sherrier, D. J., Prime, T. A. & Dupree, P. (1999) *Electrophoresis* **20**, 2027–2035.
- Boehm, A., Galvin, R. & Sickmann, A. (2004) *BMC Bioinformatics* **5**, 162.
- Shadforth, I., Dunkley, T., Lilley, K. & Bessant, C. (2005) *BMC Genomics* **6**, 145.