Published in final edited form as:

J Proteome Res. 2016 March 04; 15(3): 851–867. doi:10.1021/acs.jproteome.5b00772.

Proteome Scale-Protein Turnover Analysis Using High Resolution Mass Spectrometric Data from Stable-Isotope Labeled Plants

Kai-Ting Fan^{†,||}, Aaron K. Rendahl[§], Wen-Ping Chen^{†,‡}, Dana M. Freund^{†,‡}, William M. Gray^{†,||}, Jerry D. Cohen^{†,‡}, and Adrian D. Hegeman^{†,‡,||,*}

[†]Department of Horticultural Science, University of Minnesota, Twin Cities, Minnesota 55108, United States

[‡]Microbial and Plant Genomics Institute, University of Minnesota, Twin Cities, Minnesota 55455, United States

[§]School of Statistics, University of Minnesota, Twin Cities, Minnesota 55108, United States

^{||}Department of Plant Biology, University of Minnesota, Twin Cities, Minnesota 55455, United States

Abstract

Protein turnover is an important aspect of the regulation of cellular processes for organisms when responding to developmental or environmental cues. The measurement of protein turnover in plants, in contrast to that of rapidly growing unicellular organismal cultures, is made more complicated by the high degree of amino acid recycling, resulting in significant transient isotope incorporation distributions that must be dealt with computationally for high throughput analysis to be practical. An algorithm in R, *ProteinTurnover*, was developed to calculate protein turnover with transient stable isotope incorporation distributions in a high throughput automated manner using high resolution MS and MS/MS proteomic analysis of stable isotopically labeled plant material. *ProteinTurnover* extracts isotopic distribution information from raw MS data for peptides identified by MS/MS from data sets of either isotopic label dilution or incorporation experiments. Variable isotopic incorporation distributions were modeled using binomial and beta-binomial distributions to deconvolute the natural abundance, newly synthesized/partial-labeled, and fully labeled peptide distributions. Maximum likelihood estimation was performed to calculate the

*Corresponding Author: hegem007@umn.edu. Telephone: 1 (612) 626-3650.

Notes

The authors declare no competing financial interest.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.5b00772.

Figure S-1, Correlation by linear regression for unsupervised extraction of relative isotopomer intensities;

Figure S-2, Turnover data for the large subunit of RuBisCO in label dilution experiments (PDF) Document S-1, *ProteinTurnover* R script annotated with instructions (TXT)

Tables S-1, Turnover rates of proteins enriched from soluble, organelle, and microsomal fractions of *Arabidopsis* seedling roots; Table S-2, Maximum likelihood estimation output for 20 peptides of the RuBisCO large subunit isolated from *Arabidopsis* leaves; Table S-3, List of columns in Scaffold spectrum reports used and renamed by the algorithm; Table S-4, Elemental composition, monoisotopic mass, and average mass of all amino acids; Table S-5, Natural abundance of the elements carbon (C), hydrogen (H), nitrogen (N), oxygen (O), phosphorus (P), and sulfur (S) (XLSX)

distribution abundance proportion of old and newly synthesized peptides. The half-life or turnover rate of each peptide was calculated from changes in the distribution abundance proportions using nonlinear regression. We applied *ProteinTurnover* to obtain half-lives of proteins from enriched soluble and membrane fractions from *Arabidopsis* roots.

Graphical abstract

Keywords

protein turnover rates; proteome dynamics; stable isotope; metabolic labeling; *Arabidopsis thaliana*

INTRODUCTION

The dynamics of protein synthesis and degradation are becoming an increasingly important part of our understanding of cellular mechanisms for the regulation of organismal behavior, growth, and development across biota.^{1,2} Studies in a range of organisms including bacteria, mammals, and plants show weak correlation between steady state mRNA and protein abundances, although generally the correlation is considered to be positive.^{3–6} This observation supports the idea that widespread variation in protein synthesis and degradation rates occurs quite generally across a given proteome. The widespread occurrence and potential regulatory importance of protein dynamic changes demonstrates the need for robust approaches for the high throughput measurement of protein turnover in plants and other organisms.

Measurement of protein dynamics began over 70 years ago,⁷ yet methods that are capable of describing turnover of large numbers of proteins at a proteomic scale have only recently been described.^{8,9} Most strategies introduce some form of isotopic label into proteins and follow the appearance/disappearance of the label as it is added to, or taken away from, the biological system. Radioisotope labeled amino acids (such as ³⁵S-methionine) have been used in earlier studies for measurement of bulk protein turnover, or for individual proteins in concert with two-dimensional electrophoresis.^{10,11} For various reasons, including physiological perturbations associated with radiolabeling approaches,^{12,13} stable isotopic labeling (²H, ¹³C, ¹⁵N, ¹⁸O) strategies coupled with mass spectrometry (MS) dominate the approaches used for proteomics-scale protein turnover measurement. Studies of protein turnover or synthesis/degradation typically follow label incorporation or loss following

metabolic labeling¹⁴ where an organism is provided with labeled nutrients such as $^{13}\text{CO}_2$,^{15,16} ^{13}C -sugars,¹⁷ deuterated water,¹⁸ $^{13}\text{C}/^{15}\text{N}/^{18}\text{O}$ -amino acids,^{19,20} $^{15}\text{NH}_4^+$, $^{15}\text{NO}_3^-$ salts,²¹ or food formulated with ^{15}N -algae.^{22,23} Significant isotope effects are associated with ^2H labeling^{24,25} which severely perturbs most biological systems, limiting its usefulness to special situations where protein turnover rates are significantly faster than nutrient incorporation.¹⁸ Other nuclides such as ^{13}C , ^{15}N , and ^{18}O do not appear to cause physiological perturbations of detectable magnitude.^{26,27}

Many proteome-scale protein turnover experiments to date have been performed using rapidly growing cellular cultures that allow for facile switching of nutrient inputs and minimal problems associated with internal nutrient storage and trafficking. Multicellular or slower growing organisms present several challenges for measuring turnover that are typically not present in cell culture systems. Nutrients provided as sugars or amino acids often also function as metabolic regulatory signals in multicellular organisms, and thus, the choice of nutrient used for label introduction may be critical. Using great care to avoid physiological perturbation by labeled sugars or amino acids, others have successfully employed stable isotope labeled amino acids to label proteins for turnover experiments especially in cell culture.²⁸ Problems with this approach occur when either amino acid recycling from protein degradation or amino acid synthesis from internal nutrient stores significantly contributes to intracellular amino acid pools. Under these conditions profound systematic errors result from an underestimation of new protein synthesis due to fractional incorporation of label into newly synthesized protein from mixed labeling of precursor amino acid pools. This phenomenon is not easily detected using labeled amino acid nutrient strategies such as ^{35}S -methionine labeling or SILAC, as these strategies do not typically provide adequate definition of the degree of amino acid precursor from internal nutrient stores to correct for slow incorporation, or even indicate that there is a problem. Alternatively, metabolic labeling using $^{13}\text{CO}_2$ or $^{15}\text{NO}_3/^{15}\text{NH}_4$ provides better definition of these partial labeling phenomena and can be used to provide data to determine accurate turnover measurements. To avoid the potential problems with using labeled amino acids or sugars, our group previously used $^{13}\text{CO}_2$ and/or $^2\text{H}_2\text{O}$ metabolic labeling strategies to measure turnover rates of several proteins.^{16,18} In this study, we use $^{15}\text{NH}_4/^{15}\text{NO}_3$ labeling to study proteome turnover in *Arabidopsis thaliana* seedling roots.

To help illustrate this point, Scheme 1 shows the net rates of nutrient uptake and amino acid synthesis (k'_1), amino acid synthesis from internal nutrient stores (k'_2), protein synthesis (k'_3), and degradation (k'_4). In cases where an organism is rapidly growing and does not involve elaborate nutrient uptake and transport mechanisms, as can be accomplished with many microbial systems, nutrient uptake is directly funneled into protein synthesis so that $k'_1 + k'_3 \gg k'_2 + k'_4$. This scenario results in protein turnover mass spectral (MS) data that have a simple labeling pattern due to what is, in effect, direct incorporation of label from the growth medium into newly synthesized protein. Figure 1A shows simulated mass spectra of a peptide extracted in stable isotopic label incorporation and dilution experiments when $k'_1 + k'_3 \gg k'_2 + k'_4$. Protein turnover has been measured in many systems, including bacteria²⁹ and yeast²⁰ using growth conditions that produce this labeling pattern. This special case does not occur when there is either a significant pre-existing internal amino acid pool, amino acid synthesis from internal nutrient stores, protein recycling, or slow nutrient

uptake and transport such that amino acids derived from nonexogenous nutrient sources are also incorporated into newly synthesized proteins. On the other hand, most multicellular or slow growing biological systems including plants, animals, and quiescent or stationary microbes will deviate from the simple labeling pattern for one or several of these reasons and produce MS data like that shown in Figure 1B. Here, as label is either incorporated or diluted, one does not observe a simple intensity transition between discrete heavy and light isotopic envelopes, but instead the isotopic envelopes for peptides derived from newly synthesized protein shift in both intensity and abundance with prolonged isotopic enrichment. At infinite time the isotopic abundance approaches that of the nutrient supplied. Intermediate time points show shifting isotopic distributions for newly synthesized protein that have, until recently, made it very difficult to use MS and stable isotopes to measure protein turnover in these systems.

Despite the complexity in modeling the shifting isotopic distributions in the partially labeled samples, several tools have used approaches that allow automated protein turnover analysis with relatively simple approximations that make it possible to calculate protein turnover rates using the relative isotope abundance (RIA), which is the ratio of the sum of intensities of defined heavy isotope peaks to the sum of all isotopic peaks. For example, using a data set from labeled green algae, “Protein TurnStILE” uses the first 5 isotopic peaks of each peptide as “light” species and the sixth to last isotopic peaks noted as “heavy” species were iteratively fitted with ^{15}N % ranging from 30% to 99%.¹⁹ “ProTurnyzer”, on the other hand, relies on the ratio of the extracted intensity and calculated theoretical fraction frequency of the monoisotopic peak in order to avoid the need for complex RIA information for the precursor pool, and was applied thus for the analysis of data from mice fed with the labeled diet.³⁰ Utilizing the principle of RIA, Guan and colleagues estimated proteome-scale turnover rates in brain, liver, and blood of mice fed with ^{15}N -labeled algae with an extended model that considers the exchange kinetics between the free amino acid pool, the total protein pool, and each individual protein.^{22,31,32} Recently, an automated algorithm written by Lyon et al. took an interesting approach for peak selection. By setting the time course files in reverse chronological order, the data from the later time point (longer-labeled) serves as a template to pick the isotopic peaks for the data from the previous labeling time point (shorter-labeled).³³ This approach allows their algorithm to always pick fewer peaks in the shorter-labeled time point data sets. It also allows settings to filter out coeluted picked peaks based upon an empirically found threshold ratio of two adjacent peaks and to choose a best single scan to be used for peak picking.

In contrast to those previous published algorithms, which often emphasize how to extract accurate intensities of defined light and heavy isotopic peaks to calculate the protein turnover rates, we utilize the probability distribution of old and newly synthesized peptide’s isotopic peaks to develop a new method with an automated workflow for turnover rate calculation using metabolic stable isotope labeling. The application of the *compartment modeling* (or *relative isotope fraction*) methodology proposed by Guan et al.³² seems to be especially important for the measurement of protein turnover in mammalian systems where a significant lag in relative isotopic abundance is observed. This approach, while needed in mammalian systems to explain the much slower and delayed incorporation of label into certain organs (such as brain), does not appear to be as important in plant tissues where

compartmentalization and growth processes are fundamentally different from animals. Their approach is also based on the estimation of ^{15}N -incorporation rate, but data processing and model fitting were accomplished in a fundamentally different way. The theoretical isotope distributions from 0 to the number of nitrogen atoms were generated using the Mercury2 algorithm,³⁴ which uses a computationally efficient Fourier transform approximation instead of the expansion of a polynomial expression. Additionally, the ^{15}N distribution for each peptide was calculated from the extracted ion peak intensities fitted using a non-negative least-squares algorithm.

The approach described here uses an algorithm, written in *R* called *ProteinTurnover*, which is open-source so that it can be shared and enhanced easily by interested colleagues. The algorithm employs maximum likelihood estimation to model the isotopic distributions of both the old and the newly synthesized protein-derived distributions of peptide mass spectra using binomial and beta-binomial distributions to account for observed asymmetrical and stochastic deviations from pure binomial distributions. *ProteinTurnover* also applies an unsupervised noise reduction strategy that is important because of the increased likelihood of signal overlap between different, broad isotopic distributions.³⁵ Using the designed automated workflow, we have analyzed the proteomic turnover rates of proteins from enriched soluble, organelle, and microsomal fractions of *Arabidopsis* seedling roots in an ^{15}N -incorporation experiment.

MATERIALS AND METHODS

Plant Growth and Labeling Conditions

Arabidopsis thaliana ecotype Columbia Col-0 was used for all experiments.

1. Stable isotope label incorporation—For ^{15}N -incorporation experiments, *Arabidopsis* seeds were sterilized with 30% (v/v) bleach containing 0.1% (v/v) Triton X-100 and vernalized at 4 °C for 2 days. Seeds were germinated on a nylon filter membrane (mesh opening 20 μm , Spectrum Laboratories, cat. #146510), which was placed on the top of ATS¹⁸ agar plates. The seedlings were grown under continuous fluorescent light ($\sim 80 \mu\text{mole photon m}^{-2} \text{ s}^{-1}$) at 22 ± 2 °C. Eight-day-old seedlings along with the filter membrane were then transferred onto fresh ATS media containing 99 atom % K^{15}NO_3 and 98 atom % $\text{Ca}^{15}\text{NO}_3)_2$ (Cambridge Isotopes Laboratories, Inc., Andover, MA) and were collected at 0, 4, 8, 16, 24, 32, 40, and 48 h after transfer. Experiments were performed using a single biological replicate, and each sample contained a pool of 400–600 seedlings grown on a single plate.

2. Stable isotope label dilution—For labeling adult plants with ^{15}N , an automated hydroponic system was constructed and used for the ^{15}N -dilution experiment. *Arabidopsis* seeds were germinated and grown on rockwool blocks (Grodan, Milton, ON) in $2' \times 2'$ pots and covered with silicone rubber mats (4.7 cm $\times$ 4.7 cm, 8608K151, Extreme-Temp Textured Silicone Foam Rubber, 1/8" thick, ordered from McMASTER-CARR, Robbinsville, NJ) to restrict algal growth. Plants were irrigated by the system from the bottom with 1.5 L of full-strength Gibeaut's medium³⁶ consisting of 1.25 mM KNO_3 , 1.5 mM $\text{Ca}(\text{NO}_3)_2$, 0.75 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.5 mM KH_2PO_4 , 1 mM $\text{Na}_2\text{O}_3\text{Si} \cdot 9\text{H}_2\text{O}$, 1 mL/L

micronutrient stock solution (50 mM KCl, 50 mM H₃BO₃, 10 mM MnSO₄·H₂O, 2 mM ZnSO₄·7H₂O, 1.5 mM CuSO₄·5H₂O, 0.01 mM CoCl₂·6H₂O and 0.2 mM Na₂MoO₄·2H₂O), and 4 mL/L 20 mM Fe–EDTA) solution (most chemicals from Sigma-Aldrich). The irrigation pump (SHURflo, Model 2088–594–154, Cypress, CA) was started by a timer control (Dayton 2VJ57, Grainger, NY) four times a day. Each time lasted for 5 min. During the irrigation, the medium on the tray where plant plugs were placed drained back to the medium reservoir through two stainless open nuts with the opening 1 cm above the bottom of tray. There were three small holes (1 mm diameter) on the tray to gradually allow the medium to be completely drained back to the reservoir within 10 min. To reduce algal and bacterial growth in the medium, the tray was covered with a Plexiglas sheet and the medium was passed through a UV sterilizer (Helix Max UV 5W, AQUA MEDIC, Germany) and a 30 μ m membrane filter (Spectrum) twice a day (duration time: 5 min). The ¹⁵N labeling was achieved by replacing nitrogen salts with 99 atom % K¹⁵NO₃ and 98 atom % Ca(¹⁵NO₃)₂ (Cambridge Isotopes Laboratories, Inc., Andover, MA). After 3 weeks, plant roots in the rockwool blocks were rinsed thoroughly with running distilled water, three times with 20 mL of the unlabeled medium then whole plants were placed on small trays filled with unlabeled medium to start the dilution experiments for protein turnover analysis. The whole rinsing step was finished within 5 min. After various dilution durations (0, 2, 6, and 12 days), leaves were harvested and immediately frozen by liquid nitrogen then stored at –80 °C. The leaves of fully labeled plants were frozen directly without any rinse and used as the time 0 sample. Experiments were performed using a single biological replicate, and each sample contained a pool derived from 3 plants.

Protein Separation and Digestion

1. Stable isotope label incorporation—*Arabidopsis* seedlings were dissected into hypocotyl and cotyledon (shoot), and root tissue samples. Methods were adapted from Huttlin et al.³⁷ for the separation of soluble and membrane proteins. Briefly, root tissues were homogenized with the ice cold grinding buffer containing 290 mM sucrose, 250 mM Tris-HCl (pH 7.6), 25 mM EDTA, 5 mM DTT, 1 mM PMSF, 0.5 $\times$ protease inhibitor cocktail (Roche, Indianapolis, IN) at a ratio of 3 mL per gram of fresh weight tissue using Potter-Elvehjem grinder (Fisher Scientific, Pittsburgh, PA). After the homogenate was filtered through four layers of Miracloth, which was packed into a disposable syringe to accommodate the small volume of the sample, it was separated into soluble, organelle, and microsomal protein fractions via centrifugation. The samples were first spun at 1,500 $\times g$ for 5 min to pellet the organellar fraction, and then the microsomal proteins were separated as pellet from soluble proteins by centrifugation at 100,000 $\times g$ for 1 h. The concentrations of soluble protein fractions were measured by the Bradford method using a commercial kit from Bio-Rad (Hercules, CA), which includes bovine serum albumin fraction V as the calibration standard. Soluble proteins (300 μ g) were precipitated by addition of acetone to 80% with incubation at –20 °C overnight (or –80 °C for 30 min), recovered by centrifugation for 15 min at 16,000 $\times g$. Air-dried pellets were dissolved in 8 M urea/8 mM DTT to a protein concentration of 8 μ g/ μ L. Prior to proteolysis, the dissolved soluble protein fractions were diluted to 1 μ g/ μ L with 1 M urea/1 mM DTT/50 mM ammonium bicarbonate. Proteolysis was initiated by adding 5 mM CaCl₂ and 6 μ g of sequencing grade trypsin

(Promega, Madison, WI) to a final ratio of 1:50. The proteolysis was continued overnight (14 h) at 37 °C and terminated by addition of formic acid to 5% (v/v).

Membranous protein fractions derived from $1,500 \times g$ (organelle) and $100,000 \times g$ (microsomal) pellets were processed identically as follows. The pellets were resuspended in grinding buffer and pelleted by centrifugation at $16,000 \times g$ for 90 min. After removing the supernatant, the pellet was resuspended in 50–100 μL of 50 mM ammonium bicarbonate and homogenized by grinding pellets with plastic pestles in 1.5 mL microtubes. Protein concentration was then measured using the Bradford assay kit (Bio-Rad) and each membrane homogenate was diluted to $\sim 2 \mu\text{g}/\mu\text{L}$. The samples were reduced by addition of DTT to 2 mM and heated to 90 °C for 5 min. After the samples were cooled down to below 50 °C, 1 volume of methanol was added to the sample. The proteolysis of organelle and microsomal proteins was initiated by addition of trypsin to a final ratio of 1:50 to the sample containing 1 $\mu\text{g}/\mu\text{L}$ of protein, 50% of methanol (v/v), 1 mM DTT, and 5 mM CaCl_2 . The proteolysis reaction proceeded overnight at room temperature (~ 22 – 24 °C) with gentle rocking and was terminated by addition of formic acid to 5% (v/v). Peptides of membrane proteins were collected from the supernatant following centrifugation ($3,000 \times g$ for 1–2 min) of the overnight digestion. The tryptic peptides of soluble or membrane fractions were then purified by C_{18} solid phase extraction using OMIX C_{18} pipet tips (Agilent) and its manufacturer's protocol. After purification, peptides were dried in SpeedVac (Savant) and resuspended in 5% (v/v) acetonitrile, 0.1% formic acid for LC-MS/MS analysis.

2. Stable isotope label dilution—For the measurement of the turnover of RuBisCO large subunit, leaves harvested from three-week-old *Arabidopsis* plants were ground in liquid N_2 with a mortar and pestle and then total proteins were extracted and washed twice with ice-cold methanol containing a protease inhibitor cocktail (Roche, Indianapolis, IN), then twice with ice-cold acetone. The protein pellets after centrifugation at $14,000 \times g$ for 10 min were air-dried and resuspended in TE buffer containing 1% SDS. The protein concentration was estimated by the Bradford method (Bradford 1976) using a commercial kit from Bio-Rad (Madison, WI). Labeled protein samples were spiked with known amount of the unlabeled protein (at 4:1 ratio labeled to unlabeled) and separated by SDS-PAGE. Protein bands around 52 kDa corresponding to molecular weight of the RuBisCo large subunit were excised manually after visualizing with colloidal Coomassie G-250 stain. Excised bands were subjected to trypsin enzymatic digestion³⁸ on a ProPrep System (Genomic Solutions, Ann Arbor, MI). Briefly, protein bands were subjected to two series of dehydration and hydration steps by the addition, incubation and removal of acetonitrile followed by the addition, incubation and removal of 25 mM NH_4HCO_3 . Gel plugs were then subjected to reduction in the presence of 10 mM DTT/25 mM NH_4HCO_3 at 56 °C for 30 min. The DTT solution was aspirated and a 55 mM iodoacetamide/25 mM NH_4HCO_3 solution was added and the sample incubated for 30 min at room temperature. The iodoacetamide solution was aspirated, followed by two series of dehydration and hydration steps as above. Protein bands were then subjected to tryptic digestion using 12 ng/ μL trypsin (Sigma-Aldrich, St. Louis, MO) in 25 mM NH_4HCO_3 , 5 mM CaCl_2 at 37 °C for 10 h. The reaction was stopped with the addition of formic acid to a final concentration of 0.1% (v/v). Sample digests were manually aspirated and dispensed into 1.5 mL tubes with subsequent

extraction by addition, incubation and removal to the respective tubes of 70% acetonitrile, 0.1% formic acid. All digested extracts were evaporated *in vacuo* (SC210A SpeedVac Plus, Thermo-Savant, Asheville, NC), resuspended in LC-MS/MS loading buffer (98% H₂O, 2% acetonitrile and 0.1% formic acid), and analyzed using a QSTAR Pulsar *i* quadrupole-TOF MS system (Applied Biosystems, Foster City, CA).

LC-MS/MS Analysis

1. Stable isotope label incorporation—The peptides were analyzed using a hybrid quadrupole Orbitrap mass spectrometer (Q Exactive, Thermo Fisher Scientific, San Jose CA) with an ACQUITY UPLC BEH C₁₈ column (1.0 mm × 150 mm, 1.7 μm particle size, Waters, Milford, MA). Solvents A (0.1% (v/v) formic acid in ddH₂O) and B (0.1% (v/v) formic acid in acetonitrile) were used as mobile phases for gradient separation. The equivalent of 25 μg of soluble protein digest, about 6 μg of organelle protein digest or 30 μg of microsomal protein digest was loaded onto the column with 2% solvent B for 12 min at a flow rate of 0.1 mL/min, followed by the following gradient separation: 2 min from 2% B to 10% B, 60 min to 40% B, 1 min to 85% B and maintained for 10 min. The column was equilibrated for 15 min with 2% B prior to the next run. Data-dependent acquisition was used with similar settings to Sun et al.³⁹ with minor modifications. Full MS scans (range 350–1800 *m/z*) were acquired with 70 k resolution. The target value based on predictive automatic gain control (AGC) was 1.0×10^6 with 20 ms of maximum injection time. The 12 most intense precursor ions ($z = 2$) were sequentially fragmented in the HCD collision cell with normalized collision energy of 30%. MS/MS scans were acquired with 35k resolution and the target value was 2.0×10^5 with 120 ms of maximum injection time. The ion selection threshold of 1.0×10^4 counts and a 2.0 *m/z* isolation width in MS/MS was used. The dynamic exclusion time for precursor ion *m/z* was set to 15 s.

2. Stable isotope label dilution—The LC system (LC Packings/Dionex, Sunnyvale, CA) was interfaced with the QSTAR instrument (Applied Biosystems, Ontario, Canada), which was equipped with Protana's (Odense, Denmark) nanoelectrospray source. Peptides (0.5 μg) were eluted with a linear gradient from 0 to 35% B (0.1% formic acid in solution of 95% acetonitrile and 5% water) over 45 min, followed by 35–80% B over 2 min, and held isocratic at 80% B for 10 min. Solvent A was 0.1% formic acid in 95% water and 5% acetonitrile. Product ion spectra were collected in an information-dependent acquisition (IDA) mode, using continuous cycles of one full scan TOF MS from 400 to 1200 *m/z* (1 s) plus four product ion scans from 50 to 2000 *m/z* (2 s each). Precursor *m/z* values were selected starting with the most intense ion, using a selected quadrupole resolution of 3 Da. The rolling collision energy feature was used, which determines collision energy based on precursor *m/z* and charge state. Dynamic exclusion time for precursor ion *m/z* values was 60 s.

Protein Identification

1. Stable isotope label incorporation—All .raw files were converted to *mzXML* files by msConvert tool of ProteoWizard⁴⁰ and then converted to *mgf* format by MGF formatter (v0.1.0). OMSSA (v2.1.9)⁴¹ was used for database searching against the UniProt *Arabidopsis thaliana* database (accessed on February 2013, 33,311 sequences, <http://>

www.uniprot.org) combined with the contamination list from the cRAP database (common Repository of Adventitious Proteins, accessed on February 2013, 115 sequences, <http://www.thegpm.org/crap/>) and reversed sequences. The search parameters were: 6 ppm precursor ion mass tolerance, 0.02 m/z product ion mass tolerance, methionine oxidation as variable modification and a maximum missed cleavage of 2. The search results were then analyzed by Scaffold (v3.6.5, Proteome Software Inc., Portland, OR)⁴² to validate MS/MS based peptide and protein identifications. The results were filtered with a false discovery rate of less than 0.5% on the peptide level and 1% on the protein level with a minimum of two unique peptides required for identification. Proteins that contained similar peptides and that could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. All the above activities of data conversion and protein database searching were performed on the Galaxy-P platform (<https://galaxyp.msi.umn.edu/>), which is a Web-based data-analyzing workbench originated from the Galaxy framework^{43–45} and supported by Minnesota Supercomputing Institute of University of Minnesota (tool wrapper v0.1.0 shown on <https://bitbucket.org/galaxyp/>). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the MassIVE partner repository with the data set identifiers PXD003481, PXD003482, and PXD003483.

2. Stable isotope label dilution—MS/MS data were assigned using *ProteinPilot* (Applied Biosystems) using the TAIR9 nonredundant protein sequence database from *TAIR*. The list of identified peptides from confidently identified proteins was then saved in text format. Next, the original MS data in *WIFF* format was converted to *mzXML* format using the converter, *mzWiff*, from *The Trans-Proteomic Pipeline* developed at the *Institute for Systems Biology*, which was later included in the ProteoWizard project.⁴⁰

Turnover Rate Calculation

The algorithm in R (package *ProteinTurnover*) used in this study is available through GitHub (<https://github.com/HegemanLab/ProteinTurnover>) or upon request. To run the algorithm, the R script (included as Supporting Document S1) should be opened in the R workspace, which lists detailed instructions for use. After database dependent protein assignment, the Scaffold spectrum report (*CSV* format) and all *mzXML* files of MS data were assembled for use by the algorithm. Experiment-specific parameters were set in the command lines of R, including labeling time points (0, 4, 8, 16, 24, 32, 40, 48 h), stable isotope (“N”, as nitrogen) used for labeling, experimental design (label dilution or incorporation), peptide identification confidence threshold (80%), m/z extraction window (0.005 m/z), retention time (RT) window (30 s), and the number of extra channels (5). Other parameters for noise-removal and spectral fitting were used to set thresholds for the linear regression step (using least-squares regression model), beta-binomial options (maximum likelihood estimation, MLE, as “one” or “many”), and *alpha* options (nonlinear regression as “many”, “K”, or “log2K”).

After setting the parameters, the R script first extracts information from the MS data files using the peptide identity input file, which is reannotated as described in Supporting Information Table S-3. The RTs are converted from minutes to seconds and rounded to the

nearest second. Peptide identity scoring thresholds are set such that the script omits the analysis of peptides where the confidence level is below the threshold, or if a peptide is not identified in either the zero time point or in just in the zero time point and not at any subsequent times. The elemental composition of each peptide is then calculated by the code (Supporting Information Table S-4); currently the R script accommodates the elemental composition calculation for the peptide modifications of oxidized methionine or carbamidomethylated cysteine, which are identified as lower-case m and c in the peptide sequences. Additional modifications can easily be added to the open source *ProteinTurnover* code. For each unique protein-peptide combination at each labeling time point (rows of EIC plots), the EIC values were extracted using following information: (1) computed m/z value; (2) the identified RT of the peptide from the input file; and (3) user-defined m/z and RT window. For each peptide, computed m/z value equals $[MW + channelNum * delta] / z + 1.007277$ where the MW is the actual peptide mass (in AMU) provided in the input file, $delta$ is 0.9970348932 for nitrogen and 1.0033548378 for carbon, and $channelNum$ goes from 0 to the number of channels minus 1. The number of channels is the count of the labeled element plus user-defined number of additional channels, such as 5, to account for natural abundance isotopes of other elements. For example, peptide FLIDGFPR has 11 nitrogens, so it has 16 channels in total ($channelNum$ 0–15) if additional number of channels are set to 5 (the default is 1). The complete set of channels represents the distribution of isotopologues of the analyzed peptide at each time point. Then the code takes the total EIC count for each row, chooses the one with the maximum, and uses the EIC data corresponding to that row for that time point, so when +2, +3, or even +4 charged ions are all identified for the same peptide, only the one with the maximum signal intensity is taken into account. At this step, for each unique peptide sequence, an EIC plot is also generated including Counts with RT for each Channel at each Time Point.

Once the EIC data are extracted, the relative intensities of each peptide isotopologue peaks are obtained separately at each time point using a retention time correlation strategy. For this approach, every point of each EIC for each channel of the peptide isotopic envelope is plotted (y -axis) against each point from the EIC of the channel with the highest intensity (x -axis). A linear regression is performed on the plot and the slope of the line is the fractional intensity of each isotopic peak relative to the highest. This approach effectively sets the isotopologue with the highest intensity to a value of one, and all others as some fraction of one. If the slope is less than zero, it was set to zero. Examination of the R^2 value can also be used as a rapid sorting method for detecting overlapping signals that will interfere with the analysis. At this step, the plot of relative abundance fits for the peptide is generated indicating the relative abundance (slope from linear regression) of each isotopic channel at each time point. The isotopic distributions of old and newly synthesized peptide is calculated using binomial or beta-binomial distributions, respectively, which are then fitted to the data using maximum likelihood estimation (MLE). The natural abundance of isotopes used to calculate the isotopic envelope for all elements except the one being labeled is listed in Supporting Information Table S-5.⁴⁶ The abundance proportion of unlabeled (a) and newly synthesized/labeled peptide populations (r) (or *distribution abundance proportions*, as shown in Figure 2, the areas under the distribution), derived from the fitted distribution parameters at each time point, are used to estimate the turnover rates. For isotope label incorporation

experiments, the turnover rate constant (k) or the $\log_2$ value of turnover rate constant ($\log_2 k$) of each peptide is calculated by the nonlinear regression of the *distribution abundance proportion* of unlabeled peptide population (α_j) against time, assuming single exponential decay:

$$\alpha_j = \alpha_0 e^{-kt} \quad (1)$$

On the other hand, for isotope label dilution experiments, the turnover rate is calculated by the nonlinear regression of the *distribution abundance proportions* of newly labeled peptide population (r_j) against time, assuming it fits a single exponential growth:

$$r_j = r_0 (1 - e^{-kt}) \quad (2)$$

As the protein turnover is considered to be first order kinetics, the first-order rate constant (k) is used to calculate the half-life of the particular peptide as

$$t_{1/2} = \ln(2)/k \quad (3)$$

When the analysis used the *alpha* model setting as “ k ” or “ $\log_2 k$ ”, a corresponding standard error (*SE*) for each peptide is also generated. After obtaining the turnover results from *ProteinTurnover*, peptides are selected by applying the filtering threshold: (1) visual score of the spectral fitting (to the beta-binomial model) greater than 80; (2) standard error of turnover rate less than 10; (3) three or more labeling time points. With those selected peptides, only the k or $\log_2 k$ data of unique peptides are then averaged to obtain the protein turnover rate. The selection and final turnover rate calculation processes are done in R using standard data managing commands.

RESULTS AND DISCUSSION

New Statistical Modeling Approach for Automated High Throughput Measurement of Protein Turnover

Several modeling approaches have been previously described for high throughput measurement of protein synthesis and degradation rates, or protein turnover rates, including studies in green alga²¹ and *Arabidopsis* suspension cells.⁴⁷ Our model for label dilution experiments considers the individual peptides as coming from three distinct groups (Figure 2): (1) unlabeled plant material spiked for protein identification, (2) labeled plant material after time 0, but before time index j , which is the sampling time after being transferred to unlabeled medium, and (3) labeled plant material from time 0. We then estimated the proportion of peptides that are in each group by modeling the distribution of the number of labeled elements in peptides from each of these groups, combining with the distribution of extra neutrons in elements that cannot be labeled, and finding the parameters of interest using maximum likelihood estimation.

In practice, two types of experiments, label incorporation and label dilution, could be performed to estimate protein turnover rates. Because of the difficulty in identifying fully and partially ^{15}N -labeled peptides using current available protein search algorithm, in our label incorporation experiment to study the RuBisCO turnover rate, labeled protein samples were spiked with known amount of the unlabeled protein (at 4:1 ratio labeled to unlabeled) before separation by SDS-PAGE and proteolysis. For incorporation experiments in which unlabeled plants are transferred to labeled medium, there is no need to add additional unlabeled control to help with peptide identification since the unlabeled peptide content at time 0 is sufficient for this purpose.

In theory, within each of the three groups, if the atoms in each peptide had the same probability of being labeled, then the isotopic distribution of the number of labeled element k in a particular peptide could be modeled by a binomial distribution, with probability mass function defined as

$$\text{Bin}(n, \pi, k) = \binom{n}{k} \pi^k (1-\pi)^{n-k} \quad (4)$$

More realistically, proteins are synthesized from amino acids with *in vivo* turnover rates that vary so that labeling ratios of the labeled atoms in different amino acids are not fixed and can vary significantly.^{15,18} This difference may account for the observed asymmetrical and stochastic deviations in the shapes of isotopic distributions of the newly synthesized proteins during the period of isotope label dilution or incorporation. Additionally, the labeling ratio is changing between time 0 and time index j , so the probability of being labeled is also not the same for all peptides in the third group. To account for this, the observed peptide distributions are modeled at time 0 and between time 0 and j using beta-binomial distribution instead of binomial distributions, with the probability mass function defined as

$$\text{Beta-Bin}(n, \pi, M, k) = \binom{n}{k} \frac{B(\alpha+k, \beta+n-k)}{B(\alpha, \beta)} \quad (5)$$

where B is the beta function, α and β are two shape parameters defined as $\alpha = M * \pi$, and $\beta = M * (1 - \pi)$. This can also be thought of as a binomial distribution where the probability of success is not fixed but instead follows a beta distribution with mean π and variance $\frac{\pi(1-\pi)}{M+1}$. Since the probability of ^{15}N at each nitrogen atom is not fixed, this beta-binomial distribution should better fit the real biological situation. In other words, we applied the beta-binomial model to predict the ^{15}N -labeled peptide distribution because it is the binomial distribution in which the probability of ^{15}N (instead of ^{14}N) at each nitrogen atom is random and follows the beta distribution.

For modeling simplicity, we first assumed that the probability distribution of the number of nitrogen atoms being labeled with ^{15}N in a particular peptide at time j is binomially

distributed, where the probability of any nitrogen atom on a peptide being labeled with ^{15}N is fixed. Because the labeling ratios of different amino acids are, in fact, not at any given time point, fixed, while a particular protein has been synthesized *in vivo*, a beta-binomial model was subsequently employed as an option for modeling. The probability distribution of the number of extra isotopic labels in other elements depends on their natural abundance, and is computed by convolving together the natural abundance distribution for each type of element in turn. To obtain the probability distribution of the total number of labeled elements, the probability distribution of the ^{15}N element and the fixed natural abundance distributions are combined.

For illustration, the model incorporates an assumption that the labeled system is in a steady-state where protein synthesis and degradation rates at any observed time point are the same, so that no net change in protein concentration is observed. To build our model, let the data consists of triplets $\{j, k, f_{jk}\}$, where j indexes the time of the measurement ($0, \dots, n_j$), k is the number of extra neutrons ($0, \dots, n_{jk}$), and f_{jk} is the observed frequency of k extra neutrons at time index j . We build up to the f_{jk} by considering the outcome for each individual copy of the peptide. Let i index the copy of the peptide, and

n_{ij} = number of labeled elements in i th copy of the peptide at time j

c_{ij} = number of extra neutrons in other elements (and in elements of interest that cannot be labeled) in i th copy of the peptide at time j

$y_{ij} = n_{ij} + c_{ij}$, the total count of extra isotopic labels for the i th observation at time j

First consider the number of labeled elements n_{ij} . As described above, we model each peptide as possible members of three groups. The probability mass function $\Pr(n_{ij} = k)$ within each group is defined as follows, where N is the total number of labelable atoms.

1. The probability mass function of unlabeled plant material: $\delta_j(k) = \text{Bin}(N, \pi_N, k)$, where π_N is the natural abundance of the heavy isotope of that element, as in the Supporting Information Table S-5.
2. The probability mass function of labeled plant material from time 0 and before time index j : $\gamma_j(k) = \text{Beta-Bin}(n, \pi_j, M_j, k)$
3. The probability mass function of labeled plant material from time 0: $\gamma_0(k) = \text{Beta-Bin}(n, \pi_0, M_0, k)$

We model the proportion of peptides that are in each group by specifying α_j , the proportion of peptides that were unlabeled at time j , and r_j , the proportion of labeled peptides that were labeled after time 0 and before time index j . The overall probability that a particular peptide has k extra neutrons is

$$\begin{aligned} P_{n_{ij}} &= \Pr(n_{ij} = k) \\ &= \alpha_j \delta_j(k) + (1 - \alpha_j) ((1 - r_j) \gamma_0(k) + r_j \gamma_j(k)) \end{aligned} \quad (6)$$

Therefore, in label dilution experiments, α and r_j can be used to define a *distribution abundance proportion* of unlabeled (α_j) to newly synthesized and old/fully labeled peptides ($1 - \alpha_j$), where old/fully labeled peptide ($1 - r_j$) is degraded and substituted by new peptide (r_j) over time j .

Next, consider the number of extra neutrons c_{ij} in other elements. This is the sum of the number of extra neutrons across all the other elements, so it is computed by convolving together the natural abundance distribution for each type of element in turn. That is, $P_{c_{ij}} = \Pr(c_{ij} = k)$ is the convolution of the proportions of the various isotopes of the unlabeled elements, where the convolution of two mass functions is defined as

$$(f * g)(k) = \sum_{i=0}^k f(i)g(k-i) \quad (7)$$

Finally, to obtain the probability distribution of the total number of extra isotopic labels y_{ij} , the probability distribution of the number of labeled elements n_{ij} and the fixed natural abundance distribution (if an equal amount of unlabeled material was spiked in the label dilution experiment) are combined; as $y_{ij} = n_{ij} + c_{ij}$, this is again computed using the convolution formula, and the total probability that a particular peptide has k extra neutrons at time j is

$$p_{jk} = \Pr(y_{ij} = k) = (P_{n_{ij}} * P_{c_{ij}})(k) \quad (8)$$

The complete mass function is then proportional to $\prod_{j=0}^{n_t} \prod_{k=0}^{n_{jk}} p_{jk}^{f_{jk}}$, where N_{jk} is the number of peptides with k extra neutrons at time j . We use instead the known frequency $f_{jk} = N_{jk}/N$, where N is the total number of peptides, and the log-likelihood (omitting constants) for a given set of $\{j, k, f_{jk}\}$ triplets, where f_{jk} represents the relative abundance of peptides with k extra labels on time j , is then

$$\begin{aligned} \text{Log}(L(f|\alpha, r, \pi, M)) &\propto \log\left(\prod_{j=0}^{n_t} \prod_{k=0}^{n_{jk}} p_{jk}^{f_{jk}}\right) \\ &\propto \sum_{j=0}^{n_t} \sum_{k=0}^{n_{jk}} f_{jk} \log(p_{jk}) \end{aligned} \quad (9)$$

which is maximized over π_0 , π_j , α , r_j , and M_j to find estimates of these parameters. In practice, this can be more parameters that can be reliably fit by the data, so additional constraints may be enforced, such as forcing all M_j values to be the same, to decrease the number of fitted parameters. Also, in data sets from label incorporation experiments, the natural abundance distribution becomes the old peptide distribution and thus no isotopic

distribution with high enrichment exists in spectra at time 0, which means $1 - r_0 = 0$. Thus, the eq 6 for label incorporation experiments could be modified to

$$P_{n_{ij}} = \Pr(n_{ij} = k) \\ = \alpha_j \delta_j(k) + (1 - \alpha_j) \gamma_j(k) \quad (10)$$

Figure 2 shows examples of peptide mass spectra fitted using the ^{15}N -isotope label incorporation (Figure 2A,B) or dilution (Figure 2C) experimental models based on the above principles. To deconvolute these peptide isotopic distributions in an automated manner, isotopologue distributions were first modeled as mixed beta-binomial distributions, using a MLE statistical approach. The raw MS spectrum of a tryptic peptide of Adenosylhomocysteinase 1 from the sample incorporated with ^{15}N for 48 h and the fitted distribution was shown in Figure 2A and 2B, respectively. For incorporation experiments, as there is no additional unlabeled control, there is no π_0 defined in Figure 2B (eq 10). Figure 2C presents the spectral fitting of a tryptic peptide of the ^{15}N -labeled RuBisCO large subunit extracted from a label dilution experiment. In label dilution experiments, this mass spectrum contained three peptide distributions: fully labeled peptide (old peptide), new peptide, and a natural abundance distribution from spiked unlabeled control protein. In these three distributions, π_N represents the natural abundance of ^{15}N as defined in $\delta_f(k)$; π_j represents the average probability of ^{15}N in the newly synthesized peptide at time j as defined in $\gamma_f(k)$; and π_0 represents the average probability of ^{15}N in the fully labeled peptide as defined in $\gamma_0(k)$ at time 0. As shown in eq 6 and Figure 2C, the three proportions of those peptide distributions in label dilution experiments could be calculated as follows.

1. The proportion of unlabeled peptide distribution: α .
2. The proportion of newly synthesized (partially labeled) peptide distribution from time 0 and before time j : $(1 - \alpha) \times r$.
3. The proportion of fully labeled peptide distribution from time 0: $(1 - \alpha) \times (1 - r)$.

r is the proportion of the labeled peptide distribution that were newly labeled after time 0. In label incorporation experiments, $1 - r = 0$, so the proportions of unlabeled and labeled/newly synthesized group are defined as α and $1 - \alpha$, respectively (Figure 2B and eq 10).

Automated Workflow of Data Extraction and Analysis

We have developed a novel automated workflow using an algorithm in *R* for the calculation of proteomics-scale protein turnover rates based on the above modeling strategy (Figure 3). This algorithm can process data sets collected from both isotope label incorporation and dilution experiments. After preparing the data files and setting the parameters (Supporting Document S1, Section 3), the script will operate by the following steps: (1) Necessary information including Protein ID (UniProt access number), peptide sequence, precursor molecular weight, precursor m/z , retention time and confidence level, will be pulled out from the peptide identification file (Supporting Information Table S-3). (2) After extraction, the script calculates how many nitrogen atoms are present based on the amino acid

composition of the peptide. Each isotopic channel is assigned an m/z value calculated from the monoisotopic m/z value plus the nitrogen isotope *delta* value. Extra additional channels are also included to take account of natural abundance of other isotopes (^{13}C , ^2H , ^{18}O , etc.) (3) Extracted Ion Chromatograms (EICs) are generated for each peptide at each isotope channel and retention time. The retention time and all the isotope channels associated with it for each peptide serves as the data set for all analyses. (4) To reduce chemical noises of EIC data before maximum likelihood estimation (MLE), an additional regression analysis is performed for each isotope channel against the monoisotopic channel for each peptide. (5) The distribution of normalized intensities across all of the channels are fitted to a simple binomial or beta-binomial model for each peptide at each time point by MLE to determine the isotopic *distribution abundance proportion* of old to newly synthesized peptide. (6) A scoring system is also implemented in the script to evaluate the MLE fitting results. It can be used to set the threshold of peptide selection when calculating protein turnover rates. (7) At the end, the results contain each individual peptide's html report with EICs, regression plots, fitted isotopic distribution, visual score of the fitting is generated. Based on the nonlinear regression curve of change in *distribution abundance proportion* of old (α) or newly synthesized peptide (r) against time, peptide half-life (hr) or the first-order turnover rate constant (k) is calculated.

Retention Time Correlation for Removal of Chemical Noise

Because of the broad spectral and chromatographic width of data needed for modeling both new and old peptide isotopic distributions, deconvolution of overlapping spectral data is a significant problem. A typical peptide mixture derived from a tryptic digest of cellular proteins is very likely to contain many peptides with similar m/z values and retention times. Zhang and co-workers⁴⁸ reported a method to deconvolute overlapping signals from different peptides. However, it relies on comparison in intensity between the observed and theoretical isotopic peak profile, so when the atypical isotopic peak heights are detected, it indicates the contribution of another peptide to the same m/z isotopic peaks. Lyon and colleagues³³ presented another way of removing overlapping peaks by setting the threshold of intensity ratio between the two adjacent isotopic ions, although the threshold ratio must be fixed and determined empirically.

We adapted a method from MacCoss et al.⁴⁹ that tests MS peaks for correlation of chromatographic retention properties. The method uses linear regression to reduce chemical noise and to flag overlapping spectra for removal, or reprocessing, based on the correlation coefficient. As ^{15}N -isotope effects on peptide C_{18} HPLC chromatographic retention are insignificant, all of the peaks for every ^{15}N -isotopologue of a given peptide coelute. Using this fact, the intensity ratio for any given pair of isotopic channels will remain constant over a retention time window, except noise, regardless the change in the signal intensity. First, EICs are generated for M_0 , $M + 1$, ..., $M + n$ (n = number of atoms of the labeled element plus five added channels) isotopologue peaks over a defined chromatographic window centered on the MS/MS event from which the peptide was identified (Figure 4A). The retention window and m/z extraction width may be set by the user depending on the chromatographic and MS resolution. For every time point, each EIC (from channels M_0 to $M + n$) is plotted vs the highest intensity EIC (often the M_0 or $M + 1$) to construct a series of

plots. All of the time points from each of the isotopic EICs are plotted against their corresponding M0 time points. Linear regressions of these plots are performed to obtain slopes that correspond to the intensities of each isotopologue peak relative to that of the most intense EIC. The result is a description of all peaks in the envelope normalized to the monoisotopic peak. Figure 4B shows an example of the normalization process for each isotopic channel of a peptide of 40S ribosomal protein S25-4 at 8 time points of ^{15}N -incorporation. Finally, these relative intensity values are used to describe the isotopic distributions for each time point/peptide that are subsequently fitted to combinations of binomial or beta-binomial distributions by MLE. Efficient unsupervised exclusion of overlapping analytes may be accomplished in cases where a negative slope or a poor correlation coefficient is observed. The arrayed EIC and regression plots shown in Figure 4 also allow rapid inspection of data reduction performance. An additional example of the approach removing uncorrelated spectral signal is shown in Supporting Information Figure S-1.

We tested several different linear regression models for use in our algorithm. The *least-squares regression* model (“lm”) was found to be the most reliable and stable model and was set as the default in the algorithm. We found that the linear regression approach helps remove chemical noise and overlapping signal from low intensity coeluting species, as shown in channels 14–20 of the 6 and 12 day time points of Supporting Information Figure S-1B. In some cases, very intense overlapping signal from coeluted contaminants will be assigned as *X* in the linear regression function resulting in an inaccurate estimation of all parameters and of course the final turnover rate calculation. In most such cases, the scoring method included in the algorithm (*score.visual*) is useful to identify the presence of these intense coeluting signals. It evaluates the fitting results as comparing to the observed spectra for both overall and for each time points separately. The *score.visual* function computes the percent difference of the fitted values from the observed values, and subtracts the difference from 100. The overall and individual scores at different time points are also reported in the html output file. In our studies, coeluted contaminants usually result in low scores such as <60. Based on our experience after manually examining MLE fittings of *Arabidopsis* peptides, peptides with visual scores greater than 80 were selected for further analysis. If obviously flawed spectral fitting results are identified in the output report, users can pull out the raw data, manually remove signals from coeluted contaminants and process the data using the standalone R package (*isotopelabelling*). Another possible approach could take advantage of the similar peak shape in EICs between each isotopic ion. As each isotopic ion of the same peptide should have similar peak shape across the defined retention window in EICs, it should be easy to identify the ions of coeluted contaminants in EIC plots.

Calculation of Isotopic Enrichment and Fractional Abundance of Old and New Peptides Using Maximum Likelihood Estimation

Our algorithm employs mathematical modeling coupled with statistical analysis to (1) deconvolute overlapping peptide isotopologue distributions, and (2) to estimate isotopic enrichment and fractional abundance for those distributions representing old (pre-existing before label-shift) and newly synthesized peptides. Using a simple binomial distribution of natural abundance carbon, hydrogen, oxygen and sulfur and a variable isotopic abundance of

nitrogen isotopes, the MLE performed very well to model the complex spectra with good fit. A somewhat broader distribution than a pure binomial was typically observed for the peaks associated with newly synthesized peptides, which may be due to variation in the isotopic abundance across the sampled tissues. For this reason, a mixed stochastic/binomial distribution, or beta-binomial distribution was evaluated as a more appropriate model.

1. Binomial vs beta-binomial—As noted, the newly synthesized peptide distribution always has a broader distribution than expected for a pure binomial, and the distribution is often somewhat asymmetric with a longer tail on the heavier m/z end in label dilution experiments. This phenomenon might be due to averaging of a range of cells or cell types within a sampled tissue representing a range of ages or metabolic activities that vary significantly in net isotopic label incorporation. To correct for this stochastic component, a beta-binomial model was tested for fitting peptide isotopic distributions, which adds an additional random variable for broadening and shaping the binomial distribution (eq 4). Data sets using both ^{15}N -label incorporation and ^{15}N -label dilution were used for the test. As shown in Figure 5, the beta-binomial model clearly gives a better fit of the peptide (NPEDIPWAEAGADYVVESTGVFTDKDK) isotope distribution for these samples. This conclusion was also confirmed by comparing the fitted distributions to the observed original distributions by *chi*-square test. Both beta-binomial and binomial models have been integrated to the algorithm as user-defined parameters allowing users to decide which is most appropriate for their own data set. Judging from the results of this study and previous observations,⁵⁰ labeling with ^{15}N was found to be generally better than using $^{13}\text{CO}_2$ for the purpose of protein turnover measurements due to the fact that the resulting peptides have fewer nitrogen than carbon atoms resulting in more compact distributions that do not dilute the signal over as many peaks.¹⁶ Interestingly, the binomial model seems to be a better model for fitting ^{13}C -labeled peptide spectra because the lower degrees of freedom of this model improve the convergence stability over the much broader ^{13}C isotope distributions. In these cases, however, the fitted result for the newly synthesized peptide distributions do not adequately capture the stochastic broadening within these widespread distributions.¹⁶ With the earlier versions of the algorithm, the MLE distribution fitting approach often failed due to failure to converge when a beta-binomial model was used for ^{13}C spectra. The current version of the algorithm has been improved to reduce the convergence failure rates.

Analysis of *Arabidopsis* Proteome Turnover: ^{15}N -incorporation experiment

1. ^{15}N -Incorporation—Extracted from *Arabidopsis* root tissue, 348 proteins (1,850 peptides) of an enriched soluble fraction, 406 proteins (2,115 peptides) of an enriched organellar fraction, and 1,008 proteins (6,286 peptides) from a microsomal fraction were analyzed using our automated workflow to calculate the protein half-lives using the ^{15}N isotopic label incorporation approach. The final products generated from *ProteinTurover* have also been deposited to the ProteomeXchange Consortium via the MassIVE partner repository with the dataset identifiers PXD003481, PXD003482, and PXD003483. Each data set includes every peptide's html report, EIC plot, plots of fitted relative isotope abundance (linear regression method), fitted spectral models, and a html summary file providing individual peptide half-life (as rate of constant k or $\log_2 k$) and values of calculated parameters. Figure 6A shows the changes in isotopic distributions for the tryptic peptide

GILGYTEDDVVSTDFVGDNR of *Arabidopsis* glyceraldehyde-3-phosphate dehydrogenase C subunit 1 (GAPC1; UniProt accession ID P25858) as they occur in discrete time points after switching from growth on unlabeled- to ^{15}N -media. The panels on the left show the raw MS spectra and the panels on the right show the processed and fitted spectra with the deconvolution of old and new peptide distributions. When the beta-binomial distribution is selected for fitting, the observed distributions are fitted and parametrized by a maximum likelihood estimate into two isotopic distributions: a binomial natural abundance (old peptide) distribution and a beta binomial variable abundance (newly synthesized peptide) distribution as shown in each processed spectrum. At time 0 only natural abundance is present in the spectrum. 4–8 h after plants are transferred to the unlabeled medium, an additional peptide distribution appeared in the spectra. The mass shift (labeling ratio) of newly synthesized peptide distribution (pi) and the *distribution abundance proportion* of newly synthesized peptide (r) and old peptide distributions (α) can be derived from the fitted distribution parameters at each time point. In order to measure protein turnover in this study, we expected that the increase in the *distribution abundance proportion* of new peptide (synthesized after label change), r_j , or the decrease in the *distribution abundance proportion* of old peptide (synthesized prior to label change), α_j , over time (eq 6) should follow first order kinetics and it could be fitted by nonlinear regression to a first order curve functions (Figure 6B). In the case of GAPC1 protein, which is enriched in the soluble fraction of *Arabidopsis* seedling roots, the half-life was calculated to be 28.9 ± 0.72 h.

Newly synthesized peptide distributions also show a pronounced mass shift in the isotopic distribution during the labeling time course that can be described using the average probability of ^{15}N in the newly synthesized peptide (pi), as shown in Figure 6C. Unexpectedly high values of pi from the two peptides DDVELVAVNDPFITTEYMTYMFK and VVISAPSK were acquired at the earliest time point (4 h), although there is no obvious outlier data point in the nonlinear regression curve of α term against time (Figure 6B). It is worth noting that these two peptides contain multiple amino acid residues that can vary significantly in their turnover rates within the free amino acid pool as reported by Chen et al.¹⁶ and Yang et al.¹⁸ in *Arabidopsis* seedlings. Furthermore, it should be noted that the kinetics of the pi term reflect the rate of exchange within the free amino acid pool used for protein synthesis and is not useful for calculating protein turnover rates. In our isotope distribution model, pi is the mean of the discrete random variable, and is the probability of labeling ratio in a newly synthesized peptide, calculated from the probability mass function (eq 5 and 9). In other words, pi represents the average probability of ^{15}N chosen from the amino acid pool, which contains the mixture of ^{14}N and ^{15}N at each measured time point when the new peptide is synthesized, reflecting combined information on amino acid turnover, protein turnover, and also the recycling of ^{15}N in plants. Although values of pi could not be used to calculate peptide half-life directly, this provides a novel proteomic measurement for studies of protein metabolism. The rate of isotopic enrichment derived by fitting the pi term to an exponential rise curve can be used to determine the upper limit on the measurability of a protein's turnover rate, for as once label incorporation becomes the primary contributing step to rate limitation, the actual contribution of protein turnover becomes inaccessible via this method.

For label incorporation experiments, the algorithm can either model the α_j values using the kinetics of protein synthesis or not. To assume the change of α_j over time follows first-order reaction kinetics (eq 1), use $\alpha = k$ or $\alpha = \log_2 k$ in the R command lines. By doing this, the spectral fittings will also be affected based upon this assumption. To make no such assumption and fit each α_j separately, use $\alpha = \text{many}$ in the R command lines instead, which assumes the value α_j at time j could have many possibilities without considering any kinetics of protein synthesis and degradation. This will be closer to the raw data, but modeling the kinetics can provide additional information, including the standard error of turnover rate k or $\log_2 k$. This standard error is computed based on the theoretical nonlinear regression curve of α_j over time, and could be used to differentiate between good and bad fitting peptides as peptides with skewed spectral fitting result in huge standard errors.

In order to filter out those peptides with bad spectral fitting, based on the experience in manual inspection, peptides which were identified at fewer than 3 labeling time points, or which have visual scores less than 80 and standard errors greater than 10 were dropped out. These selection principles could be applied in the *fits.csv* file generated during the data processing by *ProteinTurnover*. With those selected peptides, the protein turnover rate was calculated by averaging peptide turnover rate. Using α model as “ $\log_2 k$ ” and M_j model as “*one*” in *ProteinTurnover*, 1845, 1998, and 6225 peptides with turnover data were generated in the enriched soluble, organellar and microsomal fractions of *Arabidopsis* seedling roots, respectively. Among those, 345, 321, and 798 peptides were selected in the soluble, organellar and microsomal fractions, respectively. As only proteins with at least 2 unique peptides were reported (Supporting Information Table S-1, turnover results further analyzed in R using general command lines for data organization) in this study, we were able to measure turnover rates of 64, 54, and 133 proteins in soluble, organellar and microsomal fractions, respectively. The distributions of these protein turnover rates ($\log_2 k$) are shown for comparison purpose as histograms in Figure 7. The median turnover rate ($\log_2 k$) is -5.31 , -5.59 , and -5.47 for the root soluble, organellar and microsomal fraction, respectively, corresponding to median protein half-lives is 27.66, 33.48, and 30.77 h. Among those proteins, Ubiquitin-NEDD8-like protein RUB1/RUB2 (At1g31340/At2g35635) exhibited the shortest half-life (7.97 h); while proteins like TRAF-like protein (At1g58270), Probable aquaporin PIP1-4 (At4g00430) and Myrosinase-binding protein-like protein-300B (At3g16440) enriched in the membrane fractions are the ones with longest half-lives (>90 h). Myrosinase 5 (Beta-glucosidase 35; At1g51470) enriched in the microsomal fraction also appeared to have slow turnover rate (half-life 78.92 h) in our study. Among these 251 unique proteins, only 3 proteins had coefficients of variation more than 0.20 (0.25 as the highest) in turnover rate $\log_2 k$. Another 6 proteins had coefficients of variation ranging from 0.11 to 0.20 while the rest had less than 0.10 in turnover rate $\log_2 k$, indicating low dispersion of those selection peptide turnover rates in general. Table 1 lists the protein turnover rate (values of $\log_2 k$) and half-lives of 10 proteins with the most identified unique peptides in each of the enriched soluble, organellar and microsomal fractions of *Arabidopsis* seedling roots. As some proteins like ATP synthase subunit beta-1/beat-2/beat-3, beta-glucosidase 23, or V-ATPase subunit A (Supporting Information Table S-1) were identified in multiple protein fractions, it is worth noting that the turnover rates of those proteins were quite similar in both organelle and microsomal fractions whereas beta-glucosidase 23, which was

identified in all three fractions, had longer half-life in the soluble fraction (55.40 h) than the organelle (35.11 h) or microsomal fraction (33.39 h).

Using this approach the intrinsic protein turnover rates will be convoluted with the rate of increase in total organismal protein and free amino acid pools as the plant grows. In bacteria or animal cell culture the turnover rates are often deconvoluted by correcting for growth rate as measured by cell number assuming homogeneous cellular protein/amino acid content. In intact plants, where a significant component of growth is accomplished by vacuolar expansion and uptake of water, a direct measurement of change in the abundances of individual proteins would be needed to deconvolute growth from each turnover rate. If one accepts that protein turnover rates are typically much faster than the plant growth rate, then the uncorrected numbers will be reasonable approximations of the true protein turnover rates. In cases where the protein turnover rate is very slow there will be a much larger plant growth contribution and correction should be considered.

2. ¹⁵N-Dilution—For the label dilution experiments, since the current protein search software (Mascot, MaxQuant, OMSSA, etc.) are designed to search for only either unlabeled and fully labeled peptides but not a mixture of fully labeled (old) and partially labeled (newly synthesized) peptides as we observed in our spectra, originally we could not obtain reliable peptide identifications for our samples from the label dilution experiments. To help with protein identification, we spiked every labeled sample at different time points with fixed amounts of unlabeled control, which would coelute with the labeled peptides. We successfully identified several common peptides of RuBisCO LS in all samples (Supporting Information Table S-2). The preliminary version of this algorithm has been used in previously published studies and half-lives of specific proteins like Cullin-Associated and Neddylation Dissociated 1 (CAND1) and Transport Inhibitor Response 1 (TIR1) were calculated prior to implementation of the automated workflow.^{18,51}

To measure the turnover rate of RuBisCO LS, total proteins were extracted from ¹⁵N-labeled *Arabidopsis* leaves 0, 2, 6, 12 days after being chased with unlabeled medium. Protein samples spiked with known amount of the unlabeled proteins at 4:1 ratio were run on 1-D gels and RuBisCO LS protein band around 52 kDa was excised and in-gel digested with trypsin before analysis on LC-MS/MS. Supporting Information Figure S-2A shows the raw MS data and fitted distributions of RuBisCO LS peptide DTDILAAFR. Twenty peptides of RuBisCO LS were analyzed by the algorithm. The nonlinear regression graphs for label loss (pi) and change in abundance fraction (r ; inset) for the complete set of observed RuBisCO LS peptides fitted simultaneously (Supporting Information Figure S-2B). The half-life of RuBisCO LS calculated using r data is 54.24 h. We found that most outliers were due to overlapping and/or low signal-to-noise spectra although we still used all data points for nonlinear regression. Our algorithm provides both individual peptide turnover rate outputs with corresponding protein ID, and thus the protein half-life could be calculated by combining data from all uniquely identified peptides for each protein (Supporting Information Table S-2).

In theory, the labeling ratio of newly synthesized peptides should shift back to the natural abundance ratio after infinite dilution with unlabeled medium. However, our results showed

that the decrease in labeling ratio (mass shift) of the newly synthesized peptide seems to be biphasic, with a rapid initial drop followed by a slow phase (plateau) as shown in Supporting Information Figure S-2B. This phenomenon might be the result of metabolically inactive cells/compartments or pools where turnover is occurring at a dramatically slower rate. This phenomenon has also been observed in kinetic flux profiling studies of metabolism, where it was at least tentatively explained by evoking metabolically inactive metabolite pools.⁵²

3. Comparison of using different parameters α and M_j —When setting the parameters before analysis, two parameters: α and M_j , could be set as fixed or variable depending on different label approaches and models of spectral fitting (eq 9). When the label dilution approach is chosen for turnover calculation and a fixed amount of spiked unlabeled control is added to the labeled samples, the *distribution abundance proportion* of unlabeled peptides to labeled peptides (α) is usually set as a fixed value during maximum likelihood estimation. However, it can be set as an unfixed variable (“*many*”) for the isotope label incorporation experiments as α becomes a variable that changes over time due to protein turnover. The unfixed α also allows the evaluation of the variation of the spiking amount of the unlabeled control in the label dilution experiments (Supporting Information Table S-2).

The α parameter can also be set as “ k ” or “ $k_{plateau}$ ”, which will affect how the spectra fit to the beta-binomial distribution (as described in the R script in the Supporting Document S1, Section 3). Using “ k ”, the spectral fit by MLE will assume the change of α over time fits the exponential decay curve and the y -intercept is always 1; while using “ $k_{plateau}$ ”, it will assume the change of α over time fits the exponential decay with a plateau. Ideally the results of using “ k ” and “*many*” should be the same if the proteome turnover follows the first order kinetics. Taking the same GAPC1 protein as example, the average first-order rate constant of turnover among all 11 GAPC1 peptides is 0.0225, so the half-life of GAPC1 protein is 30.81 ± 2.39 h. In theory, the protein half-life calculated using α parameter as “*many*” or “ k ” should be the same if the turnover follows the first-order kinetic mechanism and as expected, the calculated half-life of GAPC1 using either method yields similar results (28.9 ± 0.72 h vs 30.81 ± 2.39 h). On the other hand, using α model “ k ” or “ $\log_2 k$ ” has another advantage in facilitating the automation process of turnover rate calculation: the standard error of k or $\log_2 k$ of each peptide calculated from the fitting could be used to select peptides when calculating protein turnover rates.

In addition, as described above, the parameter M_j of the beta-binomial model can be set as a fixed or unfixed variable during MLE (as changed the setting of “p\$M.model” in the Supporting Document S1, Section 3). As shown in eq 5, M_j is used to reflect the pattern of over-dispersion. In theory M_j should be allowed to be a changing variable to reflect the fact that the labeling ratios of different amino acids are not at the same rates at different time points in the label experiments. However, if M_j is set as “*many*”, sometimes overfitting occurs, thus resulting in a poorly predictive isotopic distribution. In our studies of (*Arabidopsis* root proteome, setting M_j as “*many*” with α as “ $k_{plateau}$ ” generates similar results with M_j as “*one*” and α as “ k ” or “*many*”).

CONCLUSIONS

We described a novel approach for the measurement of protein turnover using metabolic labeling coupled with the LC-MS/MS analysis. We developed an algorithm to allow supervised or unsupervised high throughput data processing, which prepares raw spectral data, and fits and deconvolutes old and newly synthesized peptide distributions via a mathematical modeling approach and statistical maximum likelihood estimation that accommodates dynamic shifts in isotopic enrichment during turnover that has been observed in plants and other slow growing organisms. This algorithm is written in *R*, in open source code; the *R* packages run on Linux and Windows OS and have been made publically available. This algorithm can be used for protein turnover analysis of proteomic-scale data sets using ^{13}C , ^{15}N , ^{18}O , and ^2H stable isotopic labels using either label-dilution or label-incorporation strategies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

A significant portion of the preliminary proteomics data was collected at the Center for Mass Spectrometry and Proteomics at the University of Minnesota, and we thank LeeAnn Higgins, Todd Markowski, and Bruce Witthuhn for their help with sample preparation and LC-MS/MS analysis. We were also assisted by the Minnesota Supercomputing Institute in turnover data analysis and thank John Chilton and Pratik Jagtap for their help with the Galaxy-P platform. The authors also thank Thomas F. McGowan, Sanford Weisberg, for helping with the preliminary design of the algorithm and Xiao-Yuan Yang for testing the early version of the algorithm. We are grateful for funding provided by the NSF Plant Genome Research Program grants DBI-0606666, IOS-0923960, IOS-1238812, and IOS-1400818 as well as NSF grant IOS-0820940, NIH grant GM067203, the University of Minnesota Informatics Institute, and by the Gordon and Margaret Bailey Endowment for Environmental Horticulture.

ABBREVIATIONS

ATS	<i>Arabidopsis thaliana</i> salts medium
EIC	extracted ion chromatogram
LC-MS	liquid chromatography–mass spectrometry
MLE	maximum likelihood estimate
RuBisCO LS	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit

References

- Schwanhauser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. Global quantification of mammalian gene expression control. *Nature*. 2011; 473:337–342. [PubMed: 21593866]
- Behrends C, Harper JW. Constructing and decoding unconventional ubiquitin chains. *Nat Struct Mol Biol*. 2011; 18:520–528. [PubMed: 21540891]
- Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*. 2012; 13:227–232. [PubMed: 22411467]

4. Jamet E, Roujol D, San-Clemente H, Irshad M, Soubigou-Taconnat L, Renou JP, Pont-Lezica R. Cell wall biogenesis of *Arabidopsis thaliana* elongating cells: transcriptomics complements proteomics. *BMC Genomics*. 2009; 10:505. [PubMed: 19878582]
5. Ponnala L, Wang Y, Sun Q, van Wijk KJ. Correlation of mRNA and protein abundance in the developing maize leaf. *Plant J*. 2014;1–17.
6. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol*. 1999; 19:1720–1730. [PubMed: 10022859]
7. Schoenheimer, R. The dynamic state of body constituents. Harvard University Press; Cambridge, MA: 1942.
8. Doherty MK, Beynon RJ. Protein turnover on the scale of the proteome. *Expert Rev Proteomics*. 2006; 3:97–110. [PubMed: 16445354]
9. Hughes C, Krijgsveld J. Developments in quantitative mass spectrometry for the analysis of proteome dynamics. *Trends Biotechnol*. 2012; 30:668–676. [PubMed: 23107010]
10. Mosteller RD, Goldstein RV, Nishimoto KR. Metabolism of individual proteins in exponentially growing *Escherichia coli*. *J Biol Chem*. 1980; 255:2524–253. [PubMed: 6987224]
11. Larrabee KL, Phillips JO, Williams GJ, Larrabee AR. The relative rates of protein synthesis and degradation in a growing culture of *Escherichia coli*. *J Biol Chem*. 1980; 255:4125–4130. [PubMed: 6989832]
12. Marko NF, Dieffenbach PB, Yan G, Ceryak S, Howell RW, McCaffrey TA, Hu VW. Does metabolic radiolabeling stimulate the stress response? Gene expression profiling reveals differential cellular responses to internal beta vs. external gamma radiation. *FASEB J*. 2003; 17:1470–1486. [PubMed: 12890701]
13. Hu VW, Heikka DS, Dieffenbach PB, Ha L. Metabolic radiolabeling: experimental tool or Trojan horse? (35S)-Methionine induces DNA fragmentation and p53-dependent ROS production. *FASEB J*. 2001; 15:1562–1568. [PubMed: 11427488]
14. Beynon RJ, Pratt JM. Metabolic labeling of proteins for proteomics. *Mol Cell Proteomics*. 2005; 4:857–872. [PubMed: 15849272]
15. Ishihara H, Obata T, Sulpice R, Fernie AR, Stitt M. Quantifying protein synthesis and degradation in *Arabidopsis* by dynamic ¹³CO₂ labeling and analysis of enrichment in individual amino acids in their free pools and in protein. *Plant Physiol*. 2015; 168:74–93. [PubMed: 25810096]
16. Chen WP, Yang XY, Harms GL, Gray WM, Hegeman AD, Cohen JD. An automated growth enclosure for metabolic labeling of *Arabidopsis thaliana* with ¹³C-carbon dioxide - an in vivo labeling system for proteomics and metabolomics research. *Proteome Sci*. 2011; 9:9. [PubMed: 21310072]
17. Cargile BJ, Bundy JL, Grunden AM, Stephenson JL. Synthesis/degradation ratio mass spectrometry for measuring relative dynamic protein turnover. *Anal Chem*. 2004; 76:86–97. [PubMed: 14697036]
18. Yang XY, Chen WP, Rendahl AK, Hegeman AD, Gray WM, Cohen JD. Measuring the turnover rates of *Arabidopsis* proteins using deuterium oxide: an auxin signaling case study. *Plant J*. 2010; 63:680–695. [PubMed: 20525007]
19. Fuller JC, Nissen SL, Huiatt TW. Use of ¹⁸O-labelled leucine and phenylalanine to measure protein turnover in muscle cell cultures and possible futile cycling during aminoacylation. *Biochem J*. 1993; 294(Pt 2):427–433. [PubMed: 8373357]
20. Pratt JM, Petty J, Riba-garcia I, Robertson DHL, Gaskell SJ, Oliver SG, Beynon RJ. Dynamics of protein turnover, a missing dimension in proteomics. *Mol Cell Proteomics*. 2002; 1:579–591. [PubMed: 12376573]
21. Martin SF, Munagapati VS, Salvo-Chirnside E, Kerr LE, Le Bihan T. Proteome turnover in the green alga *Ostreococcus tauri* by time course ¹⁵N metabolic labeling mass spectrometry. *J Proteome Res*. 2012; 11:476–486. [PubMed: 22077659]
22. Price JC, Guan S, Burlingame A, Prusiner SB, Ghaemmaghami S. Analysis of proteome dynamics in the mouse brain. *Proc Natl Acad Sci U S A*. 2010; 107:14508–14513. [PubMed: 20699386]
23. Savas JN, Toyama BH, Xu T, Yates JR, Hetzer MW. Extremely long-lived nuclear pore proteins in the rat brain. *Science*. 2012; 335:942. [PubMed: 22300851]

24. Crespi HL, Archer SM, Katz JJ. Culture of algae and other micro-organisms in deuterium oxide. *Nature*. 1959; 184:729–730. [PubMed: 13812799]
25. Bal AK, Gross PR. Suppression of mitosis and macromolecule synthesis in onion roots by heavy water. *J Cell Biol*. 1964; 23:188–193. [PubMed: 14228511]
26. Jones PJH, Leatherdale ST. Stable isotopes in clinical research: safety reaffirmed. *Clin Sci*. 1991; 80:227–280. [PubMed: 1850682]
27. Koletzko B, Sauerwald T, Demmelmair H. Safety of stable isotope use. *Eur J Pediatr*. 1997; 156(Suppl):S12–S17. [PubMed: 9266209]
28. Schwanhaussner B, Gossen M, Dittmar G, Selbach M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics*. 2009; 9:205–209. [PubMed: 19053139]
29. Rao PK, Roxas BAP, Li Q. Determination of global protein turnover in stressed mycobacterium cells using hybrid-linear ion trap fourier transform mass spectrometry. *Anal Chem*. 2008; 80:396–406. [PubMed: 18085750]
30. Zhang Y, Reckow S, Webhofer C, Boehme M, Gormanns P, Egge-Jacobsen WM, Turck CW. Proteome scale turnover analysis in live animals using stable isotope metabolic labeling. *Anal Chem*. 2011; 83:1665–1672. [PubMed: 21314131]
31. Guan S, Price JC, Prusiner SB, Ghaemmaghami S, Burlingame AL. A data processing pipeline for mammalian proteome dynamics studies using stable isotope metabolic labeling. *Mol Cell Proteomics*. 2011; 10:M111.010728.doi: 10.1074/mcp.M111.010728.
32. Guan S, Price JC, Ghaemmaghami S, Prusiner SB, Burlingame AL. Compartment modeling for mammalian protein turnover studies by stable isotope metabolic labeling. *Anal Chem*. 2012; 84:4014–4021. [PubMed: 22444387]
33. Lyon D, Castillejo MA, Staudinger C, Weckwerth W, Wienkoop S, Egelhofer V. Automated protein turnover calculations from ^{15}N partial metabolic labeling LC/MS shotgun proteomics data. *PLoS One*. 2014; 9:e94692. [PubMed: 24736476]
34. Rockwood AL, Van Orden SL. Ultrahigh-speed calculation of isotope distributions. *Anal Chem*. 1996; 68:2027–2030. [PubMed: 21619291]
35. Hinkson IV, Elias JE. The dynamic state of protein turnover: It's about time. *Trends Cell Biol*. 2011; 21:293–303. [PubMed: 21474317]
36. Gibeault DM, Hulett J, Cramer GR, Seemann JR. Maximal biomass of Arabidopsis using a simple, low-maintenance hydroponic method and favorable environmental conditions. *Plant Physiology*. 1997; 115:317–319.
37. Huttlin EL, Hegeman AD, Harms AC, Sussman MR. Comparison of full versus partial metabolic labeling for quantitative proteomics analysis in Arabidopsis thaliana. *Mol Cell Proteomics*. 2007; 6:860–881. [PubMed: 17293592]
38. Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal Chem*. 1996; 68:850–858. [PubMed: 8779443]
39. Sun L, Zhu G, Dovichi NJ. Comparison of the LTQ-Orbitrap Velos and the Q-Exactive for proteomic analysis of 1–1000 ng RAW 264.7 cell lysate digests. *Rapid Commun Mass Spectrom*. 2013; 27:157–162. [PubMed: 23239329]
40. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egerton J, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol*. 2012; 30:918–920. [PubMed: 23051804]
41. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH. Open mass spectrometry search algorithm. *J Proteome Res*. 2004; 3:958–964. [PubMed: 15473683]
42. Searle BC. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics*. 2010; 10:1265–1269. [PubMed: 20077414]
43. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*. 2010; 11:R86. [PubMed: 20738864]
44. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, et al. Galaxy: A platform for interactive large-scale genome analysis. *Genome Res*. 2005; 15:1451–1455. [PubMed: 16169926]

45. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Galaxy Taylor J. A web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol*. 2010;1–21. [PubMed: 20373502]
46. Mook, WG. Environmental Isotopes in the Hydrological Cycle Principles and Applications, Vol. I, INTRODUCTION: Theory, Methods, Review. In: Mook, WG., editor. IHP-V Technical Documents in Hydrology. Vol. 39. UNESCO/IAEA; Paris/Vienna: 2000.
47. Li L, Nelson CJ, Solheim C, Whelan J, Millar AH. Determining degradation and synthesis rates of Arabidopsis proteins using the kinetics of progressive ^{15}N labeling of two-dimensional gel-separated protein spots. *Mol Cell Proteomics*. 2012; 11:M111.010025.doi: 10.1074/mcp.M111.010025.
48. Zhang X, Hines W, Adamec J, Asara JM, Naylor S, Regnier FE. An automated method for the analysis of stable isotope labeling data in proteomics. *J Am Soc Mass Spectrom*. 2005; 16:1181–1191. [PubMed: 15922621]
49. MacCoss MJ, Wu CC, Liu H, Sadygov R, Yates JR. A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. *Anal Chem*. 2003; 75:6912–6921. [PubMed: 14670053]
50. Nelson CJ, Huttlin EL, Hegeman AD, Harms AC, Sussman MR. Implications of ^{15}N -metabolic labeling for automated peptide identification in *Arabidopsis thaliana*. *Proteomics*. 2007; 7:1279–1292. [PubMed: 17443642]
51. Sauer MLA, Xu B, Sutton F. Metabolic labeling with stable isotope nitrogen (^{15}N) to follow amino acid and protein turnover of three plastid proteins in *Chlamydomonas reinhardtii*. *Proteome Sci*. 2014; 12:14. [PubMed: 24580857]
52. Szecowka M, Heise R, Tohge T, Nunes-Nesi A, Vosloh D, Huege J, Feil R, Lunn J, Nikoloski Z, Stitt M, et al. Metabolic fluxes in an illuminated *Arabidopsis* rosette. *Plant Cell*. 2013; 25:694–714. [PubMed: 23444331]

Figure 1.

Label incorporation or dilution over time. (A) Protein turnover when nutrient assimilation and protein synthesis are fast in organisms with very small internal unlabeled (label incorporation) or labeled (label dilution) pools. Simulated mass spectra are shown for a peptide derived from protein in two different metabolic labeling strategies. *Label incorporation* occurs when an organism grown on natural abundance media is switched to a stable isotope (such as ^{15}N) labeled media at the start of a turnover experiment; *label dilution* occurs when an organism is grown on labeled media and switched to unlabeled. Either approach allows one to differentiate between the peptide signals from preexisting and newly synthesized protein. In this special case, nutrients are effectively directly assimilated into new protein and discrete and consistent isotope distributions are observed for light and heavy isotopic envelopes. Protein turnover can be derived from the simple ratios of these distributions via a number of different methods. (B) Protein turnover when internal amino acid pools are large or have significant non-nutrient derived inputs, such as protein degradation or synthesis from stored nutrients. In this case, identical labeling experiments are performed as in Figure 1A, but because of significant internal contributions to internal

amino acid pools (or other factors) the isotopic distribution for the peptide derived from newly synthesized protein shifts over the course of the experiment. This behavior makes it necessary to model the isotopic distribution dynamics in order to derive the relative abundance of new versus preexisting protein.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2.

Modeling peptide MS data of ^{15}N -incorporation (panels A, B) or dilution (panel C) experiments. (A) The raw MS data of the entire isotopic envelope for the +2 charged ion of peptide TEFGPSQPFK of adenosylhomocysteinase 1 at retention time 30.46–31.46 min from the 48 h-incorporated sample. (B) Processed and fitted isotopic distributions of panel (A) after correlation by linear regression and MLE fitting of the binomial distribution π_N to the natural abundance and of the beta-binomial distribution π_j to the newly synthesized peptide population. In label incorporation experiments, α is the *distribution abundance proportion* of unlabeled peptide population and $1 - \alpha$ is the *distribution abundance proportions* of labeled peptide population. (C) Spectral fitting for the entire isotopic

envelope for a +3 charged ion of peptide GGLDFTKDDENVNSQPFMR of RuBisCO LS after processing as in (B). Here, r is the *distribution abundance proportions* of newly labeled peptide population after time 0. π_N is a binomial distribution with fixed natural abundance, π_0 and π_j are beta-binomial distributions also of variable abundance.

Figure 3.

Workflow of the automated analysis algorithm. Both MS data and the protein identification results have to be first converted to a specified format by users before processing. The *ProteinTurnover* package is designed to calculate the protein turnover rates. It will first extract EICs for each peptide at each isotopic channel and retention time. Then an additional regression analysis is performed for each isotope channel to reduce chemical noises of EICs before spectral fitting. A scoring system is also used to evaluate to the spectral fitting. Using maximum likelihood estimation, the proportion of old and newly synthesized peptide distribution will then be estimated. Lastly, the analyzed data files including MLE values for each peptide and the fitted histogram for each peptide/experiment as well as all EICs, regression analysis for diagnostic purposes, and calculated half-lives (or the first-order rate constant of turnover) using nonlinear-regression will be generated along with the summary html reports. Nonlinear regression was used instead of linear regression of transformed data to avoid systematic fitting bias associated with measurement error distortion during transformation. In the html report, each peptide will have an individual link to each html report, which includes the turnover results, EIC plots, and plots of relative abundance fits and model fits.

Figure 4.

Removing chemical noise by the linear regression method. (A) EICs of 18 isotopic channels (M0 to M17) at 0, 4, 8, 16, 24, 32, 40, and 48 h postincorporation times for a tryptic peptide (LITPSILSDR) of vacuolar 40S ribosomal protein S25-4. The protein was enriched in the microsomal fraction of *Arabidopsis* seedling root tissue in the ^{15}N -incorporation experiment. (B) Algorithm output for the linear correlations of each channel (M0 to M17) to the M0 channel for all eight time points applied in Panel A.

Figure 5.

Binomial vs beta-binomial distributions for the newly synthesized peptide isotopic pattern. Differences in binomial and beta-binomial fitting are apparent for the newly synthesized peak population of the isotopic envelope for peptide LVSWYDNEWGYSSR of the glyceraldehyde-3-phosphate dehydrogenase C subunit at the 48 h time point. Beta-binomial fitting results in higher visual score than binomial fitting.

Figure 6.

Turnover of glyceraldehyde-3-phosphate dehydrogenase C subunit 1 (GAPC1) in *Arabidopsis* roots using the ^{15}N -incorporation approach. Soluble proteins were obtained from *Arabidopsis* seedling roots 0, 4, 8, 16, 24, 32, 40, and 48 h after transfer to ^{15}N -medium. (A) Pileup of raw MS data (the left panel) and spectral fitting (the right panel) of one tryptic peptide, GILGYTEDDVSTDFVGDNR, of GAPC1. (B) First order decay curve for the *distribution abundance proportion* (α) over time for all observed 11 peptides of GAPC1 protein fitted by nonlinear regression analysis. The derived equation of the exponential decay is $y = (1.0062 \pm 0.0123)e^{-(0.0240 \pm 0.0006)t}$ with an R^2 value 0.9624 for this global curve fitting of all peptides. The half-life of GAPC1 calculated from *alpha* set as “many” is 28.9 ± 0.72 h. (C) First order rise plot of the average probability of ^{15}N shown in the newly synthesized peptides (π) of GAPC1 protein against time fitted by nonlinear regression analysis. The derived equation of the exponential rise is $y = (0.5589 \pm 0.0164)(1$

$-e^{-(0.2487 \pm 0.0371)x}$). R^2 is 0.7713 (Standard Error of Estimate = 0.1057) for the global fit of all 11 peptides. R^2 could be increase to 0.9002 (Standard Error of Estimate = 0.0658) if data of peptides DDVELVAVNDPFITTEYMTYMFK and VVISAPSK were exclude. As this curve reflects the change in the isotopic enrichment of the new peptide, the plot can vary significantly depending on peptide amino acid content, as the turnover for free amino acids can vary over a couple of orders of magnitude depending on the amino acid.¹⁶

Figure 7.

Distributions of protein turnover rates across different enriched fractions of roots.

Histograms show distributions in the protein $\log_2 k$ values plotted for soluble, organellar, and microosomal fraction of roots. Only proteins with at least 2 unique peptides which were identified at more than 3 labeling time points, having visual score more than 80 and standard error of $\log_2 k$ less than 10, are reported. The y-axis is the number of protein counts. The median value is indicated with an arrow. The bin width is 0.15 for all histograms.

Scheme 1.

Labeled Nutrient Uptake and Dilution through Internal Amino Acid Stores and Metabolism

Table 1

Turnover Rates of 10 Proteins with the Most Identified Peptides from Each of the Enriched Soluble, Organelle, and Mitochondrial Fractions of Arabidopsis Seedling Roots Using *alpha* Model Set as “log₂k”^a

Fraction	UniProt ID	Protein	AGI	# pep ^b	Turnover rate (log ₂ k) ^c	SD ^d	CV ^e	Half-life (hr) ^f
Soluble ^g	Q9ASR1	Putative Elongation factor EF-2	At1g56070; At1g56075	10	-4.99	0.08	0.02	22.08
	Q9SRZ6	Cytosolic isocitrate dehydrogenase [NADP]	At1g65930	9	-5.34	0.08	0.01	28.09
	P25696	Bifunctional enolase 2/transcriptional activator	At2g36530	8	-5.46	0.39	0.07	30.61
	Q9LSB4	TSA1-like protein	At3g15950	8	-5.45	0.15	0.03	30.30
	Q9SYT0	Annexin D1	At1g35720	7	-5.26	0.08	0.02	26.49
	O23255	Adenosylhomocysteinase 1	At4g13940	6	-5.46	0.13	0.02	30.48
	Q9LF98	Fructose-bisphosphate aldolase	At3g52930	6	-5.40	0.11	0.02	29.30
	O50008	5-methyltetrahydropteroyl/triglutamate-homocysteine methyltransferase 1	At5g17920	5	-5.10	0.11	0.02	23.69
	Q9XEE2	Annexin D2	At5g65020	5	-5.37	0.26	0.05	28.65
	Q9SR37	Beta-glucosidase 23	At3g09260	5	-6.32	1.42	0.23	55.40
Organelle ^h	O04310	Jacalin-related lectin 34	At3g16460	11	-5.00	0.12	0.02	22.21
	Q9SR37	Beta-glucosidase 23	At3g09260	10	-5.66	0.09	0.02	35.11
	Q1H583	GDSL esterase/lipase 22	At1g54000	10	-5.97	0.10	0.02	43.53
	P83483_P83484_Q9C5A9	ATP synthase subunit beta-1/beta-2/beta-3, mitochondrial	At5g08670/At5g08690/At5g08680	9	-5.78	0.11	0.02	38.13
	P31167	ADP, ATP carrier protein 1, mitochondrial	At3g08580	8	-5.70	0.28	0.05	36.04
	O04314	PYK10-binding protein 1	At3g16420	8	-4.84	0.15	0.03	19.84
	P92549	ATP synthase subunit alpha, mitochondrial	AtMg01190	6	-5.77	0.15	0.03	37.77
	P0DH99	Elongation factor 1-alpha 1	At1g07940	6	-5.28	0.13	0.02	26.98
	F4JPJ7_P19456	ATPase 2, plasma membrane-type	At4g30190	5	-5.22	0.47	0.09	25.82
	Q9C525	Beta-glucosidase 21	At1g66270	5	-5.93	0.29	0.05	42.32
Mitochondrial ⁱ	P83483_P83484_Q9C5A9	ATP synthase subunit beta-1/beta-2/beta-3, mitochondrial	At5g08670/At5g08690/At5g08680	18	-5.71	0.18	0.03	36.18
	P92549	ATP synthase subunit alpha, mitochondrial	AtMg01190	11	-5.67	0.10	0.02	35.25
	Q9SR37	Beta-glucosidase 23	At3g09260	11	-5.59	0.08	0.01	33.39
	O23654	V-type proton ATPase catalytic subunit A	At1g78900	11	-5.98	0.21	0.03	43.75
	P31167	ADP, ATP carrier protein 1, mitochondrial	At3g08580	9	-5.53	0.10	0.02	31.93
	Q9FT52	ATP synthase subunit d, mitochondrial	At3g52300	8	-5.70	0.11	0.02	35.93
	Q9C8G5	CSC1-like protein ERD4	At1g30360	8	-5.96	0.23	0.04	43.12

Fraction	UniProt ID	Protein	AGI	# pep	Turnover rate (log ₂ k) ^c	SD ^d	CV ^e	Half-life (hr) ^f
	O04310	Jacalin-related lectin 34	At3g16460	8	-5.04	0.10	0.02	22.81
	P51818_P55737	Heat shock protein 90-2/90-3	At5g56030/At5g56010	7	-5.30	0.10	0.02	27.31
	Q9LKR3	Mediator of RNA polymerase II transcription subunit 37a (HSP 70-1)	At5g28540	7	-4.98	0.04	0.01	21.90

^aThe entire data set is provided in Supporting Information Table S-1.

^bNumber of peptides used for calculation. Only exclusive peptides with a visual score (calculated from the *score.visual* function, which generates the percent difference of the fitted values from the observed values, and subtracts the difference from 100) greater than 80, standard error less than 10, and 3 available time points were used for calculation.

^cThe mean of log₂ k of all selected peptides.

^dThe standard deviation of mean of all selected peptide log₂ k.

^eThe coefficient of variation of log₂ k.

^fProtein half-life was calculated from the mean of the peptide turnover rate log₂ k following the equation $t_{1/2} = \ln(2)/k$ (eq 3).

^gSoluble enriched protein fraction from the differential centrifugation (1 h, 100,000 × g, supernatant) of *Arabidopsis* root tissue homogenate.

^hOrganelle enriched protein fraction from the differential centrifugation (5 min, 1,500 × g, pellet) of *Arabidopsis* root tissue homogenate.

ⁱMicrosomal enriched protein fraction from the differential centrifugation (1 h, 100,000 × g, pellet) of *Arabidopsis* root tissue homogenate.