Current Bioinformatics Challenges in Proteome Dynamics using Heavy Water-based Metabolic Labeling

Belinda Willard[1], Rovshan G Sadygov[1]* and Takhar Kasumov[2]

[1]Department of Biochemistry and Molecular Biology, Sealy Center for Molecular Medicine, University of Texas Medical Branch, Galveston, TX 77555, USA
[2]Departments of Gastroenterology and Hepatology, and $Proteomics Core, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, OH 44195, USA

Keywords: Proteome dynamics; Heavy water labeling; Stable isotope labeling; Mass spectral profile integration; High resolution and mass accuracy mass spectrometry

Protein turnover in living systems has been measured with the use of stable isotope labeled tracers for over half a century [1]. Recent advances, in mass spectrometry, sample preparation and separation sciences have made it possible for this approach to become applicable at a global proteomics level, permitting analysis of the turnover of many proteins instead of single proteins or an aggregated protein pool [2-4]. Different stable isotope based tracers, including exogenously labeled amino acids [U-^{13}C], glucose and ^{15}N-labeled diet were used to assess global proteome dynamics. Among the available stable isotope precursors, heavy water, D_2O, has advantages in safety, higher sensitivity (due to incorporation of multiple copies of ^2H into analyzed peptide) and cost [5]. In addition, the intake of heavy water leads to universal ^2H incorporation into all biomolecules, potentially permitting turnover analyses of nucleic acids, carbohydrate, or lipids and comparing them with protein turnover. Since heavy water can be administered in drinking water and does not require iv infusion, proteome dynamics studies can be conducted in free leaving organisms, including humans. The key assumptions for the heavy water-based metabolic labeling method are that the labeling of endogenous amino acids is very rapid and there is no post-synthetic labeling of a protein. Both assumptions have been tested experimentally; we have confirmed that there is no post-secretory labeling of plasma proteins in rodents [6]. In addition, we and others have demonstrated that most of the amino acids are labeled after 10-20 min of D_2O administration [7,8].

Experimental procedure for heavy water labeling starts with bolus loading of heavy water followed by regular supply of 0.5-5% D_2O enriched water for varying durations depending of the half-life of the proteins to be analyzed. As the heavy water is administered, total body water is rapidly labeled followed by extensive labeling of non-essential amino acids and minor labeling essential amino acids through de novo synthesis and/or transaminase reactions. The time course of 2H-incorporation into a protein allows the calculation of the rate of protein synthesis. Note, that longer durations of heavy water intake allow more accurate measurements of the kinetics of proteins with longer half-lives. The average protein turnover rate in a human cell line is about 20 h [9]. One of the drawbacks of the heavy water labeling method is that it cannot be applied to measure the synthesis rates of proteins with the half-life shorter than 1 hour.

Applications of heavy water labeling in combination with high-throughput proteomics is a relatively new approach [1]. Just like in other proteomic research areas, the practical successes of heavy water labeling are dependent on advances in mass spectrometry, sample preparation, separation and especially on bioinformatics tools. The bioinformatics workflow for the heavy water labeled samples starts with the peptide/protein identifications from tandem mass spectra using protein sequence databases. The identifications are performed at every time point. It incorporates data processing from mass spectral data that includes isotope envelop detection, integration of peptide profiles in the mass and time domains, determining contributions from different isotopomers, and computing the fractional protein synthesis rates (Figure 1). Several bioinformatics challenges are encountered in heavy water data analysis for software development. The problems are both objective and subjective. The major subjective problem is that, up to date, there is no "golden" data set for testing, tuning and benchmarking software. For example, for peptide identification from tandem mass spectra and protein sequence databases there are number of freely available data sets that can be used for unbiased benchmarking of peptide identification software. There are no comparable, freely available data sets from heavy water labeling experiments. Such data sets from different groups with different types of instrumentation and sample preparation/separation techniques would provide a good basis for unbiased benchmarking of the existing and the development of new software.

To the best of our knowledge, there are currently very few freely available software packages for quantification of proteome dynamics using heavy water-based metabolic labeling approach. The proprietary Mass Hunter software package (B0.4) from Agilent (Santa Clara, CA) was specially designed for the isotopic distribution analysis of peptide processed in Agilent 6520 quadrupole time-of-flight, QT of, mass spectrometers. In addition to this software being unavailable to the public, QT of instruments have relatively lower resolution (~30,000 compared to 120,000 in Orbitrap Ultima) that limits the accuracy of isotope ratio analysis. A recent publication [4] by Ping and colleagues described software for determining protein turnover ratios from heavy water labeling experiments. The software uses mass accuracy of 100 ppm and resolution of 15,000. Error rates are determined by a boosting algorithm. However, it is not clear if their software will be freely available.

To aid our heavy water based proteome dynamics studies we recently developed alternative software which is freely available at a UT website, https://ispace.utmb.edu/users/rgsadygo/Proteomics/HeavyWater/Version.1.0. This software was applied in our recent work [3] focused on comparative proteome dynamics of spatially distinct cardiac mitochondrial subpopulations. Protein turnover rates of subcarcolemma mitochondria and interfibrillar mitochondria

*Corresponding author: Rovshan G Sadygov, Department of Biochemistry and Molecular Biology, Sealy Center for Molecular Medicine, University of Texas Medical Branch, Galveston, TX 77555, USA, Tel: +1 409 772 3287; Fax: +1 409 772 9679; E-mail: rovshan.sadygov@utmb.edu
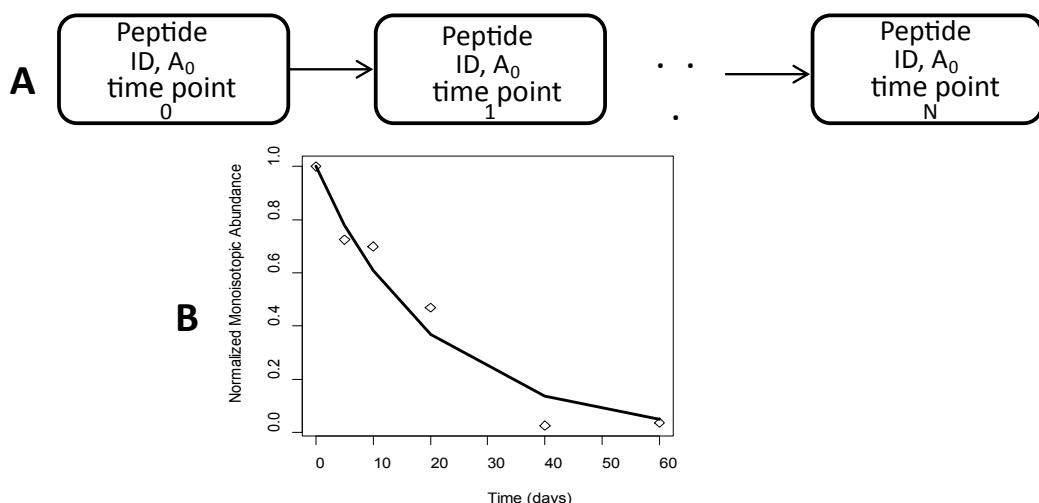
Figure 1: Time-course progression of data processing and isotopomer computation in proteome dynamics experiments with heavy water labeling. In the first step, **A.**, for every time point, a peptide is identified from using tandem mass spectra, precursor mass and a protein sequence database. To avoid statistical problems associated with missing data, only peptides that are consistently identified in all time-course experiments are used. Relative abundance values, A_0, of the monoisotopic proportions are determined from Brauman's least squares distribution. In the second step, **B.**, the monoisotopic proportions, A_0, at every time are fit to an exponential time-dependent decay function to determine peptide turnover rate.

from rat heart were computed and compared for differential protein turnover. The mass accuracy used in our software is a dynamic parameter and can be changed by the user. In our studies of a data set acquired on an Orbitrap Elite mass spectrometer, we set the mass accuracy at 10 ppm, and the resolution was 100,000. We compute the isotopomers from theoretical isotope distributions [10] generated from amino acid sequences, and experimental isotope distributions of the corresponding peptides. We emphasized in the publication that at the current state of the data processing tools, sample preparation and data acquisition, it is important to carefully choose the peptides used for quantification. The main elements to consider are non-overlapping profiles (no co-eluting species that effect mass profiles of target peptides), reproducibility (we require that a target peptide be identified in all time-course experiments), and peptide sequences that have many non-essential amino acids. Publicly and freely available software will encourage research in the proteome dynamics, as has happened in other quantitative proteomics fields. The software that works with standardized file formats (mzML [11] and mzIdentML) designed by the Proteomics Initiative Group of the Human Proteome Organization are preferable since they are easily incorporated into the workflows for protein identification and mass spectral data storage.

The bioinformatics core needed for the data processing of proteome dynamics experiments with heavy water labeling are similar to other metabolic labeling (e.g. ^{15}N-labeling[2]) experiments and can be viewed as being consisting of two major steps. In the first step, the isotope envelopes of a peptide are identified and integrated in the (elution time) and mass (precursor ion) domains. Contributions of each isotopomer to the total peptide isotope envelope are estimated using Brauman's least-squares solution [12]. The relative proportion of the monoisotopic isotopomer is determined for every time point. In the second step, the decay of the monoisotopic isotopomers at different time points is fit to a single compartmental exponential decay to determine protein's fractional synthesis rate and half-life (Figure 1).

Recently we have demonstrated that the kinetics of plasma proteins could be studied in single mice with small blood sampling. However, it is not possible to have a time-course data from the same animal, when applied to mitochondrial and other tissue proteins, as (often) at every time point the animals are sacrificed to collect samples. Therefore, it is not possible to determine within individual and between individuals variability in tissue proteome dynamics although tissue biopsies are potentially possible in larger animals. In addition, there may be experiments where a limited amount of sample is available from an organism, and in these cases technical replicates may not be possible.

The current software allows the determination of which fraction of a protein is newly made and what its half-life is. However, in many occasions it is critically important to know the absolute production rate of a protein. This type of measurements would require in addition to quantification of isotopic distribution the determination of the absolute protein abundance. Also, currently used regression analysis for the curve fitting is based on a single compartmental model. The future bioinformatics tools based on multi-compartmental kinetic analysis and quantification of absolute protein production rate would greatly advance proteome dynamics studies.

In conclusion, the currently available software for heavy water-based metabolic labeling adds a dynamic dimension to traditional proteomics and enables measurements of dynamic gene expression. However, wide spread application of proteome dynamics studies with heavy water based metabolic labeling will require solving the above mentioned questions related to the bioinformatics challenges of these studies.

Acknowledgements

References

1. Claydon AJ, Beynon R (2012) Proteome dynamics: revisiting turnover with a global perspective. Mol Cell Proteomics 11: 1551-1565.

2. Vincow ES, Merrihew G, Thomas RE, Shulman NJ, Beyer RP, et al. (2013) The PINK1-Parkin pathway promotes both mitophagy and selective respiratory chain turnover in vivo. Proc Natl Acad Sci U S A 110: 6400-6405.

3. Kasumov T, Dabkowski ER, Shekar KC, Li, L, Ribeiro RF, et al. (2013) Assessment of cardiac proteome dynamics with heavy water: slower protein synthesis rates in interfibrillar than subsarcolemmal mitochondria. Am J Physiol Heart Circ Physiol 304: H1201-H1214.

4. Kim TY, Wang D, Kim AK, Lau E, Lin AJ, et al. (2012) Metabolic labeling reveals proteome dynamics of mouse mitochondria. Mol Cell Proteomics 11: 1586-1594.

5. Dufner D, Previs SF (2003) Measuring in vivo metabolism using heavy water. Curr Opin Clin Nutr Metab Care 6: 511-517.

6. Kasumov T, Willard B, Li L, Li M, Conger H, et al. (2013) 2H2O-based high-density lipoprotein turnover method for the assessment of dynamic high-density lipoprotein function in mice. Arterioscler Thromb Vasc Biol 33: 1994-2003.

7. Li L, Willard B, Rachdaoui N, Kirwan JP, Sadygov RG, et al. (2012) Plasma proteome dynamics: analysis of lipoproteins and acute phase response proteins with 2H2O metabolic labeling. Mol Cell Proteomics 11: M111.

8. Belloto E, Diraison F, Basset A, Allain G, Abdallah P, et al. (2007) Determination of protein replacement rates by deuterated water: validation of underlying assumptions. Am J Physiol Endocrinol Metab 292: E1340-E1347.

9. Boisvert FM, Ahmad Y, GierliÅ„ski M, CharriÃ¨re F, Lamont D, et al. (2012) A quantitative spatial proteomics analysis of proteome turnover in human cells. Mol Cell Proteomics 11: M111.

10. Sadygov RG, Zhao Y, Haidacher SJ, Starkey JM, Tilton RG, et al. (2010) Using power spectrum analysis to evaluate (18)O-water labeling data acquired from low resolution mass spectrometers. J Proteome Res 9: 4306-4312.

11. Gilski MJ, Sadygov RG (2011) Comparison of Programmatic Approaches for Efficient Accessing to mzML Files. J Data Mining Genomics Proteomics 2.

12. Brauman JI (1966) Least Squares Analysis and Simplification of Multi-Isotope Mass Spectra. Analytical Chemistry 38: 607-610.