



Published in final edited form as:

J Proteome Res. 2009 January ; 8(1): 290–299. doi:10.1021/pr800467r.

Large-Scale Multiplexed Quantitative Discovery Proteomics Enabled by the Use of an ^{18}O -Labeled “Universal” Reference Sample

Wei-Jun Qian[†], Tao Liu[†], Vladislav A. Petyuk[†], Marina A. Gritsenko[†], Brianne O. Petritis[†], Ashoka D. Polpitiya[†], Amit Kaushal[‡], Wenzhong Xiao[‡], Celeste C. Finnerty[§], Marc G. Jeschke[§], Navdeep Jaitly[†], Matthew E. Monroe[†], Ronald J. Moore[†], Lyle L. Moldawer^{||}, Ronald W. Davis[‡], Ronald G. Tompkins[⊥], David N. Herndon[§], David G. Camp II[†], and Richard D. Smith [on behalf of the Inflammation and the Host Response to Injury Large Scale Collaborative Research Program]^{*,†,||}

[†]Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352

[‡]Stanford Genome Technology Center, Stanford University School of Medicine, Palo Alto, California 94304

[§]Department of Surgery, University of Texas Medical Branch, Shriners Burns Hospital, Galveston, Texas 77550

^{||}Department of Surgery, University of Florida College of Medicine, Gainesville, Florida 32610

[⊥]Department of Surgery, Shriners Burn Center and Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114

Abstract

The quantitative comparison of protein abundances across a large number of biological or patient samples represents an important proteomics challenge that needs to be addressed for proteomics discovery applications. Herein, we describe a strategy that incorporates a stable isotope ^{18}O -labeled “universal” reference sample as a comprehensive set of internal standards for analyzing large sample sets quantitatively. As a pooled sample, the ^{18}O -labeled “universal” reference sample is spiked into each individually processed unlabeled biological sample and the peptide/protein abundances are quantified based on $^{16}\text{O}/^{18}\text{O}$ isotopic peptide pair abundance ratios that compare each unlabeled sample to the identical reference sample. This approach also allows for the direct application of label-

© XXXX American Chemical Society

*To whom correspondence should be addressed. Dr. Richard D. Smith, Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, P.O. Box 999, MSIN: K8-98, Richland, WA 99352. E-mail: rds@pnl.gov.

[†]Additional participating investigators in the Large Scale Collaborative Research Program entitled, “Inflammation and the Host Response to Injury”: Henry V. Baker, Ph.D., Ulysses Balis, M.D., Paul Bankey, M.D., Timothy R. Billiar, M.D., Bernard H. Brownstein, Ph.D., Steven E. Calvano, Ph.D., Irshad H. Chaudry, Ph.D., J. Perren Cobb, M.D., Joseph Cuschieri, M.D., Asit K. De, Ph.D., Constance Elson, Ph.D., Bradley Freeman, M.D., Richard L. Gamelli, M.D., Nicole S. Gibran, M.D., Brian G. Harbrecht, M.D., Douglas L. Hayden, M.A., Laura Hennessy, R.N., Jureta W. Horton, Ph.D., Jeffrey Johnson, M.D., Matthew B. Klein, M.D., James A. Lederer, Ph.D., Stephen F. Lowry, M.D., Ronald V. Maier, M.D., John A. Mannick, M.D., Philip H. Mason, Ph.D., Grace P. McDonald-Smith, M.Ed., Carol L. Miller-Graziano, Ph.D., Michael N. Mindrinos, Ph.D., Joseph P. Minei, M.D., Ernest E. Moore, M.D., Avery B. Nathens, M.D., Ph.D., M.P.H., Grant E. O’Keefe, M.D., M.P.H., Laurence G. Rahme, Ph.D., Daniel G. Remick, Jr., M.D., David A. Schoenfeld, Ph.D., Michael B. Shapiro, M.D., Geoffrey M. Silver, M.D., John Storey, Ph.D., Robert Tibshirani, Ph.D., Mehmet Toner, Ph.D., H. Shaw Warren, M.D., Michael A. West, M.D.

Supporting Information Available: Supplemental Tables are provided for the complete lists of unique peptides and unique proteins along with relative abundances, as well as the detailed data for labeling and label-free comparisons presented in Figure 6. This material is available free of charge via the Internet at <http://pubs.acs.org>.

free quantitation across the sample set simultaneously along with the labeling-approach (i.e., dual-quantitation) since each biological sample is unlabeled except for the labeled reference sample that is used as internal standards. The effectiveness of this approach for large-scale quantitative proteomics is demonstrated by its application to a set of 18 plasma samples from severe burn patients. When immunoaffinity depletion and cysteinyl-peptide enrichment-based fractionation with high resolution LC-MS measurements were combined, a total of 312 plasma proteins were confidently identified and quantified with a minimum of two unique peptides per protein. The isotope labeling data was directly compared with the label-free ^{16}O -MS intensity data extracted from the same data sets. The results showed that the ^{18}O reference-based labeling approach had significantly better quantitative precision compared to the label-free approach. The relative abundance differences determined by the two approaches also displayed strong correlation, illustrating the complementary nature of the two quantitative methods. The simplicity of including the ^{18}O -reference for accurate quantitation makes this strategy especially attractive when a large number of biological samples are involved in a study where label-free quantitation may be problematic, for example, due to issues associated with instrument platform robustness. The approach will also be useful for more effectively discovering subtle abundance changes in broad systems biology studies.

Keywords

human plasma; proteomics; ^{18}O labeling; Reference; LC-MS; accurate mass and time tag; label-free quantitation; isotope-labeling

Introduction

Proteomics-based technologies have increasingly been applied to the study of disease-related clinical samples (e.g., human blood serum/plasma, proximal fluids, and disease tissues) for the purpose of identifying novel disease-specific protein biomarkers, gaining better understanding of disease processes, and discovering novel protein targets for therapeutic interventions and drug developments.¹ While there are many important analytical challenges for performing effective clinical proteomics applications, clinical discovery efforts have been further enhanced by recent advances in liquid separations and MS technologies that provide high sensitivity and broad dynamic range of measurements that allow for the detection of low-abundance proteins, as well as advances in informatics/statistics that have enabled large-scale data analysis.²

One unique aspect of clinical proteomics applications is the necessity of analyzing a relatively large number of patient samples to address the large patient-to-patient variations that are typically observed. To discover statistically confident candidate protein biomarkers among large biological variations, it will be essential to have reliable quantitative approaches that allow the direct comparison of protein abundances across a cohort of patient samples. To date, the most broadly applied approaches for quantitative proteomics measurements have been stable isotope labeling approaches due to their higher accuracy for relative quantitation.³ These methods include metabolic labeling,^{4–6} chemical labeling of specific functional groups using reagents such as isotope-coded affinity tags (ICAT)⁷ and iTRAQ,^{8,9} and enzymatic ^{18}O -labeling of peptide C-termini.^{10–12} These labeling techniques are designed for detecting accurate changes in pairwise comparisons or for comparison across a few conditions; however, in clinical applications, it remains challenging to compare relative protein abundances across a large number of clinical samples using isotopic labeling methods.

More recently, there has been significant interest in the application of “label-free” direct MS intensity-based quantitation for comparing relative protein abundances due to its greater flexibility for comparative analyses and simpler sample processing procedures as compared to labeling approaches.^{13–18} Although it is generally accepted as less accurate compared to the

stable isotope labeling approaches,³ the label-free method should, in principle, allow the direct comparison of relative protein abundances across any number of patient samples. With the proper application of data normalization to mitigate variations in instrumental performance between analyses, an accurate comparative quantitation can be achieved for both biological and clinical applications.^{13–18} One of the significant challenges for the label-free approach is its dependence on the reproducibility of LC-MS performance and the robustness of the platform. Issues with reproducibility could be especially problematic for large-scale biological studies involving hundreds of samples that require an extended period of analysis time where the drift in platform performance becomes more obvious. In contrast, the stable isotope labeling approaches depend much less on platform reproducibility because the relative abundances are measured by the relative ratios of isotopic peptide pairs, which are coeluting from the separations and have the same ionization efficiency independent of the instrumental variation.

To enable stable isotope labeling-based quantitation for large sample sets, we devised a strategy that incorporates an ¹⁸O-labeled “universal” reference sample as a comprehensive set of internal standards spiked into each individually processed unlabeled ¹⁶O biological sample. The peptide/protein abundances are quantified based on ¹⁶O/¹⁸O isotopic peptide pair ratios that compare each unlabeled biological sample to the identical labeled reference. In addition, this strategy allows for the simultaneous application of label-free quantitation and isotope-labeling based quantitation (i.e., dual-quantitation) so that the quantitative results from the two methods can be directly compared and integrated. The effectiveness of the strategy was demonstrated by analyzing a set of 18 severe burn patient plasma samples via LC-MS proteome profiling. The results revealed better precision using the ¹⁸O-reference based isotope labeling approach compared to the label-free approach with results from both approaches showing a strong correlation of the observed relative protein abundances.

Experimental Section

Human Plasma Patient Samples

Eighteen blood plasma samples from individual human burn patients were used in this study. These samples were part of a large-scale proteomics study of clinical outcomes for severely burned patients. All samples were supplied by the Department of Surgery at the University of Florida College of Medicine, which serves as the sample collection and coordination site for a multicentered clinical study (Inflammation and the Host Response to Injury). Approval for the conduct of this programmatic research was obtained from the Institutional Review Boards of the University of Texas Medical Branch Galveston, University of Texas Southwestern, University of Washington, Massachusetts’s General Hospital, the University of Florida College of Medicine, and the Pacific Northwest National Laboratory in accordance with federal regulations. A total of 300 μ L of plasma from each subject was applied toward proteomics analysis and their initial protein concentrations were determined by a BCA Protein Assay (Pierce, Rockford, IL). A reference sample was generated by pooling 100 μ L aliquots of each patient’s plasma. Unless otherwise noted, all protein sample processing was performed at 4 °C.

Immunoaffinity Depletion

All patient plasma samples and the pooled reference sample were individually depleted of their 12 most abundant plasma proteins (albumin, IgG, α 1-antitrypsin, IgA, IgM, transferrin, haptoglobin, α 1-acid glycoprotein, α 2-macroglobulin, apolipoprotein A-I, apolipoprotein A-II, and fibrinogen) in a single step using the a prepacked Seppro IgY-12 affinity LC-5 column (loading capacity of 65 μ L of human plasma) (GenWay Biotech, San Diego, CA; now known as ProteomeLab IgY-12, Beckman Coulter, Fullerton, CA) on an Agilent 1100 series HPLC system (Agilent, Palo Alto, CA) per the manufacturer’s instruction.¹⁹ Either 65 or 130 μ L

(depending on its initial patient plasma protein concentration) of patient plasma was individually depleted and only the flow-through fractions were collected for the individual patient samples. A total of ~1.5 mL of reference plasma sample was depleted using the same procedure and the flow-through portions were pooled. The flow-through portion for each sample was then individually concentrated in iCON concentrators with 9 kDa molecular weight cutoffs (Pierce) followed by buffer exchange with 50 mM NH_4HCO_3 per the manufacturer's instructions. Protein concentration was then measured using a BCA protein assay (Pierce).

Plasma Protein Digestion

Protein samples were denatured and reduced in 50 mM NH_4HCO_3 buffer (pH 8.2), 8 M urea, and 10 mM dithiothreitol (DTT) for 1 h at 37 °C. The resulting protein mixture was diluted 10-fold with 50 mM NH_4HCO_3 before sequencing grade modified porcine trypsin (Promega, Madison, WI) was added at a trypsin/protein ratio of 1:50 (w/w). The sample was incubated for 5 h at 37 °C. The digested samples were loaded on a 1-mL SPE C18 column (Supelco, Bellefonte, PA) and washed with 4 mL of 0.1% trifluoroacetic acid (TFA)/5% acetonitrile (ACN). Peptides were eluted from the SPE column with 1 mL of 0.1% TFA/80% ACN and lyophilized. The resulting peptide samples were then reconstituted in 25 mM NH_4HCO_3 and the residual trypsin activity was quenched by boiling the samples for 10 min and immediately placing the samples on ice for 30 min. The resulting peptide amount for each sample was again measured by a BCA protein assay (Pierce).

Trypsin-Catalyzed ^{18}O -Labeling of the Reference Sample

Trypsin-catalyzed ^{18}O -labeling for preparing the reference sample was carried out as previously described.¹¹ After residual trypsin activity was quenched via boiling followed by immediately placing the samples on ice for 30 min, the peptide sample was lyophilized to dryness and initially reconstituted in 60 μL of acetonitrile, followed by the addition of 600 μL of 50 mM NH_4HCO_3 in ^{18}O -enriched water (95%, ISOTECH, Miamisburg, OH). Then, 6 μL of 1 M CaCl_2 and 30 μL of immobilized trypsin (Applied Biosystems, Foster City, CA) were added to the digests and the samples were mixed continuously for 5 h at 30 °C. After labeling, the sample was acidified by adding 6 μL of formic acid and the supernatant was collected after centrifuging the samples for 5 min at 15 000g. The labeled sample was lyophilized and reconstituted in 25 mM NH_4HCO_3 with its peptide concentration measured by a BCA assay. The labeled reference sample was then divided into 18 identical aliquots in 18 individual tubes and an equal amount of peptides from each patient sample was added to each tube to form 18 patient/reference mixed samples.

Fractionation via Cysteinyl-Peptide Enrichment

The 18 patient/reference mixed samples were further fractionated into Cys- and Non-Cys-peptide fractions using the procedures as previously described.²⁰ Briefly, the tryptic digest was reduced with 5 mM DTT in Tris buffer, and cysteinyl-peptides were captured by Thiopropyl Sepharose 6B thiol-affinity resin ($4 \times 100 \mu\text{L}$; Amersham Biosciences) following incubation of the reduced peptides with the resin. The Non-Cys-peptide supernatant portion was collected. The resin was further washed to remove any nonspecifically bound peptides. The captured cysteinyl-peptides were then released by incubation with 20 mM DTT for 30 min at room temperature and the released peptides were then alkylated with 80 mM iodoacetamide. Both the eluted cysteinyl-peptide samples and the unbound, non-cysteinyl-peptide samples were desalted using a SPE C18 column and then lyophilized. Both fractions were then reconstituted in 25 mM NH_4HCO_3 with their concentrations measured by a BCA assay.

Reversed-Phase Capillary LC-MS Analyses

Both the Cys- and Non-Cys-peptide samples from individual burn patients were analyzed using a fully automated, custom-built, two-column capillary LC system²¹ coupled online using an in-house-manufactured ESI interface to an 11.5 T Fourier transform ion cyclotron resonance (FTICR) mass spectrometer. All of the Cys-samples were analyzed on a single capillary column with the Non-Cys-samples analyzed on the other column of an automated dual-column LC-system. The capillary columns were made by slurry packing 5 μm Jupiter C₁₈ bonded particles (Phenomenex, Torrance, CA) into a 65 cm long, 150 μm i.d. fused-silica capillary (Polymicro Technologies, Phoenix, AZ). The mobile phase consisted of 0.2% acetic acid and 0.05% trifluoroacetic acid (TFA) in water (A) and 0.1% TFA in 90% acetonitrile/10% water (B). Ten-microliter aliquots of each peptide sample with concentrations of 0.1 $\mu\text{g}/\mu\text{L}$ were injected onto the reversed-phase column for LC-FTICR analysis. The mobile phase was held at 100% A for 20 min followed by a nonlinear exponential gradient elution generated by increasing the mobile-phase composition to ~70% B over 150 min using a stainless steel mixing chamber. The LC-FTICR was configured and operated as described elsewhere.²²

LC-MS Data Processing

The LC-FTICR data sets were automatically analyzed using an in-house developed software package that includes tools such as ICR2LS and VIPER.²³ The initial analysis of raw LC-MS data involved a mass transformation, or deisotoping, step using ICR2LS, which generated a text file report for each LC-MS data set including both the monoisotopic masses and the corresponding intensities for all detected species for each mass spectrum. Then, each data set was processed by the feature matching tool (VIPER) for peptide identification and quantification. VIPER displays the data in a two-dimensional mass and LC-elution time format. The feature matching process included “distinct feature” (i.e., a peak with unique mass and elution time) finding, searching for ¹⁶O/¹⁸O feature pairs, computing abundance ratios for pairs of features, an intensity report for all detected features, normalizing LC elution times via alignment to a database, and feature identification. The ¹⁶O/¹⁸O abundance ratio calculation has been previously described.¹¹ The details of VIPER peak matching have also been described previously.^{23,24} A robust LC-MS alignment algorithm was incorporated into VIPER for correcting any variations in mass and elution times.²⁵ Feature identification was performed by matching the accurately measured masses and normalized elution time (NET) values of each detected feature to a pre-established human plasma proteome accurate mass and time (AMT) tag database. The plasma AMT tag databases were generated from the combined results of several comprehensive LC-MS/MS profiling investigations,¹¹ including a recent profiling of the human trauma patient plasma proteome.²⁰ The process of generating the AMT tag database and the criteria of peptide inclusion in the database have been described previously in detail.¹¹ The AMT tag database essentially serves as a “look-up” table for LC-MS feature identifications. In this work, a dynamic modification of 4.0085 Da was applied during the peak (or feature) matching process to enable both unlabeled peptides and ¹⁸O-labeled peptides to be identified simultaneously. The peptide sequences of a given feature or pair of features were assigned when the measured mass and NET for each given feature matched the calculated mass and NET of a peptide in the AMT tag database within a 4 ppm mass error and 2% NET error (the mass error tolerance was later refined to 2.5 ppm, as described in Results). For each identified peptide, both the label-free ¹⁶O-abundance and the relative ¹⁶O/¹⁸O abundance ratio are reported if the feature is paired. The details for computing the ¹⁶O/¹⁸O abundance ratios using VIPER have been previously described.¹¹ The final identified peptide list was then used to generate a nonredundant protein group list using Protein-Prophet²⁶ and all of the peptides were annotated based on protein group ID for downstream data normalization and for rolling-up peptide level abundance information to the protein level. The identified peptides were further excluded for calculating relative abundance ratios if the peptide was not observed in at

least four of the 18 subjects. Peptides were removed from downstream data analysis if they were shared by more than three protein groups.

Data Normalization and Relative Quantitation

To enable direct comparison of labeled data and label-free data, both $^{16}\text{O}/^{18}\text{O}$ isotopic abundance ratios (patient/reference) and direct ^{16}O abundances were used as measures of relative peptide abundance. Here, we will use R_{ij} to indicate peptide $^{16}\text{O}/^{18}\text{O}$ ratios and A_{ij} to indicate direct ^{16}O abundances for peptides, where ij indicates peptide j in a given LC-MS data set i . To remove any systematic biases potentially introduced from sample processing or LC-MS analyses, a global normalization procedure similar to that used for microarray normalization was applied to relative peptide abundances or abundance ratios across LC-MS data sets.^{27,28} Prior to normalization, both $^{16}\text{O}/^{18}\text{O}$ abundance ratios R_{ij} and direct ^{16}O -abundances A_{ij} were transformed in Log_2 format as Log_2R_{ij} and Log_2A_{ij} , respectively. For $^{16}\text{O}/^{18}\text{O}$ ratio data, the normalized peptide abundance ratio $\text{Log}_2R^*_{ij}$ was computed as $\text{Log}_2R^*_{ij} = \text{Log}_2R_{ij} - \text{median}_j(\text{Log}_2R_{ij})$, which basically centers the median Log_2R_{ij} to zero for each data set i . For label-free ^{16}O abundance Log_2A_{ij} normalization, a linear regression approach was applied to peptide abundances across different LC-MS data sets. For this method, an ^{16}O -reference data set M_j was first generated by taking the median abundance of each peptide across all LC-MS data sets as $M_j = \text{median}_i(\text{Log}_2A_{ij})$. The normalized peptide abundances $\text{Log}_2A^*_{ij}$ were then computed for each LC-MS data set by linearly regressing Log_2A_{ij} against the reference data set M_j and adjusting the slope and intercept.

Following data normalization across LC-MS data sets, we further converted the normalized label-free data $\text{Log}_2A^*_{ij}$ into a patient/reference ratio format similar to the $^{16}\text{O}/^{18}\text{O}$ ratio format by subtracting M_j from $\text{Log}_2A^*_{ij}$. Then, for each quantified protein, a rescaling procedure for peptide profiles across data sets was performed for both the label-free and labeled data. This procedure scales all peptide profiles originating from the same protein to the same level using a simple scaling factor for each peptide obtained relative to a chosen reference peptide for the given protein. The peptide with the most observations was chosen as the reference peptide for a given protein and a scaling factor for every other peptide was calculated as the median ratio for the common data points between the peptide to be scaled and the reference peptide. A similar normalization and rescaling procedure has been recently described for label-free quantitation.¹⁷ Relative protein abundance ratios were calculated by averaging the rescaled relative peptide abundance ratios for each data set. All the data normalization and rescaling procedures were performed using DANTE, a recently developed software tool for quantitative analysis of omics data.²⁹

Results

The Quantitation Strategy

The quantitation strategy is based on the concept of using a stable isotope labeled “universal” reference sample and adding it to each individually processed biological or patient sample. This “universal” reference sample serves as a comprehensive set of internal standards for enabling the comparison of relative peptide/protein abundances across any number of patient samples based on isotopic abundance ratios. In effect, the reference sample acts as a “bridge” to facilitate cross-sample comparison based on isotopic labeled pair information (Figure 1). This concept is similar to the culture-derived isotope tags (CDITs) approach reported by Ishihama et al. for quantitative comparison of two different tissue samples that were otherwise difficult to label.³⁰ The use of the labeled reference sample effectively addresses the challenge that isotope labeling approaches typically cannot be applied to a large number of samples due to the labor and cost involved in labeling many samples. Ideally, the reference sample should have sufficient similarity in peptide/protein composition to each of the patient samples, which

makes pooling across all patient samples to generate such a reference sample a good method. In this work, postdigestion trypsin-catalyzed ^{18}O labeling is the ideal choice for the reference sample because the ^{18}O labeling can be applied to any type of cell, biofluid, or tissue sample and all tryptic peptides will be labeled at the carboxyl terminus of each peptide.¹¹ All patient samples are individually processed and digested before mixing with an equal amount of the labeled reference peptide sample. Each of the final mixed samples to be analyzed by LC-MS contains the unlabeled ^{16}O -peptides from individual patients plus labeled ^{18}O -peptides from the reference sample. For each detected LC-MS feature, the quantitative information can be reported as the $^{16}\text{O}/^{18}\text{O}$ abundance ratio if the feature is paired as well as the direct ^{16}O -abundance for the feature. Therefore, the $^{16}\text{O}/^{18}\text{O}$ isotopic pair abundance ratio-based quantitation for the peptide pairs can be applied simultaneously with the label-free quantitation for all ^{16}O -peptides for comparing the relative peptide/protein abundances across the patient samples.

Proteomics Analysis of Individual Plasma Samples

In this work, the quantitation strategy is applied to the analysis of 18 human plasma samples from severe burn patients. For enhancing the detection of low-abundance plasma proteins, we integrated immunoaffinity depletion and cysteinyl-peptide enrichment-based fractionation²⁰ into the overall analysis scheme. All individual patient samples and the pooled reference sample were initially subjected to immunoaffinity depletion of the top 12 abundant proteins prior to trypsin digestion. After the mixing of each patient peptide sample with the labeled reference peptide sample, the mixed sample was fractionated into Cys- and Non-Cys-fractions based on the cysteinyl-peptide enrichment procedure. Both the Cys- and Non-Cys-samples were individually analyzed by LC-MS and peptides were identified based on the AMT tag approach.^{11,31} The identifications and abundance data for the two fractions were combined for each patient to achieve overall improved proteome coverage.

Figure 2A,B shows the distribution of average mass errors and average NET errors for all initially matched ^{16}O -peptides within a 4 ppm mass error and 2% NET error against the preestablished human plasma AMT tag database. As shown, a majority of peptides were identified within a mass error of 2.5 ppm and a NET error of 1%. As previously reported,³² the background (or random) match level can be used to estimate the false discovery rate (FDR) for peptide identifications, and similarly, the FDR can be estimated by performing peak matching to a 11 Da-shifted AMT tag database. When both approaches were used, the FDR was estimated to be ~5.5% at the unique peptide level that is comparable to the recently reported level.³² Following the final filtering with 2.5 ppm mass error, a total of 3848 different peptides were identified, corresponding to a total of 312 proteins confidently identified with two or more unique peptides, while an additional 284 proteins were identified with just one unique peptide. Only those peptides with at least four observations out of the 18 patient samples were included for calculating relative abundance ratios with the detailed data listed in Supplemental Table I and Supplemental Table II in Supporting Information. Because of the overlap of ^{16}O and ^{18}O isotopic envelopes, it is often difficult to find pairs if the relative abundance difference between the pair is >10-fold, although moderate abundance differences can be accurately quantified with this labeling approach.³³ In the case of large abundance differences, the peptides were identified as either ^{16}O -peptides or ^{18}O -peptides without the definite observation of the other pair's member. Indeed, only ~60% of peptides (2351 out of 3934) and ~78% of proteins (244 out of 312) were identified by isotopic pairs. The results show that the label-free approach generally provides higher proteome coverage compared to the $^{16}\text{O}/^{18}\text{O}$ labeling approach.

Since an identical amount of labeled reference sample was spiked into each patient sample, peptides in the labeled reference sample serve as a set of comprehensive internal standards for evaluating the technical reproducibility of LC-MS analyses across the study set. Figure 2C

shows the reproducibility of the ^{18}O -labeled peptide abundances observed among a subset of 10 samples as illustrated by both scatter plots (bottom left) and pairwise Pearson correlation coefficients (top right) for detected peptide intensities. Good technical reproducibility was observed across 18 patient samples based on the spiked reference sample, with an average correlation coefficient of 0.95 ± 0.01 . Comparable reproducibility for label-free processing replicates with the identical immunoaffinity depletion and digestion procedures has also been reported, where an average correlation coefficient of 0.94 ± 0.02 was observed for nine processing replicates.² The results observed for both the label-free processing replicate data and the labeled-reference data supports the overall reproducibility of this label-reference based approach for quantitation. As anticipated, much larger patient-to-patient variations were observed based on the average correlation coefficient of 0.75 ± 0.12 for ^{16}O -peptide abundances among patients (data not shown). Such large variations in human patients reinforces the need for analyzing relatively large numbers of patient samples in order to gain sufficient statistical power for identifying candidate biomarkers for diseases such as cancer.²

Comparison between Isotopic Labeling and Label-Free Quantitation

To facilitate better comparison and integration between the stable isotope labeled and label-free data, we have designed a data analysis process as illustrated in Figure 3 to transform both the stable isotope labeled and label-free data in a similar format at both the peptide and protein levels. Each sample analyzed by LC-MS contained an equal mass of peptides from a given patient sample (^{16}O -Pat) and peptides from the labeled reference sample (^{18}O -Ref). For stable isotope labeled data, the relative abundance of each peptide was automatically reported as an $^{16}\text{O}/^{18}\text{O}$ abundance ratio, representing the peptide abundance from the given patient relative to that from the labeled ^{18}O -reference (^{16}O -Pat/ ^{18}O -Ref). For label-free data, the peptide abundance was initially reported as direct ^{16}O -abundance. Prior to normalization, we generated a ^{16}O -reference data set by taking the median abundance value for each peptide across all 18 patients. All LC-MS data sets were first normalized against the ^{16}O -reference data set and then the ^{16}O -abundances for peptides in each data set were converted into relative abundance ratios by dividing the normalized abundances in patient samples against those in the ^{16}O -reference data set (Pat/ ^{16}O -Ref). Following such conversion, the stable isotope labeled data and label-free data are in comparable relative abundance formats (i.e., patient relative to a reference). The only difference is that the labeled data actually uses a labeled reference sample (^{18}O -ref) and the label-free data uses a label-free reference data set that are actually median abundances of ^{16}O peptides (^{16}O -Ref). The two references are conceptually similar since they both represent average signals across the 18 patients. The relative peptide abundance ratios were then rescaled to put all peptide profiles at the same level for a given protein before rolling-up to the protein level by averaging.

The nearly identical data format for label-free and stable isotope labeled data enabled the direct comparison between the relative precisions in quantitation for the two approaches. We first examined the consistency of peptide profiles for a given protein obtained between the two approaches. In general, we observed that the relative abundance profiles of labeled data were more closely correlated for different peptides originating from a given protein compared to the label-free data. Figure 4 shows an example of the relative peptide abundance profiles across the 18 patient samples, where the seven common peptides originating from the C-reactive protein were shown for both the labeled and label-free data. As shown, the isotopic pair data (Figure 4A) provided significantly better consistency for different peptide abundance profiles or better precision in relative abundance measurements compared to the label-free data (Figure 4B); however, more missing data points were typically observed in the stable isotope labeled pair data compared to label-free data. Overall, the averaged protein abundance profiles between the stable isotope labeled and label-free data were well-correlated.

The quantitation performance was further compared based on the measurements of the relative abundance differences between two samples. Figure 5 shows the relative abundance changes in Log_2Ratio to compare two selected patient samples based on the labeled (Figure 5A) and label-free (Figure 5B) data (only proteins with two or more peptides are shown). The standard deviations for each protein (derived from different peptide abundance ratios from the same protein) are also shown. The labeled data showed much tighter standard deviations for the majority of proteins compared to the label-free data, again confirming better precision in quantitation for the labeling approach. Nevertheless, the label-free approach quantified more proteins from the same two samples compared to labeled paired data. Figure 5C further compares the distribution of coefficient of variances (CV) values for the quantified proteins in the 18 samples where the CV values were calculated based on the variations among the different peptides originated from a given protein. A majority of proteins were quantified with 5–20% CV based on the stable isotope labeling approach; while a majority of the label-free quantified proteins have CV ranging from 20% to 50%.

Correlation in Protein Abundance Differences

We next examined the degree of concordance for measuring the relative protein abundance differences between biological samples for these two approaches. For this purpose, we assessed the correlation for relative protein abundance differences between two selected patients as well as between two groups of patients (9 females vs 9 males) as shown in Figure 6 (detailed ratio data available in Supplemental Table III in Supporting Information). The set of patients was divided into the male and female groups purely for the purpose of method illustration since the degree of variation in terms of clinical outcomes and severity of burn for these patients are much more significant factors than the sex difference.

As shown, relatively high Pearson correlation coefficients for protein abundance ratios obtained between the two approaches were observed for comparing either two patients or two groups of patients. The different degrees of missing abundance data across the patient samples probably are the main reason for the worse correlation observed for the male/female group comparison. Overall, only a small percentage of proteins showed disagreement in their relative protein abundance differences as measured by the two approaches (Figure 6), presumably due to differential variations introduced for each approach or the different degrees of “missing data” from the two approaches; however, those proteins displaying good concordance provide higher confidence in quantitation.

Discussion

The ability to perform reliable, quantitative measurements of relative protein abundances across many biological samples remains a key challenge for proteomics discovery applications. To overcome the difficulty of applying stable isotope labeling approaches to large sets of biological samples, we introduce a strategy that incorporates an ^{18}O -labeled “universal” reference sample as a comprehensive set of internal standards for large-scale quantitation. The labeled reference is spiked into each individually processed (and unlabeled) patient sample so that every sample can be quantified based on $^{16}\text{O}/^{18}\text{O}$ ratios against the same reference sample. The principle of this strategy should allow the application of the isotope labeling strategy to large-scale biological studies for cross-comparison among any number of samples as long as sufficient ^{18}O -labeled “universal” reference sample is available for spiking into each biological sample. Cross-comparison among biological data sets acquired at different times (e.g., months or years apart) or with different LC-MS instrument platforms will be feasible based on the stable isotope dilution concept, provided that the proteins are commonly identified. This aspect of large-scale cross-comparison among many samples is one of the current limitations for both labeling and nonlabeling MS based quantitation, although such robust comparison is extremely

important for large-scale discovery proteomics where studies will require extended analysis times for the large number of samples. With this labeled-reference based strategy, such a requirement can be effectively addressed.

To make the optimal utility of the labeled “universal” reference sample as comprehensive internal standards for all biological samples to be analyzed, the reference sample needs to have sufficient complexity to ideally contain all species presented in any given biological samples. In this situation, every peptide present in a given biological sample will ideally be paired with an ^{18}O -labeled peptide within the reference sample and its relative abundance can be quantified based on $^{16}\text{O}/^{18}\text{O}$ ratio. Therefore, it is generally not desirable to only include control or healthy samples for the reference because one would not be able to quantify those proteins uniquely present in patients by the labeled reference if patient samples are not included. A simple solution would be to generate a pooled sample from all biological samples (i.e., include both control and disease) so that the “universal” reference sample contains peptides from all of the samples to be studied.

The ^{18}O -labeled “universal” reference strategy also has the advantage of allowing simultaneous applications of stable isotope labeling and label-free approaches to the same sample set without reducing the throughput of LC-MS analyses. This advantage allowed us to investigate the performance of this ^{18}O -reference based quantitative approach against the label-free data through the analysis of 18 human plasma samples from severe burn patients. The results demonstrated that the ^{18}O -reference enabled isotope labeling-based quantitation is effective by clearly offering better precision in quantifying relative abundances than the label-free approach, while the label-free data provided better proteome coverage due to the limited dynamic range of isotopic pair measurements as well as the computational issue with finding paired LC-MS features. Nevertheless, the advantage of label-free quantitation can also be fully utilized along with isotope labeling quantitation with this ^{18}O -labeled “universal” reference strategy. This strategy does have the limitation of reduced sample loading for LC-MS analysis for the mixed $^{16}\text{O}/^{18}\text{O}$ sample as compared to a label-free sample similar to all stable isotope labeling approaches; however, we have observed that the 2-fold variation in sample loading for high sensitivity LC-MS applications has a minimal effect on the total number of peptide identifications when the sample loading was close to optimal.

In addition, the use of one common labeled “universal” reference sample for quantitation is advantageous in several specific aspects, including (1) only one labeling preparation is required, which significantly reduces overall sample preparation time and potential variations from individual labeling sample manipulations, as well as the overall experimental cost; and (2) the common reference serves as a set of comprehensive internal standards, which provides a unique means for evaluating LC-MS platform reproducibility and for identifying analyses of “bad” quality across large-scale proteomics studies.

Although the isotope labeling data showed much better precision than the label-free data based on CV values, the relative abundance differences when comparing either two patients or two groups of patients showed very good concordance between the two approaches, suggesting that both approaches can provide relatively reliable measurements of the relative protein abundance differences following averaging peptide abundance data into the protein level. Similarly good correlation between labeled data and nonlabeling data was also previously reported.^{34,35} Fundamentally, the two approaches should be complementary due to the different sources of errors associated with each approach. For label-free quantitation, all steps of sample processing, electrospray ionization performance, and mass spectrometry will contribute to the overall variations observed. However, for isotope labeling measurements, the factors that negatively affect the overall reproducibility are much more limited, and are primarily related to sample processing prior to the mixing with the labeled reference and ratio

computation errors. Therefore, the simultaneous integration of the two quantitative approaches into a single strategy enhances the overall confidence of the quantitative data by providing cross-validation between the two approaches. Differences between the two approaches were also observed for some proteins due to the different variations introduced in the process for each approach, as well as the different levels of “missing data” for the two approaches. In general, the labeled data should have better confidence due to its higher accuracy; however, the labeled results typically have more missing data for a given protein (since we require that both versions of the peptide be detected as a pair in this work).

On the basis of our knowledge, this is the first report illustrating the large-scale isotope labeling-based quantitation for 18 individual samples by incorporating a single ^{18}O -labeled “universal” reference sample into each biological sample. We anticipate that the advantages of this ^{18}O -reference enabled quantitation strategy will allow it to be applied as a general approach for large-scale proteomics discovery applications as well as longitudinal biological applications. The robustness of isotope-labeling based quantitation will make this strategy especially attractive when a study involves a large number of samples where label-free quantitation may encounter limitations due to issues associated with platform robustness. The better precision of the labeling approach will also make this strategy useful for discovering subtle abundance changes in large-scale systems biology studies. Moreover, this strategy is flexible for coupling any peptide level fractionation technique such as strong cation exchange-based fractionation to address the dynamic range challenge for low-abundance protein detection because the ^{18}O -reference sample is incorporated into individual sample prior to peptide-level fractionation and the effective quantitation based isotope labeling can be achieved as previously demonstrated.¹¹ As illustrated in this study, the coupling of immunodepletion, cysteinyl-peptide enrichment-based fractionation, and 1D LC-MS analyses with the quantitation strategy has allowed confident quantification of 312 plasma proteins with at least two different peptides per protein. We have previously reported that typically ~170 plasma proteins were identified using IgY-12 depletion and 1D LC-MS analysis.² The cysteinyl-peptide enrichment-based fractionation contributed to this improvement in protein identifications. The focus of this report has been mainly on the proof-of-principle demonstration of the ^{18}O -labeled “universal” reference enabled quantitative methodology based on a limited sample set; this approach is currently being applied to a large set of clinical burn patient plasma samples in our laboratory to identify differentially regulated proteins correlating to patient outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

NET, normalized elution time; AMT, accurate mass and time; FTICR, Fourier transform ion cyclotron resonance.

Acknowledgment

Portions of this research were supported by the National Institute of General Medical Sciences (NIGMS, Large Scale Collaborative Research Grants U54 GM-62119-02 and T32 GM-008256), the NIH National Center for Research Resources (RR18522), the Entertainment Industry Foundation (EIF), and the EIF Women’s Cancer Research Fund to the Breast Cancer Biomarker Discovery Consortium, and the Environmental Molecular Science Laboratory [a national scientific user facility sponsored by the U.S. Department of Energy (DOE) Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory (PNNL)]. PNNL is operated by Battelle Memorial Institute for the DOE under contract DE-AC06-76RLO-1830.

References

1. Hanash S. Disease proteomics. *Nature* 2003;422(6928):226–232. [PubMed: 12634796]
2. Qian WJ, Jacobs JM, Liu T, Camp DG II, Smith RD. Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications. *Mol. Cell. Proteomics* 2006;5(10):1727–1744. [PubMed: 16887931]
3. Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol* 2005;1(5):252–262. [PubMed: 16408053]
4. Pasa-Tolic L, Jensen PK, Ison GA, Lipton MS, Peden KK, Martinovic S, Tolic N, Bruce JE, Smith RD. High throughput proteome-wide precision measurements of protein expression using mass spectrometry. *J. Am. Chem. Soc* 1999;121:7949–7950.
5. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. U.S.A* 1999;96:6591–6596. [PubMed: 10359756]
6. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 2002;1(5):376–386. [PubMed: 12118079]
7. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol* 1999;17(10):994–999. [PubMed: 10504701]
8. Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA, White FM. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell. Proteomics* 2005;4(9):1240–1250. [PubMed: 15951569]
9. DeSouza L, Diehl G, Rodrigues MJ, Guo J, Romaschin AD, Colgan TJ, Siu KW. Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. *J. Proteome Res* 2005;4(2):377–386. [PubMed: 15822913]
10. Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C. Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem* 2001;73(13):2836–2842. [PubMed: 11467524]
11. Qian WJ, Monroe ME, Liu T, Jacobs JM, Ison GA, Shen Y, Moore RJ, Ison DJ, Zhang R, Calvano SE, Lowry SF, Xiao W, Moldawer LL, Davis RW, Tompkins RG, Camp DG, Smith RD. Quantitative proteome analysis of human plasma following in vivo lipopolysaccharide administration using $^{16}\text{O}/^{18}\text{O}$ labeling and the accurate mass and time tag approach. *Mol. Cell. Proteomics* 2005;4(5):700–709. [PubMed: 15753121]
12. Mason CJ, Therneau TM, Eckel-Passow JE, Johnson KL, Oberg AL, Olson JE, Nair KS, Muddiman DC, Bergen HR III. A method for automatically interpreting mass spectra of ^{18}O -labeled isotopic clusters. *Mol. Cell. Proteomics* 2007;6(2):305–318. [PubMed: 17068186]
13. Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Norton S, Kumar P, Anderle M, Becker CH. Quantification of proteins and metabolites by mass spectrometry without isotope labeling or spiked standards. *Anal. Chem* 2003;75:4818–4826. [PubMed: 14674459]
14. Jaffe JD, Mani DR, Leptos KC, Church GM, Gillette MA, Carr SA. PEPpeR: A platform for experimental proteomic pattern recognition. *Mol. Cell. Proteomics* 2006;5:1927–1941. [PubMed: 16857664]
15. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol* 2006;24(8):971–983. [PubMed: 16900146]
16. Le Bihan T, Goh T, Stewart II, Salter AM, Bukhman YV, Dharsee M, Ewing R, Wisniewski JR. Differential analysis of membrane proteins in mouse fore- and hindbrain using a label-free approach. *J. Proteome Res* 2006;5(10):2701–2710. [PubMed: 17022641]
17. Wang G, Wu WW, Zeng W, Chou CL, Shen RF. Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility, linearity, and application with complex proteomes. *J. Proteome Res* 2006;5(5):1214–1223. [PubMed: 16674111]
18. Roy S, Josephson SA, Fridlyand J, Karch J, Kadoch C, Karrim J, Damon L, Treseler P, Kunwar S, Shuman MA, Jones T, Becker CH, Schulman H, Rubenstein JL. Protein biomarker identification in the CSF of patients with CNS lymphoma. *J. Clin. Oncol* 2008;26(1):96–105. [PubMed: 18056677]

19. Liu T, Qian WJ, Mottaz HM, Gritsenko MA, Norbeck AD, Moore RJ, Purvine SO, Camp DG II, Smith RD. Evaluation of multiprotein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. *Mol. Cell. Proteomics* 2006;5(11):2167–2174. [PubMed: 16854842]
20. Liu T, Qian WJ, Gritsenko MA, Xiao W, Moldawer LL, Kaushal A, Monroe ME, Varnum SM, Moore RJ, Purvine SO, Maier RV, Davis RW, Tompkins RG, Camp DG II, Smith RD. High dynamic range characterization of the trauma patient plasma proteome. *Mol. Cell. Proteomics* 2006;5(10):1899–1913. [PubMed: 16684767]
21. Shen Y, Zhao R, Belov ME, Conrads TP, erson GA, Tang K, Pasa-Tolic L, Veenstra TD, Lipton MS, Smith RD. Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Anal. Chem* 2001;73:1766–1775. [PubMed: 11338590]
22. Belov ME, erson GA, Wingerd MA, Udseth HR, Tang K, Prior DC, Swanson KR, Buschbach MA, Strittmatter EF, Moore RJ, Smith RD. An automated high performance capillary liquid chromatography-Fourier transform ion cyclotron resonance mass spectrometer for high-throughput proteomics. *J. Am. Soc. Mass. Spectrom* 2004;15:212–232. [PubMed: 14766289]
23. Zimmer JS, Monroe ME, Qian WJ, Smith RD. Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom. Rev* 2006;25(3):450–482. [PubMed: 16429408]
24. Monroe ME, Tolic N, Jaitly N, Shaw JL, Adkins JN, Smith RD. VIPER: an advanced software package to support high-throughput LC-MS peptide identification. *Bioinformatics* 2007;23(15):2021–2023. [PubMed: 17545182]
25. Jaitly N, Monroe ME, Petyuk VA, Clauss TR, Adkins JN, Smith RD. Robust algorithm for alignment of liquid chromatography-mass spectrometry analyses in an accurate mass and time tag data analysis pipeline. *Anal. Chem* 2006;78(21):7397–7409. [PubMed: 17073405]
26. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem* 2003;75(17):4646–4658. [PubMed: 14632076]
27. Callister SJ, Barry RC, Adkins JN, Johnson ET, Qian W, Webb-Robertson BM, Smith RD, Lipton MS. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res* 2006;5(2):277–286. [PubMed: 16457593]
28. Park T, Yi SG, Kang SH, Lee S, Lee YS, Simon R. Evaluation of normalization methods for microarray data. *BMC Bioinf* 2003;4:33.
29. Polpitiya AD, Qian WJ, Jaitly N, Petyuk VA, Adkins JN, Camp DG II, erson GA, Smith RD. DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 2008;24(13):1556–1558. [PubMed: 18453552]
30. Ishihama Y, Sato T, Tabata T, Miyamoto N, Sagane K, Nagasu T, Oda Y. Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. *Nat. Biotechnol* 2005;23(5):617–621. [PubMed: 15834404]
31. Qian WJ, Camp DG II, Smith RD. High-throughput proteomics using Fourier transform ion cyclotron resonance mass spectrometry. *Expert Rev. Proteomics* 2004;1(1):87–95. [PubMed: 15966802]
32. Petyuk VA, Qian WJ, Chin MH, Wang H, Livesay EA, Monroe ME, Adkins JN, Jaitly N, erson DJ, Camp DG II, Smith DJ, Smith RD. Spatial mapping of protein abundances in the mouse brain by voxelation integrated with high-throughput liquid chromatography-mass spectrometry. *Genome Res* 2007;17(3):328–336. [PubMed: 17255552]
33. Liu T, Qian WJ, Strittmatter EF, Camp DG, erson GA, Thrall BD, Smith RD. High throughput comparative proteome analysis using a quantitative cysteinyl-peptide enrichment technology. *Anal. Chem* 2004;76:5345–5353. [PubMed: 15362891]
34. Usaite R, Wohlschlegel J, Venable JD, Park SK, Nielsen J, Olsson L, Yates JR III. Characterization of global yeast quantitative proteome data generated from the wild-type and glucose repression *saccharomyces cerevisiae* strains: the comparison of two quantitative methods. *J. Proteome Res* 2008;7(1):266–275. [PubMed: 18173223]
35. Kim YJ, Zhan P, Feild B, Ruben SM, He T. Reproducibility assessment of relative quantitation strategies for LC-MS based proteomics. *Anal. Chem* 2007;79(15):5651–5658. [PubMed: 17580949]

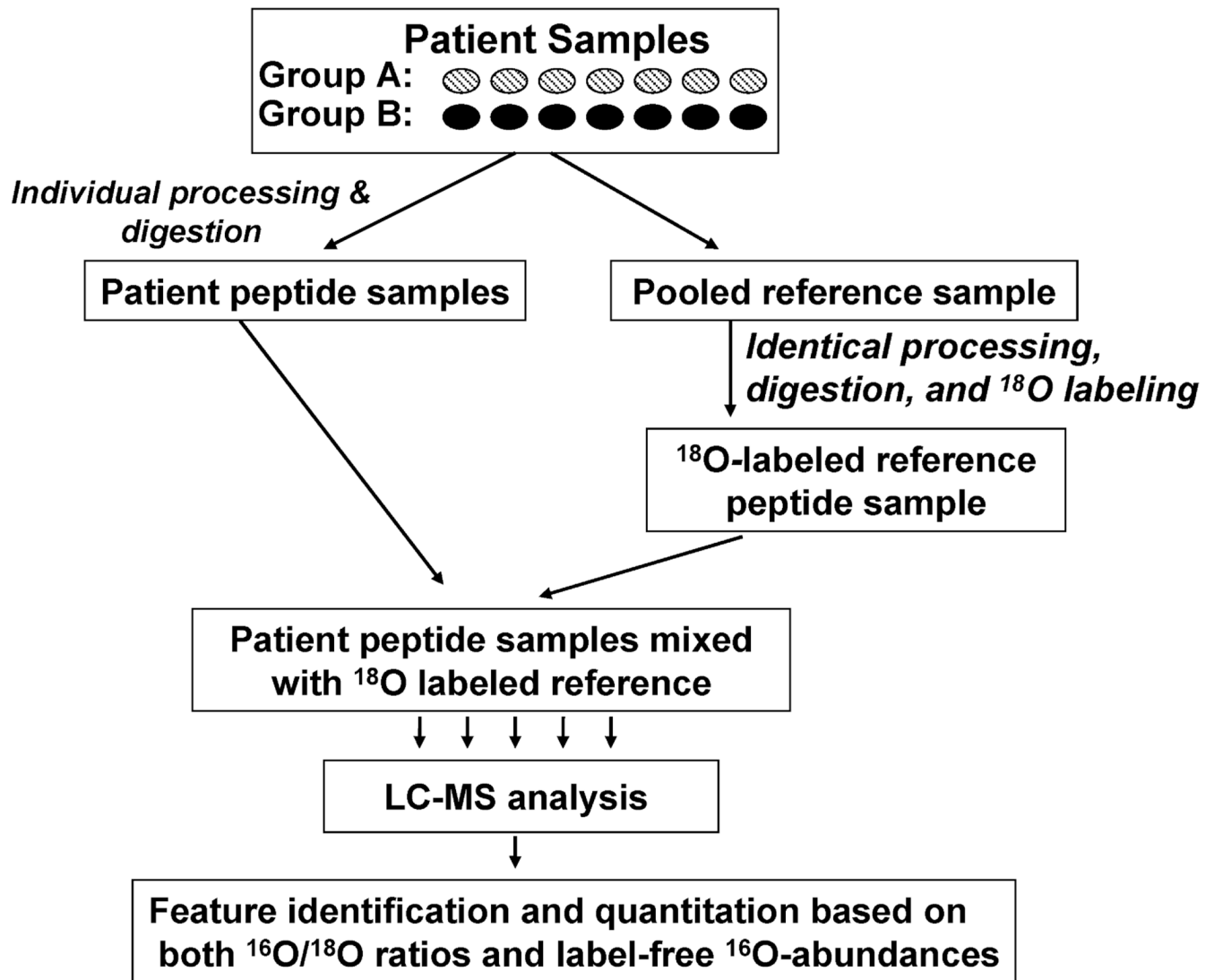


Figure 1.

The quantitation strategy. A reference sample was generated by pooling aliquots from all patient samples. All samples were individually processed into peptides. The reference sample was ^{18}O -labeled and an identical aliquot of reference sample was spiked into each of the patient samples prior to LC-MS analysis.

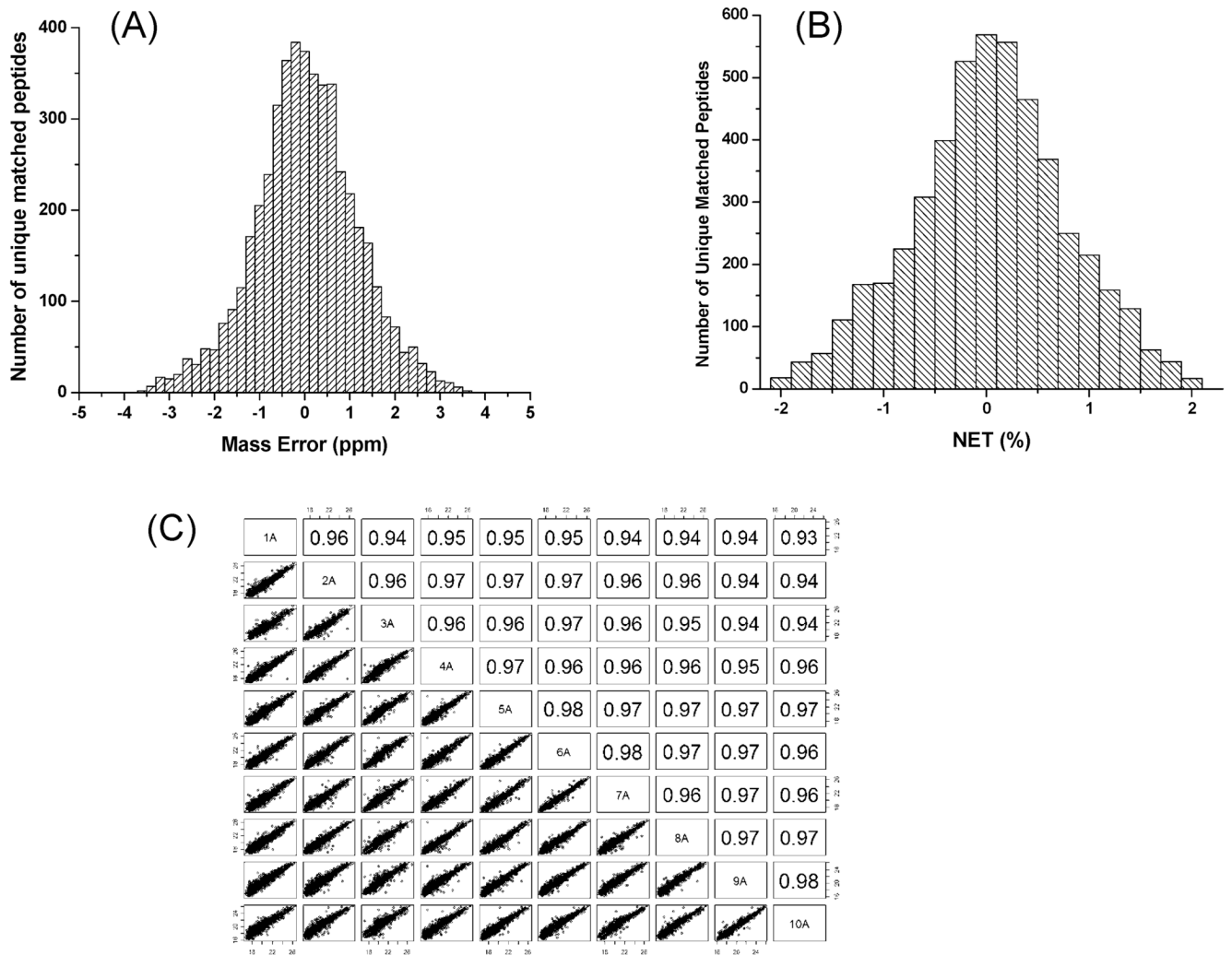
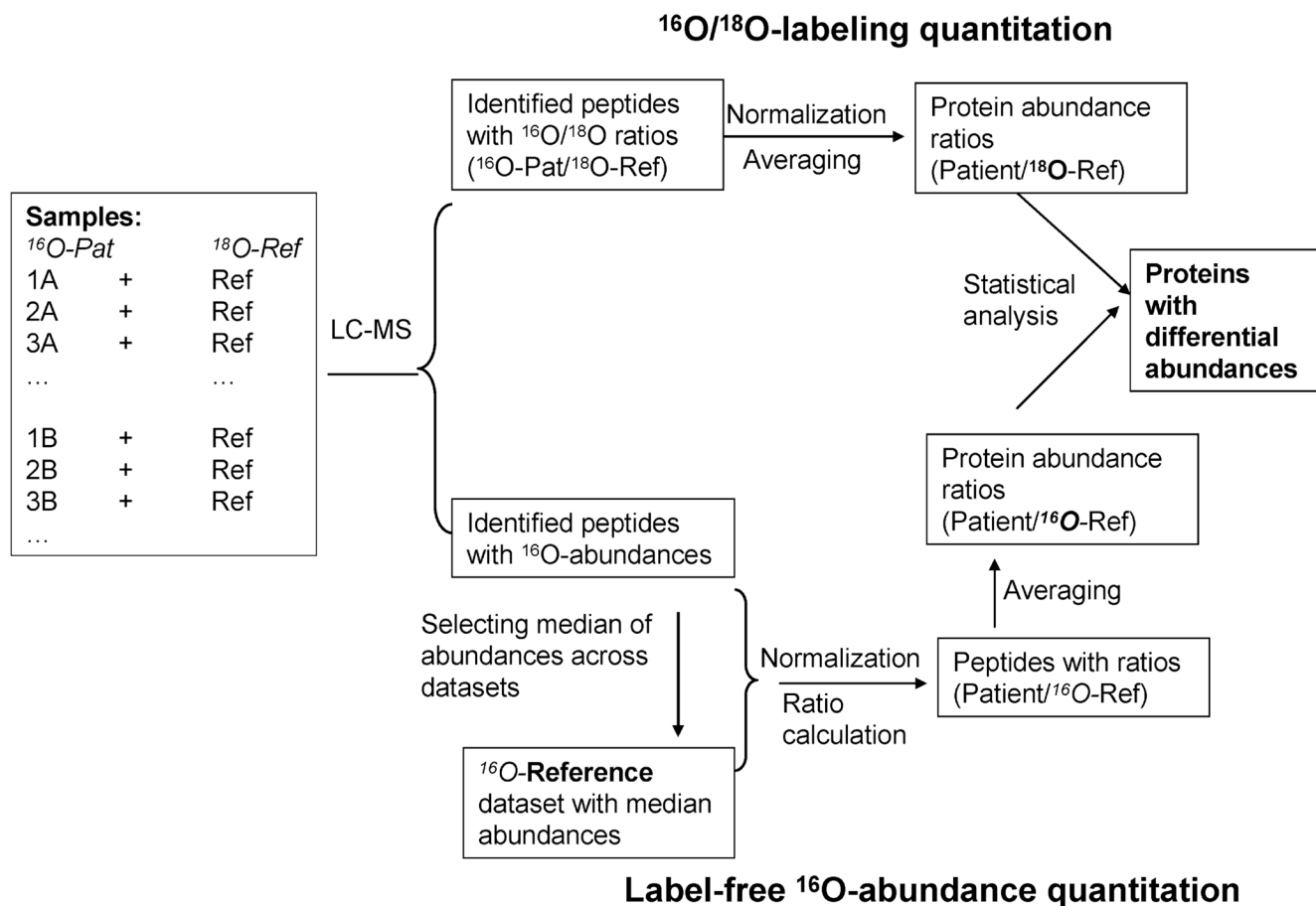
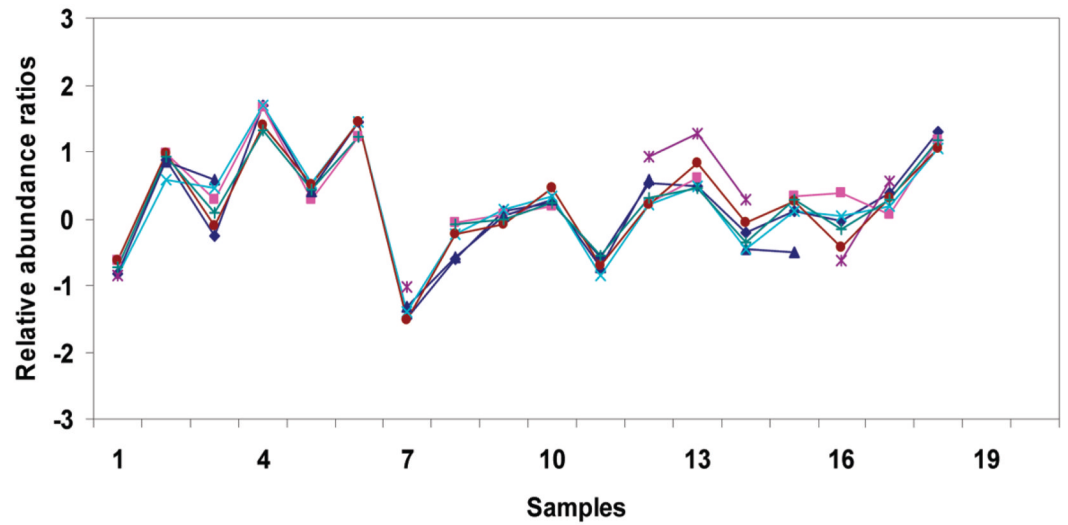


Figure 2. Histogram of mass error distribution (A), NET error distribution (B), and the reproducibility of ^{18}O -reference peptide abundances (C). The pairwise correlation of reference peptide abundances in C were shown in scatter plot style and as Pearson correlation coefficients for 10 selected samples.

**Figure 3.**

Data processing scheme for the dual-quantitation data. All patient samples to be analyzed by LC-MS contained equal amounts of ^{16}O -peptides and ^{18}O -reference peptides. For labeling quantitation, the relative peptide abundances for labeled peptides were automatically reported as $^{16}\text{O}/^{18}\text{O}$ ratios. Following normalization, protein abundance ratios were generated by averaging peptide abundance ratios within a given protein. For label-free quantitation, a ^{16}O -reference data set was computed using the median abundance for each peptide across the patients. All data sets were normalized against the reference prior to the conversion to a relative ratio format similar to the labeled data. Similar transformation of protein abundance ratios for label-free and labeled data was applied for direct comparison of the two approaches.

(A)
 $^{16}\text{O}/^{18}\text{O}$ data



(B)
Label-free
 ^{16}O data

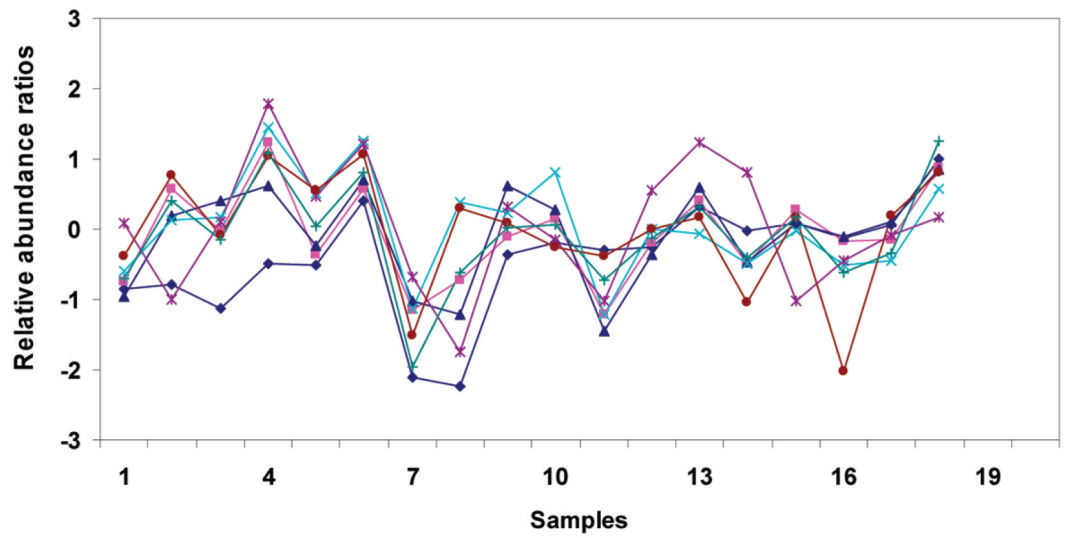
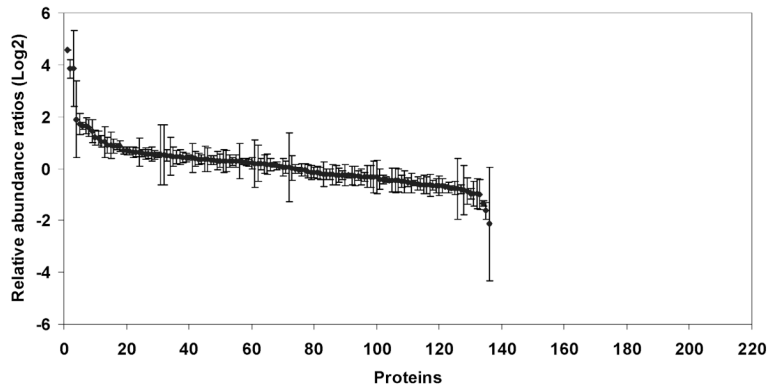
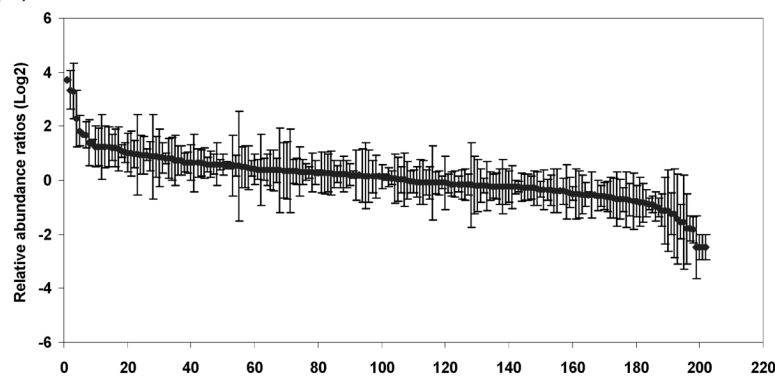
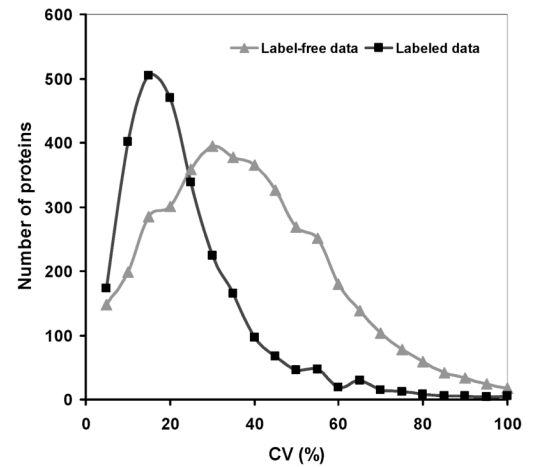


Figure 4. Peptide abundance profiles for C-reactive protein. Seven peptides quantitated commonly by both the label-free and labeling approaches were plotted. The relative abundance ratios are displayed in Log2 transformed format.

(A) $^{16}\text{O}/^{18}\text{O}$ data(B) Label-free ^{16}O data

(C)

**Figure 5.**

Precision of quantitation for the labeling and label-free data. In panels A and B, relative protein abundance ratios comparing two selected patients for all proteins with two or more peptides were plotted with standard deviations of peptide abundance ratios shown for each protein. (C) The distribution of proteins in CV ranges. CV values were calculated based on the standard deviations of multiple peptide abundance ratios belonging to the same protein relative to the protein abundance ratio for a given data set. The CV values for all proteins in individual samples were calculated to generate this plot.

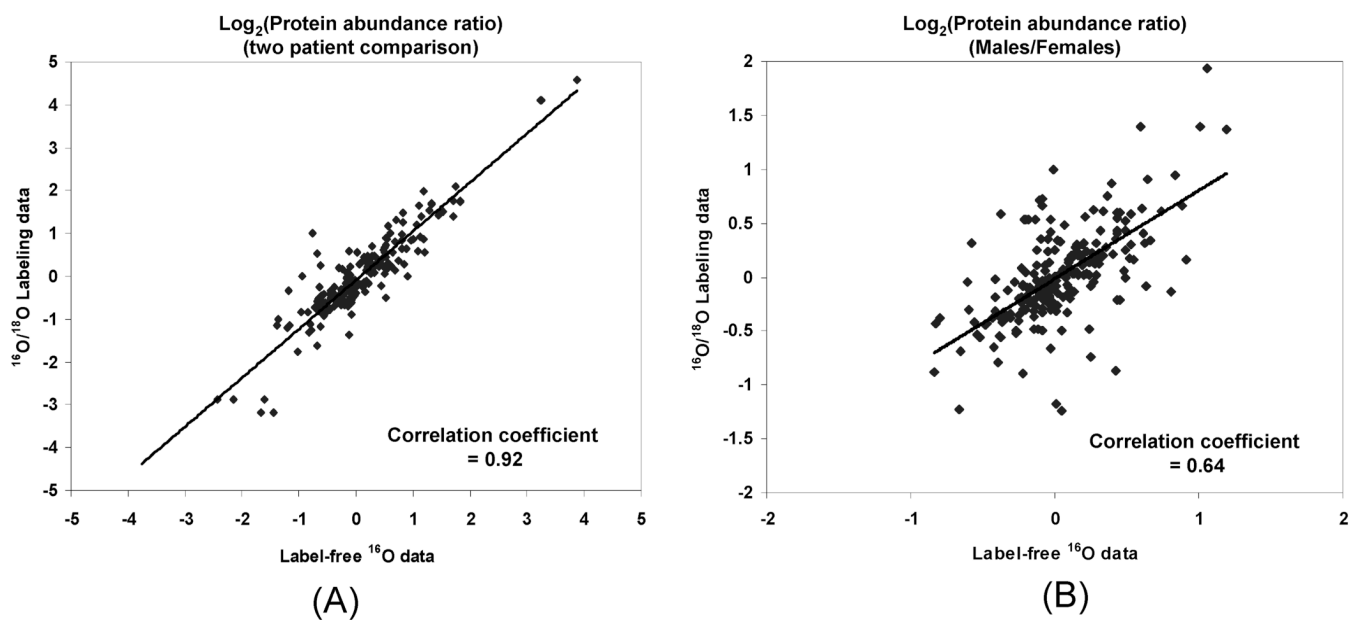


Figure 6. Correlation of relative protein abundance differences between label-free data and the labeled data. (A) Relative protein abundance differences comparing two selected patients. (B) Relative abundance differences comparing two groups of patients (9 females vs 9 males). Log₂(Protein abundance ratio) values derived from label-free data are plotted on the *x*-axis, while values from labeled data are on the *y*-axis.