

# Localization of Organelle Proteins by Isotope Tagging (LOPIT)\*<sup>§</sup>

T. P. J. Dunkley<sup>‡</sup>, R. Watson<sup>§</sup>, J. L. Griffin<sup>‡</sup>, P. Dupree<sup>‡</sup>, and K. S. Lilley<sup>‡¶</sup>

**We describe a proteomics method for determining the subcellular localization of membrane proteins. Organelles are partially separated using centrifugation through self-generating density gradients. Proteins from each organelle co-fractionate and therefore exhibit similar distributions in the gradient. Protein distributions can be determined through a series of pair-wise comparisons of gradient fractions, using cleavable ICAT to enable relative quantitation of protein levels by MS. The localization of novel proteins is determined using multivariate data analysis techniques to match their distributions to those of proteins that are known to reside in specific organelles. Using this approach, we have simultaneously demonstrated the localization of membrane proteins in both the endoplasmic reticulum and the Golgi apparatus in *Arabidopsis*. Localization of organelle proteins by isotope tagging is a new tool for high-throughput protein localization, which is applicable to a wide range of research areas such as the study of organelle function and protein trafficking. *Molecular & Cellular Proteomics* 3:1128–1134, 2004.**

Determining the subcellular localization of novel proteins is an important step toward elucidating their role in the cell, because proteins are spatially organized according to their function (1). The enrichment of subcellular compartments followed by the identification of their protein contents by proteomics is a powerful method for rapid protein localization. However, the precise localization of proteins has been hindered by difficulties in preparing pure organelles. This is essential, because confident localization of a protein to a specific organelle requires that the preparation is free of contamination from other organelle types (2). This is particularly problematic in the case of the endomembrane system, owing to the similar densities of its component organelles and the continuous cycling of membranes and proteins between these compartments. One solution to this problem is the use of analytical rather than preparative centrifugation. Analytical centrifugation is an established method for assigning individ-

ual proteins to subcellular compartments that have eluded purification and refers to the analysis of protein distributions in density gradients, as opposed to preparative centrifugation, which is the analysis of single organelle-enriched fractions (3, 4). A recent proteomic study of the human centrosome used this approach to distinguish contaminants from genuine components of this complex (5).

We have developed a proteomics technique for the localization of integral membrane proteins to compartments of the endomembrane system in *Arabidopsis thaliana* that is based upon analytical centrifugation and hence is not dependent on the production of pure organelles. The technique involves partial separation of organelles by density gradient centrifugation followed by the analysis of protein distributions in the gradient by ICAT and MS. Multivariate data analysis techniques are then used to group proteins according to their distributions and hence localizations. In this study, we demonstrate that the localization of organelle proteins by isotope tagging (LOPIT)<sup>1</sup> can be used to discriminate Golgi, endoplasmic reticulum (ER), plasma membrane (PM), and mitochondrial/plastid proteins. We validate this technique by confirming the localization of a set of predicted ER and Golgi membrane proteins.

## EXPERIMENTAL PROCEDURES

**Membrane Fractionation**—*A. thaliana* liquid callus cultures expressing a myc-tagged version of gtl-6 were established and maintained as described (6, 7). Approximately 60 g of callus were homogenized in homogenization buffer (0.25 M sucrose, 10 mM HEPES-NaOH, pH 7.4, 1 mM EDTA, 1 mM DTT) at 4 °C, using a polytron with a 94 PTA 10EC head attachment (Kinematica, Littav, Switzerland) for two pulses of 7 s. The homogenate was then centrifuged twice for 5 min at 2,200 × *g*. The supernatant was loaded onto a 6-ml 18% iodixanol cushion and centrifuged at 100,000 × *g* for 2 h in a SW28 rotor (Beckman Coulter, Fullerton, CA). Crude membranes were then collected from the interface and adjusted to 16% iodixanol. These membranes were then fractionated according to their density by centrifuging at 350,000 × *g* for 3 h in a VTi65.1 rotor (Beckman Coulter), during which time an iodixanol gradient is generated. Fractions of 0.5 ml were harvested from the top of the gradient using an Auto Densi-flow collection device (Labconco Corporation, Kansas City, MO).

**Western Blot Analysis**—Membranes from each density gradient

From the <sup>‡</sup>Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, United Kingdom; and <sup>§</sup>Applied Biosystems, Warrington, Cheshire WA3 7QH, United Kingdom

Received, June 17, 2004, and in revised form, August 4, 2004

Published, MCP Papers in Press, August 4, 2004, DOI 10.1074/mcp.T400009-MCP200

<sup>1</sup> The abbreviations used are: LOPIT, localization of organelle proteins by isotope tagging; ER, endoplasmic reticulum; PM, plasma membrane; PCA, principal components analysis; PC, principal component; PLS-DA, partial least squares discriminant analysis; GFP, green fluorescent protein.

fraction were diluted with 0.8 ml of cold deionized water, pelleted by centrifugation at  $100,000 \times g$  for 25 min in a TLA 100.3 rotor (Beckman Coulter) and resuspended in 600  $\mu$ l of 4% SDS sample buffer. Then 7  $\mu$ l of each fraction was heated at 60 °C for 3 min and then resolved on a 12% SDS-PAGE gel, processed for Western analysis according to Wee *et al.* (8), and probed with rabbit anti-myc (A14, Santa Cruz Biotechnology, Santa Cruz, CA) and anti-Sec12 antibodies (a gift from S. Mogelsvang).

**Membrane Protein Preparation and Labeling**—To remove soluble and membrane-associated proteins, 0.8 ml of cold 160 mM  $\text{Na}_2\text{CO}_3$  was added to each density gradient fraction. After 30 min on ice, the washed membranes were pelleted as above. After removing the supernatant, the pellet was washed with 1 ml of cold de-ionized water and then re-pelleted by centrifuging at  $100,000 \times g$  for 10 min. The second supernatant was removed and the membrane pellets solubilized in 100  $\mu$ l of labeling buffer (50 mM Tris-HCl, pH 8.5, 8 M urea, 2% Triton X-100, 0.1% SDS). The protein concentration of each fraction was then determined using the DC Protein Assay kit (Bio-Rad, Hercules, CA). Using cleavable ICAT reagents (Applied Biosystems, Framingham, MA), six pair-wise comparisons were performed on the first 12 fractions collected, where 1 is the least dense fraction and 12 the most dense: 1 *versus* 4, 2 *versus* 5, 3 *versus* 6, 4 *versus* 7, 5 *versus* 8 and 7 *versus* 11 and 12 (which were pooled). For each comparison, the lighter fraction was labeled with the light ICAT reagent and the denser fraction with the heavy ICAT reagent. Labeling was performed according to the standard Applied Biosystems ICAT protocol ([www.appliedbiosystems.com](http://www.appliedbiosystems.com)) with the following exceptions: labeling buffer (50 mM Tris-HCl, pH 8.5, 8 M urea, 2% Triton X-100, 0.1% SDS) was used in place of denaturing buffer (50 mM Tris-HCl, pH 8.5, 0.1% SDS); reduction was performed for 30 min at 20 °C; labeling was performed at 20 °C; excess ICAT reagents were quenched by the addition of 10 mM cysteine after labeling and then incubated at 20 °C for 30 min before pooling the samples; prior to digestion the pooled samples were diluted to 2 ml with 25 mM Tris-HCl, pH 8.5, 10  $\mu$ g of modified trypsin were then added (Promega, Madison, WI); prior to the cation exchange step samples were diluted to 25 ml with 10 mM  $\text{KH}_2\text{PO}_4$ , pH 3, 25% ACN. After cleavage samples were resuspended in 25  $\mu$ l of 2% ACN, 0.1% TFA.

**Analysis of Peptides by MS**—Peptide separation by LC-MS/MS was performed using an Ultimate Nano-LC system (Dionex, Sunnyvale, CA) and a QSTAR XL Q-TOF hybrid mass spectrometer (Applied Biosystems). Samples (6.4  $\mu$ l) were delivered to the system using a FAMOS autosampler (Dionex) at 30  $\mu$ l/min, and the peptides were trapped onto a PepMap C18 pre-column (5 mm  $\times$  300  $\mu$ m i.d.; Dionex). Peptides were then eluted onto the PepMap C18 analytical column (15 cm  $\times$  75  $\mu$ m i.d.; Dionex) at  $\sim$ 100 nl/min and separated using a 4-h gradient of 5–32% ACN. The QSTAR XL was operated in information-dependent acquisition mode, in which a 1-s TOF MS scan from 350–1,600  $m/z$ , was performed, followed by 3-s product ion scans from 65–1,600  $m/z$  on the two most intense doubly or triply charged ions. Each sample was analyzed three times to improve the quantitation accuracy and identification confidence.

**Peptide Identification and Quantification**—MS data files were analyzed using ProCAT software version 1.0 service pack 2.0 (Applied Biosystems) and searched against the MIPS *Arabidopsis* protein database ([ftpmips.gsf.de/crest/arabiprot/version\\_110903](http://ftpmips.gsf.de/crest/arabiprot/version_110903) containing 26643 entries). Peptides identified with less than 90% confidence were excluded. Multiple peptide matches resulting from the same MS/MS data were removed using Microsoft Access with the queries included in the ProCAT software. To further increase identification stringency, the proteins identified using ProCAT were verified by analyzing the MS data using MASCOT Daemon version 2.0.01 (Matrix Science, London, United Kingdom). Proteins with scores above 33 were identified with 95% confidence. Proteins positively identified by

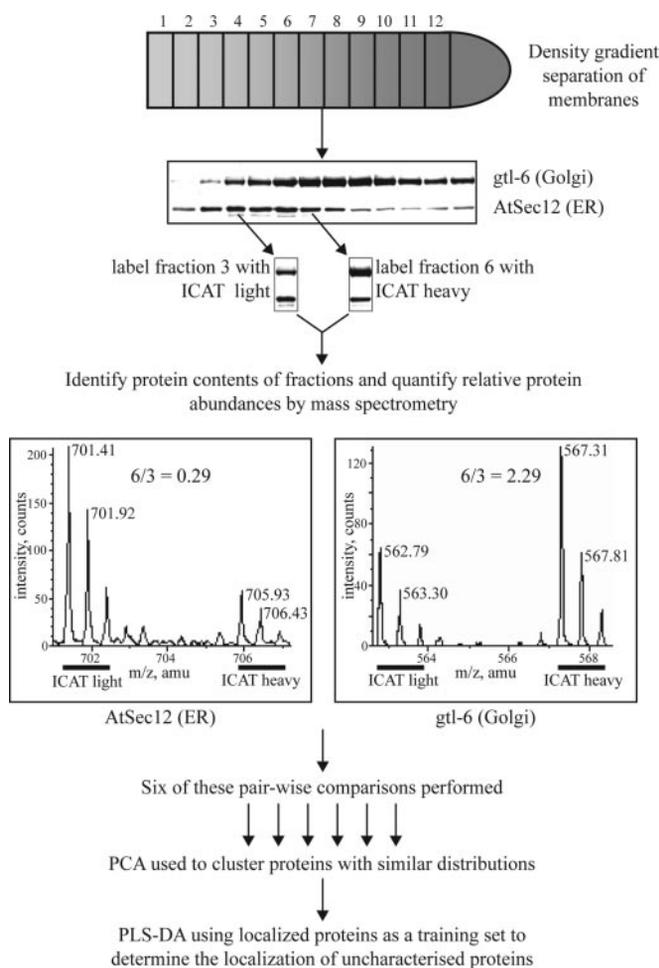
ProCAT but with MASCOT scores below this threshold were discarded. In addition, the 14 proteins with MASCOT scores below 40 were verified by manual inspection of the corresponding MS/MS spectra. ProCAT automatically measures the ratio of the heavy-labeled peptide to the light-labeled peptide. The heavy-to-light ratio for each protein is the mean ratio of its component peptides from the three MS runs. Only proteins with quantification information from at least four of the six ICAT comparisons were included in the final results table. Prediction of trans-membrane helices was performed using the web-based TMHMM software ([www.cbs.dtu.dk/services/TMHMM-2.0/](http://www.cbs.dtu.dk/services/TMHMM-2.0/)) (9).

**Multivariate Data Analysis**—The logged ICAT ratios obtained for each protein in the initial experiment and the repeat were imported into the SIMCA 10.0 package (Umetrics, Umea, Sweden) and pre-processed with mean centered scaling. The data was analyzed by principle components analysis (PCA) for the two experiments, and the proteins represented in two-dimensional scores plots. This approach summarizes the correlated variation in the dataset in an unsupervised manner, allowing the visualization of clusters within the dataset.

To measure the significance of these clusters in terms of the proteins of known sub-compartmentalization, the 12 known ER, Golgi, and mitochondrial/plastid proteins identified in the first experiment were then assigned to three classes and used to build a partial least squares-discriminant analysis (PLS-DA) model. This is a supervised pattern recognition approach that uses the class membership information (*i.e.* where a protein is localized) to maximize the separation between these proteins. The success of this approach was measured using a goodness of fit algorithm that produced a score ( $Q^2$ ) where 0 is no better than chance and 1 is the theoretical maximum for a component that explains all the variation in the data and correctly classifies the dataset. This model was then used to predict the localization of the remaining 158 proteins. The PLS-DA results consist of three values for each protein that indicate how well that protein fits the three localizations, where values closest to one indicate a good fit with the pattern associated with that localization. The lowest scoring known organelle resident in each class was used as the threshold for assigning confident organelle localizations to proteins. The thresholds were therefore 0.93 for the Golgi, 0.88 for the ER, and 0.83 for the mitochondria/plastid class.

## RESULTS

**Basis of the LOPIT Technique**—The basis of the LOPIT technique is outlined in Fig. 1. Organelles were prepared from *Arabidopsis* callus and partially separated according to their density using self-generating iodixanol gradients. Iodixanol is an alternative media to sucrose, which has the advantage of being able to form density gradients from homogenous solutions within a few hours of high-speed centrifugation. These self-generating gradients are extremely reproducible and simple to manipulate because their shape depends only on the starting iodixanol concentration and the centrifugation time (10, 11). To determine the distribution of the ER and Golgi apparatus, gradient fractions were analyzed by Western blot, using antibodies against the ER marker AtSec12 and the Golgi marker gtl-6. The density gradient centrifugation only partially separated these organelles. Furthermore, proteomic analysis of the peak Golgi fraction, by SDS-PAGE followed by LC-MS/MS, resulted in the identification of ER and Golgi proteins, as well as numerous proteins from other compartments including the PM, mitochondria, and plastids, making it impossible to



**FIG. 1. LOPIT workflow.** Organelles from *Arabidopsis* callus were partially separated by centrifugation through self-generating iodixanol density gradients. Golgi and ER membrane distributions were determined using Western blots with anti-myc antibodies to detect the Golgi marker gtl-6-myc and anti-sec12 antibodies to detect the ER marker AtSec12. Protein distributions were determined by performing six pair-wise comparisons across the gradient in which the lighter fraction was labeled with ICAT light and the denser fraction with ICAT heavy. Samples were then pooled, digested, and the labeled peptides avidin-affinity purified. The ICAT-labeled peptides were then analyzed by LC-MS/MS. Proteins were assigned to organelles by matching their distributions to those of proteins that have known subcellular localizations. This was achieved using the multivariate data analysis techniques, PCA and PLS-DA.

assign localizations to novel proteins (data not shown). In contrast, LOPIT involves the analysis of protein distribution across the gradients rather than the cataloguing of single fractions. This enables proteins from different organelles to be distinguished even if their distributions overlap. A comparison of fractions three and six, for example, could be used to distinguish ER and Golgi residents because the Golgi is more abundant in fraction six than in fraction three, while the ER exhibits the opposite distribution. By comparing the distributions of novel proteins to those of known organelle residents using the multivariate data analysis techniques PCA and PLS-

DA, the subcellular localization of novel proteins can be determined.

**Determination of Protein Identities and Distributions**—To determine the distributions of the proteins in the density gradient fractions, we utilized the cleavable ICAT reagents from Applied Biosystems. ICAT enables the relative quantitation of protein levels in two samples by MS (12). In total, six pair-wise comparisons were performed across the density gradient. The integral membrane proteins in each fraction were enriched by subjecting the membranes to a high pH wash, which removes the majority of soluble and peripheral membrane proteins (13). The proteins were then labeled with cleavable ICAT reagents, which specifically react with cysteine sulfhydryls. In each comparison, the least dense fraction was labeled with the light ICAT reagent and the denser fraction with the heavy reagent. The two fractions in each comparison were then pooled, and protein digestion followed by avidin affinity purification of the labeled peptides was performed. Each sample was then analyzed by LC-MS/MS, resulting in the confident identification and quantification of 170 proteins (see supplemental information). Of these proteins, 70% were predicted to be integral membrane proteins with 28% having two or more trans-membrane spans, re-confirming the applicability of ICAT technology to the analysis of hydrophobic integral membrane proteins (14). Thirteen of the proteins identified have been previously localized to specific organelles in *Arabidopsis*, while a further 15 had predicted subcellular locations based on homology to proteins in other plants, animals, or yeast. These proteins are listed in Table I. For the purpose of validating the LOPIT technique, the remainder of this study focuses on this set of 28 proteins. A detailed discussion of the remaining proteins is beyond the scope of this article and will be reported elsewhere.

**Clustering of Proteins According to Their Localization Using PCA**—To validate the LOPIT technique, the localization of the predicted ER and Golgi proteins, listed in Table I, was attempted. Manual analysis of the large datasets generated in these experiments is time consuming and error prone. Therefore, PCA was used to condense the dataset and identify patterns within it (15). In this case, PCA was used to reduce the six variable dataset (the six pair-wise comparisons) to a two-variable scores plot (Fig. 2). Principal component (PC) 1 (the x-axis in Fig. 2) is a best-fit vector that represents the greatest correlated variation in the dataset. PC2 (the y-axis in Fig. 2) represents the greatest variation at right angles to PC1, and is also the next most amount of correlated variation in the dataset. The scores for the two PCs for an individual protein describe the position of that protein in relation to these two new axes and the other proteins contributing to the model. Proteins that have similar distributions across the density gradient will cluster with one another in the PC scores plot. Although the two PCs represent variation across the total group analyzed, PC1 can be loosely defined as a measure of the likelihood a protein is either Golgi, mitochondria, or plastid

TABLE I  
Localization of proteins to the ER and Golgi by LOPIT

The 28 proteins identified with known or predicted localizations are shown. A PLS-DA model was built based on the known ER, Golgi, and mitochondrial/plastid proteins. This model was then used to rank the remaining proteins according to how well they fit these three localizations. Values closest to 1 indicate a good fit. Proteins with Golgi values above 0.93 were assigned to the Golgi, and proteins with ER values above 0.88 were assigned to the ER. The highest PLS-DA result for each protein is shown in bold. Asterisks indicate proteins that fall just below the threshold for confident assignment to the Golgi.

Name	Accession no.	Predicted localization (Ref.)	PLS-DA results		
			Golgi	ER	Mit/plastid
Proteins assigned to the Golgi					
Glycosyltransferase family 29	At1g08660	Golgi (36)	<b>1.06</b>	-0.09	0.03
Glycosyltransferase family 34 (AtGT5)	At1g74380	Golgi (37)	<b>1.19</b>	-0.28	0.09
Glycosyltransferase family 8 (quasimodo1)	At3g25140	Golgi (38)	<b>1.25</b>	-0.34	0.09
Glycosyltransferase family 8	At3g61130	Golgi (36)	<b>0.96</b>	-0.07	0.11
Glycosyltransferase family 47	At4g38040	Golgi (36)	<b>1.11</b>	-0.06	-0.05
Proteins assigned to the ER					
Obtusifoliol 14-demethylase (CYP51)	At1g11680	ER (40)	0.04	<b>1.07</b>	-0.11
$\beta$ -ketoacyl reductase (glossy8)	At1g67730	ER (41)	0.11	<b>0.89</b>	0.00
Cinnamate-4-hydroxylase (AtC4H)	At2g30490	ER (43)	-0.05	<b>1.08</b>	-0.03
Protein disulfide isomerase	At2g32920	ER (44)	-0.48	<b>1.37</b>	0.11
Serine C-palmitoyltransferase family (LCB1 homolog)	At4g36480	ER (45)	0.11	<b>0.96</b>	-0.07
Proteins not assigned to the ER or Golgi					
Plasma membrane proton ATPase	At2g18960	PM (35)	-0.14	0.53	<b>0.61</b>
NADH dehydrogenase family (ndb4)	At2g20800	Mit. (46)	0.16	-0.03	<b>0.88</b>
TOC75 homolog	At3g46740	Plastid (47, 48)	0.39	-0.60	<b>1.22</b>
Glycosyltransferase family 2 *	At3g07330	Golgi (36)	<b>0.78</b>	0.07	0.15
Apyrase *	At5g18280	Golgi (39)	<b>0.90</b>	-0.01	0.11
Ribophorin I	At1g76400	ER (42)	0.43	<b>0.63</b>	-0.06
Training set of proteins used to build the PLS-DA model					
Fucosyltransferase	At1g49710	Golgi (22, 23)	<b>1.06</b>	-0.12	0.06
Glycosyltransferase family 34 (gtl-6)	At2g22900	Golgi (7)	<b>0.99</b>	0.08	-0.07
N-acetyl glucosaminyltransferase	At4g38240	Golgi (23, 24)	<b>0.93</b>	-0.04	0.11
ER-type $\text{Ca}^{2+}$ -pumping ATPase (ECA1)	At1g07810	ER (25)	0.11	<b>0.96</b>	-0.08
AtSec12	At2g01470	ER (26)	0.06	<b>0.99</b>	-0.05
NADH-cytochrome $b_5$ reductase	At5g17770	ER (27)	-0.17	<b>0.88</b>	0.29
Calnexin	At5g61790	ER (28)	0.14	<b>1.00</b>	-0.14
Mitochondrial ADP,ATP carrier protein 1 (AAC1)	At3g08580	Mit. (29)	0.17	-0.01	<b>0.84</b>
Mitochondrial phosphate transporter (AT5)	At5g14040	Mit. (30, 31)	0.18	-0.13	<b>0.95</b>
Dicarboxylate/tricarboxylate carrier (AtDTC)	At5g19760	Mit. (32)	-0.06	0.04	<b>1.01</b>
Long chain acyl-CoA synthetase 9 (LACS9)	At1g77590	Plastid (33)	0.08	0.08	<b>0.83</b>
Chloroplast NAD-dependent malate dehydrogenase	At3g47520	Plastid (34)	0.00	-0.10	<b>1.10</b>

localized (positive score in PC1) or found localized in the ER (negative score in PC1). In a similar manner, PC2 distinguishes mitochondrial and plastid proteins (positive scores) from Golgi proteins (negative scores).

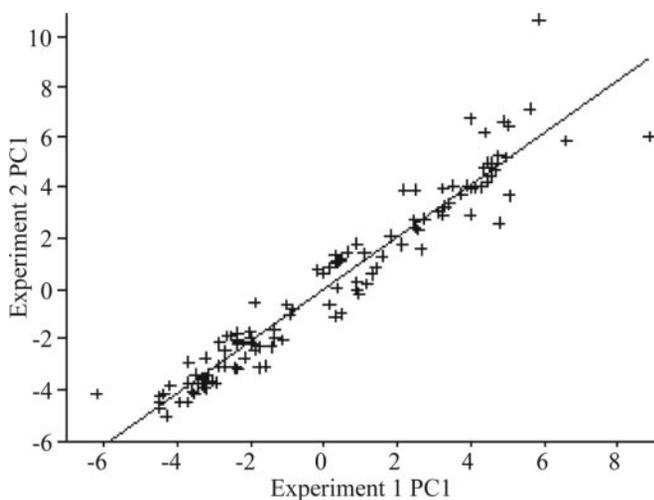
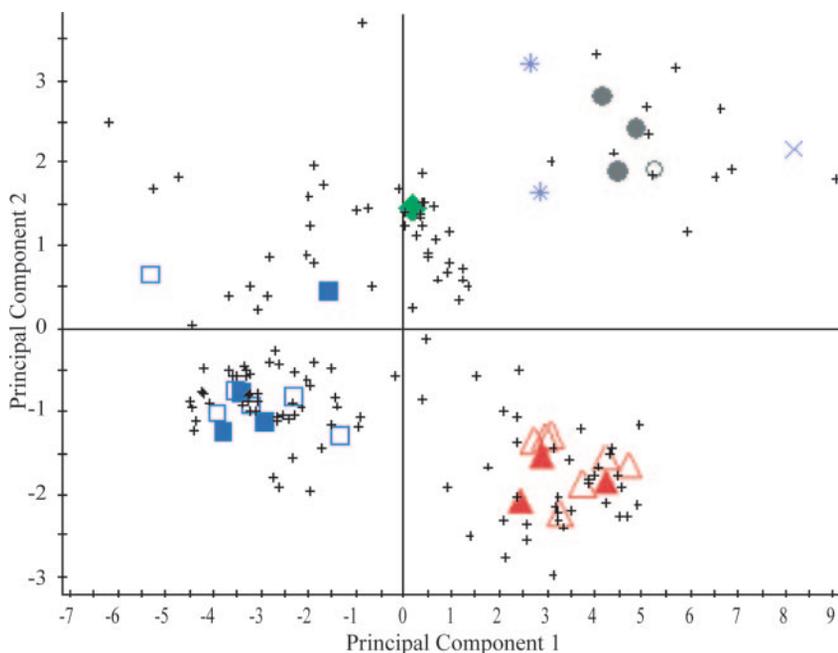
Fig. 2 clearly shows that the known ER and Golgi proteins form clusters that are distinct from one another and from the mitochondrial, plastid, and PM proteins. In this analysis, the predicted ER and Golgi proteins also cluster tightly with the known residents of their respective organelles, thus supporting the predicted localizations of these proteins. These results demonstrate that using this method it is possible to distinguish proteins that are localized to different subcellular compartments, without the need to purify individual organelles. We can also begin to assign the unknown proteins to subcellular compartments, because the majority of them cluster with the known organelle residents. The remaining proteins, which do not cluster with the marker proteins, may be residents of organelles for which no known resident proteins have been

identified, such as the endosome. Alternatively, these proteins may be localized to multiple compartments.

In order to assess whether protein clustering was reproducible, the experiment was repeated using a second membrane preparation. A total of 114 of the proteins that were identified in the original experiment were also found in the repeat. The distributions of these proteins, in the two experiments, were again analyzed using PCA. Fig. 3 shows the PC1 scores obtained for each protein in the two repeats of the experiment. In each case, PC1 represents the maximum amount of correlated variation in the two PCA models. In both models, PC1 distinguished the combined group of the Golgi, mitochondrial, and plastid proteins from the ER proteins. The strong correlation observed between the original scores and those obtained in the repeat indicates that the clustering, and therefore the localization, of proteins by LOPIT is reproducible.

*Confirmation of the Localization of the Predicted ER and Golgi Proteins Using PLS-DA—Visual clustering of proteins*

**FIG. 2. PCA scores plot showing clustering of proteins according to their distributions.** The logged ICAT ratios obtained for each protein were imported into the SIMCA 10.0 package and pre-processed with mean centered scaling. The data was then analyzed by PCA and the proteins represented in two-dimensional scores plots. Proteins that cluster in the scores plot exhibit similar distributions in the density gradient. ▲, Golgi; ■, ER; ◆, PM; ●, mitochondria; \*, plastid; △, predicted Golgi; □, predicted ER; ○, predicted mitochondria; ×, predicted plastid; +, unknown localization.



**FIG. 3. PCA scores plot showing the reproducibility of protein clustering.** The LOPIT experiment was carried out using two independent membrane preparations. The gradient distributions of the 116 proteins identified in both experiments were analyzed by PCA, and the PC1 scores obtained for each protein in the two experiments were compared. PC1 represents the maximum amount of correlated variation in the two PCA models.

based on their positions in the PC scores plot is a useful method for initial data analysis. However, the assignment of novel proteins to organelles requires a discriminate approach that can predict class membership. As a result, PLS-DA, which is a regression extension of PCA, was used. PLS-DA is a technique that can be used to predict the class membership of observations in a large multidimensional dataset based on a training set of observations with known class membership (16). In this case, the 12 known ER, Golgi, and mitochondrial/plastid proteins identified in the first experiment were as-

signed to three classes and used to build a PLS-DA model ( $Q^2 = 0.909$ ), which was used to determine the localization of the remaining 158 proteins (see supplemental information). The PLS-DA scores for the proteins in the training (known localization) and test (predicted localization) sets are listed in Table I. Using the stringent cut-off described above, LOPIT confirmed the Golgi localization of five of the seven predicted Golgi proteins and five of the six predicted ER proteins. The remaining two predicted Golgi proteins are also very likely to be Golgi localized, given their high scores for Golgi localization, which fall just below the threshold, and their low scores for the other compartments. Ribophorin I, the sole predicted ER protein whose localization was not confirmed, has a high ER score of 0.63 but also a reasonable Golgi score of 0.43. This protein may be present in a different subdomain of the ER than the proteins in the training and test sets. ER membrane proteins continuously escape from the ER to the Golgi and are retrieved in retrograde vesicles (17). The magnitude of the ribophorin I Golgi score could therefore reflect the steady state distribution of this protein, with the majority resident in the ER and a minority in the Golgi awaiting retrieval.

#### DISCUSSION

In this study, 170 proteins were identified and their density gradient distributions determined. Of these, a subset of 28 proteins, with known or predicted localizations, was used to validate the LOPIT technique. Fig. 2 shows the identified proteins clustered according to their distributions and hence localizations. The clear separation of the known ER and Golgi proteins, in this plot, demonstrates that LOPIT can be used to discriminate between the residents of these two organelles. The ER and the Golgi apparatus have very similar densities. In addition, proteins from the ER continuously escape to the

Golgi, so that even if an ER-free Golgi preparation could be obtained, the identification of ER resident proteins is inevitable (18). Previous Golgi proteome studies in rat liver have identified the ER as the major contaminant organelle (19, 20). Therefore, the ability to distinguish Golgi resident proteins from ER contaminants using LOPIT represents a significant advance in the localization of Golgi proteins by organelle proteomics.

To demonstrate the ability of LOPIT to determine the subcellular localization of previously uncharacterized proteins, a PLS-DA model was built using the proteins with known localizations. Using this model, the test set of 16 proteins with predicted localization were then assigned to the ER, Golgi, or the mitochondria/plastid categories. This resulted in the confident assignment of five proteins to the ER and five proteins to the Golgi, none of which had previously been localized in *Arabidopsis*. The subcellular localization of these proteins individually by microscopy would require substantial work because either antibodies would need to be raised or green fluorescent protein (GFP) fusions constructed. In addition, these methods have serious limitations in that the specificity of antibodies is often difficult to determine. Moreover, both the addition of GFP and its ectopic expression can affect the localization of a protein (21). LOPIT avoids these problems because the precise nature of the protein identifications by MS can be determined by examining the MASCOT and ProICAT search results, which indicate when peptides are identified that correspond to multiple protein isoforms. LOPIT also involves the analysis of proteins that are expressed at their endogenous levels and in their native forms, thus avoiding the possibility of protein mislocalization.

ICAT was chosen as the quantification technique used in this study because the reagents are commercially available and because the ICAT protocol was readily adapted to enable the labeling of hydrophobic membrane proteins. An alternative approach to protein quantification is the analysis of sample digests in consecutive LC-MS/MS runs and the subsequent comparisons of the peak intensities obtained in each experiment. ICAT has several advantages over this approach: samples are analyzed simultaneously so changes in LC conditions or electrospray efficiency that occur over time do not impair the quantification quality; samples are also prepared simultaneously so sample losses during digestion, and any pre-fractionation, will effect both samples equally; the isolation of cysteine-containing peptides, in the ICAT protocol, reduces sample complexity, thereby increasing the likelihood of identifying low-abundance proteins and reducing the length of the LC runs required to achieve sufficient peptide separation. In addition, the cost of isotopic labeling is partially offset by the fact that two samples are analyzed simultaneously, thus halving the MS time required for analysis.

Further refinements in the LOPIT methodology will include the use of two-dimensional LC to improve peptide separation prior to MS. This will result in the identification of an increased

number of low-abundance proteins, which should include markers for organelles that were not included in this study, such as the endosome and the vacuole. In addition, the gradients used for membrane separation could be manipulated with the aim of resolving endomembrane subcomponents or any poorly resolved organelles. For this reason, the use of self-generating iodixanol density gradients is advantageous because their shape and range can be manipulated by adjusting the initial iodixanol concentration and modifying the centrifugation time or speed (10, 11).

In conclusion, we have developed and validated a proteomics technique for the localization of organelle proteins. This method has been used to localize proteins to the ER and the Golgi apparatus. To our knowledge, this is the first example of a proteomic method that can discriminate between proteins resident in these organelles. LOPIT provides a new tool that will enable preliminary functional assignment to uncharacterized proteins, based on their determined localization. LOPIT can also be used for the accurate determination of an organelle's protein complement, which will facilitate the investigation of organelle function and protein trafficking.

*Acknowledgments*—We thank Svenja Hester and Julie Howard for assistance with mass spectrometry, Zhinong Zhang for maintaining the *Arabidopsis* callus lines, and Chris Barton for critical reading of this manuscript.

\* This work was supported by a BBSRC CASE award to T.P.J.D. from BBSRC and Applied Biosystems. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this manuscript (available at <http://www.mcponline.org/>) contains supplemental material.

¶ To whom correspondence should be addressed: Cambridge Centre for Proteomics, Department of Biochemistry, Building O, Downing Site, Cambridge CB2 1QW, United Kingdom. Tel.: 44-(0)1223-765255; Fax: 44-(0)1223-333345; E-mail: k.s.lilley@bioc.cam.ac.uk.

## REFERENCES

1. Dreger, M. (2003) Subcellular proteomics. *Mass Spectrom. Rev.* **22**, 27–56
2. Dunkley, T. P., Dupree, P., Watson, R. B., and Lilley, K. S. (2004) The use of isotope-coded affinity tags (ICAT) to study organelle proteomes in *Arabidopsis thaliana*. *Biochem. Soc. Trans.* **32**, 520–523
3. de Duve, C. (1971) Tissue fractionation. *J. Cell Biol.* **50**, 20D–55D
4. Dunphy, W. G., and Rothman, J. E. (1983) Compartmentation of asparagine-linked oligosaccharide processing in the Golgi-apparatus. *J. Cell Biol.* **97**, 270–275
5. Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A., and Mann, M. (2003) Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**, 570–574
6. Sherrier, D. J., Prime, T. A., and Dupree, P. (1999) Glycosylphosphatidylinositol-anchored cell-surface proteins from *Arabidopsis*. *Electrophoresis* **20**, 2027–2035
7. Mogelsvang, S. (1999) A family of *Arabidopsis thaliana* Golgi glycosyltransferases. Ph.D. thesis, University of Cambridge
8. Wee, E. G. T., Sherrier, D. J., Prime, T. A., and Dupree, P. (1998) Targeting of active sialyltransferase to the plant Golgi apparatus. *Plant Cell* **10**, 1759–1768
9. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov mod-

- el: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580
10. Ford, T., Graham, J., and Rickwood, D. (1994) Iodixanol—a nonionic isosmotic centrifugation medium for the formation of self-generated gradients. *Anal. Biochem.* **220**, 360–366
  11. Graham, J., Ford, T., and Rickwood, D. (1994) The preparation of subcellular organelles from mouse-liver in self-generated gradients of iodixanol. *Anal. Biochem.* **220**, 367–373
  12. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999
  13. Fujiki, Y., Hubbard, A. L., Fowler, S., and Lazarow, P. B. (1982) Isolation of intracellular membranes by means of sodium-carbonate treatment—Application to endoplasmic-reticulum. *J. Cell Biol.* **93**, 97–102
  14. Han, D. K., Eng J., Zhou, H. L., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19**, 946–951
  15. Lindon, J. C., Holmes, E., and Nicholson, J. K. (2001) Pattern recognition methods and applications in biomedical magnetic resonance. *Progr. Nuclear Magn. Reson. Spectr.* **39**, 1–40
  16. Karp, N. A., Griffin, J. L., and Lilley, K. S. (2004) Application of partial least squares discriminant analysis to two dimensional difference gel studies in expression proteomics. *Proteomics*, in press.
  17. Vitale, A., and Denecke, J. (1999) The endoplasmic reticulum—Gateway of the secretory pathway. *Plant Cell* **11**, 615–628
  18. Neumann, U., Brandizzi, F., and Hawes, C. (2003) Protein transport in plant cells: in and out of the Golgi. *Ann. Botany* **92**, 167–180
  19. Bell, A. W., Ward, M. A., Blackstock, W. P., Freeman, H. N. M., Choudhary, J. S., Lewis, A. P., Chotai, D., Fazel, A., Gushue, J. N., Paiement, J., Palcy, S., Chevet, E., Lafreniere-Roula, M., Solari, R., Thomas, D. Y., Rowley, A., and Bergeron, J. J. M., (2001) Proteomics characterization of abundant Golgi membrane proteins. *J. Biol. Chem.*, **276**, 5152–5165
  20. Wu, C. C., MacCoss, M. J., Mardones, G., Finnigan, C., Mogelsvang, S., Yates, J. R. 3rd, and Howell, K. E. (2004) Organellar proteomics reveals Golgi arginine dimethylation. *Mol. Biol. Cell*, **15**, 2907–2919
  21. Lisenbee, C. S., Karnik, S. K., and Trelease, R. N. (2003) Overexpression and mislocalization of a tail-anchored GFP redefines the identity of peroxisomal ER. *Traffic* **4**, 491–501
  22. Bakker, H., Schijlen, E., de Vries, T., Schiphorst, W., Jordi, W., Lommen, A., Bosch, D., and van Die, I. (2001) Plant members of the alpha 1 -> 3/4-fucosyltransferase gene family encode an alpha 1 -> 4-fucosyltransferase, potentially involved in Lewis(x) biosynthesis, and two core alpha 1 -> 3-fucosyltransferases. *FEBS Lett.* **507**, 307–312
  23. Kornfeld, R., and Kornfeld, S. (1985) Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**, 631–664
  24. Bakker, H., Lommen, A., Jordi, W., Stiekema, W., and Bosch, D. (1999) An *Arabidopsis thaliana* cDNA complements the *N*-acetylglucosaminyltransferase I deficiency of CHO Lec1 cells. *Biochem. Biophys. Res. Commun.* **261**, 829–832
  25. Liang, F., Cunningham, K. W., Harper, J. F., and Sze, H. (1997) ECA1 complements yeast mutants defective in  $Ca^{2+}$  pumps and encodes an endoplasmic reticulum-type  $Ca^{2+}$ -ATPase in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 8579–8584
  26. BarPeled, M., and Raikhel, N. V. (1997) Characterization of AtSEC12 and AtSAR1—Proteins likely involved in endoplasmic reticulum and Golgi transport. *Plant Physiol.* **114**, 315–324
  27. Fukuchi-Mizutani, M., Mizutani, M., Tanaka, Y., Kusumi, T., and Ohta, D. (1999) Microsomal electron transfer in higher plants: Cloning and heterologous expression of NADH-cytochrome  $b_5$  reductase from *Arabidopsis*. *Plant Physiol.* **119**, 353–361
  28. Huang, L. Q., Franklin, A. E., and Hoffman, N. E. (1993) Primary structure and characterization of an *Arabidopsis thaliana* calnexin-like protein. *J. Biol. Chem.* **268**, 6560–6566
  29. Haferkamp, I., Hackstein, J. H. P., Voncken, F. G. J., Schmit, G., and Tjaden, J. (2002) Functional integration of mitochondrial and hydrogenosomal ADP/ATP carriers in the *Escherichia coli* membrane reveals different biochemical characteristics for plants, mammals and anaerobic chytrids. *Eur. J. Biochem.* **269**, 3172–3181
  30. Takabatake, R., Hata, S., Taniguchi, M., Kouchi, H., Sugiyama, T., and Izui, K. (1999) Isolation and characterization of cDNAs encoding mitochondrial phosphate transporters in soybean, maize, rice, and *Arabidopsis*. *Plant Mol. Biol.* **40**, 479–486
  31. Hamel, P., Saint-Georges, Y., de Pinto, B., Lachacinski, N., Altamura, N., and Dujardin, G. (2004) Redundancy in the function of mitochondrial phosphate transport in *Saccharomyces cerevisiae* and *Arabidopsis thaliana*. *Mol. Microbiol.* **51**, 307–317
  32. Picault, N., Palmieri, L., Pisano, I., Hodges, M., and Palmieri, F. (2002) Identification of a novel transporter for dicarboxylates and tricarboxylates in plant mitochondria—Bacterial expression, reconstitution, functional characterization, and tissue distribution. *J. Biol. Chem.* **277**, 24204–24211
  33. Schnurr, J. A., Shockey, J. M., de Boer, G. J., and Browse, J. A. (2002) Fatty acid export from the chloroplast. Molecular characterization of a major plastidial acyl-coenzyme A synthetase from *Arabidopsis*. *Plant Physiol.* **129**, 1700–1709
  34. Berkemeyer, M., Scheibe, R., and Ocheretina, O. (1998) A novel, non-redox-regulated NAD-dependent malate dehydrogenase from chloroplasts of *Arabidopsis thaliana* L. *J. Biol. Chem.* **273**, 27927–27933
  35. Harper, J. F., Surowy, T. K., and Sussman, M. R. (1989) Molecular-cloning and sequence of cDNA-encoding the plasma-membrane proton pump ( $H^+$ -ATPase) of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 1234–1238
  36. Henrissat, B., Coutinho, P. M., and Davies, G. J. (2001) A census of carbohydrate-active enzymes in the genome of *Arabidopsis thaliana*. *Plant Mol. Biol.* **47**, 55–72
  37. Faik, A., Price, N. J., Raikhel, N. V., and Keegstra, K. (2002) An *Arabidopsis* gene encoding an alpha-xylosyltransferase involved in xyloglucan biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7797–7802
  38. Bouton, S., Leboeuf, E., Mouille, G., Leydecker, M. T., Talbotec, J., Granier, F., Lahaye, M., Hofte, H., and Truong, H. N. (2002) Quasimodo1 encodes a putative membrane-bound glycosyltransferase required for normal pectin synthesis and cell adhesion in *Arabidopsis*. *Plant Cell* **14**, 2577–2590
  39. Abejion, C., Yanagisawa, K., Mandon, E. C., Hausler, A., Moremen, K., Hirschberg, C. B., and Robbins, P. W. (1993) Guanosine diphosphatase is required for protein and sphingolipid glycosylation in the Golgi lumen of *Saccharomyces cerevisiae*. *J. Cell Biol.* **122**, 307–323
  40. Kushihiro, M., Nakano, T., Sato, K., Yamagishi, K., Asami, T., Nakano, A., Takatsuto, S., Fujioka, S., Ebizuka, Y., and Yoshida, S. (2001) Obtusifoliol 14 alpha-demethylase (CYP51) antisense *Arabidopsis* shows slow growth and long life. *Biochem. Biophys. Res. Commun.* **285**, 98–104
  41. Xu, X. J., Dietrich, C. R., Lessire, R., Nikolau, B. J., and Schnable, P. S. (2002) The endoplasmic reticulum-associated maize GL8 protein is a component of the acyl-coenzyme A elongase involved in the production of cuticular waxes. *Plant Physiol.* **128**, 924–934
  42. Harnikort, V., Prakash, K., Marcantonio, E., Colman, D. R., Rosenfeld, M. G., Adesnik, M., Sabatini, D. D., and Kreibich, G. (1987) Isolation and characterization of cDNA clones for rat ribophorin-I—Complete coding sequence and *in vitro* synthesis and insertion of the encoded product into endoplasmic-reticulum membranes. *J. Cell Biol.* **104**, 855–863
  43. Ro, D. K., Mah, N., Ellis, B. E., and Douglas, C. J. (2001) Functional characterization and subcellular localization of poplar (*Populus trichocarpa* x *Populus deltoides*) cinnamate 4-hydroxylase. *Plant Physiol.* **126**, 317–329
  44. Sevier, C. S., and Kaiser, C. A. (2002) Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell. Biol.* **3**, 836–847
  45. Hanada, K. (2003) Serine palmitoyltransferase, a key enzyme of sphingolipid metabolism. *Biochim. Biophys. Acta Mol. Cell. Biol. Lipids* **1632**, 16–30
  46. Michalecka, A. M., Svensson, A. S., Johansson, F. I., Agius, S. C., Johanson, U., Brennicke, A., Binder, S., and Rasmusson, A. G. (2003) *Arabidopsis* genes encoding mitochondrial type II NAD(P)H dehydrogenases have different evolutionary origin and show distinct responses to light. *Plant Physiol.* **133**, 642–652
  47. Schnell, D. J., Kessler, F., and Blobel, G. (1994) Isolation of components of the chloroplast protein import machinery. *Science* **266**, 1007–1012
  48. Hinnah, S. C., Hill, K., Wagner, R., Schlicher, T., and Soll, J. (1997) Reconstitution of a chloroplast protein import channel. *EMBO J.* **16**, 7351–7360