A Method for Calculating $^{16}O/^{18}O$ Peptide Ion Ratios for the Relative Quantification of Proteomes

Kenneth L. Johnson and David C. Muddiman

Department of Biochemistry and Molecular Biology, W. M. Keck FT-ICR Mass Spectrometry Laboratory, Mayo Clinic College of Medicine, Rochester, Minnesota, USA

A method is described for the identification and relative quantification of proteomes using accurate mass tags (AMT) generated by nLC-dual ESI-FT-ICR-MS on a 7T instrument in conjunction with stable isotope labeling using $^{16}O/^{18}O$ ratios. AMTs were used for putative peptide identification, followed by confirmation of peptide identity by tandem mass spectrometry. For a combined set of 58 tryptic peptides from bovine serum albumin (BSA) and human transferrin, a mean mass measurement accuracy of 1.9 ppm ±0.94 ppm ($CIM_{99\%}$) was obtained. This subset of tryptic peptides was used to measure $^{16}O/^{18}O$ ratios of 0.36 ± 0.09 ($CIM_{99\%}$) for BSA (μ = 0.33) and 1.48 ± 0.47 ($CIM_{99\%}$) for transferrin (μ = 1.0) using a method for calculating $^{16}O/^{18}O$ ratios from overlapping isotopic multiplets arising from mixtures of ^{16}O, $^{18}O_1$, and $^{18}O_2$ labeled C-termini. The model amino acid averagine was used to calculate a representative molecular formula for estimating and subtracting the contributions of naturally occurring isotopes solely as a function of peptide molecular weight. The method was tested against simulated composite $^{16}O/^{18}O$ spectra where peptide molecular weight, $^{16}O/^{18}O$ ratio, $^{18}O_1/^{18}O_2$ ratios, and number of sulfur atoms were varied. Relative errors of 20% or less were incurred when the $^{16}O/^{18}O$ ratios were less than three, even for peptides where the number of sulfur atoms was over- or under-estimated. These data demonstrate that for biomarker discovery, it is advantageous to label the proteome representing the disease state with ^{18}O; and the method is not sensitive to variations in $^{18}O_1/^{18}O_2$ ratio. This approach allows a comprehensive differentiation of expression levels and tentative identification via AMTs, followed by targeted analysis of over- and under-expressed peptides using tandem mass spectrometry, for applications such as the discovery of disease biomarkers. (J Am Soc Mass Spectrom 2004, 15, 437–445) © 2004 American Society for Mass Spectrometry

The characterization of differences between the expressed protein complement, or proteome, of related biological states is a recurring, fundamental task at the center of many research programs in the life sciences. The ability to detect and quantify significant changes in protein expression levels is essential to both discovery (e.g., identification of biomarkers) and hypothesis-driven research. The analytical challenges arise from the high degree of sample complexity and wide dynamic range of proteins that are present in tissues and biofluids. As a common example serum, the most highly sampled biological fluid, has been estimated to contain as many as 10^6 different proteins present at concentrations ranging over 10 orders of magnitude [1].

A variety of analytical approaches have been reported that utilize bottom-up proteomics and stable isotope labeling to perform relative quantification of proteins [2–6]. These methods can be broadly classified as either (1) metabolic, where the isotope label is incorporated during protein synthesis, (2) amino acid specific, where the stable isotope label is applied only to peptides containing a specific amino acid, such as cysteines with the ICAT method, and (3) global labeling methods where the label is applied to every peptide within the complex mixture. Metabolic labeling is only possible in a limited number of circumstances, while amino acid specific methods such as ICAT have the advantage of reducing sample complexity but have disadvantages of discriminating against proteins with low number of cysteines, and being cost-prohibitive especially for matrices such as serum that have a high total protein content where the majority of the total protein content is comprised of a small number of highly abundant proteins.

One method of global labeling inserts an isotopic label via the molecule of water that is incorporated into

Published online February 3, 2004

Address reprint requests to Dr. D. C. Muddiman, W. M. Keck FT-ICR Mass Spectrometry Laboratory, Medical Sciences Bldg. 3-115, Mayo Clinic College of Medicine, 200 First Street S.W., Rochester, MN 55905, USA. E-mail: muddiman.david@mayo.edu

1044-0305/04/$30.00
doi:10.1016/j.jasms.2003.11.016

Received September 18, 2003
Revised November 8, 2003
Accepted November 12, 2003

peptides during cleavage of amide backbones by serine proteases [4, 7–13]. By using $H_2^{18}O$ during proteolysis, ^{18}O is incorporated into the carboxy-terminus of peptides formed during proteolysis. Prior to its use for relative quantification, the insertion of an ^{18}O molecule during proteolysis has been utilized to produce internal standards for absolute quantification [9], identify C-terminal peptides [8], distinguish y-type ions from b-type ions in tandem mass spectrometry sequencing of peptides [12, 14], and identify sites of N-linked glycosylation after enzymatic removal of the glycans [13].

When using ^{18}O for relative quantification of proteomes, a proteome sample representing the control state is digested in $H_2^{16}O$, and the diseased state is digested in the presence of $H_2^{18}O$; subsequently, the two samples are mixed. The relative abundances of ^{16}O and ^{18}O species for each tryptic peptide are used to quantify the relative protein abundance between each proteome.

Differential labeling using ^{18}O water results in overlapping isotope patterns where signal from the ^{18}O label is superimposed with naturally occurring isotopes from the non-labeled sample. Typically, any response from the ^{13}C content of the unlabeled sample is subtracted from the abundance of the isotope peak containing ^{18}O signal after identifying the peptide sequence (most commonly by tandem mass spectrometry) and calculating its ^{13}C contribution to the signal based upon the empirical formula of the identified peptide. This limits the quantification process to peptides that have been identified, which is a significantly reduced subset of the total number of isotope pairs present in a data set.

This report describes a method to decouple the quantification process from peptide identification by estimating the naturally occurring isotope component solely from peptide molecular weight. The isotopic multiplets observed in $^{16}O/^{18}O$ labeling experiments are reduced into their three contributing components representing two biological states: A ^{16}O component which represents the control state, and the $^{18}O_1$ and $^{18}O_2$ components which are summed to represent the disease state. The precision and accuracy of our approach are reported and discussed for a model system using bovine serum albumin (BSA) and human apotransferrin.

Experimental

Preparation of Samples

Bovine serum albumin (BSA), apo-transferrin, DTT, and iodoacetamide were purchased from Sigma Chemical (St. Louis, MO). Stock solutions of BSA and transferrin were prepared at 5 mg/mL and 2 mg/mL, respectively, in 100 mM ammonium bicarbonate, pH 7.0. For the $^{16}O/^{18}O$ labeling experiments, two samples were prepared; the ^{16}O sample consisted of 50 μg each of BSA and transferrin, the ^{18}O sample consisted of 150 μg of BSA and 50 μg of transferrin for an expected $^{16}O/^{18}O$ ratio of 1:3 for BSA and 1:1 for transferrin. Each sample was evaporated to dryness, resuspended in 50 μL of 6 M urea (in 50 mM ammonium bicarbonate, pH 8.4), and sonicated for five min. The samples were reduced with DTT for 30 min at 37 °C, and alkylated with iodoacetamide for 60 min at room temperature in the dark. Reagents for reduction and alkylation were prepared in ^{16}O water. After reduction and alkylation, the samples were evaporated to dryness on a vacuum centrifuge.

Before proteolytic digestion, the two samples were reconstituted in 150 μL of deionized ^{16}O water and ^{18}O water, respectively, using a vortex mixer and brief sonication. ^{18}O water (≥95% purity) was obtained from ICON Stable Isotopes, Summit, NJ. The two samples were digested overnight with TPCK-treated trypsin (Promega, Madison, WI) at 37 °C, using an enzyme: substrate ratio of 1:50 (w : w). Digestion was terminated by the addition of 10 μL of 1% formic acid/0.2% heptafluorobutyric acid (HFBA) and samples were frozen at −20 °C.

Nano-Scale Liquid Chromatography (nLC)

nLC was performed using a 24 cm long by 100 μm i.d. IntegraFrit column (NewObjective, Woburn, MA) packed with Magic C_{18}Aq, 5 μm, 200 Å particles (Michrom BioResources, Auburn, CA). Mobile phase A was water/acetonitrile/n-propanol/formic acid (98/1/1/0.2 by volume), while mobile phase B was acetonitrile/n-propanol/water/formic acid (80/10/10/0.2 by volume). A nLC flow rate of 0.4 μL/min was split from HPLC pumps flowing at 20 μL/min. A gradient from 0 to 50% B over 40 min was used. A reversed phase pre-column (0.3 mm i.d. by 5 mm long, Dionex, Sunnyvale, CA) was used to pre-concentrate and desalt samples within the sampling loop of a 10-port valve (VICI, Houston, TX) prior to switching the valve to place the precolumn in-line with the nano-scale column during development of the gradient. The pre-column was also packed with Magic C_{18}Aq, 5 μm, 200 Å.

nLC-Dual ESI-FTICR-MS Accurate Mass Measurements

nLC-dual ESI-FTICR-MS accurate mass data were obtained on an IonSpec FT-ICR mass spectrometer (IonSpec, Irvine, CA) with a 7 tesla actively shielded magnet, and equipped with a dual sprayer nano-ESI source compatible with coupling to LC that has been described elsewhere [15]. This second generation dual ESI source, building on our success with an earlier design [16, 17], allowed faster (30 ms versus 300 ms) computer-controlled switching between two independent spray emitters, one spraying the nLC eluant, the other spraying a solution containing 1 μM each of bradykinin and glucagon. This allowed for internal mass calibration of all spectra since sample ions and mass calibrant ions were in the ICR cell simultaneously. Each 1024 K word transient was Fourier transformed to the frequency domain after being zero-filled twice and processed with

a Blackman windowing function. The *m/z* axis was then mass calibrated using the monoisotopic peaks from the doubly charged ion of bradykinin (*m/z* 530.7898), and the quadruply and triply charged ions of glucagon at *m/z* 871.1612 and *m/z* 1161.2125, respectively.

nLC-Q-TOF-MS/MS Measurements

nLC-MS/MS measurements were performed on a Q-TOF 2 mass spectrometer (Waters, Milford, MA) with automatic selection of up to 4 precursor ions from every MS survey scan. Tandem mass spectra were processed by MassLynx software (version 3.5) and searched against the NCBI non-redundant database (Feb. 03, ftp.ncbi.nih.gov) using Sequest software (ThermoFinnigan, San Jose, CA).

Results and Discussion

In order to characterize the use of $^{16}O/^{18}O$ ion ratios as a method for quantification of relative protein expression levels, two samples, hereafter referred to as the light and heavy sample, were prepared for $^{16}O/^{18}O$ labeling. Each sample contained known amounts of BSA and transferrin. The two samples were digested with trypsin as described in the Methods section using $H_2^{16}O$ and $H_2^{18}O$ (≥95% purity), respectively. The expected relative abundances between the two samples for the two proteins were 1:3 for BSA and 1 : 1 for transferrin. The tryptic peptides were analyzed by nLC-dual ESI-FT-ICR-MS for accurate mass measurements of peptides and determination of $^{16}O/^{18}O$ ion ratios.

Peptides were identified on the basis of both accurate mass tags (AMT) [18] from the 7T FT-ICR instrument and tandem mass spectrometry data from the Q-TOF 2. From the set of peptides from BSA and transferrin that were identified by both techniques, a mean mass accuracy ± the confidence interval of the mean at the 99% confidence level ($CIM_{99\%}$) of 1.9 ppm ± 0.94 was obtained from 60 measurements.

Figure 1a shows a portion of a FTICR mass spectrum containing the triply charged ion at *m/z* 627.6459 from the BSA tryptic peptide RPCFSALTPDETYVPK, (M_r = 1879.916). This peptide was measured with a mass accuracy of 1.0 ppm using the dual ESI source and internal mass calibration. The spectrum illustrates a type of composite spectrum that is obtained from a $^{16}O/^{18}O$ labeling experiment. The isotopic peaks annotated as A, A + 2, and A + 4 are the peaks used in relative quantification. The non-annotated peaks are primarily composed of $^{13}C_1$ contributions related to the annotated peaks.

The monoisotopic peak, labeled A, primarily represents the light sample, although a contribution to its abundance originates from the heavy sample due to the ≤5% impurity of ^{16}O present in ^{18}O water when only one oxygen has been exchanged (vide infra). The relative abundance of the A + 2 peak is a composite of signals arising from the heavy sample due to variable incorporation of one ^{18}O atom(s), and a predictable abundance arising from the naturally occurring isotopes (e.g., ^{13}C, ^{34}S) of the light sample.

Likewise, the signal at the A + 4 peak is also a composite, with contributions from three entities: (1) From the heavy sample where two ^{18}O atoms have been incorporated, (2) from the heavy sample, as the naturally occurring A + 2 contribution from one ^{18}O atom incorporation, and finally (3) a contribution from the light sample via its naturally occurring A + 4 isotope.

The primary contributors to naturally occurring isotopes 2 and 4 mass units higher than the monoisotopic signal are $^{13}C_2$ and ^{34}S, and $^{13}C_4$, $^{13}C_2^{34}S$, and $^{34}S_2$, respectively. For simplicity, we will refer to A + 2 and A + 4 signals arising from naturally occurring isotopes as the $^{13}C_2$ and $^{13}C_4$ responses.

Another complication to calculation of the $^{16}O/^{18}O$ ratio comes from impurities of ^{16}O present in ^{18}O water. The purity of ^{18}O water used in this study was reported by the vendor to be ≥95%. Assuming 95% purity, the insertion of a water molecule that occurs during the initial hydrolysis of a polypeptide bond will introduce a 5% response into the ^{16}O channel from the heavy sample, and reduce the $^{18}O_1$ response by 5%.

The incorporation of two ^{18}O atoms, resulting in a 4 da mass increase, has been widely reported and can be desirable for more easily separating the light sample from the heavy sample. Methods have been reported for driving this enzyme-mediated oxygen substitution, or EMOS, reaction toward incorporation of two ^{18}O atoms [19–21]. For this to happen, a tryptic peptide must re-associate with the protease, where another oxygen atom from water can be exchanged for one of the two oxygens in the free carboxy terminus of the peptide. Due to the resonance of the carboxyl oxygens at pH 8, this second event has equal chances of replacing the recently incorporated ^{18}O or the other ^{16}O atom. Therefore, to accomplish a quantitative 4 da mass shift (quantitative incorporation of two ^{18}O atoms), multiple EMOS events must occur for every tryptic peptide.

After the initial enzyme hydrolysis which generates a 50:50 distribution of ^{16}O and ^{18}O for the two carboxyl oxygens, four additional EMOS events would be expected to increase the probability that both oxygen atoms had been exchanged to 97%. At that point the overall distribution of ^{16}O, $^{18}O_1$, and $^{18}O_2$ species within the heavy sample would be limited by the purity of the ^{18}O water. If that condition were met, then the contribution to the ^{16}O channel would be negligible: 0.25% as described by $(a + b)^2$, where a is the abundance of ^{16}O (0.05) and b is the abundance of ^{18}O (0.95) in the $H_2^{18}O$. The term a^2 represents the relative contribution to the ^{16}O signal, 2ab represents the contribution to the $^{18}O_1$ signal, and b^2 represents the contribution to the $^{18}O_2$ signal.

In our experience, given high sample complexity (e.g., such as a tryptic digest of serum), together with the variability in kinetics of re-association with the enzyme for different peptide sequences, it is common to

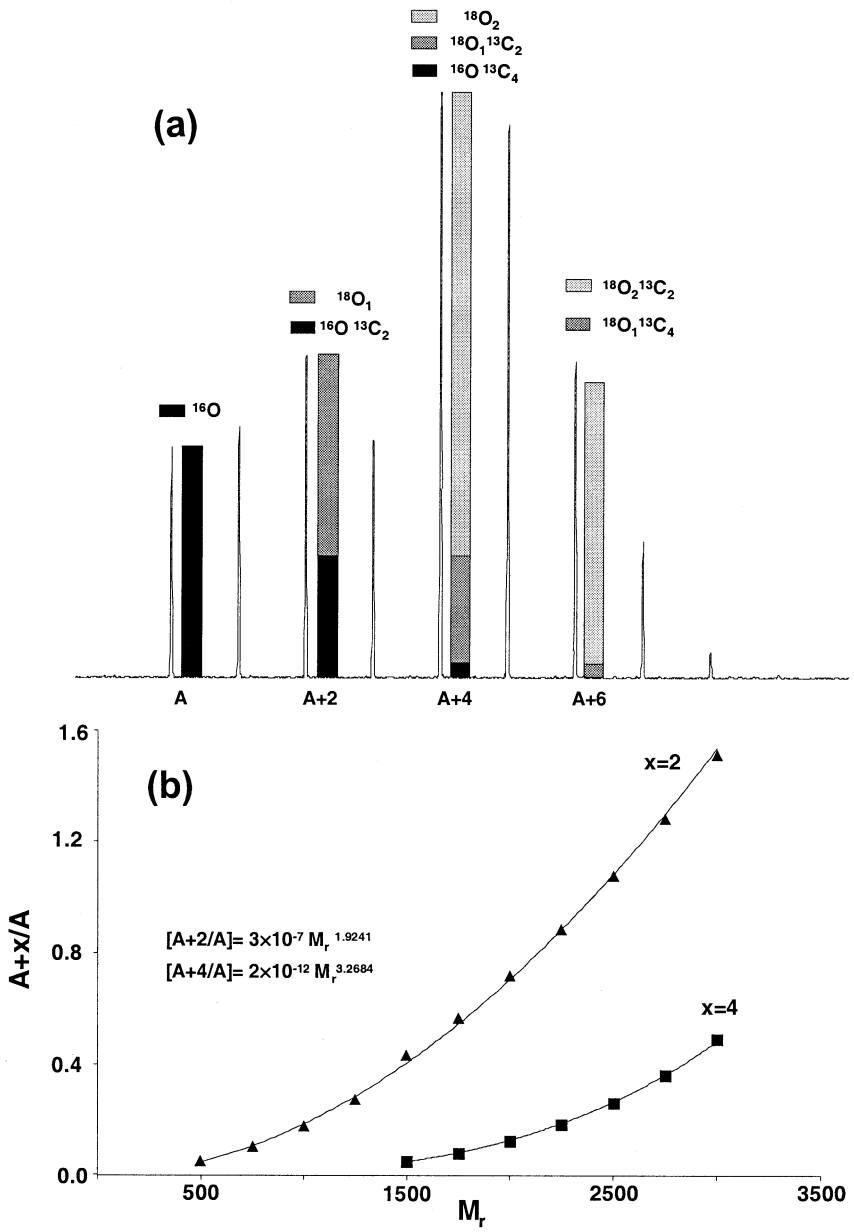

Figure 1. (**a**) FTICR mass spectrum of a $^{16}O/^{18}O$-labeled triply charged ion from the BSA tryptic peptide RPCFSALTPDETYVPK. The histograms show the relative contributions from ^{16}O and ^{18}O labeled species to the A, A + 2, and A + 4 isotopic peaks. (**b**) Plot of naturally occurring A + 2 and A + 4 relative abundance versus peptide molecular weight. A + 2 and A + 4 relative abundances were calculated using the model amino acid averagine to generate chemical formulas for the isotope modeling software within MassLynx V4.0.

see a significant variability in the ratio of $^{18}O_1$ to $^{18}O_2$. Therefore, we applied a correction of 5% to the ^{16}O and $^{18}O_1$ ion abundances due to contribution by the heavy sample to the ^{16}O signal, and we define the absolute abundances of the contributing components to the $^{16}O/^{18}O$ ratio as:

$$^{16}O=[A]-(0.05\times{}^{18}O_1) \quad (1)$$

$$^{18}O_1=[A+2]-{}^{16}O{}^{13}C_2+(0.05\times{}^{16}O) \quad (2)$$

$$^{18}O_2=[A+4]-{}^{18}O_1{}^{13}C_2-{}^{16}O{}^{18}C_4 \quad (3)$$

Calculation of Naturally Occurring Isotope Distributions for Peptides

The abundances of naturally occurring stable isotopes are well characterized and algorithms for calculating the relative abundances of these isotopes have been reported [22, 23]. Corrections to $^{16}O/^{18}O$ ratios based

upon these calculations have also been applied in relative quantification methods [4, 13]. However, all of these methods are predicated on knowing the amino acid sequence or composition, and using its chemical formula to calculate the relative abundances of the naturally occurring isotopes. This approach requires peptide identification, primarily done from tandem mass spectral data and database searching, before relative quantification can be performed.

We have developed a method that calculates the relative abundances of naturally occurring stable isotopes solely as a function of peptide molecular weight. Thus, relative quantification can be done without prior knowledge of sequence. This offers the advantage of allowing sample differentiation and putative peptide identification, via relative quantification and AMTs, independently of tandem mass spectrometry. Subsequently, tandem mass spectrometry can be performed on a targeted subset of tryptic peptides to augment the accurate mass tag for confident peptide identification.

The method uses the computational amino acid averagine, described by Senko et al., that was derived by calculating an average amino acid from all proteins in the PIR database [24]. Averagine has a chemical formula of $C_{4.9384}H_{7.7583}N_{1.3577}O_{1.4773}S_{0.0417}$ and an average molecular weight of 111.1254 Da. Using averagine as implemented in ICR-2LS [25], we generated a table of molecular formulas and their respective monoisotopic masses for a series of average chemical masses ranging from 500–3000 Da. We used the isotope modeling software in MassLynx version 4.0 to calculate theoretical abundances of the A + 2 and A + 4 ions relative to A, and plotted relative ion abundance versus monoisotopic mass. These data were fit using a power function to generate equations that describe the relative abundances for the naturally occurring A + 2 and A + 4 ions as a function of peptide molecular weight. The results of this are shown in Figure 1b where the relative abundance of the naturally occurring A + 2 ion (A + 2/A) over an M_r range of 500–3000 is described by eq 4:

$$[A + 2/A] = 3 \times 10^{-7} M_r^{1.9241} (R^2 = 0.9989) \quad (4)$$

and the relative abundance of the naturally occurring A + 4 ion (A + 4/A) over an M_r range of 1500–3000 is described by eq 5:

$$[A + 4/A] = 2 \times 10^{-12} M_r^{3.2684} (R^2 = 0.9993) \quad (5)$$

The lower molecular weight range for this curve was limited to 1500 since the contribution to the A + 4 channel by naturally occurring isotopes is less than 5% below 1500 Da.

Eqs 4 and 5 were used to restate the abundances of ^{16}O, $^{18}O_1$, and $^{18}O_2$ terms (eqs 1–3) as:

$$^{16}O = [A] - (0.05 \times {}^{18}O_1) \quad (6)$$

$$^{18}O_1 = [A + 2] - (^{16}O \times (3 \times 10^{-7} M_r^{1.9241})) + (0.05 \times {}^{16}O) \quad (7)$$

$$^{18}O_2 = [A + 4] - (^{18}O_1 \times (3 \times 10^{-7} M_r^{1.9241})) - (^{16}O \times (2 \times 10^{-12} M_r^{3.2684})) \quad (8)$$

Calculation of $^{16}O/^{18}O$ ratios for BSA and Transferrin Tryptic Peptides

For calculation of $^{16}O/^{18}O$ ratios we made the reasonable assumption that molar responses of the ^{16}O and ^{18}O peptides are equivalent, and that the ^{18}O abundance is described by the sum of ion abundances from the $^{18}O_1$ and $^{18}O_2$ species.

Figure 2a and b compares the results obtained using averagine-based chemical formulas versus using sequence-specific molecular formulas to calculate $^{16}O/^{18}O$ ratios for BSA (1 : 3 ratio) and transferrin (1 : 1 ratio) for peptides identified by both tandem mass spectrometry and accurate mass measurements. The slope of the lines indicate a strong correlation between the two methods for calculating naturally occurring contributions to the A + 2 and A + 4 ion abundances; the data are described by the lines $y = 0.983X + 0.0022$, $R^2 = 0.998$ for BSA ($^{16}O/^{18}O$ ratio of 0.33) and $y = 0.965X + 0.0053$, $R^2 = 0.997$ for transferrin ($^{16}O/^{18}O$ ratio of 1.0). From the slope of each line, we can see that using averagine to calculate A + 2 relative abundance gave $^{16}O/^{18}O$ ratios within 5% of ratios obtained when naturally occurring A + 2 and A + 4 ratios were calculated using their specific amino acid sequence. Using a 99% confidence interval, the y-intercepts for both BSA and transferrin pass through the origin. Moreover, the deviations did not appear to be biased as a function of peptide molecular weight or $^{18}O_1/^{18}O_2$ ratio (data not shown).

For BSA, the mean $^{16}O/^{18}O$ ratio obtained from 26 peptides was 0.36 ± 0.09 ($CIM_{99\%}$). For transferrin, the mean $^{16}O/^{18}O$ ratio obtained from 32 peptides was 1.48 ± 0.47 ($CIM_{99\%}$). For each protein, the mean value ± the confidence interval of the mean at the 99% confidence interval, contained the expected relative expression ratio. In addition, Figure 2c and d show that most of the deviation of ion ratios from the accepted value occurred at lower ion abundances. When the transferrin data set is recalculated for all data points with ion abundance greater than an arbitrary value of 1.0, the mean ± $CIM_{99\%}$ is 1.25 ± 0.32.

To further characterize the ability of averagine to calculate naturally occurring contributions to the A + 2 and A + 4 ion abundances, we simulated composite $^{16}O/^{18}O_1/^{18}O_2$ spectra for tryptic peptides from BSA and transferrin. Initially, naturally occurring isotope abundances were calculated from peptide sequences with molecular weights of approximately 1000, 1500, and 2500 using the isotope modeling tool of the Micromass software (MassLynx version 4.0). Isotope distribu-

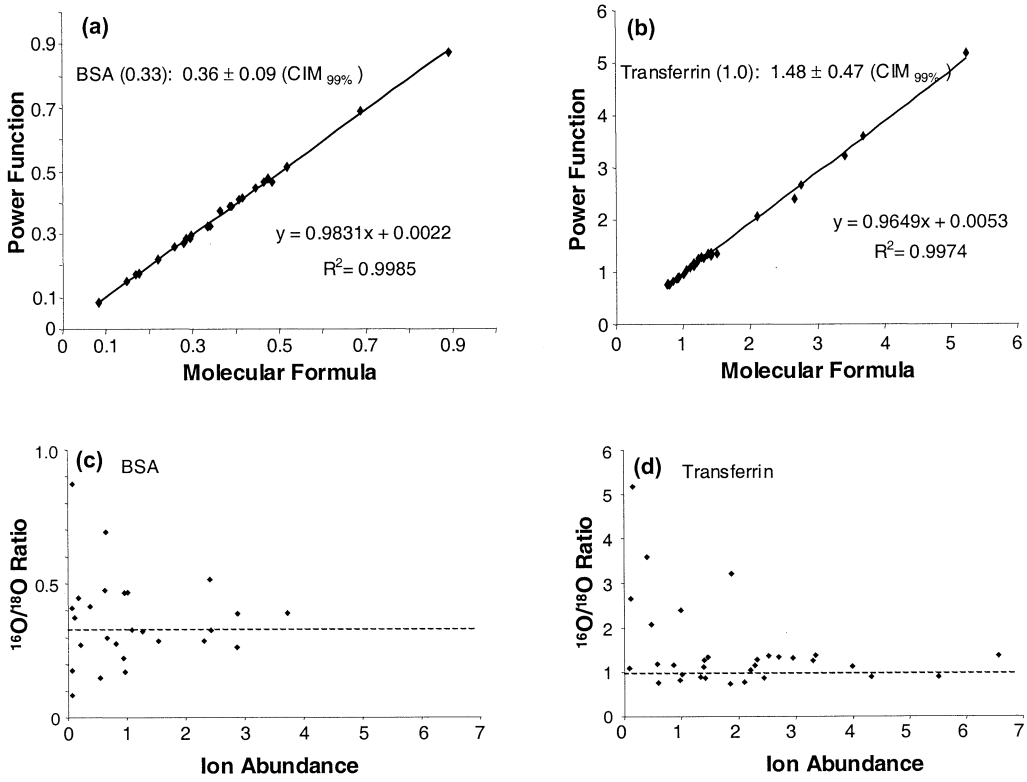

Figure 2. Plot of $^{16}O/^{18}O$ ion ratios calculated using averagine (Y-axis) versus from specific chemical formulas (x-axis) for (**a**) BSA and (**b**) transferrin; and $^{16}O/^{18}O$ ion ratios measured plotted versus ion abundance (amplitude) for (**c**) BSA and (**d**) transferrin. The dashed lines depict the expected $^{16}O/^{18}O$ ratios for these proteins (μ = 0.33 for BSA and 1.0 for transferrin).

tions for $^{18}O_1$- and $^{18}O_2$-labeled equivalents were then created by duplicating the relative abundances of the naturally occurring isotopes shifted by 2 and 4 mass units. Composite spectra were created by summing the channels contributed by the ^{16}O, $^{18}O_1$, and $^{18}O_2$ isotope distributions.

Additionally, the relative abundances of the ^{16}O, $^{18}O_1$, and $^{18}O_2$ isotope distributions were scaled using two input parameters: The first parameter specified the overall $^{16}O/^{18}O$ ratio, and the second parameter varied the relative abundance of the $^{18}O_1$ and $^{18}O_2$ forms for the heavy sample. An example is shown in Figure 3 for the transferrin tryptic peptide YLGEEYVK, ($[M + H^+]^+$ = 1000.503, $C_{47}H_{69}N_9O_{15}$) using a $^{16}O/^{18}O$ ratio of 1.0 and a fractional value of 0.5 for the $^{18}O_1$ abundance (i.e., a 1:1 distribution between $^{18}O_1$ and $^{18}O_2$). We tested our averagine-based method for calculating $^{16}O/^{18}O$ ratios against composite spectra generated from sequences at molecular weight ranges from 1000 to 2500 mass units. The $^{16}O/^{18}O$ ratio was varied from 0.1 to 10, while using fractional values of 1, 0.5, and 0 for the relative abundance of $^{18}O_1$ versus $^{18}O_2$.

The data in Figure 4 summarizes errors in $^{16}O/^{18}O$ ratios that are introduced by averagine-based empirical formulas solely based on molecular weight. The relative error (the percent difference between calculated values and prescribed values) varied between 0 and 35% over the range of molecular weights that were tested. From this data the greatest errors were observed when the $^{16}O/^{18}O$ ratio is substantially greater than 1, since it is more difficult to discern small additions to the A + 2 or

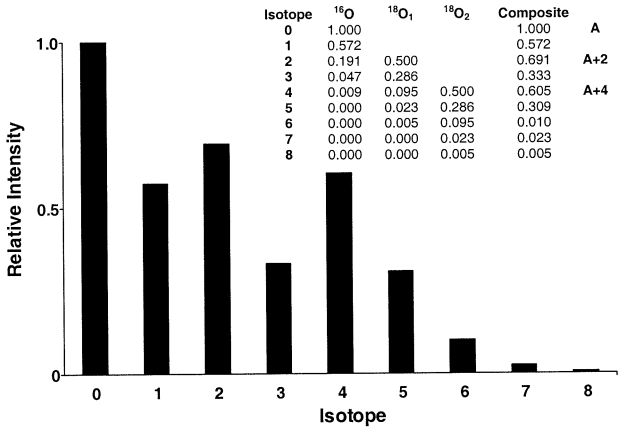

Figure 3. A composite $^{16}O/^{18}O$ spectrum formed from relative isotope abundances calculated for $C_{47}H_{69}N_9O_{15}$, the chemical formula for the tryptic peptide YLGEEYVK, ($[M + H^+]^+$ = 1000.503, $C_{47}H_{69}N_9O_{15}$). The composite spectrum was formed using a $^{16}O/^{18}O$ ratio of 1.0 and a $^{18}O_1/^{18}O_2$ distribution of 0.5 ($^{18}O_1/(^{18}O_1+^{18}O_2)$) as shown in the inset table. The calculated $^{16}O/^{18}O$ ratio using our averagine-based method was 0.987, a −1.3% error from actual.

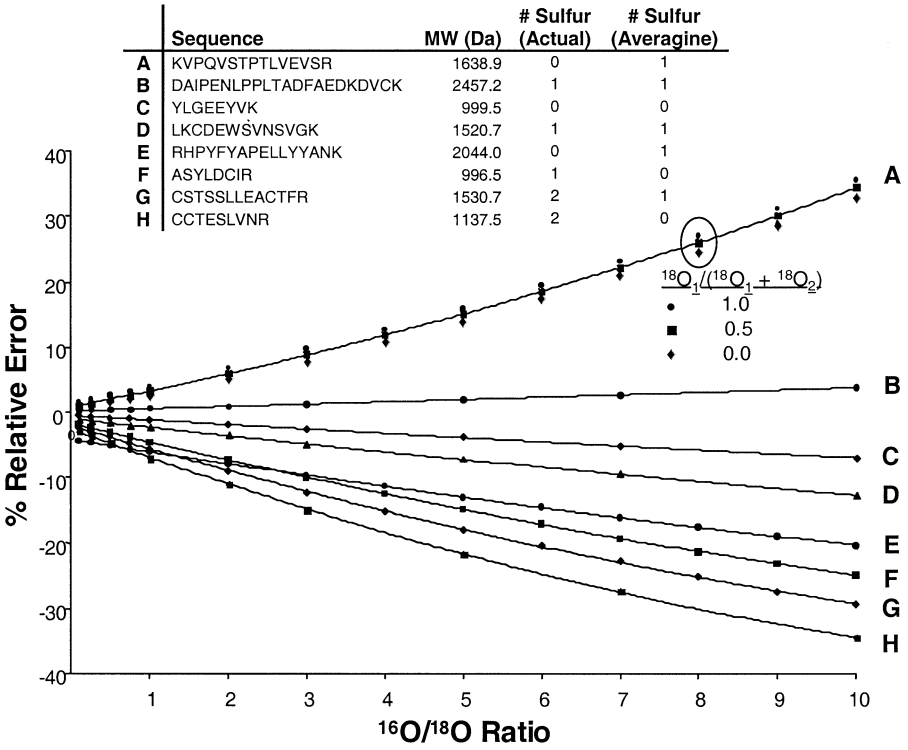

Figure 4. Plot of relative errors incurred using averagine to calculate $^{16}O/^{18}O$ ion ratios for eight peptides, as a function of $^{16}O/^{18}O$ ratio, peptide molecular weight and sulfur content. The eight data series are described in the table by peptide sequence, molecular weight, and actual sulfur content versus the sulfur content predicted by averagine. Data series A shows the relative error introduced by varying the distribution between $^{18}O_1$ and $^{18}O_2$.

A + 4 ion channels. These data show that the ^{18}O label is optimally applied to the sample state where information on up-regulated proteins is being sought (e.g., for disease biomarker discovery, the ^{18}O label would preferentially be applied to the diseased sample rather than the control).

Another source of significant error occurs when the number of sulfur atoms from cysteine or methionine residues differs from the number of sulfur atoms predicted by averagine. Using averagine to generate representative chemical formulas for peptides based only on peptide molecular weight, the transition from zero sulfur atoms to one sulfur atom occurs at approximately 1335 Da; this point of sulfur incorporation is observed in Figure 1b as an inflection in the A + 2 curve between 1250 and 1500 Da.

For example, if the calculated $^{16}O/^{18}O$ ratio is <5, and there is a mismatch of one sulfur atom between the averagine model and actual peptide sequence, an approximate error of $\pm 20\%$ will be introduced into the $^{16}O/^{18}O$ ratio; if the $^{16}O/^{18}O$ ratio is less than one, the relative error decreases to less than 10%. The largest errors will occur for peptides at lower mass containing two sulfur atoms. As an extreme example, the BSA peptide CCAADDK, 838 Da (as carbamidomethylated cysteines), and containing two sulfur atoms, incurs an error of -42% for a $^{16}O/^{18}O$ ratio of 10 and when all of the ^{18}O label is expressed as $^{18}O_1$. However, for a $^{16}O/^{18}O$ ratio of 1.0, the error ranges from 10–12%; and even at a $^{16}O/^{18}O$ value of 3, the error is 19–21% depending on $^{18}O_1/^{18}O_2$ distribution. These errors are in the same range of uncertainties that have been reported for the ICAT method [26].

Yao et al, have recently reported a method to drive proteolysis toward a complete incorporation of two molecules of ^{18}O [19], and have also demonstrated the effectiveness of the labeling method for a low molecular weight fraction of serum [27]. While there may be an advantage to incorporate two molecules of ^{18}O in analyses where isotopes are not fully resolved, the distribution between $^{18}O_1$ and $^{18}O_2$ introduces less error into the overall $^{16}O/^{18}O$ ratio calculation than when the ratio is greater than 3, or when mis-estimation of sulfur content occurs.

Even with complete incorporation of $^{18}O_2$, a need for calculating contributions from naturally occurring isotopes still exists when one considers that peptides with molecular weight of 2000 have a naturally occurring A + 4/A ratio of 0.12; and for a 2500 mass unit peptide, the naturally occurring A + 4/A ratio increases to 0.26.

Conclusions

We have described a method for the putative identification and relative quantification of protein

expression levels that combines the use of accurate mass tags and stable isotope labeled ion ratios generated from nLC-dual ESI-FT-ICR-MS to calculate $^{16}O/^{18}O$ ratios without prior knowledge of amino acid composition of the peptides being quantified. This allows comprehensive differentiation of expression levels by $^{16}O/^{18}O$ ratios, and tentative peptide identification using accurate mass tags, followed by targeted tandem mass spectrometry on peptides of interest wherein the accurate mass tag does not result in a unique protein identification. Using a model system consisting of BSA and transferrin, we obtained a mean MMA of 1.9 ppm ±0.94 ppm ($CIM_{99\%}$, $n = 58$) on a 7T instrument. The same data set gave a mean $^{16}O/^{18}O$ ratio of 0.36 ± 0.09 ($CIM_{99\%}$) for 26 peptides from BSA ($\mu = 0.33$) and 1.48 ± 0.47 ($CIM_{99\%}$) for 32 peptides from transferrin ($\mu = 1.0$). Our data suggests that in the context of detecting proteins with elevated relative expression levels, such as in biomarker discovery, there is an advantage to labeling the disease sample with the ^{18}O label and the control sample with ^{16}O. The use of averagine-based calculations to estimate naturally occurring isotopes should also prove effective for other isotope-coded experiments where overlapping isotopic distributions are present.

While we are pursuing the use of FT-ICR-MS for the discovery of biomarkers because of the technique's advantages in sensitivity, mass resolving power, mass measurement accuracy, and dynamic range, the method reported here for calculating $^{16}O/^{18}O$ ratios is also applicable to data from any mass analyzer capable of resolving and accurately recording the abundances of the contributing isotopes.

Acknowledgments

The authors gratefully acknowledge the financial support from the W. M. Keck Foundation and the Mayo Clinic College of Medicine.

References

1. Anderson, N. L.; Anderson, N. G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol. Cell Proteom.* **2002,** *1,* 845–867.
2. Gygi, S. P.; Rist, B.; Gerver, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. *Nat. Biotechnol.* **1999,** *17,* 994–999.
3. Veenstra, T. D.; Martinovic, S.; Anderson, G. A.; Pasa-Tolic, L.; Smith, R. D. Proteome Analysis Using Selective Incorporation of Isotopically Labeled Amino Acids. *J. Am. Soc. Mass Spectrom.* **2000,** *11,* 78–82.
4. Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic 18O-Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus. *Anal. Chem.* **2001,** *73,* 2836–2842.
5. Regnier, F. E.; Riggs, L.; Zhang, R.; Xiong, L.; Liu, P.; Chakraborty, A.; Seeley, E.; Sioma, C.; Thompson, R. A. Comparative Proteomics Based on Stable Isotope Labeling and Affinity Selection. *J. Mass Spectrom.* **2002,** *37,* 133–145.
6. Goshe, M. B.; Smith, R. D. Stable Isotope-Coded Proteomic Mass Spectrometry. *Curr. Opin. Biotechnol.* **2003,** *14,* 101–109.
7. Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D. Protease-Catalyzed Incorporation of 18O into Peptide Fragments and Its Application for Protein Sequencing by Electrospray and Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Electrophoresis* **1996,** *17,* 945–953.
8. Rose, K.; Savoy, L.; Simona, M. G.; Offord, R. E.; Wingfield, P. C-Terminal Peptide Identification by Fast Atom Bombardment. *Biochem. J.* **1998,** *250,* 253–259.
9. Desiderio, D. M.; Kai, M. Preparation of Stable Isotope-Incorporated Peptide Internal Standards for Field Desorption Mass Spectrometry Quantification of Peptides in Biologic Tissue. *Biomed. Mass Spectrom.* **1983,** *10,* 471–479.
10. Wang, Y. K.; Ma, Z.; Quinn, D. F.; Fu, E. W. Inverse 18O-Labeling Mass Spectrometry for the Rapid Identification of Marker/Target Proteins. *Anal. Chem.* **2001,** *73,* 3742–3750.
11. Mirgorodskaya, O. A.; Kozmin, Y. P.; Titov, M. I.; Körner, R.; Sönksen, C. P.; Roepstorff, P. Quantitation of Peptides and Proteins by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Using 18O-Labeled Internal Standards. *Rapid Commun. Mass Spectrom.* **2000,** *14,* 1226–1232.
12. Uttenweiler-Joseph, S.; Neubauer, G.; Christoforidis, S.; Zerial, M.; Wilm, M. Automated De Novo Sequencing of Proteins Using the Differential Scanning Technique. *Proteomics* **2001,** *1,* 668–682.
13. Kuster, B.; Mann, M. 18O-Labeling of N-Glycosylation Sites to Improve the Identification of Gel-Separated Glycoproteins Using Peptide Mass Mapping and Database Searching. *Anal. Chem.* **1999,** *71,* 1431–1440.
14. Shevchenko, A.; Chernushevich, I.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. Rapid "de novo" Peptide Sequencing by a Combination of Nanoelectrospray, Isotopic Labeling, and a Quadrupole/Time-of-Flight Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **1997,** *11,* 1015–1024.
15. Nepomuceno, A.I.; Muddiman, D. C.; Bergen, III, H. R.; Craighead, J. R.; Burke, M. J.; Caskey, P. E.; Allan, J. A. Dual Electrospray Ionization Source for Confident Generation of Accurate Mass Tags Using Liquid Chromatography Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2003,** *75,* 3411–3418.
16. Flora, J. W.; Hannis, J. C.; Muddiman, D. C. High Mass Accuracy of Product Ions Produced by SORI-CID Using a Dual Electrospray Ionization Source Coupled with FTICR Mass Spectrometry. *Anal. Chem.* **2001,** *73,* 1247–1251.
17. Hannis, J. C.; Muddiman, D. C. A Dual Electrospray Ionization Source Combined with Hexapole Accumulation to Achieve High Mass Accuracy of Biopolymers in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2000,** *11,* 876–883.
18. Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R. An Accurate Mass Tag Strategy for Quantitative and High-Throughput Proteome Measurements. *Proteomics* **2002,** *2,* 513–523.
19. Yao, X.; Afonso, C.; Fenselau, C. Dissection of Proteolytic 18O-Labeling: Endoprotease-Catalyzed 16O to 18O Exchange of Truncated Peptide Substrates. *J. Proteome Res.* **2003,** *2,* 147–152.
20. Reynolds, K. J.; Yao, X. ; Fenselau, C. Proteolytic 18O-Labeling for Comparative Proteomics: Evaluation of Endoprotease Glu-C as the Catalytic Agent. *J. Proteome Res.* **2002,** *1,* 27–33.

21. Liu, P.; Regnier, F. E. An Isotope Coding Strategy for Proteomics Involving both Amine and Carboxyl Group Labeling. *J. Proteome Res.* **2002,** *1,* 443–450.
22. Rockwood, A. L.; VanOrden, S. L. Ultrahigh-Speed Calculation of Isotope Distributions. *Anal. Chem.* **1996,** *68,* 2027–2030.
23. Yergey, J. A. A General Approach to Calculating Isotopic Distributions for Mass Spectrometry. *Int. J. Mass Spectrom. Ion Phys.* **1983,** *52,* 337–349.
24. Senko, M. W.; Beu, S. C.; McLafferty, F. W. Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions. *J. Am. Society Mass Spectrom.* **1995,** *6,* 229–233.
25. Anderson, G. A., Bruce J. E., Smith R. D. ICR-2LS, version 2.18; Richland, WA, 1996.
26. Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. Quantitative Profiling of Differentiation-Induced Microsomal Proteins Using Isotope-Coded Affinity Tags and Mass Spectrometry. *Nat. Biotechnol.* **2001,** *19,* 946–951.
27. Heller, M.; Mattou, H.; Menzel, C.; Yao, X. Trypsin Catalyzed 16O to 18O Exchange for Comparative Proteomics: Tandem Mass Spectrometry Comparison Using MALDI-TOF, ESI-QTOF, and ESI-Ion Trap Mass Spectrometers. *J. Am. Soc. Mass Spectrom.* **2003,** *14,* 704–718.