Published in final edited form as:

J Proteome Res. 2018 November 02; 17(11): 3740–3748. doi:10.1021/acs.jproteome.8b00417.

d2ome, Software for in Vivo Protein Turnover Analysis Using Heavy Water Labeling and LC-MS, Reveals Alterations of Hepatic Proteome Dynamics in a Mouse Model of NAFLD

Rovshan G. Sadygov^{*,†}, Jayant Avva^{†,§}, Mahbubur Rahman^{†,||}, Kwangwon Lee[‡], Sergei Ilchenko[‡], Takhar Kasumov[‡], and Ahmad Borzou[†]

[†] Department of Biochemistry and Molecular Biology, Sealy Center for Molecular Medicine, The University of Texas Medical Branch, 301 University Blvd., Galveston, Texas 77555, United States

[‡] Department of Pharmaceutical Sciences, Northeast Ohio Medical University, Rootstown, Ohio 44272, United States

Abstract

Metabolic labeling with heavy water followed by LC-MS is a high throughput approach to study proteostasis in vivo. Advances in mass spectrometry and sample processing have allowed consistent detection of thousands of proteins at multiple time points. However, freely available automated bioinformatics tools to analyze and extract protein decay rate constants are lacking. Here, we describe d2ome—a robust, automated software solution for in vivo protein turnover analysis. d2ome is highly scalable, uses innovative approaches to nonlinear fitting, implements Grubbs' outlier detection and removal, uses weighted-averaging of replicates, applies a data dependent elution time windowing, and uses mass accuracy in peak detection. Here, we discuss the application of d2ome in a comparative study of protein turnover in the livers of normal vs Western diet-fed LDLR^{-/-} mice (mouse model of nonalcoholic fatty liver disease), which contained 256 LC-MS experiments. The study revealed reduced stability of 40S ribosomal protein subunits in the Western diet-fed mice.

Graphical abstract

*Corresponding Author rgsadygo@utmb.edu. Phone: (409)772-3287.

§Present Addresses 870 Dundas Street East, #103, Toronto, Ontario M4M 1R3, Canada.

|| VizExplorer, San Diego, California 92127, United States.

Author Contributions

RGS designed the software components, developed and implemented the peptide information parsing and integration, the models for time course evaluations, parameter learning, software integration, and tested the software. JA was involved in early manuscript editing and in the discussions. MR was involved in MPE calculations and incorporation of NNLS. AB designed and implemented Grubbs' outlier detection and removal, worked on software integration and extensive testing. The North East Ohio Medical University group performed the in vivo labeling of NAFLD mice and provided the mass spectral data: TK designed and performed the metabolic labeling study, KL helped with animal experiments and prepared samples, and SI performed LC-MS analysis.

The authors declare no competing financial interest.

Keywords

in vivo protein turnover; proteome dynamics; metabolic labeling; nonlinear least-squares modeling; NAFLD; peak detection and integration; protein half-life; UPR; 40S ribosomal proteins; isotopomer quantification

INTRODUCTION

The cellular proteome is a complex microcosm of structural and regulatory networks that is continuously controlled and modified to meet the dynamic needs of the cell.¹ The proteostasis network maintains proteins in the appropriate abundance, folding state, concentration, and location. The network includes protein translational machinery, a chaperone-based folding system, and an autophagy-and proteasome-based degradome. Disruptions in proteostasis are associated with multiple diseases² such as neurological disorders and diseases associated with metabolic syndrome.

Metabolic stable-isotope labeling followed by liquid chromatography and high resolution and accuracy mass spectrometry (LC-MS) has become a powerful tool for high throughput studies of protein turnover in vivo.^{3–6} A traditional protein turnover study using proteomics involves steady-state administration of a tracer precursor, sample collection at different time points, isolation of proteins, and LC-MS analysis of tryptic peptides.⁷ High-resolution MS allows the analysis of isotopomers of peptides (both endogenous mass and heavier peptides that are enriched with labeled amino acids) from a large number of proteins in a single run, and enables the assessment of proteome dynamics. Methods have been developed to extract protein turnover rates from the time course incorporation curves of heavy isotopes.^{8,9}

Several isotope labeled precursors have been used for in vivo metabolic labeling¹⁰ studies. They can be divided into two groups—prelabeled amino acids and in situ labeling of amino acids using other tracer precursors, including ¹³C-glucose, ¹⁵N-alginate, or ²H₂O. Among these, the ²H (deuterium) labeling using heavy water is attractive because of its simplicity (²H₂O is provided in drinking water), rapid equilibration of water in all cell compartments of body, and cost-effectiveness. However, the data analysis from heavy water labeling is relatively more complex, because labeling experiments achieve 10% or less deuterium enrichment in the body water. Higher concentrations of deuterium could be toxic¹¹. The incomplete labeling leads to a more complicated mixture of overlapping natural and labeled isotope profiles. Bioinformatics tools deconvolve this overlap to extract the relative abundance of the deuterium labeled species.

Two bioinformatics tools have been reported to integrate and automate data analysis from heavy water labeling experiments. Dueterater¹² is publicly available Python-based software. ProTurn⁴ is the other algorithm for protein turnover estimation. It is a Java-based application. While the techniques used in the algorithm have been described,⁴ the code is not freely available. Here, we report on our freely available software, d2ome, to process data sets from heavy water labeling experiments. d2ome works with the Human Proteome Organizations' file standards for spectral data (mzML) and database search results (mzIdentML). It has automated peak detection (based on mass accuracy) and integration, extracts relative abundances of labeled peptides, and uses time course modeling to determine protein turnover rates. We have used several innovative approaches to data types that are specific to the protein decay rates. Thus, d2ome makes use of parameter transformation in the nonlinear regression fit, and produces only non-negative values for the rate constants. It uses the abundance-based weighted averaging of the isotopomer measurements to reduce sensitivity to noisy signals. To more faithfully represent the elution window of peptides, d2ome uses information about all elution times that triggered peptide fragmentation and MS/MS analysis. d2ome has no preset number of experiments that can be analyzed in a single study. To reduce the standard deviation due to variability in quantifying different peptides, d2ome implements Grubbs' outlier detection and removal.¹³ These features, combined with its conformity with the HUPO data formats, make d2ome a valuable tool for measuring protein half-lives from in vivo studies using heavy water metabolic labeling.

We applied d2ome to a comparative study of LDLR^{-/-} mice, a diet-induced model of nonalcoholic fatty liver disease (NAFLD). We analyzed hundreds of experiments at six time points. Comparing the data sets of normal diet (ND) and Western diet (WD) fed mice and using the protein interaction database, STRING,¹⁴ we identified several complexes with reduced stability in WD fed mice. In addition to increased degradation of hepatic mitochondrial proteins, we found that 40S ribosomal complex protein stability was reduced in WD fed mice. The 40S ribosomal complex may be related to the PERK arm of the unfolded protein response (UPR). The UPR is known to be dysregulated in NAFLD.¹⁵ However, our findings of reduced stability of the 40S ribosomal complex subunits are new. We also present a large scale comparison of the rate constants computed using our algorithm with those from published data on the mice heart proteome (calculated by ProTurn) using the same identifications and spectral inputs.⁴ It is the first large-scale comparison of proteome dynamics using different bioinformatics solutions.

METHODS

Animal experiments and measurements of body water enrichment (in deuterium) are described in the Supporting Information.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jproteo-me.8b00417](https://doi.org/10.1021/acs.jproteo-me.8b00417).

Supporting notes, figures, and tables (PDF)

Supporting Information 1: List of quantified proteins and their characteristics, WD fed mice (XLSX)

Supporting Information 2: List of quantified proteins and their characteristics, ND fed mice (XLSX)

Supporting Information 3: List of proteins which showed significant changes of degradation rate constants in transition from normal liver to fatty liver (XLSX)

Protein Fractionation and Sample Preparation for Proteomics Analysis

Hepatic proteins (30 μg) were fractionated by SDS-PAGE using AnyKD gel (BioRad, Hercules, CA) for 45 min at 180 V. After staining, gel bands corresponding to specific molecular weight protein fractions were cut and processed for proteomic analyses. Extracted gel bands were reduced with dithiothreitol (DTT), alkylated with iodoacetamide and digested with trypsin (Promega, Madison, WI) at 37 °C overnight. The tryptic peptides were extracted from polyacrylamide gel bands in a sonication bath with 100 μL of 40% acetonitrile with 0.1% trifluoroacetic acid (TFA) and 100 μL of 70% acetonitrile with 0.1% TFA, consecutively. These extracts were combined, dried in a SpeedVac, and analyzed by nanospray LC-MS/MS after reconstitution in 0.1% formic acid.

LC-MS/MS Analysis

A solution containing the tryptic peptides equivalent to 0.1 μg protein was analyzed by Ultimate 3000 UHPLC (Thermo Scientific, CA) coupled online to Q Exactive Plus Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific, CA). The samples were first loaded on an Acclaim PepMap100 precolumn (300 $\mu\text{m} \times 5 \text{ mm}$, C18, 5 μm , 100 Å, Thermo Fisher Scientific) for desalting, and then directed to an Acclaim PepMap RSLC reverse phase nanocolumn (75 $\mu\text{m} \times 15 \text{ cm}$, C18, 2 μm , 100 Å, Thermo Fisher Scientific) at 300 nL/min with mobile phase A (0.1% formic acid in water) and B (20% water in acetonitrile with 0.1% formic acid). For a chromatographic fractionation of tryptic peptides, a stepwise gradient was employed with 2% of mobile phase B. After 4 min of desalting, mobile phase B was linearly increased to 40% in 100 min. Mobile phase B was then ramped to 90% in 5 min and held at 90% B for 10 min. Subsequently, mobile phase B was decreased to 2% for 2 min and equilibrated for 13 min at 2%.

Mass spectrometry analysis was performed at m/z 380–1300 (MS) with 70 000 resolution (200 m/z). MS/MS spectra were collected in data-dependent acquisition mode for the 12 most abundant product ions with an isolation window of 1.5 m/z and 17 500 resolution (200 m/z). MS and MS/MS spectra were acquired for 100 ms with the automatic gain control (AGC) target set at 1.0×10^6 and 2.0×10^4 ions for MS and MS/MS scans, respectively. Higher-energy collisional dissociation (HCD) was performed at normalized collision energy of 25%. Dynamic exclusion was enabled for a duration of 17 s.

THEORETICAL METHODS AND APPROACHES

Peptide/Protein Identification from Tandem Mass Spectrometry

We used the Mascot¹⁶ database search engine to identify peptides/proteins from their tandem mass spectra. The following parameters were used for searches of tandem mass spectra of LDLR^{-/-} mice: precursor mass accuracy set to 15 ppm; fragment ion mass accuracy set to 0.6 Da; carbamido-methylation of Cys was the fixed modification; oxidation of Met and acetylation of Lys were set as dynamic modifications. Trypsin specificity of peptides was used and up to 2 missed cleavages were allowed. The SwissProt database (downloaded in May 2017) and mouse taxonomy were used. The false discovery rate (FDR) of peptide-spectrum match was controlled by using the decoy database approach. The FDR was set at 5%.

Isotopomer Quantification

Two quantities, relative isotope abundance (RIA) and molecular percent enrichment (MPE), have been used to quantify label incorporation into peptides in heavy water metabolic labeling. d2ome provides the isotope integration for up to 6 isotopes, and from these data any of the quantities can be calculated. The RIA at time point t is determined as

$$\text{RIA}(t) = \frac{\sum_{j=1}^5 M_{j(t)}/}{\sum_{j=0}^5 M_{j(t)}}$$

where $M_j(t)$ is the intensity of the j -th isotope at the time point t . $\text{RIA}(t)$ is related to the normalized abundance of the monoisotopic peak, $I_0(t)$ (which is used in eq 1).

$$I_0(t) = 1 - \text{RIA}(t)$$

The MPE takes account of the isotope numbers, and it is defined as

$$\text{MPE}(t) = \frac{\sum_{j=1}^5 j M_{j(t)}/}{\sum_{j=0}^5 M_{j(t)}}$$

As it is seen from the definition, MPE can be considered an expectation number of isotopes defined on the experimental distribution of the isotopomers. The modeling is done on the quantity termed net labeling ($\text{NET}(t)$). It is defined as the difference between the MPE at the current labeling time point, and the MPE of the natural isotope distribution (before the start of heavy water labeling):

$$\text{NET}(t) = \frac{\sum_{j=1}^5 j M_{j(t)}/}{\sum_{j=0}^5 M_{j(t)}} - \frac{\sum_{j=1}^5 j M_{j(0)}/}{\sum_{j=0}^5 M_{j(0)}}$$

As a default, we used the $I_0(t)$ for nonlinear regression and rate constant determination (see above). However, in addition, we also provided MPE data in the output, as it is informative of stable isotope incorporation. Relative isotope abundance can be extracted either from relative abundance of the mono-isotopic peak, or by using non-negative least-squares (NNLS) fit of the theoretical isotopomers (natural isotope distribution) to experimental distribution. The default is the relative isotope abundance of the monoisotopic peak. NNLS calculation is also provided in the algorithm.

Technical Replicates

When replicates are available, different techniques for combining signals can be applied. We compared two techniques: arithmetic averaging and intensity based averaging of relative abundance of the monoisotopic peak. In the latter approach, algorithm incorporates all replicates for each time point and assigns them weights based on the signal intensity. If there are n replicates, then the weight of the i -th measurement is calculated as

$$w_i = I_0^i / \sum_{j=1}^n I_0^j$$

where I_0^j is the integrated ion chromatogram of the monoisotope in the j -th replicate. The weighted average for the monoisotope was calculated as

$$I_0 = \sum_{j=1}^n w_j I_0^j$$

The averaged intensities for other isotopes were calculated similarly. In Figure S1, we show a comparison of fractional abundance of the monoisotopic peak of the peptide sequences “SWNETFHAR” of murine mitochondrial protein ATP synthase subunit d (ATPH5). The time course of abundance calculated using intensity-based averaging is monotonic, which allows a better fit to the decay function.

Calculations of the Degradation Rate Constant

The degradation rate constant, k , can be calculated using changes in individual isotopomers or a function of isotopomers.¹⁷ When a single isotope peak is used for calculations, the motivation is to choose the isotopomer with the largest signal-to-noise ratio which will be observed at every time point of stable isotope labeling. The normalized abundance (sum of abundances of six isotopomers is set equal to 1) of the monoisotopic peak, I_0 , satisfies these conditions. If we assume the body water enrichment level p , and equilibration between the body water enrichment and the pool of precursor amino acids, in a one-compartment model, the time course behavior of $I_0(t)$ can be modeled as¹⁸

$$I_0(t) = I_0^{\text{asympt}} - (I_0(0) - I_0^{\text{asympt}}) \times e^{-kt} \quad (1)$$

where $I_0(0)$ is the normalized abundance of the monoisotope at the time 0 (before the start of the heavy water labeling). To generate the theoretical isotopes, we employed a model for fast calculation of isotopes.¹⁹ $I_0(t)$ is normalized monoisotope intensity obtained in the experiment at the labeling time t , I_0^{asympt} is the asymptotic value of the monoisotopic peak (normalized) at the asymptote—e.g., the expected value of monoisotopic peak when all exchangeable hydrogen atoms in a peptide (protein) have achieved equilibrium with the body water enrichment levels. This is a theoretical value, and to calculate it we need to estimate the number of exchangeable hydrogens per amino acid. These numbers have been obtained from the literature^{20,21} and are hard coded into d2ome. The value of I_0^{asympt} is^{4,18}

$$I_0^{\text{asympt}} = (1 - p)^N I_0$$

where N is the number of exchangeable hydrogens in the peptide. To obtain the parameter k , the degradation rate constant, we minimized the residual sum of squares with model in eq 1.

Note that the formula is affected by the variations in body water enrichment levels that can possibly occur at different time points. The body water enrichment enters eq 1 via I_0^{asympt} . To theoretically fit the formula, the single value of I_0^{asympt} should be used. This can be achieved by rearranging the eq 1: moving the time point specific asymptote to the left side of the equation and adding to both sides of the equation a term corresponding to asymptotic enrichment expected at the body water enrichment level of the last time point that was experimentally measured.

We illustrate the effect of deuterium incorporation into a peptide in the example peptide sequence, “VPAIYGVDTTR”, from murine Carbamoyl-phosphate transferase. Figure S2 shows the precursor in full mass scan, MS, and its fragment ions in MS/MS. In Figure S3, we show the isotope envelope of this peptide at different time points after the start of metabolic labeling. The peptide has 16 exchangeable hydrogens. At 3.35% body water enrichment, the average number of deuteriums incorporated into one peptide was 0.56. The experimental isotopomer distribution of the peptide changes with deuterium incorporation, Figure S3. In Figure S4, we show the relative abundance of the monoisotopic peak for all time points and for ND and WD fed animals for this peptide. The figure shows that the relative abundance of the monoisotopic peak of WD fed mice reached its plateau by the end of the labeling duration (21 days). The curve for the ND fed mice was still in “decay” at the end of the labeling experiment. The figure shows qualitatively that decay rate constant is larger in the WD fed mice than in ND fed mice. As described above, these time course data are used to extract the peptide degradation rate constants quantitatively using a nonlinear regression.

We note that due to the different number of exchangeable hydrogens, peptide sequences of a same protein will have different time course profiles. The peptides with a smaller number of exchangeable hydrogens will have plateau levels (for relative abundance of the monoisotopic peak) higher than those with a larger number of exchangeable hydrogens. This is shown in Figure S5, where we show incorporation time course data for different peptides of the murine protein, ER chaperone, BiP (also known as glucose regulated protein 78 kDa). The number of exchangeable hydrogens of peptides of this protein are presented in Table S1.

Parameter Estimation

We used a nonlinear regression to fit the expression in eq 1 to the data and learn the parameters in the model. The optimization algorithm is the Broyden—Fletcher—Goldfarb—Shanno algorithm (BFGS).²² This algorithm optimizes the sum of squared residuals and uses a gradient of the sum with respect to the parameters. In general, BFGS searches for the parameters to fit the data in an unrestricted space, e.g., on the whole parameter axis. However, for the particular type of applications used in this study, there are restrictions on the parameters. The rate constants cannot be negative in all models; in two- and three-parameter models the starting and asymptotic enrichments cannot be larger than 1. To account for these restrictions, we introduced respective reparameterizations. Thus, the rate constant is reparameterized as an exponential function of a parameter, θ .

$$k = e^{-\theta}$$

θ is unrestricted, while k is now naturally restricted to non-negative values. A logistic function is to reparametrize I_0^{asympt} in the three-parameter model to restrict it only between 0 and 1,

$$I_0^{\text{asympt}} = 1 / \left(1 + e^{-\theta_1} \right)$$

while the parameter θ_1 is unrestricted and its optimized values is obtained from the BFGS algorithm.

Outlier Filtering Using the Grubbs' Approach

Estimation of a protein's half-life in LC—MS experiments has several components including peptide identification, peak detection and integration, and isotope incorporation. Multiple sources of variability originating from different sources may affect the results. For example, biological variability can affect several aspects, such as protein half-lives, isotope distributions (different individuals may have differing half-lives and basal distributions of isotopomers); variabilities are introduced via fluctuations in mass spectral intensity measurements, coelutions creating overlapping profiles, algorithms for peptide identifications, etc. These and other sources of errors will result in variability of half-lives of peptides of a protein. Protein half-life is obtained as a median of its peptides. To improve the robustness, we implemented a Grubbs' outlier¹³ detection algorithm (described in the Supporting Information) to filter out the outliers. The threshold p -value is a two-sided level of significance at 0.1.

The summary of the data processing workflow in d2ome is shown in Figure S6. The figure shows inputs (spectral data sets and protein/peptide identification results), processing steps (data filtering, peak detection and integration, determining relative isotope abundance, nonlinear regression, and statistical processing of outliers), and the results (quantified proteins, their rate constants, number of unique peptides, standard deviations).

RESULTS AND DISCUSSIONS

C57BL/6J Mice

For comparison of ProTurn and d2ome quantification and rate estimation, we used freely available data sets of the heart proteome of C57 mice.⁴ There were two separate data sets in this study corresponding to C57 mice with the normal heart and isoproterenol-induced hypertrophic heart. The normal heart data set had 127, and the hypertrophic heart data set had 129 LC-MS experiments obtained at seven time points. ProTurn quantified 1896 and 2018 proteins from normal and hypertrophic heart conditions, respectively. To remove the variabilities associated with peptide/protein identifications, and to compare only peptide quantifications, we used exactly the same database search identification results (and the MS/MS scans) that were reported in the original study.⁴ As mentioned above, d2ome can work with replicates, but there were no replicates reported in the final results, therefore no replicates at the peptide level were used by us. Figure 1 shows the scatter plot of the rate constant calculations from the two algorithms for the normal heart (Figure 1A) and induced

hypertrophic (Figure 1B) heart. The rate constants for the proteins that were quantified by three or more peptides are shown in the figures. Figure S7A,B show the scatter plot of rate constant calculations from two algorithms for all C57 mice heart proteins that were quantified by ProTurn in the original study.

The hypertrophic heart data set contained 2004 common proteins. The Pearson correlation between the unfiltered rate constants from ProTurn and d2ome was 0.33. When we required filtering based on the number of distinct peptides, the correlation coefficients were 0.57, 0.76, and 0.88 for proteins quantified with at least 2, 3, and 4 distinct peptides. The slope of linear regression of rate constants for proteins with at least three distinct peptides was 1.02, and intercept was 0.01.

For the normal heart data set, the Pearson correlation coefficient was 0.91 for proteins quantified by more than three distinct peptides. 818 proteins passed this filter. The correlation coefficient of proteins quantified by ten or more peptides (the list comprised 271 heart proteins) was 0.98. The slope of the linear regression of rate constants from d2ome on those to ProTurn (for proteins quantified by more than three distinct peptides) was 0.85, and the corresponding intercept was 0.01. A similar value for the slope (0.82) and intercept (0.01) were obtained for regression of the d2ome rate constants on ProTurn rate constants after filtering of proteins to retain only those with more than 10 distinct peptides.

Analyzing the structure of the rate constant values for proteins identified by three or fewer distinct peptides, we observed that they were enriched in rate constant values that were not appropriate to the experimental setup. For example, in the group of proteins quantified by fewer than four peptides, there were 57 proteins whose rate constants were larger than 0.6 day^{-1} . While in the group of proteins quantified by four or more peptides, there were only 16 proteins with such rate constants. We note that the label incorporation measurements started 1 day after initiation of the labeling. Therefore, proteins whose half-lives are shorter than 1 day (corresponds approximately to the rate constant of 0.6 day^{-1}) do not have enough experimental points in the linear portion of the label incorporation curve for accurate rate constant estimations.

For comparison of rates from the two approaches, we also computed the percent differences of the rates for each protein: $100 \times \text{abs}(R_{\text{ProTurn}} - R_{\text{d2ome}})/R_{\text{ProTurn}}$. R_{ProTurn} and R_{d2ome} are protein degradation rate constants computed by ProTurn and d2ome, respectively. The histogram of the distributions for data filtered based on the number of distinct peptides are shown in Figure S8A-C. For unfiltered data, for 87% of proteins the percent difference was less than 30%. 93% of proteins quantified by at least three distinct peptides had less than 30% percent difference. Similar to the conclusion from the Pearson correlations, the more consistent results are observed for rate constants of proteins identified by two or more distinct peptides.

A previous study showed good reproducibility of protein rate constant calculations from heavy water labeling and LC-MS technology for 27 proteins.¹² The Spearman's correlation from two studies using heavy water labeling was 0.90. The slope of the intercept of regression was 0.81. These good characteristics of comparison were achieved for 27

common proteins in the two studies. Here, we did a direct comparison of two bioinformatics solutions on a set of data with a few thousand proteins.

To evaluate d2ome, we also compared the rate constants from the heart proteome study (normal heart) with those from amino acid labeling.¹⁰ The comparison was done on turnover rates of different animals in two different studies with different labeling precursors (both studies used LC-MS for quantification). The plot of the rate constants is shown in Figure S9. These experiments were carried out at different sites with different mice, using different labeling precursors (labeled valine or deuterium in heavy water). Despite the differences, the results from the two studies agreed well, as is seen in the figure.

LDLR^{-/-} Mice Hepatic Proteins under ND and WD Conditions

There were a total of 127 LC-MS experiments (raw files) of ND, and 131 experiments of the WD fed mice. From this data using the database search conditions listed in the Methods section, we identified and quantified 1944 proteins in the ND and 2102 proteins in the WD fed mice. The results are in the Supporting Information (Supporting Information 1 and 2). Of these, 1735 proteins were common to both conditions. After applying our outlier detection and removal algorithm (Grubbs' outlier test), requiring at least two distinct peptides per protein and coefficient of variation (CV) of less than 0.3, the ND data set had 965, and the WD data set had 1007 proteins. There were 511 proteins common to both data sets and quantified by at least three distinct peptides. The results are summarized in Table S2. The scatter plot of the rate constants for the two diet conditions is shown in Figure S10. For most of the proteins the rate constants in the WD fed mice were faster than those for the ND fed mouse. The scatter plot showing the abundance levels of proteins is presented in Figure S11.

To identify statistically significant changes in protein rate constants, we applied stringent thresholds. It was required that only proteins with total abundances larger than 10^9 , CV less than 0.3, number of distinct peptides more than two, and p -value (as determined using two-sample t test and adjusted for multiple hypothesis testing using Benjamini-Hochberg's false discovery rate²³) less than 0.01 passed the comparison criteria. In addition, it was required that the rate constant changes be at least 25%. These filters resulted in a list of 333 proteins. The list is presented in the Supporting Information (Supporting Information 3). We have used the STRING¹⁴ database to analyze this list and infer biological meanings of the observed changes. Among the biological processes, the most enriched were the metabolic processes (small molecule, lipid, fatty acid metabolic processes). Molecular functions were enriched for catalytic activity and oxidoreductase activity. For the evidence of protein—protein interactions (STRING), we required only those with experimental validations (text mining, literature and database-based, and other associations were excluded). In addition, we required a high confidence value (scores greater than 0.7) for the protein—protein interactions. The resulting connected subnetworks are shown in Figure 2. In the figure, we kept only proteins with at least one interacting protein in the list. As can be seen in the figure, most of the proteins with changed stability (difference in rate constants), cluster into five subnetworks of subunits of complexes. These are subunits of NADH dehydrogenase (mitochondrial complex I), cytochrome c oxidase (complex IV), ATP synthase (complex V),

the proteasome complex, and ribosomal proteins (the network with the largest number of nodes and edges). A large portion of the ribosomal network are the proteins of the 40S ribosome. The first two complexes (NADH dehydrogenase and cytochrome c oxidase complexes) are known to be dysregulated in NAFLD.²⁴ Even though a previous study reported increased abundance of some 40S ribosomal proteins in a diet-induced hepatic steatosis,²⁵ the reduced stability of 40S ribosomal protein subunits has not previously been reported. In Figure 3, we show relative monoisotopic peak abundances as a function of time, and the nonlinear fit to the data for ribosomal proteins, Rps18 and Rps25. The different protein subunits of the 40S ribosome showed consistency in half-lives under both conditions of the diet. Thus, for proteins (quantified by at least three peptides and CV less than 0.3) in ND fed mice, the average half-lives and standard deviations were 9.8 days, and 1.1 days, respectively. Corresponding values for the 40S ribosomal proteins of the mice fed WD were 6.2 days and 0.8 days. The number of proteins that were quantified by three or more peptides in each group were 15 (ND) and 16 (WD). The *p*-value of the *t* test comparing the means of the distributions was less than 10^{-7} . The 95% confidence interval of the *t* test statistic was (2.7, 4.4). The value of the test statistic, 8.5, was outside of this interval. The boxplots of the protein half-lives are shown in Figure 4. In addition to the significant changes in the stability of the 40S ribosomal proteins induced by the diet, the boxplot also shows that the proteins undergo coherent turnover under both conditions. Note that for 40S ribosomal proteins, the changes in proteome dynamics are detectable in protein half-lives. An oft used protein metric, protein abundance, does not differentiate between the two diet conditions of LDLR^{-/-} mice. This is demonstrated in Figure 5, which shows the boxplots of the protein abundances in ND and WD fed mice. The *p*-value of the *t* test comparing the means of the log transformed abundances was 0.84 (nonsignificant).

Decreased stability of the 40S ribosomal subunits in WD fed mice is complementary to the PERK arm of UPR that is known to be dysregulated in NAFLD.¹⁵ It is known that NAFLD associated disruption of proteostasis involves PERK, an ER membrane bound protein kinase that phosphorylates eukaryotic initiation factor 2 subunit α ; eIF2 α .²⁶ Nonphosphorylated eIF2 α is required for formation of preinitiation complex (PIC) for protein synthesis. As a result, this post-translational modification (PTM) leads to attenuation of protein synthesis for many genes. PIC includes eukaryotic initiation factors and the ribosomal 40S subunit. We found that the proteins of the 40S subunit have increased degradation rate constants. The phosphorylation of eIF2 α and decreased stability of 40S ribosomal proteins impair the formation and functioning of PIC. Our findings indicate that in addition to phosphorylation, protein synthesis attenuation is contributed to by decreased stability of ribosomal proteins. In addition, we determined reduced stability of the folding proteins HSPA8 and DnaJB4 (CV for these proteins is larger than 0.3 in WD fed mice). HSPA8 is a member of heat shock 70 family of proteins and its functions include facilitation of proper folding of newly translated and misfolded proteins.²⁷ It contributes to apoptosis, autophagy and protein homeostasis via its chaperoning activity.¹⁵ This protein participates cotranslation-ally in folding of nascent peptides (along with another protein DnaJB4, which also shows reduced stability). These proteins have not previously been implicated in pathology of NAFLD. The reduced stabilities of ribosomal and nascent peptide folding proteins indicate a complementary effect to that of PERK. We summarized these proposed relationships between the PERK arm of

UPR and decreased stability of the 40S ribosomal proteins in Figure S12. Figure S13 combines the effects of the increased instabilities of all proteins that were observed in this study.

We note that previous software reported their results using samples with body water enrichment at 5%.⁴ In this study (LDLR^{-/-} mice), the body water enrichment of deuterium was around 3%. The ability to work with lower body water enrichment of deuterium is an important characteristic of the sensitivity of our algorithm. In addition, low body water enrichment levels are suggested for in vivo studies involving human subjects.²⁸

CONCLUSION

We developed a fully automated, computational tool for protein half-life estimations from heavy water metabolic labeling experiments. The algorithm performs peak detection and integration using the chromatographic elution profiles of peptides identified in tandem mass spectrometry. The peak profile is determined by identifying peak start, apex and elution. The end points are obtained via noise estimations. Theoretically computed isotope distributions are used to model the incorporation at the initial and asymptotic time points. Multiple measures of deuterium incorporation such as molar percent enrichment, relative isotope abundance, and relative isotope fraction are produced. The time course of isotope labeling is fitted to a nonlinear decay function, and the parameters are obtained by minimizing the differences between theoretical predictions and experimental observations using the BFGS algorithm. We tested the utility of this software to study liver proteome dynamics in LDLR^{-/-} mice and found that a WD induces increased degradation of hepatic proteins involved in energy metabolism, 40S ribosomal proteins, subunits of the proteasome, and mitochondrial proteins of oxidative phosphorylation complexes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Research reported in this publication was supported in part by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM112044. The authors would like to thank Heather Lander, PhD, for language editing and proofreading of the manuscript.

ABBREVIATIONS

CV	coefficient of variation
HUPO	Human Proteome Organization
LC-MS	liquid chromatography—mass spectrometry
LDLR	low-density lipoprotein receptor
MPE	molar percent enrichment
NAFLD	nonalcoholic fatty liver disease

ND	normal diet
RIA	relative isotope abundance
UPR	unfolded protein response
WD	Western diet

REFERENCES

- (1). Harper JW; Bennett EJ Proteome complexity and the forces that drive proteome imbalance. *Nature* 2016, 537, 328–338. [PubMed: 27629639]
- (2). Labbadia J; Morimoto RI The biology of proteostasis in aging and disease. *Annu. Rev. Biochem.* 2015, 84, 435–464. [PubMed: 25784053]
- (3). Kasumov T; Dabkowski ER; Shekar KC; Li L; Ribeiro RF Jr.; Walsh K; Previs SF; Sadygov RG; Willard B; Stanley WC Assessment of cardiac proteome dynamics with heavy water: slower protein synthesis rates in interfibrillar than subsarcolemmal mitochondria. *Am. J. Physiol Heart Circ. Physiol* 2013, 304, H1201–H1214. [PubMed: 23457012]
- (4). Lau E; Cao Q; Ng DC; Bleakley BJ; Dincer TU; Bot BM; Wang D; Liem DA; Lam MP; Ge J; Ping P A large dataset of protein dynamics in the mammalian heart proteome. *Sci. Data* 2016, 3, 160015. [PubMed: 26977904]
- (5). Vincow ES; Merrihew G; Thomas RE; Shulman NJ; Beyer RP; MacCoss MJ; Pallanck LJ The PINK1-Parkin pathway promotes both mitophagy and selective respiratory chain turnover in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 2013, 110, 6400–6405. [PubMed: 23509287]
- (6). Rauniyar N; McClatchy DB; Yates JR 3rd. Stable isotope labeling of mammals (SILAM) for in vivo quantitative proteomic analysis. *Methods* 2013, 61, 260–268. [PubMed: 23523555]
- (7). Li L; Bebek G; Previs SF; Smith JD; Sadygov RG; McCullough AJ; Willard B; Kasumov T Proteome Dynamics Reveals Pro-Inflammatory Remodeling of Plasma Proteome in a Mouse Model of NAFLD. *J. Proteome Res.* 2016, 15, 3388–3404. [PubMed: 27439437]
- (8). Hsieh EJ; Shulman NJ; Dai DF; Vincow ES; Karunadharma PP; Pallanck L; Rabinovitch PS; MacCoss MJ Topograph, a software platform for precursor enrichment corrected global protein turnover measurements. *Mol. Cell. Proteomics* 2012, 11, 1468–1474. [PubMed: 22865922]
- (9). 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.
- (10). Claydon AJ; Thom MD; Hurst JL; Beynon RJ Protein turnover: measurement of proteome dynamics by whole animal metabolic labelling with stable isotope labelled amino acids. *Proteomics* 2012, 12, 1194–1206. [PubMed: 22577021]
- (11). Busch R; Siah IM; Gee TA; Hellerstein MK Heavy water labeling of DNA for measurement of cell proliferation and recruitment during primary murine lymph node responses against model antigens. *J. Immunol. Methods* 2008, 337, 24–34. [PubMed: 18590913]
- (12). Naylor BC; Porter MT; Wilson E; Herring A; Lofthouse S; Hannemann A; Piccolo SR; Rockwood AL; Price JC Deuterater: a tool for quantifying peptide isotope precision and kinetic proteomics. *Bioinformatics* 2017, 33, 1514–1520. [PubMed: 28093409]
- (13). Jensen DR; Ramirez DE Outlier Detection Under Star-Contoured Errors. *J. Stat Theory Pract* 2015, 9, 850–867.
- (14). Szklarczyk D; Morris JH; Cook H; Kuhn M; Wyder S; Simonovic M; Santos A; Doncheva NT; Roth A; Bork P; Jensen LJ; von Mering C The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 2017, 45, D362–D368. [PubMed: 27924014]
- (15). Baiceanu A; Mesdom P; Lagouge M; Foulle F Endoplasmic reticulum proteostasis in hepatic steatosis. *Nat. Rev. Endocrinol.* 2016, 12, 710–722. [PubMed: 27516341]

- (16). Perkins DN; Pappin DJ; Creasy DM; Cottrell JS Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, 20, 3551–3567. [PubMed: 10612281]
- (17). Rahman M; Previs SF; Kasumov T; Sadygov RG Gaussian Process Modeling of Protein Turnover. *J. Proteome Res.* 2016, 15, 2115–2122. [PubMed: 27229456]
- (18). Kasumov T; Ilchenko S; Li L; Rachdaoui N; Sadygov RF; Willard B; McCullough AJ; Previs S Measuring protein synthesis using metabolic (2)H labeling, high-resolution mass spectrometry, and an algorithm. *Anal. Biochem.* 2011, 412, 47–55. [PubMed: 21256107]
- (19). Sadygov RG Poisson Model To Generate Isotope Distribution for Biomolecules. *J. Proteome Res.* 2018, 17, 751–758. [PubMed: 29202576]
- (20). Price JC; Khambatta CF; Li KW; Bruss MD; Shankaran M; Dalidd M; Floreani NA; Roberts LS; Turner SM; Holmes WE; Hellerstein MK The effect of long term calorie restriction on in vivo hepatic proteostasis: a novel combination of dynamic and quantitative proteomics. *Mol. Cell. Proteomics* 2012, 11, 1801–1814. [PubMed: 22984287]
- (21). Busch R; Kim YK; Neese RA; Schade-Serin V; Collins M; Awada M; Gardner JL; Beysen C; Marino ME; Misell LM; Hellerstein MK Measurement of protein turnover rates by heavy water labeling of nonessential amino acids. *Biochim. Biophys. Acta, Gen. Subj.* 2006, 1760, 730–744.
- (22). Byrd RH; Lu PH; Nocedal J; Zhu CY A Limited Memory Algorithm for Bound Constrained Optimization. *Siam J. Sci. Comput* 1995, 16, 1190–1208.
- (23). Benjamini Y; Hochberg Y Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc.* 1995, 57, 289–300.
- (24). Perez-Carreras M; Del HP; Martin MA; Rubio JC; Martin A; Castellano G; Colina F; Arenas J; Solis-Herruzo JA Defective hepatic mitochondrial respiratory chain in patients with nonalcoholic steatohepatitis. *Hepatology* 2003, 38, 999–1007. [PubMed: 14512887]
- (25). Khan SA; Wollaston-Hayden EE; Markowski TW; Higgins L; Mashek DG Quantitative analysis of the murine lipid droplet-associated proteome during diet-induced hepatic steatosis. *J. Lipid Res.* 2015, 56, 2260–2272. [PubMed: 26416795]
- (26). Hetz C; Chevet E; Oakes SA Proteostasis control by the unfolded protein response. *Nat. Cell Biol.* 2015, 17, 829–838. [PubMed: 26123108]
- (27). Mayer MP; Bukau B Hsp70 chaperones: cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* 2005, 62, 670–684. [PubMed: 15770419]
- (28). Neese RA; Misell LM; Turner S; Chu A; Kim J; Cesar D; Hoh R; Antelo F; Strawford A; McCune JM; Christiansen M; Hellerstein MK Measurement in vivo of proliferation rates of slow turnover cells by 2H2O labeling of the deoxyribose moiety of DNA. *Proc. Natl. Acad. Sci. U. S. A.* 2002, 99, 15345–15350. [PubMed: 12424339]

Figure 1. Scatter plot of protein rate constants computed using ProTurn (y axis) and d2ome (x axis) for the C57 mice used for control (A) and (B) heart hypertrophic mice.⁴ The data were limited to those proteins and peptides that passed the quality metrics thresholds used in ProTurn. In addition, it was required that each protein had three or more peptide identifications. A similar plot for unfiltered data is shown in Figure S1A,B. The axes' range was limited to 0.69 day^{-1} , which corresponds to the rate constant of the half-life of 1 day. The points are colored from red to blue based on the log2 transformed abundance of the proteins (computed as the sum of abundances of all peptides of a protein).

Figure 2. STRING protein interaction network of the statistically significant interactions of proteins with changed stability (rate constants) in ND and WD fed LDLR^{-/-} mice.

Figure 3. Relative abundance of the monoisotopic peak and nonlinear fits to five peptides of (A) Rps18 (murine 40S ribosomal protein S18) obtained from liver samples of ND fed mice, (B) Rps18 for WD fed mice, (C) Rps25 (murine 40S ribosomal protein S25) obtained from liver samples of ND fed mice, (D) Rps25 for WD fed mice.

Figure 4. Boxplot of the half-lives of the protein subunits of 40S ribosome quantified by at least three peptides in ND and WD fed LDLR^{-/-} mice. The *p*-value of the difference of the means of the two distributions was highly significant and equal to 3.4×10^{-8} .

Figure 5. Boxplot of the $\ln$ transformed abundances of the protein subunits of 40S ribosome quantified by at least three peptides in ND and WD fed LDLR^{-/-} mice. The means of the abundances of the proteins of the two diet conditions were similar. The p -value from the t test was 0.84.