Comparison of Full Versus Partial Metabolic Labeling for Quantitative Proteomics Analysis in Arabidopsis thaliana*

Edward L. Huttlin‡§¶¶, Adrian D. Hegeman‡§📸, Amy C. Harms§‡‡, and Michael R. Sussman‡ §§

In recent years a variety of quantitative proteomics techniques have been developed, allowing characterization of changes in protein abundance in a variety of organisms under various biological conditions. Because it allows excellent control for error at all steps in sample preparation and analysis, full metabolic labeling using 15N has emerged as an important strategy for quantitative proteomics, having been applied in a variety of organisms from yeast to Arabidopsis and even rats. However, challenges associated with complete replacement of 14N with 15N can make its application in many complex eukaryotic systems impractical on a routine basis. Extending a concept proposed by Whitelegge et al. (Whitelegge, J. P., Katz, J. E., Pihakari, K. A., Hale, R., Aguilera, R., Gomez, S. M., Faull, K. F., Vavilin, D., and Vermaas, W. (2004) Subtle modification of isotope ratio proteomics; an integrated strategy for expression proteomics. Phytochemistry 65, 1507–1515), we investigate an alternative strategy for quantitative proteomics that relies upon the subtle changes in isotopic envelope shape that result from partial metabolic labeling to compare relative abundances of labeled and unlabeled peptides in complex mixtures. We present a novel algorithm for the automated quantitative analysis of partial incorporation samples via LC-MS. We then compare the performance of partial metabolic labeling with traditional full metabolic labeling for quantification of controlled mixtures of labeled and unlabeled Arabidopsis peptides. Finally we evaluate the performance of each technique for comparison of light-versus dark-grown Arabidopsis with respect to reproducibility and numbers of peptide and protein identifications under more realistic experimental conditions. Overall full metabolic labeling and partial metabolic labeling prove to be comparable with respect to dynamic range, accuracy, and reproducibility, although partial metabolic labeling consistently allows quantification of a higher percentage of peptide observations across the dynamic range. This difference is especially pronounced at extreme ratios. Ultimately both full metabolic labeling and partial metabolic labeling prove to be well suited for quantitative proteomics characterization. Molecular & Cellular Proteomics 6:860–881, 2007.

Over the past several years as the field of proteomics has matured, a great emphasis has been placed on the development of robust techniques for comparison of protein abundances among biological samples (1). Many of the most successful strategies developed to date have involved introduction of an isotope-coded tag that can be used to differentiate among the samples to be compared. Although a variety of approaches may be used to introduce isotopic labels, the fundamental strategy for each technique is the same: following introduction of either heavy or light forms of the isotopic tag, these samples may then be combined and analyzed simultaneously with each sample acting as an internal control for the other. By comparing the intensities of heavy and light labeled forms in the combined sample, the relative abundance of each peptide can be determined, and the relative abundance of each corresponding parent protein can be inferred.

In general, strategies for isotope-assisted quantitative proteomics can be separated into two groups, depending on whether the isotopic tag is incorporated in vitro during sample preparation or in vivo during growth of the model organisms. Each approach has its relative strengths and weaknesses. Some of the most widely applicable techniques for quantitative proteomics involve introduction of an isotopic tag in vitro via either chemical or enzymatic means. Although these approaches are highly versatile and may be readily applied for comparison of proteins from virtually any source, they are also particularly prone to effects of sample handling error because the relatively late introduction of the isotopic tag requires independent sample processing through highly variable steps such as protein extraction before the samples can be mixed. Common chemical isotopic labeling strategies include ICAT (2) and isobaric tag for relative and absolute quantification (iTRAQ) (3), which differentially label sulfhydryls (cysteine) and primary amines (lysine and N termini), respectively. The most common enzymatic means is introduction of an isotopic label through tryptic digestion of protein samples in either H2 16O or H2 18O prior to mixing (4).

In contrast, in vivo isotopic labeling involves the introduction of an isotopic tag into an organism through its food or medium. As the organism grows it consumes the label, naturally incorporating it into its proteins. To date, no obvious deleterious effects have been observed following isotopic labeling of any model organism with 15N or 13C even when high isotope abundance is achieved. However, introduction of an isotopic label in this way can be challenging for some
larger organisms and is most suited for use with small model organisms that can be quickly grown on a defined medium or tightly controlled diet. These techniques can be especially useful for experiments involving extensive sample preparation. Because they are labeled with either heavy or light tags during growth, both experimental and control samples may be combined immediately at harvest or sacrifice prior to all steps in sample preparation including tissue homogenization, protein extraction, protein digestion, and all kinds of fractionation in addition to mass spectrometric analysis. Thus in vivo isotopic labeling provides the ideal internal control for all steps in a proteomics experiment. Cells grown in culture are often specifically labeled with essential amino acids provided in their medium via a technique called stable isotope labeling by amino acids in cell culture (SILAC) (5). Alternatively in cases where use of a labeled amino acid is impractical, organisms can be labeled through growth on medium containing $^{15}$N or $^{13}$C in place of their natural abundance counterparts via full metabolic labeling.

Full metabolic labeling, most often with $^{15}$N, has been successfully used for quantitative proteomics characterization of numerous prokaryotic species as well as many eukaryotes including Saccharomyces cerevisiae (6), Caenorhabditis elegans (7), Drosophila melanogaster (7), Arabidopsis thaliana (8), and Rattus norvegicus (9). For a recent review describing the application of metabolic labeling with stable isotopes to a variety of experimental systems, see Beynon and Pratt (10). This review describes not only the use of static metabolic labeling for comparison of protein abundances as considered here but also details the use of stable isotopes to monitor protein turnover via changes in incorporation of selected isotopes over time.

Several variations on metabolic labeling have been published in recent years. In one case, researchers combined the use of both $^{13}$C and $^{15}$N labeling to allow simultaneous comparison of three biological samples while also providing elemental constraints to aid peptide identification (11). Additionally full $^{15}$N metabolic labeling has been applied for top-down proteomics (12) and has been adapted for use in conjunction with two-dimensional gel electrophoresis (13). Finally as an indication of its reliability, metabolic labeling has served as the standard by which other proposed quantitative proteomics techniques including DIGE (14) and spectral counting (15) have been evaluated. A key challenge for any quantitative proteomics experiment involving stable isotopes is the development of an effective automated system for determination of peptide ratios from LC-MS analyses. Because the mass difference between labeled and unlabeled peptides is dependent on the number of labeled atoms in the heavy peptide and thus varies with amino acid sequence, the automated analysis of data incorporating full metabolic labeling can be especially challenging. An algorithm for automated analysis of quantitative proteomics data involving stable isotope labeling has been published by MacCoss et al. (16) that addresses many of these challenges and specifically allows interpretation of $^{15}$N-labeled samples. The key features of this algorithm are detailed in Fig. 1 as adapted for our own metabolic labeling experiments.

The most fundamental challenge for full metabolic labeling concerns the efficient growth of an organism entirely on $^{15}$N. Although this strategy has been used in many organisms with no obvious detrimental effects on health, the difficulty associated with this labeling varies greatly from organism to organism. For most prokaryotes and some eukaryotes such as yeast, $^{15}$N metabolic labeling may be performed very efficiently and economically. The only labeled nitrogen sources necessary are inorganic salts that are relatively inexpensive. Furthermore their small size and rapid growth easily allow essentially complete (reagent-limited) $^{15}$N incorporation, generally in excess of 98%. Complex eukaryotes such as C. elegans and Drosophila may be labeled to completion indirectly through feeding on labeled Escherichia coli and yeast, respectively (7).

In contrast, labeling of larger eukaryotes can be considerably more difficult, requiring either unnatural growth conditions or considerable expense to achieve complete $^{15}$N labeling. For example, the model plant Arabidopsis has recently been labeled with $^{15}$N for quantitative proteomics to $>$98% incorporation (8). However, to achieve complete labeling the plants were grown in liquid culture, essentially submerged in medium. Because these growth conditions differ significantly from the natural growth conditions of the plant, they may limit to some extent the biological questions to which complete metabolic labeling may be applied in Arabidopsis. Similarly metabolic labeling has been achieved in mammals including rats by feeding a $^{15}$N-enriched diet. Relatively long growth times and large amounts of labeled food are required for labeling of mammals, thus leading to considerable expense. Even after extended periods, the extent of $^{15}$N incorporation varies greatly from tissue to tissue in rats (9); this can preclude use of metabolic labeling for study of some biological questions. These challenges would likely be far greater for labeling of other larger mammals. Clearly the requirement for complete replacement of a selected atom with its heavy isotope makes full metabolic labeling difficult to apply in many biological experiments.

Recently Whitelegge et al. (17) presented an alternative strategy to use isotopic labeling for quantitative proteomics. They observed that when the relative abundance of a selected heavy isotope such as $^{15}$N or $^{13}$C is increased slightly over natural abundance, the result is an isotopic envelope with distinct shape (see Fig. 2). Although the isotopic envelopes for both natural abundance and partially labeled forms of the same peptide are easily distinguished, they should both be compatible with existing MS/MS data acquisition and sequencing software if the change in $^{15}$N or $^{13}$C abundance is small enough. Their key observation was that when natural abundance and partially labeled forms of the same peptide are combined, the shape of the resulting composite isotopic envelope can be used to determine the relative amounts of each peptide form that are present in the mixed sample. Similar types of measurement approaches involving overlapping envelopes have been used in the analysis of biopolymerization via mass isotopomer distribution analysis (MIDA) (18). Thus, quantitative proteomics information can be extracted through partial rather than full metabolic labeling.
Because label would still be incorporated during growth, partial metabolic labeling provides the key benefit of full metabolic labeling: excellent internal control throughout all steps in sample preparation and analysis. Partial incorporation would be much less expensive to use in larger, more complex organisms and would be amenable to a wider variety of growth conditions due to its significantly reduced isotopic enrichment requirements. Furthermore, the dramatically reduced expense for enriched food would allow reasonable use of metabolic labeling for studies involving more statistically meaningful sample sizes. Finally, labeling could be more easily maintained over longer time periods potentially encompassing multiple generations. This could even allow effective stable isotope labeling of tissues with slow protein turnover rates in mammals and other large organisms.

Despite these potential benefits, a number of challenges currently prevent use of partial metabolic labeling in quantitative proteomics experiments. First, no approaches for automated interpretation of data from partial labeling experiments have been developed. More fundamentally, the effectiveness of partial metabolic labeling as a quantitative technique has not been investigated. Although full metabolic labeling involves comparison of two distinct isotopic envelopes, partial metabolic labeling would require the interpretation of much more subtle changes in the shape of a single composite envelope. Because it relies upon such subtle measurements, one might expect partial metabolic labeling to display less

---

**Fig. 1.** Full metabolic labeling quantification scheme. Quantifying relative abundance of natural abundance and ~100% 15N-labeled peptides involves several discrete processing steps that are accomplished sequentially for an entire dataset using Mathematica scripts. An example of a single peptide pair is given in this figure to demonstrate each of the following steps. 1, all quantified peak pairs must have an amino acid sequence determined by MS/MS and database searches. Sequencing accommodates MS/MS data collected from either or both the heavy and/or light peaks by searching the database twice using heavy or light amino acid mass definitions. 2, the number of nitrogen atoms may be calculated from the amino acid sequence yielding the spacing between the monoisotopic peaks of the peptide pair. 3, EICs are generated from MS data for both heavy and light monoisotopic peaks using a 0.25-Da mass range over a 5-min chromatographic window centered on the time of the MS/MS switching event that resulted in each heavy or light peptide identification. 4, the EICs are fitted to Gaussian distribution within ±1 of the mean m/z and within 10% of the monoisotopic mass. The ratio of heavy to light monoisotopic peaks is then calculated for each scan in which the peptide pair is visible, providing a single point. Assuming a good match in chromatographic peak shape for heavy and light EICs, these points fall in a line the slope of which is the ratio of heavy to light peptides. The correlation coefficient from the linear regression provides an indication of the quality of the quantification for the peptide in question.

**Fig. 2.** Comparison of natural abundance and 5–6% 15N-labeled peptide isotopic envelopes observed for the same peptide species observed in labeled and unlabeled samples. The identities of these isotopomer peptide species were confirmed by MS/MS (not shown).

Because label would still be incorporated during growth, partial metabolic labeling provides the key benefit of full metabolic labeling: excellent internal control throughout all steps in sample preparation and analysis. Partial incorporation would be much less expensive to use in larger, more complex organisms and would be amenable to a wider variety of growth conditions due to its significantly reduced isotopic enrichment requirements. Furthermore, the dramatically reduced expense for enriched food would allow reasonable use of metabolic labeling for studies involving more statistically meaningful sample sizes. Finally, labeling could be more easily maintained over longer time periods potentially encompassing multiple generations. This could even allow effective stable isotope labeling of tissues with slow protein turnover rates in mammals and other large organisms.

Despite these potential benefits, a number of challenges currently prevent use of partial metabolic labeling in quantitative proteomics experiments. First, no approaches for automated interpretation of data from partial labeling experiments have been developed. More fundamentally, the effectiveness of partial metabolic labeling as a quantitative technique has not been investigated. Although full metabolic labeling involves comparison of two distinct isotopic envelopes, partial metabolic labeling would require the interpretation of much more subtle changes in the shape of a single composite envelope. Because it relies upon such subtle measurements, one might expect partial metabolic labeling to display less...
precision and thus to provide a semiquantitative rather than
quantitative indication of changes in protein abundance. To
date this has not been investigated.

We present a novel algorithm for the automated interpreta-
tion of individual isotopic envelopes to extract quantitative
information via partial metabolic labeling. Furthermore we
compare this new approach with standard full 15N metabolic
labeling. First we analyze mixtures of labeled and unlabeled
Arabidopsis peptides using both techniques to assess the
consistency, dynamic range, and reproducibility of each tech-
nique under controlled conditions. Additionally we use both
full and partial metabolic labeling to compare protein expres-
sion in light- versus dark-grown Arabidopsis. Thus we evalu-
ate the performance of both techniques under conditions
more closely resembling those of a typical biological experi-
ment. Finally we explore the potential of partial metabolic
labeling for use in future quantitative proteomics experiments.

EXPERIMENTAL PROCEDURES

Materials

Unless specified, all chemicals were purchased from Sigma-
Aldrich.

Full and Partial Incorporation Control Experiments: Plant Growth
and Harvest

Arabidopsis seedlings (Columbia ecotype; Lehle) were grown in
liquid culture using media containing Murashige and Skoog salts and
MES (2.5 mM; pH 5.7) with 1% sucrose. Labeled media were prepared
to 6% 15N (“partial” labeling), or >98% 15N (“full” labeling) by substi-
tuting 15NH4NO3 and K15NO3 for natural abundance salts to the
appropriate proportions. Plants were grown at room temperature with
constant shading. All plants used for the control experiments were
grown under continuous light.

After 10 days of growth plant tissue was recovered, spun to remove
excess water, and weighed. Natural abundance, partial, and full 15N-
labeled plants were harvested and prepared separately to allow sub-
sequent combination of samples at a range of ratios. All subsequent
steps up to digestion were performed on ice or in a cold room
(4–8 °C). Each sample was combined with grinding buffer (250 mM
Tris-HCl, pH 7.6, 290 mM sucrose, 25 mM EDTA, 1 mM DTT, 1 mM
PMSF, 1 μg/ml pepstatin, 1 μg/ml E64, 100 μM 1,10-phenanthroline)
at a ratio of 3 ml/g of tissue. Tissue was homogenized using a mortar
and pestle followed by a Polytron (Brinkman). Following filtration
through four layers of Miracloth (Calbiochem) the homogenate was
processed identically as follows. Insoluble material was pelleted by centrifugation in a microcen-
trifuge for 2 min at 500 × g to pellet debris. Samples were then
spun for 5 min at 1500 × g to pellet organellar material. Microsomal
proteins were isolated from soluble proteins by centrifugation for 120
min at −16,000 × g.

Digestion of Combined Light- and Dark-treated Samples

Soluble protein fractions were processed and digested as de-
scribed for the preparation of control mixtures above. Membranous
protein fractions derived from the 1500 × g (organellar) and
100,000 × g (microsomal) pellets were prepared identically as fol-

tows. Insoluble material was pelleted by centrifugation in a microcen-
trifuge at 16,000 × g for 90 min. The supernatant was removed, and
the pellet was resuspended in 50 mM ammonium bicarbonate and
homogenized using a Potter-Elvehjem grinder. Protein concentrations were then measured using a BCA protein assay kit (Pierce), and 0.4
mg of each membrane homogenate was diluted to 2 mg/ml protein
concentration with 50 mM ammonium bicarbonate. The samples were
adjusted to 2 mM DTT, heated to 90 °C for 5 min, and allowed to cool
to below 50 °C prior to addition of 1 volume of MeOH. Proteolysis was
started by addition of 8 μg of sequencing grade trypsin (Promega) to
each 400-μl reaction containing 1 mg/ml protein, 50% MeOH (v/v), 1
mM DTT to a final ratio of trypsin to protein of 1:50. Proteolysis was
allowed to proceed overnight (14 h) at −22–23 °C with gentle rocking
and was terminated by addition of formic acid to 5% (v/v); the result-
ing peptides were purified by C18 solid phase extraction using
SPEC-PT-C18 units (Varian) using the manufacturer’s protocol.

Mass Spectrometric Analysis

After digestion and desalting, samples containing ~10 μg of pep-
tides were analyzed using a QTOF-2 mass spectrometer (Micromass)
and an HP 1100 HPLC instrument (Agilent). Peptides were separated
using home-packed fused silica columns (100 μm × 11 cm) contain-
ing Eclipse C18 resin (Agilent). Samples were separated via reversed
phase chromatography using buffers A (0.1% (v/v) formic acid in
water) and B (0.1% (v/v) formic acid in 100% acetonitrile) at a con-
stant flow rate of 500 nl/min. After loading each sample in 5% B,
peptides were eluted at 500 nl/min with the following gradient: 5–12% B
in 10 min, 12–50% B over 105 min, 50–60% B over 5 min, 60–100% B in 5 min, hold at 100% B for 5 min, and return to 5% B
over 5 min. MS and MS/MS spectra were collected in data-dependent
mode with one 5.5-s sequencing attempt following every MS spect-
rum to maintain regular MS sampling over the chromatographic time
frame to provide information needed for quantification. Precursor ions

Comparison of Full and Partial Metabolic Labeling

Plants were grown in liquid culture as described above. “Light-
grown” plants were grown under continuous light, whereas “dark-
grown” plants were kept in complete darkness wrapped in aluminum
foil throughout growth. After 10 days growth plants were harvested.
Natural abundance and either full or partial 15N-labeled plants were
combined at harvest to known ratios by dry mass and processed
together. The partial incorporation samples were homogenized as
described above to produce soluble, organellar, and microsomal
protein fractions. Due to smaller amounts of material, the full incor-
poration samples were prepared using a slightly different protocol.
Following tissue homogenization in grinding buffer with a mortar and
pestle, the full incorporation protein samples were spun in a micro-
centrifuge for 2 min at 500 × g to pellet debris. Samples were then
spun for 5 min at 1500 × g to pellet organellar material. Microsomal
proteins were isolated from soluble proteins by centrifugation for 120
min at −16,000 × g.

Mass Spectrometric Analysis

After digestion and desalting, samples containing ~10 μg of pep-
tides were analyzed using a QTOF-2 mass spectrometer (Micromass)
and an HP 1100 HPLC instrument (Agilent). Peptides were separated
using home-packed fused silica columns (100 μm × 11 cm) contain-
ing Eclipse C18 resin (Agilent). Samples were separated via reversed
phase chromatography using buffers A (0.1% (v/v) formic acid in
water) and B (0.1% (v/v) formic acid in 100% acetonitrile) at a con-
stant flow rate of 500 nl/min. After loading each sample in 5% B,
peptides were eluted at 500 nl/min with the following gradient: 5–12% B
in 10 min, 12–50% B over 105 min, 50–60% B over 5 min, 60–100% B in 5 min, hold at 100% B for 5 min, and return to 5% B
over 5 min. MS and MS/MS spectra were collected in data-dependent
mode with one 5.5-s sequencing attempt following every MS spect-
rum to maintain regular MS sampling over the chromatographic time
frame to provide information needed for quantification. Precursor ions

Molecular & Cellular Proteomics 6.5 863
Comparison of Full and Partial Metabolic Labeling

were selected for MS/MS fragmentation within a ±3 m/z window. Dynamic exclusion was applied for peaks within 1.2 Da over a 120-s period following each sequencing attempt.

Protein Identification

Peak lists were generated from each LC-MS analysis using Protein Lynx Global Server 2.1.5 (Micromass). A seven-point Savitzky-Golay smooth was applied twice to each MS spectrum followed by background subtraction using a 35% threshold with a first order polynomial. MS product spectra were similarly smoothed twice via the seven-point Savitzky-Golay method, although the adaptive algorithm was used for background subtraction. No deisotoping was performed.

Peptides were identified using Mascot 2.0 (19) with the following search parameters: tryptic digestion with up to two missed cleavages, variable methionine oxidation, and variable N-terminal acetylation of proteins. Although Mascot searches of full 15N incorporation data were performed using a 0.2-Da mass tolerance for both MS and MS/MS, partial incorporation data used a 1.2-Da mass tolerance for both MS and MS/MS to account for occasional problems our peak list generation software encountered interpreting unusual isotopic envelopes. All partial incorporation identifications were subsequently filtered to remove any peptides for which the mass error was not within 0.2 Da of either the monoisotopic peak or first isotope for the predicted envelope. Searches of full incorporation sample data were performed twice, first using natural abundance mass definitions and second using masses calculated for 100% 15N. Mascot .DAT files were parsed with Java and Perl scripts that used tools provided in the msParser 1.22 Toolkit (Matrix Science).

All Mascot searches were performed against a composite database containing the forward and reversed sequences for all predicted proteins in the Arabidopsis genome (Version 4, The Institute for Genomic Research, www.tigr.org, released June 2003) as well as the sequences of common contaminant proteins including porcine trypsin and several human keratins (final database size, 57,973 entries). To ensure comparable and defined levels of confidence in peptide identifications for both partial and full incorporation datasets, a reverse database strategy was used to determine 1% false positive thresholds for peptides after separation by charge state. For peptides identified from the full incorporation samples, minimum Mascot scores of 44, 24, 13, and 10 were required for singly, doubly, triply, and quadruply charged peptides, respectively. Minimum thresholds for partial incorporation samples were 58, 44, 38, and 37 for singly, doubly, triply and quadruply charged peptides, respectively. The large difference in score thresholds between the partial and full incorporation datasets is due to the wider MS mass tolerances used for partial incorporation searches. When the size of each dataset and numbers of associated reversed database hits are considered, the expected false positive rates (average ± S.D.) are 1.42 ± 0.3% for full incorporation and 1.33 ± 0.3% for partial incorporation (20). Because the expected uncertainty in false positive rate is relatively small and similar between both datasets, the peptide identifications they contain can be compared.

Calculation of 15N Incorporation

Full Metabolic Labeling—The mean level of 15N incorporation achieved in the fully labeled plants was calculated based on the incorporations observed for those peptides identified by Mascot in the 1:1 control mixture of labeled and unlabeled Arabidopsis. Incorporations for each individual peptide were calculated in an automated fashion using an adaptation of an algorithm from MacCoss et al. (21). All calculations were performed via a series of scripts written in Mathematica 5.2 (Wolfram Research), although EIC extraction was performed with a program written in Visual C++ (Microsoft Visual Studio 6.0).

Quantitative Analysis via Full 15N Metabolic Labeling

Ratios describing the relative abundance of heavy and light forms of each peptide were calculated for all peptides matched to forward protein sequences with scores above the specified Mascot 1% false positive thresholds. First, extracted ion chromatograms (EICs) were generated corresponding to the monoisotopic peaks from both light and heavy forms of each peptide identified from the observed m/z for the sequenced peptide and its predicted nitrogen count. The EICs included values within ±0.125 Da of the predicted m/z and encompassed 2.5 min on either side of the MS/MS sequencing event from which that peptide was identified. The EIC corresponding to the monoisotopic peak from the sequenced peptide (either heavy or light) was then fitted to a Gaussian distribution to determine its mean and S.D. Only MS scans within 1 S.D. of the mean were included for peptide quantification.

Relative quantification of each peptide pair was performed via least squares regression analysis as in MacCoss et al. (16). When the intensity of the heavy monoisotopic peak is plotted with respect to light monoisotopic peak intensity all points should fall on a line whose slope indicates the ratio of heavy to light monoisotopic peak intensities and whose intercept is related to the relative noise between heavy and light m/z values. The slope and intercept for each peptide was determined via linear least squares regression with the correlation coefficient providing an indication of the quality of the fit. Ratios reflecting the relative abundance of heavy and light forms of each peptide were then derived from these slopes through application of a correction factor to account for differences in the proportion of each isotopic envelope represented by the heavy and light monoisotopic peaks. These ratios were only accepted if the correlation coefficient exceeded 0.8. These calculations were predominantly performed using scripts written in Mathematica 5.2 (Wolfram Research), although EIC extraction was performed with a program written in Visual C++ (Microsoft Visual Studio 6.0).

Quantitative Analysis via Partial 15N Metabolic Labeling

Ratios indicating the relative contributions of light and partially labeled peptides to each composite envelope were derived via a novel algorithm that incorporates elements of both the algorithm for

1 The abbreviation used is: EIC, extracted ion chromatogram.
calculating $^{15}$N incorporation and the algorithm for determining ratios in full incorporation samples (described above). Ratios were calculated for all peptides identified by Mascot that were matched with forward protein sequences and had scores above the minimum 1% false positive thresholds as described earlier.

**Description of the Observed Composite Envelope**—First extracted ion chromatograms were produced for the monoisotopic peak and the first eight isotopes of each observed composite envelope. Each chromatogram included values with ±0.125 Da of the predicted $m/z$ and included all scans within 2.5 min of the sequencing event from which that peptide was identified. The monoisotopic EIC was then fitted to a Gaussian distribution to identify the mean and S.D. associated with the chromatographic peak. Only those points within 1 S.D. of the mean were considered in further calculations. In cases where the monoisotopic distribution failed to fit a Gaussian distribution, only points within ±1.0 min of the observed MS/MS sequencing event were used for the analysis.

Next the ratio of each peak within the isotopic envelope was compared with the monoisotopic peak using linear regression in a manner analogous to the approach used to compare heavy and light monoisotopic peaks within paired envelopes in the full incorporation case. These regressions provide slopes that indicate the height of each isotope with respect to the monoisotopic peak. A minimum correlation of 0.7 was required to accept a ratio for each peak; peaks not meeting this restriction were ignored for further analysis. The result is a description of all peaks in the envelope normalized with respect to the monoisotopic peak.

**Calculation of Predicted Composite Envelopes**—Because each peptide has been sequenced, its elemental formula is known. Furthermore having measured the $^{15}$N incorporations achieved for both labeled and unlabeled samples, the shape of the isotopic envelope that would be produced by the labeled or unlabeled peptides individually can be calculated via binomial expansion (23). These predicted light and heavy spectra were combined at a range of ratios (1:99, 2:98, 3:97, ..., 97:3, 98:2, 99:1) covering 4 orders of magnitude to form a library of composite spectra whose ratio (heavy:light) was known. Distributions representing only natural abundance and only partially labeled peptides were included as well for cases when the contribution of one sample or the other was minimal. Each of these predicted spectra was normalized with respect to the monoisotopic peak to match the description of the observed spectrum obtained previously.

**Quantification by Identifying the Best Match**—To determine the relative contributions of heavy to light envelopes, each observed composite spectrum was compared with the entire library of normalized predicted spectra using linear regression. The predicted spectrum that best matched the observed spectrum as judged by correlation was used to estimate the ratio of heavy to light peptides in the original sample. A minimum correlation of 0.8 was required to accept each ratio. This approach for finding the value of some variable based on comparison of an observed spectrum to a library of predicted spectra encompassing a range of values for the variable in question is similar to the approach used to estimate $^{15}$N incorporation as discussed previously. Calculations were performed via a series of scripts written in Mathematica 5.2 (Wolfram Research), and EIC extraction was performed using a program written in Visual C++ (Microsoft Visual Studio 6.0).

**Light and Dark Experiment: Identification and Quantification of Proteins**

The comparison of full and partial labeling quantification was performed on a simple highly perturbed biological system (plant grown in 24 h light versus dark) to evaluate the methodologies at the intact protein level. A reciprocal labeling scheme was used for both full and partial labeling approaches to control for label-specific effects due to either differences in isotope or reagent quality. Mascot database searches were performed, described under “Protein Identification,” prior to quantification as both techniques require defined molecular formulas for each peptide quantified. Relative abundance ratios (labeled over unlabeled) were derived as described above for each MS/MS event for two reciprocally labeled datasets for both full and partial labeling techniques (four datasets total). Using a combination of Mathematica scripts and Excel spreadsheet manipulation, the following data manipulation steps were accomplished. First, reverse database and keratin peptides were removed from each dataset, and the remaining peptide sequences were filtered to include only amino acid sequences unique to single protein species. Second, ratios with correlation coefficients less than 0.8 were excluded from the analysis, and values outside of the apparent linear range of the technique were sorted into bins at each extreme and designated greater than 10-fold and less than 0.1-fold. Then the ratio measurements for each dataset were normalized to the median of the remaining ratios in each dataset to correct for deviations from normal (1.0) in the labeled to unlabeled plant material mixing ratio. The peptide ratios were then combined by calculating the mean value for all of the filtered and normalized peptide ratio measurements for each protein species. Positive identification of proteins from extracts of plants grown with and without illumination required a minimum of two unique peptide sequences with acceptable Mascot scores as discussed above (using a 1% false positive rate score cutoff for each charge state via decoy database strategy).

**RESULTS AND DISCUSSION**

**Determination of Optimal $^{15}$N Incorporation for Partial Metabolic Labeling**—For partial metabolic labeling to be used as a quantitative technique, the isotopic envelopes of the labeled peptides must be sufficiently distinct from their unlabeled counterparts that when combined their relative contributions to the resulting composite envelope may be readily distinguished. Yet the perturbation in isotopic envelope shape must be subtle enough that existing data-dependent MS/MS acquisition and peptide identification software may still be used. Thus for optimal performance, one must select conditions for isotopic enrichment that will provide sufficiently distinct envelope shapes between labeled and unlabeled peptides while not interfering significantly with peptide sequencing and identification.

Previously Whitelegge et al. (17) suggested that raising the $^{13}$C incorporation to around 0.75–1.5% above natural abundance would be useful for relative quantification of tryptic peptides based on subtle differences in envelope shape. However, controlled metabolic labeling of intact plants such as *Arabidopsis* with $^{13}$C may be problematic given their ability to obtain carbon from the air. Furthermore having demonstrated an easy and efficient approach for complete $^{15}$N labeling of *Arabidopsis*, we opted to apply the concept to $^{15}$N labeling instead. The extent to which a given change in isotopic enrichment alters the envelope shape for a peptide depends on how many atoms of that element it contains. Because nitrogen is less abundant in peptides than carbon, a larger change in isotopic enrichment is required to achieve a similar modification of envelope shape. Our working assump-
Comparison of Full and Partial Metabolic Labeling

tion is that the optimal $^{15}$N incorporation for quantification by partial metabolic labeling will be that incorporation at which the height of the $M + 1$ ion is greatest. At this incorporation, labeled and unlabeled forms of each peptide should have distinctly different isotope envelope shapes, yet both labeled and unlabeled forms should be compatible with existing peak extraction and peptide sequencing software. Using calculus we can then calculate exactly what this incorporation should be.

The height of the $M + 1$ peak in an isotopic distribution depends predominantly on the numbers of carbons and nitrogens in a peptide. Because the $^{13}$C contribution will be fixed in our labeling regime, we need only consider the nitrogen contribution to the second isotope resulting from presence of a single $^{15}$N, which can be described by Equation 1 in terms of the total number of nitrogens in the peptide ($N$) and the $^{15}$N incorporation ($I$), derived from the binomial distribution.

$$P(I^{15}N) = \binom{N}{1/N - 1} (1 - I)^{N-1} I$$

(Eq. 1)

To find the $^{15}$N incorporation that maximizes this function, we take the derivative of the expression with respect to $^{15}$N incorporation (Equation 2). Then we set the derivative equal to 0 and solve for $I$ (Equation 3).

$$\frac{dP}{dI} = N(1 - I)^{N-1} - (N - 1)N(1 - I)^{N-2}I$$

(Eq. 2)

$$I = \frac{1}{N} \text{ when } \frac{dP}{dI} = 0$$

(Eq. 3)

It can be proven that this critical point is in fact a maximum by evaluating the second derivative for $I = 1/N$ (not shown).

Now we need an estimate of the average number of nitrogens in the peptides we are likely to identify. Using data from an extensive survey of the Arabidopsis proteome we completed recently (8), we determined that the median nitrogen count among identifiable peptides in a tryptic digest of Arabidopsis is 18 for an optimal $^{15}$N incorporation of 5.6%.

Given the observed variability of nitrogen counts among tryptic peptides, a $^{15}$N incorporation between 5 and 6% should be appropriate. Although a range of $^{15}$N incorporation values may be suitable for this kind of experiment, it is essential that the observed $^{15}$N incorporation is consistent within a particular sample for all labeled proteins. Thus, labeling must be continued long enough to achieve a steady-state, equilibrium level of incorporation if quantitative measurements are to be made.

Plotted in Fig. 3 are the expected isotopic distributions for a typical tryptic peptide after combination of its natural abundance and ~6% $^{15}$N-labeled envelopes at various proportions. After normalization of the distributions to the monoisotopic peaks, large differences in the $M + 1$, $M + 2$, etc. peaks become obvious. These significant changes suggest that consideration of the entire peptide envelope will be important for distinguishing varying combinations of light and heavy envelopes.

A Novel Algorithm for Automated Peptide Quantification via Partial Metabolic Labeling—In their initial investigation of partial metabolic labeling as an approach for quantitative proteomics, Whitelegge et al. (17) proposed that subtle modifications in isotope envelope shape could be used for relative quantification of peptides in complex mixtures. The key challenge for application of this technique on a proteomics scale is development of an automated method for determination of the relative contributions of labeled and unlabeled forms of each peptide to each observed composite envelope. We have developed a novel algorithm incorporating elements of the correlative algorithm for analysis of full metabolic labeling developed by MacCoss et al. (16) as well as aspects of an algorithm used previously for automated calculation of isotope enrichment (21). This algorithm, diagrammed in Fig. 4, converts the observed envelope from each sequenced peptide into a list of ratios normalized to the monoisotopic peak. This normalized representation of the observed envelope is then compared with a library of predicted normalized envelopes representing various combinations of labeled and unlabeled peptides to find the optimum match. A more detailed description follows, assuming that average $^{15}$N incorporations for labeled and unlabeled samples have been determined previously.

Data processing for partial metabolic labeling using our algorithm proceeds in several steps. First, peptides are identified based on their fragmentation patterns using Mascot or

Fig. 3. Simulations of the added isotopic distributions arising from various mixtures of natural abundance and 6% $^{15}$N-labeled peptide. Five sets of peak intensities were generated including the calculated isotopic distributions for natural abundance and 6% $^{15}$N-labeled tryptic peptide (FAIPVFADLAK) and three mixtures of the two distributions. The “M” indicates the monoisotopic peak of the envelopes to which all five sets of composite envelopes were normalized.
Partial metabolic labeling quantification scheme. Quantifying relative abundance of natural abundance and 5–6% $^{15}$N-labeled peptides is accomplished by analyzing the shape of their combined overlapping envelopes. The process is slightly more complicated than that for the analysis of samples of near full enrichment but also requires several discrete processing steps accomplished sequentially for an entire dataset using Mathematica scripts. 1, amino acid sequences must be determined from MS/MS data to allow calculation of envelope libraries. 2, EICs are generated for the monoisotopic M (a), M + 1 (b), ... M + 8 (i) isotopic peaks over a 5-min chromatographic window centered on the MS/MS switching event. 3, the monoisotopic EIC (a) is fitted to a Gaussian distribution, and all of the time points within ±1 σ of the mean from each of the M + 1 (b), M + 2 (c) through M + 8 (i) EICs are plotted against their corresponding monoisotopic time points (a). Linear regressions of these plots (b versus a, c versus a, d versus a, etc.) provide slopes that correspond to the intensities of the M + 1 (b) through M + 8 (i) peaks relative to the monoisotopic peak intensity. These relative intensity values are used for subsequent steps only if the correlation coefficients ($r^2$) are greater than 0.7. The result is a description of all peaks in the envelope normalized to the monoisotopic peak. 4, the natural abundance and $^{15}$N-labeled isotopic distributions are calculated using the elemental composition for the peptide from step 1. These are mixed together in single percentage increments starting with one pure envelope and finishing with the other to generate a library of mixed isotopic envelopes. 5, the best matching mixed envelope and its corresponding relative abundance ratio are selected from the library by linear correlation of the monoisotopic peak-normalized peak heights b through h to the same calculated values from each mixed envelope in the library. A correlation coefficient greater than 0.8 is required for a successful match. In this example, only peaks b through f gave acceptable correlation coefficients in step 3 and provided ample information to select the 45% light to 55% heavy mixture with the highest correlation coefficient.
other similar search engines (Fig. 4, part 1). Once identified, we automatically know their elemental composition. This information will be useful for calculating what the composite envelope would look like for various combinations of labeled and unlabeled peptides.

Next the observed composite distribution for each peptide is defined. Extracted ion chromatograms are generated corresponding to all isotopes within the composite envelope for each peptide (Fig. 4, part 2). Each of these EICs is then compared with the monoisotopic EIC via linear regression, similar to MacCoss et al. (16) (Fig. 4, part 3). These regression analyses return lines whose slopes reflect the intensities of each peak in the isotopic envelope normalized with respect to the monoisotopic peak. Furthermore the correlation coefficients for each regression provide an indication of the quality of the fit and allow exclusion of particular isotopes that are disrupted by noise.

Because we know the elemental composition of each peptide and we know the $^{15}N$ incorporations for both labeled and unlabeled samples, we can predict the shapes of each of these envelopes through binomial expansion. We can then combine these distributions in a range of different proportions, creating a library of predicted envelopes of known heavy to light ratios (Fig. 4, part 4). Note that these composite distributions have a distinct shape from the distributions that result when the $^{15}N$ is uniformly distributed throughout all forms of the same peptide in a sample. In practice we included ratios across 4 orders of magnitude in our library for each peptide, ranging from 100:1 to 1:100. We also included the envelopes expected from the labeled or unlabeled peptides alone to account for cases where the peptide was only detectable in one sample or the other. All isotopic peaks within each distribution in the library are then normalized with respect to the monoisotopic peak for ease of comparison with the observed envelope described earlier.

Finally the observed distribution is compared with each composite envelope in the library via linear regression (Fig. 4, part 5). The best match within the library of predicted envelopes is then selected based on the resulting correlation coefficient and is used as an estimate of the ratio between labeled and unlabeled peptides in the original sample. This comparison of an observed isotopic envelope with a library of predicted envelopes is conceptually similar to an approach used previously for determination of $^{15}N$ incorporation in metabolically labeled peptides (21).

In practice a strong correlation is generally seen for the best match between the theoretical and experimental distributions, and these correlations fall quickly for the other theoretical spectra. However, the sharpness of this decline varies depending on the relative contributions of the heavy and light envelopes to the composite distribution. Plotted in Supplemental Fig. 1 are the distributions of correlation coefficients resulting from quantification of a selected peptide at a variety of ratios of heavy to light forms. When the mixing ratio is close to 1:1, the correlation distributions tend to be smooth with a peak at the best matching theoretical spectrum. However, as the ratio approaches all unlabeled or all labeled forms, the distributions of correlation coefficients then assume a sigmoidal shape with the correlation approaching 1.0 asymptotically as the mixing ratio approaches 0 or infinity, respectively.

Although computationally quite involved, this approach has some positive features. First because the observed isotopic envelopes are defined via linear regression, they are relatively tolerant of noise. Additionally because several peaks within each isotopic envelope are used for quantification, our ability to determine relative contributions of labeled and unlabeled peptides should be maximized.

Comparison of Full Versus Partial Metabolic Labeling—Although both full metabolic labeling and partial metabolic labeling are intended to provide a relative quantitative comparison of two biological samples, they differ in a number of key respects. Typical MS spectra from combined labeled and unlabeled forms of a selected peptide are plotted in Fig. 5, part A, either using partial metabolic labeling (top) or full metabolic labeling (bottom). Most fundamentally, whereas full metabolic labeling involves the comparison of two completely separate isotopic envelopes for quantification of each peptide, partial incorporation involves interpretation of the shape of a single composite envelope. As a result spectral complexity is dramatically reduced when partial metabolic labeling is used in the sense that each peptide is represented by only a single envelope rather than two. This is especially clearly illustrated in Fig. 5, part B, where MS spectra from analysis of similar biological samples are compared with either partial or full metabolic labeling. Note that envelopes from at least four chemical species are clearly visible in the partial incorporation sample. As expected, twice as many envelopes are visible in the full incorporation example with brackets to indicate labeled and unlabeled forms of each peptide. Note that peptide iv is almost completely obscured.

This difference in spectral complexity has important implications for peptide and protein identification. During a typical data-dependent MS/MS experiment the mass spectrometer collects MS/MS data on a series of peaks observed in each MS precursor spectrum in decreasing order of intensity. When each peptide is represented by two distinct envelopes, as in the case of full metabolic labeling, the instrument may tend to sequence both peptides in each pair rather than proceeding to lower intensity species elsewhere in the same MS spectrum. This redundancy in peptide sequencing could ultimately reduce the numbers of unique peptides and proteins identified in a given run, especially for complex samples. Because partial metabolic labeling results in a single envelope per peptide, it should not have this problem. Although this issue would be expected to affect mass spectrometers of all types, its effects are likely to be most pronounced on instruments with relatively slow duty cycles.

The use of full versus partial metabolic labeling also has
implications for peptide database searching using Mascot or other similar search engines. First whereas full incorporation requires two independent searches against either $^{14}$N-amino acid mass definitions or $^{15}$N-amino acid mass definitions, partial incorporation requires only a single search against $^{14}$N-amino acid mass definitions. This eliminates the possibility for duplicate assignments of particular MS/MS spectra and avoids the subtle differences in confidence of peptide identifications that have been documented between natural abundance and fully $^{15}$N-labeled samples (8). Implications for peak extraction prior to peptide database searching also exist for these labeling regimes. In our experience full metabolic labeling has generally been compatible with peak extraction algorithms designed for natural abundance samples as long as complete $^{15}$N labeling is achieved. However, pilot experiments indicated occasional problems with monoisotopic peak identification in the partial incorporation samples. These were addressed after the fact using wider mass tolerances for database searching with subsequent filtration steps to find misidentified peaks.

Partial and full metabolic labeling may also be expected to differ with respect to quantification. First, because full metabolic labeling involves the comparison of two independent envelopes whereas partial metabolic labeling requires the detection of subtle changes in shape for a single envelope, one might expect quantification via full metabolic labeling to be more accurate and to display a greater dynamic range. Second, whereas full metabolic labeling returns peptide ratios as continuous values, our current implementation for partial metabolic labeling returns discrete ratios. Depending on the spacing between ratios, this could introduce additional error in relative quantification. However, provided the spacing between composite spectra in the theoretical library considered for each peptide is small compared with the magnitude of other sources of error, this should not seriously compromise this approach. It is worth noting that statistical approaches must be adapted to accommodate the discrete data from partial metabolic labeling experiments. Finally these two approaches are likely to differ with respect to how they deal with extreme changes where a peptide is present in only labeled or unlabeled form. For full metabolic labeling these cases involve comparison of a defined chemical species with a signal that is often obscured by noise, resulting in poor quality quantification of this peptide produce two distinct isotopic envelopes in the case of full metabolic labeling, a single composite envelope results in partial metabolic labeling. Part B, MS spectra containing several peptides of similar m/z are plotted from both partial and full metabolic labeling. Both spectra are from 1:1 mixtures of labeled and unlabeled Arabidopsis peptides. Individual peptides are labeled (i–iv), and pairs of matching heavy and light envelopes are bracketed for total metabolic labeling. Note that use of partial metabolic labeling can significantly simplify MS spectra, allowing detection of peptides that could be obscured with total metabolic labeling (peptide iv).
Comparison of Full and Partial Metabolic Labeling

Mixtures of labeled (either >98% $^{15}$N or ~6% $^{15}$N) and unlabeled (natural abundance) Arabidopsis peptides were combined at a variety of ratios and analyzed via tandem mass spectrometry with quantitative characterization via either partial or full metabolic labeling. A single LC-MS/MS analysis was conducted for each mixture of labeled and unlabeled peptides combined at the ratios specified in the first column. All successful Mascot sequencing events were matched to peptides in the Arabidopsis protein sequence database with scores that exceeded 1% false positive thresholds established by searching against a concatenated forward-reverse protein sequence database (“No. Scans ID”). From this set of successful sequencing events a list of unique peptide sequences was derived, compressing multiple observations of the same peptide sequence within a given run into a single observation (“No. Peps ID”). For a peptide or MS/MS sequencing event to be considered identified and quantified, it had to return a peptide sequence with a Mascot score exceeding the appropriate 1% false positive threshold and had to return a ratio via either the full or partial incorporation technique with a minimum correlation of 0.8 for that particular measurement (“No. Peps Quant.”). Finally the percentage of identified peptides that were also successfully quantified is listed.

<table>
<thead>
<tr>
<th>Ratio (Heavy/Light)</th>
<th>Full Incorporation</th>
<th>Partial Incorporation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infinite 100</td>
<td>223</td>
<td>209</td>
</tr>
<tr>
<td>12</td>
<td>198</td>
<td>185</td>
</tr>
<tr>
<td>0.6</td>
<td>212</td>
<td>198</td>
</tr>
<tr>
<td>0.3</td>
<td>193</td>
<td>177</td>
</tr>
<tr>
<td>0.2</td>
<td>195</td>
<td>175</td>
</tr>
<tr>
<td>0.1</td>
<td>173</td>
<td>143</td>
</tr>
<tr>
<td>0.05</td>
<td>201</td>
<td>167</td>
</tr>
<tr>
<td>0.033</td>
<td>186</td>
<td>168</td>
</tr>
<tr>
<td>0.17</td>
<td>216</td>
<td>201</td>
</tr>
<tr>
<td>0.08</td>
<td>206</td>
<td>195</td>
</tr>
<tr>
<td>0.03</td>
<td>217</td>
<td>205</td>
</tr>
<tr>
<td>0.0</td>
<td>281</td>
<td>255</td>
</tr>
</tbody>
</table>

Quantitative Analysis of Controlled Peptide Mixtures: Overview—To evaluate the quantitative accuracy of both partial and full metabolic labeling across a wide dynamic range, Arabidopsis plants were grown in liquid culture containing either natural abundance, partially labeled, or fully labeled media. Each population of plants was then harvested, homogenized, and digested separately. The resulting peptides were then combined at known ratios based on total protein concentration covering a range from 100:1 to 1:100 (labeled: unlabeled), and the resulting mixed samples were characterized on a Q-TOF mass spectrometer. A summary of peptide identifications from these experiments is given in Table I, while the median ratios for all peptide mixtures are plotted in Fig. 6. A complete list of all peptides identified in these LC-MS analyses including observed ratios is provided in Supplemental Table 1.

Controlled Peptide Mixtures: Comparison of Peptide Identification and Quantification—Summarized in Table I are numbers of MS/MS acquisitions with assignable amino acid sequences, peptide identifications, and quantifications for both partial and full metabolic labeling following analysis of known mixtures of Arabidopsis peptides. Looking first at the number of successful MS/MS sequencing events, a couple of trends are apparent. First full metabolic labeling seems to allow more successful sequencing events across the entire range of pep-
Comparison of Full and Partial Metabolic Labeling

Overview of Quantification Accuracy: Median Ratios—Having considered the total numbers of peptides identified and quantified by each technique, we now need to consider the accuracy of quantification. As an initial examination of accuracy, we consider the median ratios observed for all peptides in each peptide mixture. These values are plotted in Fig. 6. Upon examination it appears that both approaches are returning median mixing ratios that roughly correspond to the intended mixing ratios. Both approaches generally do well for ratios near 1:1, although full incorporation may perform slightly better overall. Although partial incorporation does quite well for ratios near 1:1, it starts to overestimate ratios at more extreme values. In contrast, the full incorporation sample may underestimate large changes.

Based on these results it appears that both partial and full metabolic labeling allow reasonable quantification within the range between 12:1 and 1:12 (labeled:unlabeled). Outside this range both approaches provide values that are of more qualitative than quantitative value. Although these observations would still be very useful in a biological sense, providing evidence for large changes in protein abundance, they are of less use as quantitative measurements. Thus, we will limit our further quantitative evaluation of each technique to those measurements falling within the range between 12:1 and 1:12.

It is worth noting that this quantitative range is consistent with the distributions of correlation coefficients produced via the partial incorporation approach when an observed distribution is compared with a library of predicted spectra at varying ratios of labeled to unlabeled forms (see Supplemental Fig. 1). Between 12:1 and 1:12 these distributions reach smooth peaks at the appropriate ratio. However, outside this range these distributions begin to level off, eventually approaching correlations of 1.0 asymptotically as the mixing ratio approaches either 0 or infinity. This is consistent with our observation that partial metabolic labeling provides relatively consistent ratios near 1:1 but that these values are of more qualitative than quantitative value at extreme ratios.

Evaluation of Quantification for Individual Peptides—The preceding quantitative comparison of full versus partial metabolic labeling has been based upon median ratio measurements that combine multiple observations of many different peptides to provide a general indication of the performance of each algorithm across multiple samples. However, we are not so much interested in the average performance of each algorithm overall but rather in the consistency of measurements from each algorithm across multiple measurements of individual peptides. We anticipate significant variation between the labeled and unlabeled peptide populations mixed for these experiments due to separate handling through growth, harvest, protein extraction, and digestion. For this reason the best way to characterize the precision of each technique considers ratios obtained for individual peptides as this is independent of variation in sample preparation and only reflects errors in the preparation of the controlled mixtures. In

Several interesting trends emerge when we consider total numbers of peptides that were identified and quantified by each technique. First numbers of peptides and scans resulting in successful quantification show a variable trend across a range of peptide ratios. In general, the greatest percentage of peptides is successfully quantified via full metabolic labeling at ratios close to 1:1. As the ratios become more extreme, the proportion of successfully quantified peptides drops dramatically. In contrast, the percentage of peptides that are successfully quantified by partial incorporation does not vary as the peptide ratio is changed and remains well above 90% across the entire dynamic range. Even near 1:1 when full metabolic labeling is at its best, partial metabolic labeling still outperforms it by more than 10%. Ultimately although full metabolic labeling produces more peptide identifications, partial incorporation performs better overall by successfully identifying and quantifying a larger number of peptides across the entire dynamic range.

![Normalized Median Ratios](image)

**Fig. 6.** Normalized median ratios for known mixtures of Arabidopsis peptides. The median ratios observed for each mixture of Arabidopsis peptides via full metabolic labeling are represented by gray boxes, while median ratios obtained by partial metabolic labeling for similar mixtures are plotted as white boxes. The target ratios are represented by black bars. The median ratios were normalized to the mixing ratio observed for the 1:1 mixture for both full and partially labeled samples. All ratios are in the form labeled:unlabeled.
this way one can evaluate the distribution of error associated with each approach across the full dynamic range.

We can use regression analysis on a peptide by peptide basis to evaluate the performance of both partial and full metabolic labeling with respect to accuracy and consistency of quantification. Considering the two techniques separately, we plot the observed ratios for each peptide versus the ratio at which the labeled and unlabeled samples were combined. If measurements for that particular peptide are consistent across multiple ratios, these points will fall on a line whose slope is the ratio of labeled to unlabeled peptide in the 1:1 mixture. The resulting correlation coefficients will provide an indication of how consistent individual peptide measurements are across varying ratios, whereas the residuals observed will provide an indication of any deviation in calculated ratios across the dynamic range associated with each technique.

Example plots of observed versus predicted ratios obtained via either partial or full metabolic labeling are given for selected peptides in Fig. 7, part A. These plots illustrate the very clear linear response of both full and partial metabolic labeling across a range of peptide ratios between 12:1 and 1:12 (labeled:unlabeled). As expected, the slopes observed for each peptide vary significantly. The slope deviations are most likely due to differences in sample handling. When the standard peptide mixtures were prepared, the labeled and unlabeled peptides were processed separately until immediately prior to MS analysis. This allowed easy generation of consistent mixtures of peptides across a range of ratios. However, differences in any prior step in sample preparation including protein extraction, digestion, or desalting could lead to peptide- or protein-specific variation between labeled and unlabeled samples. Although these factors may increase the variability when median peptide measurements at each mixing ratio are considered (as in Fig. 6), we may still judge consistency of measurements for specific peptides across multiple mixing ratios. Because the same labeled and unlabeled samples were used to prepare all control peptide mixtures, differences in ratio from sample to sample for each peptide should still be proportional to the anticipated change in ratio. The slope of the resulting regression line simply reflects the variation between the original samples for each peptide prior to mixing.

By using linear regression to analyze the results for each peptide individually, we can evaluate the performance of these algorithms independently of any biases in the original labeled and unlabeled peptide populations that were used.

First we can examine the distributions of correlation coefficients to evaluate the consistency of measurements across the entire dataset on the basis of individual peptides. To accomplish this, a regression analysis is performed for all peptides observed and quantified in more than three mixtures via either partial or full metabolic labeling. Fig. 7, part B, presents a histogram of correlation coefficients obtained via either partial or full metabolic labeling. Quite encouragingly, the vast majority of correlation coefficients are above 0.9 for both techniques, and nearly all correlation coefficients are above 0.8. Only a few peptides show correlations less than 0.8 for either technique, although there are more peptides with poor correlations in the partial incorporation dataset. This suggests very good consistency in general for our measure-
These same data can also be plotted differently to allow better comparison of the cross-sectional distribution of these errors (see Fig. 8, part B). Presented are the numbers of observations obtained from either technique, binned by the relative error observed. Viewed from this perspective there is also little difference in error distribution between partial and full metabolic labeling, and both techniques give measurements with small relative errors.

Although the previous two error assessments show little difference between the two techniques, when the distribution of error is plotted as a cumulative percentage with respect to the relative error observed, a subtle yet significant difference is revealed (see Fig. 8, part B, inset). The curve for partial incorporation consistently lags behind the curve for full incorporation. For full metabolic labeling we find that 95% of measurements give errors less than about 0.7. However, for partial incorporation we must go as high as a relative error of 1.3 to include 95% of all measurements. Thus it appears that partial metabolic labeling is associated with somewhat greater relative error than full incorporation, although the difference is subtle.

Conclusions: Quantitative Evaluations of Partial and Full Metabolic Labeling—Overall our characterization of full and partial metabolic labeling thus far indicates comparable performance. Perhaps surprisingly given our initial expectations, both techniques display comparable dynamic ranges, providing reasonably consistent measurements from 1:12 to 1:1 (heavy:light). Beyond this range both techniques identify changes in peptide abundance, but these become more qualitative observations than quantitative measurements. Comparison of two distinct envelopes in full incorporation does not seem to provide a significant advantage over partial incorporation with respect to dynamic range where changes in a single isotopic envelope are measured. Within the range from 12:1 to 1:12 both techniques perform comparably with respect to consistency, although somewhat smaller error is associated with full incorporation. Finally although more peptide identifications were obtained with full metabolic labeling than with partial metabolic labeling, a greater proportion of peptides were then successfully quantified by partial incorporation. As a result, more peptides were successfully identified and quantified by partial metabolic labeling. As anticipated, these differences are most pronounced at relatively extreme ratios where numbers of peptide quantifications via full metabolic labeling decline rapidly, whereas numbers of peptide quantifications via partial metabolic labeling remain constant.

Although each approach has its advantages and disadvantages, neither is clearly superior as a result of this analysis.

Application of Partial and Full Metabolic Labeling to Compare Light- Versus Dark-grown Arabidopsis—Until now our comparison of partial and full metabolic labeling has focused on characterization of identical mixtures of peptides from Arabidopsis combined at known ratios of labeled to unlabeled forms for peptide identification and quantification. Although
this approach provides a good indication of the accuracy of peptide quantification across a wide range of ratios, it differs significantly from typical experimental designs that would be used in practice for quantitative proteomics analysis of biological samples. Thus we must also compare the performance of both partial and full metabolic labeling when used for the characterization of differences in a biological sample.

We have chosen a comparison of light- versus dark-grown Arabidopsis seedlings as a test for both partial and full metabolic labeling under more typical experimental conditions. This particular biological comparison is well suited for evaluation of these quantitative techniques for several reasons. First, metabolic labeling of Arabidopsis is easily controlled, allowing consistent labeling of plants to either 6 or >98% 15N incorporation. Additionally, the light versus dark growth conditions should lead to significant, reproducible changes in expression for many proteins over a wide dynamic range. Finally, the effects of light versus dark for growth of Arabidopsis are fairly well understood; thus any changes observed by either partial or full metabolic labeling may be validated based on our understanding of plant physiology.

Shown in Fig. 9 is our experimental design for characterization of light- versus dark-grown Arabidopsis via partial and full metabolic labeling. The illustrated reciprocal experimental design is important because it controls for any side effects of the labeling process including differences in preparations of labeled and unlabeled media as well as any isotope effects that, although unlikely, could be observed. We achieved

![Diagram of experimental design](image-url)
98.2% $^{15}$N enrichment for full incorporation and 5.2% $^{15}$N enrichment for partial incorporation as calculated using methods published previously (21).

Light- Versus Dark-grown Arabidopsis: Comparison of Peptide Identifications—A complete list of all peptide observations made in these experiments is provided in Supplemental Table 2. A summary of all proteins observed with average ratios is provided in Supplemental Table 3. Shown in Table II is a summary of the numbers of peptides identified and quantified in each replicate of this analysis. These numbers appear to confirm a number of trends observed with the control peptide mixtures discussed earlier (Table I). First a greater number of MS/MS spectra are assigned peptide identifications at the 1% false positive rate with full metabolic labeling than with partial metabolic labeling. This trend continues when redundant peptide identifications are removed and when only peptides that are uniquely assignable to a single protein are considered. Yet consistently a smaller proportion of peptides are assigned ratios via full metabolic labeling so that partial incorporation produces greater numbers of peptides that were identified and quantified. Ultimately although both techniques produce similar numbers of protein identifications, partial metabolic labeling allows identification and quantification of a greater number of proteins.

Interestingly we also see that more peptide and protein identifications are obtained by either full or partial incorporation when the light-treated sample is grown under natural abundance conditions. This is likely because growth in the light leads to a general up-regulation of protein expression; thus the most identifications are produced when a strong peptide signal is present in the natural abundance sample. This effect is more pronounced for partial metabolic labeling than for full metabolic labeling. Further optimization of peak list generation and database searching for metabolically labeled samples may be necessary to address this systematic bias, especially for analysis of samples in which there are large differences in protein abundance.

**Light- Versus Dark-grown Arabidopsis: a Quantitative Comparison—** Plotted in Fig. 10 are histograms showing the distributions of proteins based on their median-normalized ratios for each replicate measured via either partial or full metabolic labeling. All ratios plotted are in the form labeled:unlabeled; thus reciprocal experiments would be expected to give reciprocal plots. Notice first the general symmetry between datasets obtained via both techniques when the label is reversed. As mentioned earlier, growth in the light leads to a general up-regulation in protein expression, producing skewed distributions. These trends are also generally consistent between partially and fully labeled samples, although some differences are seen. First both datasets from partial metabolic labeling contain significantly more quantified proteins than were produced by full metabolic labeling. Thus a higher degree of symmetry is seen between partial metabolic labeling replicates due to the more extensively populated distributions. Most significantly, although all distributions show a general increase in protein expression with plant growth in the light,
far more proteins were identified by partial incorporation showing especially dramatic changes in expression. Proteins exhibiting these large changes were identified in far lower numbers via full metabolic labeling. As was anticipated, our algorithm for analysis of partial metabolic labeling is especially well suited to analysis of large changes in protein abundance.

Although the previous plots indicate that our replicates were generally consistent for both partial and full metabolic labeling, we must again consider how consistent these datasets are at the level of individual protein measurements to assess reproducibility. This can be assessed by examining those proteins that were identified and quantified in both replicates via either partial or full metabolic labeling. The products of the normalized ratios (labeled:unlabeled) that are observed for each protein via each technique should equal 1.0 if measurements are equivalent. Deviation from 1.0 would be a strong indicator of label-dependent bias due to either reagent quality differences or isotope effects. Systematic differences in spectrometric response factors between labeled and unlabeled species would also produce deviations from unity.

By plotting the distribution of these products for both partial and full metabolic labeling we can assess the reproducibility of both techniques. Histograms of these products are plotted in Fig. 11. Both show that the majority of proteins give products near 1.0 (equivalent to 0 on a log scale), although there is some variability for both labeling techniques. The range of products observed is fairly consistent for both techniques. Overall little difference in reproducibility is seen between full and partial metabolic labeling.

**Light- Versus Dark-grown Arabidopsis: Comparison of Individual Protein Measurements**—Proteins that could be identified and quantified in both replicates for both full and partial metabolic labeling experiments are given in Table III. For inclusion in this table each protein needed to be identified by at least two unique peptide sequences for each replicate, and at least one of those peptides was required to provide a ratio with a good correlation. Additionally we have only listed those proteins whose mean ratio changes fell within the 2 orders of magnitude linear range (labeled:unlabeled) for all replicates. Examining the proteins listed, most give fairly similar measurements across all four replicates. For example, isocitrate dehydrogenase and elongation factor 2 consistently show relatively little change in abundance, producing ratios near 1.0 in all cases. In general our observations of changing proteins are consistent as well, although in some cases the magnitudes of these changes vary somewhat from replicate to replicate. For example, the photosystem II oxygen-evolving complex 23 is consistently more abundant in light-grown plants across all replicates, although the magnitude of that change ranges somewhat, from 2- to 5-fold. Similarly we
Comparison of Full and Partial Metabolic Labeling

Table III

Comparison of ratio measurements from proteins identified in all four light versus dark experiments

Proteins included in this table were identified with a minimum of two unique peptides in each experiment. The Mowse scores (Score), percent coverage (Covrg.), number of unique peptide sequences followed parenthetically by the number of peptide observations with ratios (Peps.), and the mean protein ratio and S.D. (Ratio) are presented in two columns (full versus partial metabolic labeling) and two rows (forward and reciprocal labeling) for each protein. The reciprocal ratio information for both labeling approaches has been provided as 1/ratio values so that greater variation (e.g., for catalase 3), results are generally consistent across replicates. The variability we observe among biological replicates via both full and partial metabolic labeling is comparable to that observed in several previously published studies using metabolic labeling (9, 14, 24).

Additionally because we incorporated limited subcellular fractionation into our experimental design, we can compare measurements of selected proteins across multiple fractions. Observation of the same protein across multiple subcellular fractions could occur for several reasons. First the protein in question could exist in multiple cellular locations under different biological conditions. This could potentially lead to observation of different ratios for the same protein across multiple fractions if its subcellular localization is influenced by the biological perturbation in question. Additionally in the case of highly abundant proteins, there could be significant carryover from fraction to fraction. We observed both a myrosinase-associated protein (At1g20620) and a glycosyl hydrolase family 1 protein in the organellar and microsomal fractions. Interestingly both appeared to be predominantly localized in the organellar fraction: both were identified and quantified by organellar fraction: both were identified and quantified by

Although Table III provides an indication of consistency across replicates for those proteins showing ratio changes in the quantifiable range (1:12 to 12:1) for these techniques, Table IV displays those proteins whose ratios fell outside of the quantifiable range (1:12 to 12:1) for these techniques.
Comparison of Full and Partial Metabolic Labeling

**Table IV**

Proteins identified at the abundance measurement extremes

Proteins included in this table were identified with light:dark ratios \( >10 \) and a minimum of two unique peptides in both partial incorporation experiments (there were none \( <0.1 \) meeting the second requirement). As in Table III the Mowse scores (Score), percent coverage (Covrg.), number of unique peptide sequences followed parenthetically by the number of peptide observations with ratios (Peps.), and the mean protein ratio and S.D. (Ratio) are presented in two columns (full versus partial metabolic labeling) and two rows (forward and reciprocal labeling) for each protein. The reciprocal ratio information for both labeling approaches has been provided as 1/ratio values so that all values presented in the table represent light:dark ratios to facilitate comparison across experiments. RuBiSOC, ribulose-bisphosphate carboxylase/oxygenase.

<table>
<thead>
<tr>
<th>At Number</th>
<th>Protein Description</th>
<th>Light/Dark</th>
<th>Partial Incorporation (5-6%)</th>
<th>Full Incorporation (100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1g61520.1</td>
<td>chlorophyll A-B binding protein</td>
<td>1500g Forward</td>
<td>175 15.7% 3(3) &gt;10</td>
<td>123 14.6% 2(0) --</td>
</tr>
<tr>
<td>At1g15820.1</td>
<td>chlorophyll A-B binding protein</td>
<td>1500g Forward</td>
<td>103 9.0% 2(2) &gt;10</td>
<td>93 3.7% 1(0) --</td>
</tr>
<tr>
<td>At2g39730.1</td>
<td>RuBisCO activase</td>
<td>100000g Reciprocal</td>
<td>376 10.3% 3(2) &gt;10</td>
<td>401 17.9% 5(1) 8.4</td>
</tr>
<tr>
<td>At4g10340.1</td>
<td>chlorophyll A-B binding protein</td>
<td>100000g Reciprocal</td>
<td>177 12.6% 3(3) &gt;10</td>
<td>55 17.7% 1(0) --</td>
</tr>
<tr>
<td>At1g15820.1</td>
<td>chlorophyll A-B binding protein</td>
<td>1500g Forward</td>
<td>304 33.5% 3(3) &gt;10</td>
<td>378 31.0% 3(0) &gt;10</td>
</tr>
<tr>
<td>At4g10340.1</td>
<td>chlorophyll A-B binding protein</td>
<td>100000g Reciprocal</td>
<td>523 33.5% 6(10) 5.65(±9.72)</td>
<td>385 39.5% 6(3) 4.52 (±0.62)</td>
</tr>
</tbody>
</table>

this range. Although significantly more error is associated with the ratios determined for these proteins by either technique, these observations can still be of biological use as qualitative indications of changes in protein abundance.

As suggested by Fig. 10, partial metabolic labeling and full metabolic labeling deal very differently with quantification of these very large changes in protein abundance. All of these proteins were observed in all replicates with multiple peptides giving strong protein scores. Although partial metabolic labeling allows assignment of very large -fold changes for all of these proteins, for more than half of these cases full metabolic labeling failed to return any ratio at all. Even for those proteins for which full metabolic labeling did provide ratios, these were generally the average of significantly fewer peptide observations. This is a consequence of the fundamentally different way in which each of these algorithms measures large changes in abundance. In the full incorporation case, we are comparing a very large envelope from one sample with another envelope that is either absent or obscured by the background. These cases give rise to poor quantification as indicated by their poor correlation value and are thus ignored. Without extensive visual inspection, identification of these large changes via full metabolic labeling is quite difficult. In contrast, a single envelope is observed for the partial incorporation case whether that envelope comes entirely from one sample or the other or whether it is a mixture of the two. Thus our measurements of extreme changes in protein expression are not hindered by the absence of a peptide in one sample or the other. This can be a key advantage when studying biological processes that are expected to produce very large changes in protein abundance.

In our particular experiment the ability to observe very large changes in protein expression is essential to fully characterize the biological system. When plants are grown for long periods in the dark, there is a general decrease in protein abundance in the chloroplast because the photosynthetic apparatus is serving little purpose. Although some of these changes fell in the range where quantification is possible, many of these changes were more extreme. Several chlorophyll-binding proteins, an oxygen-evolving enhancer complex, and an uncharacterized expressed protein showed very large changes in expression. Several of these proteins were unquantified via full metabolic labeling, whereas all were quantified via partial metabolic labeling. Thus for this particular experiment partial incorporation provides the more complete characterization of this biological system and provides us with important observations that would have been ignored with full metabolic labeling.

**Light- Versus Dark-grown Arabidopsis: Biological Interpretation**—Our comparison of protein abundance in dark- versus light-grown Arabidopsis reveals several trends that are of potential biological interest. Tables III and IV list average abundance ratios for selected proteins as discussed previously. Supplemental Table 3 lists all proteins identified in each replicate, while Supplemental Table 4 provides a comparison of protein scores, coverage, and abundance ratios observed.
for each protein across all four replicates. As expected, a large number of proteins associated with the chloroplast show significantly higher protein abundance when grown in the light. These include several chlorophyll A-B-binding proteins as well as components of photosystems I and II.

More interestingly, we have observed changes in abundance for several proteins that may be linked to plant defense and stress response. The glycosyl hydrolase family 1 protein (At3g09260, also known as PYK-10) shows a consistent up-regulation during growth in the dark. This protein is a myrosinase and has been shown to be the primary component of endoplasmic reticulum bodies, which appear to play roles in stress response in Brassicaceae (25, 26). Similarly we see up-regulation of the jacalin lectin family protein (At3g16420, also known as PYK-10-binding protein 1, or PBP-1) in the dark. This protein appears to play a role in activation of PYK-10 (27). Up-regulation of this protein as well as other highly similar proteins (At3g16460 and At3g16470) could be an indication that they act to regulate PYK-10. Changes in the abundance of several myrosinase-associated proteins (At3g14210 and At1g54010) are consistent with this biological effect as well. One myrosinase-associated protein (At3g14210; gi:15231805) has also been independently shown to respond to cold stress (28). An up-regulation of proteins associated with stress response following extended growth in the dark is not surprising as plants grow significantly more poorly under these unusual growth conditions.

Additionally we have identified a largely uncharacterized protein (At3g46780) whose expression appears to be significantly up-regulated upon treatment with light. Although its function is unknown, its amino acid sequence shows some similarity to a flavin reductase-related protein (At2g34460) via BLAST analysis against all predicted Arabidopsis proteins (score = 143) (29, 30). Its strong response to growth under light is perhaps not surprising as its expression has been observed previously in the thylakoid membranes and its amino acid sequence may contain a chloroplast transit peptide (31). Our present observation of this protein in the organellar fraction is consistent with this result. Further investigation of the effects of light versus dark growth conditions on any of these proteins will be required to establish their biological significance.

Conclusions—Although full metabolic labeling has become an important tool for quantitative proteomics analyses, its application is limited to those model systems for which complete substitution of $^{15}$N for $^{14}$N can be achieved efficiently and economically. Although it has been proposed that quantitative measurements could be made using much lower incorporations of $^{15}$N that would be achievable in more model systems under a wider array of experimental conditions, until now no automated systems for quantitative proteomics analysis have been implemented. Furthermore no comparisons have been made between the performance of this partial labeling technique and the more standard full metabolic labeling.

We have designed and used an algorithm for automated quantitative analysis of proteomics samples through comparison of natural abundance and 5–6% $^{14}$N metabolic labeling. Additionally we have compared this algorithm with traditional full metabolic labeling for characterization of both control samples of Arabidopsis peptides combined across a range of ratios and for comparison of light- and dark-grown Arabidopsis samples. These experiments have allowed us to evaluate both techniques with respect to accuracy and dynamic range of quantification as well as numbers of peptide and protein identifications across a wide dynamic range.

Generally both full metabolic labeling and partial metabolic labeling provide ratios of comparable accuracy. Both display similar dynamic ranges, providing consistent quantification for ratios between ~12:1 and 1:12 (labeled:unlabeled). Error associated with each is relatively constant over this range, although measurements obtained by partial metabolic labeling seem to display slightly greater uncertainty. Both appear equally consistent across replicates for characterization of biological samples.

With respect to numbers of peptide identifications, full metabolic labeling appears to be more consistent across a wide dynamic range. However, partial incorporation consistently allows successful quantification of a higher proportion of these peptide identifications such that more peptides and proteins are identified and quantified by partial incorporation. This difference is most pronounced at extreme ratios. This effect is seen not only with controlled mixtures of peptides but also for peptides and proteins identified in biological samples exhibiting wide changes in protein abundance as in the comparison of light- versus dark-grown Arabidopsis.

The most significant difference between partial and full metabolic labeling appears to be their ability to measure large differences in protein abundance. This difference proved to be significant for our comparison of light- versus dark-grown Arabidopsis. Ultimately partial metabolic labeling allowed the more complete biological comparison of this system, identifying several proteins whose expression changed significantly under conditions where full metabolic labeling failed to provide reliable quantitative information. However, in quantitative proteomics experiments we are generally comparing much more subtle biological effects than those that we selected. For the kinds of ratios that are typically observed in most quantitative proteomics experiments either full or partial metabolic labeling would likely perform equally well.

In final analysis both partial metabolic labeling and full metabolic labeling appear to be excellent tools for quantitative proteomics experiments. Just as full metabolic labeling has allowed characterization of changes in protein abundance for selected model organisms, partial metabolic labeling should extend these capabilities to a much wider array of model organisms and experimental systems. Finally partial metabolic labeling should improve our ability to characterize biological systems that are expected to produce especially large
Comparison of Full and Partial Metabolic Labeling

changes. Through application of both full and partial metabolic labeling techniques in appropriate circumstances, the advantages of metabolic labeling, including good peptide and protein quantification with excellent control for variability in sample handling and analysis, should be accessible under a much wider array of experimental conditions.

Supplemental Material—The following supplemental materials are available. 1) Supplemental Table 1 lists all peptide identifications used for quantitative analysis of controlled mixtures of Arabidopsis peptides. 2) Supplemental Table 2 lists all peptides identified in all four replicates of the light- versus dark-grown Arabidopsis experiment. 3) Supplemental Table 3 lists all proteins identified and quantified in all replicates of the light- versus dark-grown Arabidopsis experiment. 4) Supplemental Table 4 lists the Mascot scores, coverage, and abundance ratios observed across each of four replicates of the light- versus dark-grown Arabidopsis experiment separated by protein identifier (At number) for easy comparison. 5) Supplemental Fig. 1 is a plot of the distributions of correlation coefficients obtained during quantification of a selected peptide via partial metabolic labeling. Distributions are presented from quantification of this peptide across a wide variety of mixing ratios. 6) Annotated MS/MS spectra are provided for all single peptide protein identifications from all experiments. These are provided as pdf files and separated by experiment and labeling technique. Spectra from the initial analyses of Arabidopsis peptide mixtures at varying ratios are contained in files called MSMS_Partial_PepMix.pdf and MSMS_Full_PepMix.pdf. Spectra from the light and dark comparisons are contained in files called MSMS_Partial_LightDark.pdf and MSMS_Full_LightDark.pdf. 7) Scripts written in Mathematica to perform the described quantitative analyses via either full or partial metabolic labeling are available upon request.

Acknowledgments—We thank all members of the University of Wisconsin-Madison Biotechnology Center Mass Spectrometry facility (Dr. Gregory Barrett-Wilt, James Brown, and Grzegorz Sabat) for technical advice and helpful conversations throughout this project. Additionally we thank Dr. Clark J. Nelson for assistance generating annotated MS/MS spectra for supplemental material.

* This work was supported in part by National Science Foundation Grant MCB-0448369 and National Institute of Health Grant 4 R33 DK070297. 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.

¶ Both authors contributed equally to this work.

§ Supported by the University of Wisconsin-Madison Biotechnology Training Program (National Institutes of Health Grant 5 T32 GM08349). To whom correspondence may be addressed: Biotechnology Center, University of Wisconsin, 425 Henry Mall, Madison, WI 53706. E-mail: elhuttlin@wisc.edu.

** To whom correspondence may be addressed: Biotechnology Center, University of Wisconsin, 425 Henry Mall, Madison, WI 53706. E-mail: hegerman@biochem.wisc.edu.

†† To whom correspondence may be addressed: Biotechnology Center, University of Wisconsin, 425 Henry Mall, Madison, WI 53706.
E-mail: harms@biotech.wisc.edu.

§§ To whom correspondence may be addressed: Biotechnology Center, University of Wisconsin, 425 Henry Mall, Madison, WI 53706.
E-mail: msussman@wisc.edu.

REFERENCES

bases using mass spectrometry data. Electrophoresis 20, 3551–3567
Prediction of error associated with false positive rate determination for
peptide identification in large-scale proteomics experiments using a
combined reverse and forward peptide sequence database strategy. J.
Proteome Res. 6, 392–398
Measurement of the isotopic enrichment of stable isotope-labeled pro-
teins using high resolution mass spectra of peptides. Anal. Chem. 77,
7646–7653
isotopes in the atmosphere and compressed gas from various sources.
isotope abundance ratios in mass spectrometry. J. Chem. Educ. 49,
610–612
of quantitative proteomic analyses of complex biological mixtures by mul-
tidimensional protein identification technology. Anal. Chem. 75,
5054–5061
seedling and root specific gene and promoter from Arabidopsis thaliana.
Plant Sci. 161, 337–346
novel ER-derived compartment, the ER body, selectively accumulates a
β-glucosidase with an ER-retention signal in Arabidopsis. Plant J. 33,
493–502
ER-body-localized β-glucosidase via a cytosolic binding partner in dam-
aged tissues of Arabidopsis thaliana. Plant Cell Physiol. 46, 1140–1148
Arabidopsis nuclear proteome and its response to cold stress. Plant J. 36,
652–663
29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990)
Basic local alignment search tool. J. Mol. Biol. 215, 403–404
functions of the thylakoid membrane proteome of Arabidopsis thaliana
revealed by a simple, fast and versatile fractionation strategy. J. Biol.
Chem. 279, 49367–49383