
An Automated Method for the Analysis of Stable Isotope Labeling Data in Proteomics

Xiang Zhang, Wade Hines, Jiri Adamec, John M. Asara,
and Stephen Naylor

Beyond Genomics Inc., Waltham, Massachusetts, USA

Fred E. Regnier

Department of Chemistry, Purdue University, West Lafayette, Indiana, USA

An algorithm is presented for the generation of a reliable peptide component peak table from liquid chromatography-mass spectrometry (LC-MS) and subsequent quantitative analysis of stable isotope coded peptide samples. The method uses chemical noise filtering, charge state fitting, and deisotoping toward improved analysis of complex peptide samples. Overlapping peptide signals in mass spectra were deconvoluted by correlation with modeled peptide isotopic peak profiles. Isotopic peak profiles for peptides were generated in silico from a protein database producing reference model distributions. Doublets of heavy and light labeled peak clusters were identified and compared to provide differential quantification of pairs of stable isotope coded peptides. Algorithms were evaluated using peptides from digests of a single protein and a seven-protein mixture that had been differentially coded with stable isotope labeling agents and mixed in known ratios. The experimental results correlated well with known mixing ratios. (J Am Soc Mass Spectrom 2005, 16, 1181–1191) © 2005 American Society for Mass Spectrometry

Mass spectrometry has long relied on stable isotope coding strategies to make up for the lack of a mass spectrometric equivalent of Beers' Law. Multiple variant strategies targeted to differential quantification of peptides offer an increasingly diverse set of tools for tackling a wide range of problems ranging across the bioanalytical spectrum from detailed explorations of macromolecular complexes, compositional changes in subcellular organelles, and "high throughput" discovery including biomarker discovery [1, 2].

Both in vitro and in vivo labeling methods have been championed. With in vitro labeling, such as the GIST [3–6], ICAT [7, 8], and ^{18}O [9–11] coding methods, experimental samples are differentially coded with isotopically distinct labeling agents (after, before, or during proteolytic digestion, respectively) and then mixed. The ratio of concentration of peptides is determined from the observed ratio of their isotopic isoforms in mass spectra. With in vivo [12–14] metabolic labeling, small animals and cell cultures are fed on isotopically distinct diets. At some stage of bioanalytical sample preparation before mass spectrometric analysis, samples are mixed. Peptide concentration ratios are again

reflected in observed ratios of paired isotopically enriched and normal component peaks.

Both of these techniques rely on the observation and principle that molecules with the same chemical structure, differing only in isotopic composition, will compete equally and respond equally if analyzed simultaneously within a mass spectrometer. While the overall response factor of any specific analyte may be a complex function of experimental conditions, including competitive ion suppression, the response factor is the same for the isotopically labeled and unlabeled component and so cancels when comparing the ratio of their observed mass spectrometric signal.

Several software packages such as MSQuant [15], ReLEx [16], and ASAPRatio [17] have been developed to find quantitative information about proteins and peptides. However, robust translation of the simple principle into practice remains a challenge. Not least among these challenges is high fidelity data reduction from raw mass spectral data through peak tables and onto peptide (and protein) specific differentials. The objective of the work described here is to present an algorithm that quantitatively deconvolutes overlapping mass spectral peaks producing an accurate table of component peaks, identifies pairs of differentially labeled components, and calculates the ratio of stable isotope coded peptide pairs. The algorithm includes chemical noise filtering, peak detection, charge elucidation, deisotoping, doublet recognition, and doublet quantification. A C++ program, GISTool, was devel-

Published online May 26, 2005

Address reprint requests to Dr. X. Zhang, Bindley Bioscience Center, Purdue University, 475 Stadium Mall Drive, West Lafayette, IN 47907, USA.
E-mail: zhang100@purdue.edu

oped to implement the algorithm. The effectiveness of the algorithm was evaluated in the analysis of GIST labeled peptides from a single protein and from protein mixtures.

Experimental

Materials and Methods

Bovine serum albumin, horse heart myoglobin, chicken lysozyme, bovine β -lactoglobulin, bovine carbonic anhydrase II, glucose oxidase from *Aspergillus niger*, N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid] (HEPES), 4-vinylpyridine, urea, N-hydroxysuccinimide, ($^1\text{H}_6$) acetic anhydride, and ($^2\text{H}_6$) acetic anhydride were purchased from Sigma Aldrich (St. Louis, MO). Human serotransferrin was supplied by Calbiochem (San Diego, CA). Sequence grade modified trypsin was obtained from Promega (Madison, WI). (2-Carboxyethyl) phosphine (TCEP) was purchased from Pierce (Rockford, IL). Succinimidyl-($^{12}\text{C}_3$)-propionate and succinimidyl-($^{13}\text{C}_3$)-propionate were synthesized by Chemical Diversity Laboratories (San Diego, CA). HPLC grade trifluoroacetic acid and acetonitrile were obtained from Burdick and Jackson (Muskegon, MI). LC-MS data were acquired with a Waters/Micromass Q-TOF API-US instrument using a $360\ \mu\text{m}$ o.d. \times $75\ \mu\text{m}$ i.d. microcapillary column from New Objective Inc. (Woburn, MA) self-packed to 10 cm long with $10\ \mu\text{m}$ C_{18} from YMC (Kyoto, Japan). A C_{18} symmetry trap cartridge was purchased from Waters Corporation (Milford, MA) and used to desalt samples online. Double-deionized water ($\text{H}_2\text{O}_{\text{dd}}$) was produced by a Milli-Q Plus system.

Proteolysis

Bovine serum albumin, human serotransferrin, and a seven-protein mixture consisting of bovine serum albumin, human serotransferrin, horse myoglobin, chicken lysozyme, bovine β -lactoglobulin, bovine carbonic anhydrase II, and glucose oxidase from *Aspergillus niger* were dissolved at a concentration of 1 mg/ml in 1 ml of 20 mM HEPES buffer (pH 8.0) containing 8 M urea and 10 mM TCEP (to reduce the disulfide bonds). Samples were incubated for 1 h at room temperature ($25\ ^\circ\text{C}$) and 4-vinyl-pyridine was added to a final concentration of 20 mM. Reactions proceeded for an additional hour at room-temperature. Samples were diluted 4-fold with 20 mM HEPES (pH 8.0) containing 7 mM CaCl_2 (5 mM final concentration). Sequencing grade trypsin (2%) was added and reaction mixtures were incubated at room temperature overnight.

Succinimidyl-($^1\text{H}_3$)-Acetate and Succinimidyl-($^2\text{H}_3$)-Acetate Synthesis

Succinimidyl-($^1\text{H}_3$)-acetate and succinimidyl-($^2\text{H}_3$)-acetate were synthesized according to the literature. Briefly, 1.77 g of N-hydroxysuccinimide was added to either 4.72 g of

($^1\text{H}_6$) acetic anhydride, or to ($^2\text{H}_6$) acetic anhydride (both are liquid) and the mixture was stirred at room-temperature for 15 h. Crystals were harvested by filtration using a fritted funnel, extensively washed with hexane and dried under vacuum.

Peptide Acylation (Labeling)

Succinimidyl-($^{12}\text{C}_3$)-propionate (light propionate), succinimidyl-($^{13}\text{C}_3$)-propionate (heavy propionate), succinimidyl-($^1\text{H}_3$)-acetate (light acetate), or succinimidyl-($^2\text{H}_3$)-acetate (heavy acetate) (50 mM final concentration) were individually added to $100\ \mu\text{l}$ of a tryptic digestion mixture and incubated for 3 h at room temperature. Light and heavy labeled forms of peptide mixtures were mixed in a 1:1 or 3:1 ratio.

Results and Discussion

The Algorithm

An outline of the processing is presented in Figure 1. LC-MS spectra are preprocessed into centroided data recording mass spectral peaks as m/z and intensity values in a series of scans. Scans sample component peaks in the chromatographic time scale. Each LC-MS dataset is recorded in a separate CDF file. Raw centroid data is processed in a filtering step to distinguish coherent signal from noise. Sets of centroids that are potentially related as isotope clusters are identified and passed on to charge and isotope fitting procedures. Initial best fit charge state assignments are determined for local clusters of peaks. From these, peptide mass can be calculated and an appropriate isotope distribution

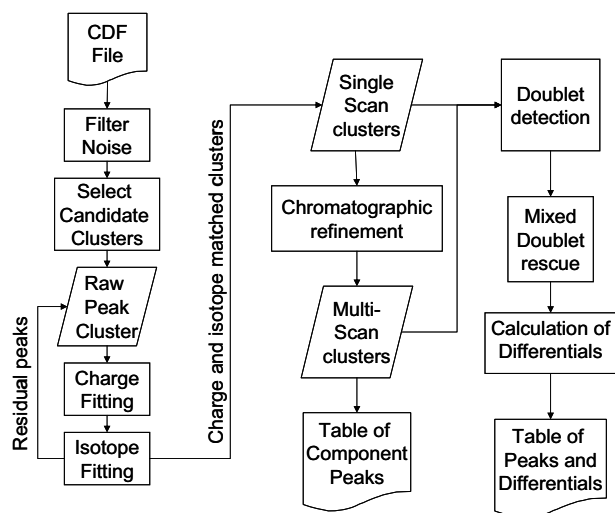


Figure 1. Process flow chart. LC-MS data in the form of a netCDF file is processed through noise filtering and a peak cluster resolving phase that resolves isotope clusters and charge state. The single scan peaks can be resolved chromatographically and dumped as a component peak table or further processed for quantification of doublets. An alternate path evaluates doublets per scan and consolidates into consensus differentials.

model is used for fitting and subsequent detection of additional components within the local peak cluster. The resulting table of monoisotopic component peaks is validated in the chromatographic dimension. If a mixture of light and heavy labeled peptides is being analyzed, potential doublet pairs are found, graded and relative abundances are determined. An alternative pathway fits doublets directly from single scan cluster peaks.

The production of a component peak table from charge and isotope fitted centroids, including deconvolution of overlapping peaks, is a key alternative role for this algorithm.

Spectral Input

LC/MS data can be acquired in profile mode or in centroid mode. If the LC/MS data are acquired in profile mode, the data need to be subsequently converted to centroid data using the instrumental software. Centroid data is effectively a series of peaks in three principle dimensions: a center in the m/z dimension, an intensity in the abundance dimension, and a retention time in the chromatographic dimension. The data files encoding peak centroids were converted into netCDF format which then served as input data for the software described below.

Removing Chemical Noise

Background chemical noise is present as peak centroids in the neighborhood of coherent signals. The noise signals are significantly less intense than “true” signals in most cases but are observed to increase in the neighborhood of large signals. Therefore a noise filter is varied according the local signals in the manner described below.

An initial noise level (ξ) is chosen as follows. The entire spectrum is separated into multiple equal segments along the m/z axis. Within each segment, all peaks are compared to that segment’s base peak and a peak intensity distribution is recorded and smoothed with a Savitzky-Golay filter [18]. An intensity cutoff is chosen where the curve fitted to the distribution has a slope of 1 representing a threshold for a rapidly increasing number of peaks. Figure 2 shows a typical example of this distribution. An initial local noise is assigned at the intensity where the slope of the curve is 1 by default. The program also allows users to define the criterion for determining the initial noise. The relative intensity of the initial noise is then converted back into an absolute value and assigned to ξ for the specific segment. While the default cutoff choice applied is somewhat arbitrary, our experience supports that more relaxed thresholds fail to increase the number of coherent doublet signals that can be observed.

After the assignment of the initial noise threshold for each segment, a second pass filter is applied within the local ranges of peptide isotope clusters. The algorithm

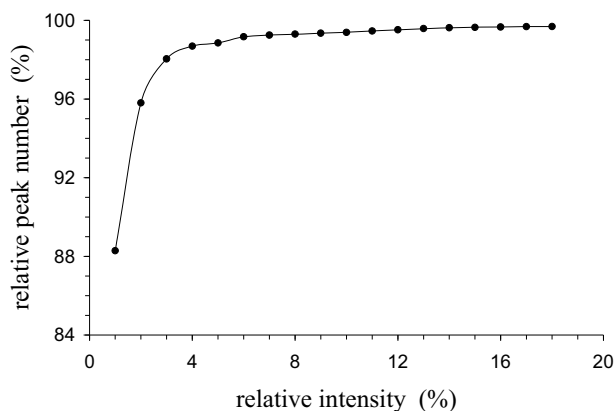


Figure 2. Determination of initial chemical noise level for each segment. Peak intensity is normalized to the most intense peak. Relative number of peaks refers to the number of peaks that have intensity less than the relative intensity.

begins with minimum m/z centroid recorded in the spectrum and proceeds by challenging centroids with successively larger m/z values. The first peak whose intensity exceeds the initial noise level of ξ is assigned as a local base peak. A local noise level is established as the maximum of the initial noise level and 15% of the intensity of this local base peak. This percentage of a large peak permits at least two higher m/z isotopic peaks to pass the intensity challenge. A local search region is also established ranging up +1 m/z unit from the local base peak. This width insures that the next larger isotope peak will be considered, regardless of the charge state. The algorithm then challenges successively larger m/z centroids until either (1) a peak more intense than the current local base peak is found or, (2) the peak being tested exceeds the local search region of the current local base peak. Any peak less intense than the local noise level is considered to be chemical noise and ignored. If a peak having a larger intensity than the current local base peak is detected, the current local base peak is reset to this larger peak. The local noise level is then recalculated according to the newly assigned local base peak and the local search region is updated. If the second condition is met, the noise level is reset to the initial value ξ of the segment. The algorithm continues until all peaks in a spectrum have been challenged.

The following example illustrates how the local base peak and local search region are used. Relative to a specific base peak ($m/z = \chi$), the local search region is defined as $(\chi, \chi + 1.0 + \epsilon)$, where ϵ is a matching tolerance in m/z for identifying peaks belonging to the same isotope cluster. This insures at least one higher m/z isotopic component falls within the considered range. For a peptide that has a charge state of 2+ and the isotopic peak intensity ranks are $M0 < M1 > M2 > M3 > M4 > M5$, the program considers M0 to be the initial local base peak supposing the intensity of M0 is higher than ξ . The local search region relating to the base peak M0 will cover peaks M0, M1, M2, and any

peak between. The local noise level is defined as 0.15 M0. Any peak between M0 and M1 with a peak height less than the local noise level will be considered chemical noise and be deleted.

The program then moves to peak M1. The peak height of M1 will be compared with the height of M0. Because $M1 > M0$, M1 now is defined as the base peak and the local search region will be updated to cover peaks M1, M2, and M3. This routine of the newly defined local search region will expire after peak M3. Thus, any peak following M3 with a peak height larger than ξ will be considered as the latest base peak and variables are updated again. Suppose M4 was the latest base peak, the latest local search region would then cover peaks M4, M5, and any other peaks in this range. If M4 is less intense than ξ , the program will search forward until it finds a peak that is more intense than ξ and then repeat the steps described above.

Chromatographic information was also used to differentiate between chemical noise and low intensity peptide signals. Peptides generally elute from chromatography columns over a span of 15–60 s. When a low intensity peptide peak overlaps with an intense peak from another peptide in a mass spectrum, it can appear to be noise. To rescue the small peptide peaks in the overlapping peak cluster, the peaks were searched in adjacent scans to determine whether they were part of a persistent chromatographic profile. Peaks that would have been discarded as noise are retained if these peaks with the same m/z continuously appear in greater than 80% of scans within an empirically chosen peak width parameter.

It occasionally occurs with commercial data systems for mass spectrometers that a single isotopic peak is split into two or more centroids while processing the raw data. It is necessary to detect this phenomenon and combine the split peaks. The program was developed to address this problem by starting with the peak of lowest m/z and searching forward in m/z to detect all peaks that have the same m/z , within a specified accuracy. If peaks that meet this criterion are detected, they are removed from the spectrum and an integrated peak is added to the spectrum to replace them. The intensity of the integrated peak is calculated as the sum of the intensities of the deleted peaks. The m/z of the integrated peak has the same value as that of the most intense peak among the deleted peaks.

Deisotoping

The term “deisotoping” is used here in reference to the recognition and stripping of signals from higher mass natural isotope peaks in mass spectra while leaving the monoisotopic peak as the representative signal. The monoisotopic peak (M0) contains only the most common isotopes (all ^{12}C , ^{14}N , ^{16}O , ^{32}S , and ^1H) while the heavier peaks in the cluster (M1, M2 etc.) successively include higher mass isotopes. For instance, the M1 peak is a composite peak including from ions with a single

^{13}C isotope or a single ^{15}N isotope while the M2 peak includes contributions from ions with two ^{13}C isotopes, one ^{13}C , and one ^{15}N or one ^{18}O . Removal of these higher mass natural isotope peaks is frequently used to simplify spectra. There are two major steps associated with deisotoping. One is ion charge state recognition [19]; the other is correlation of experimentally measured isotopic peak clusters with theoretically predicted isotopic peak profiles. The initial charge assignment relies on the spacing of peaks in the m/z dimension. Peak intensities are used in a subsequent step to address the potential of overlapping signals from multiple peptides. These algorithms were designed for analysis of data acquired with high-resolution mass spectrometers, such as TOF based instruments. The procedure for charge state determination is as follows.

1. Starting from peaks with the lowest m/z in a mass spectrum, a group of unprocessed peaks that overlap each other within $1.0 + \varepsilon$ is found, where ε is the maximum measurement error allowed for fitted isotopic peaks.
2. The most intense peak in the selected peak group is found and used as the base peak. The maximum allowed charge state (z_{max}) is calculated from the spacing between the base peak and its nearest neighbor.
3. Alternate charge states are tested with this group of peaks for integer charges between 1 and the maximum allowed charge state determined above. Theoretical m/z values for each considered charge are calculated with the base peak as the initial reference. The score of the charge state z is calculated using the empirically chosen formula $\omega_z = (\sum S_i/N_i)/(n_z + 1)$, where S_i/N_i is signal to chemical noise ratio of peak j and n_z is the number of peaks that don't match the predicted m/z .
4. The most probable modeled charge state of a peak group, or partial peak group, is assigned based on the charge state that gives the highest total score for ω_z . The quality of charge state assignment can also be evaluated with $\log(\omega_m/\omega_s)$, where ω_m is the score of the most probable charge state and ω_s is the score of the second most probable charge state.
5. Peaks that have been assigned a charge are identified as processed peaks. This initially assigned charge state of these processed peaks is recorded as z_m . The experimental peak profile of the processed peaks is fitted to a model isotopic peak profile, which will be described below. The quality of the fit is evaluated by $r^2 = \sum(I_{e,i} - I_{t,i})^2$, where $I_{e,i}$ is the experimental peak height and $I_{t,i}$ is the expected peak height calculated from the isotopic peak profile of the generic peptide and intensity of the monoisotopic peak of the processed peak group. Peaks whose intensities are significantly higher than the expected intensities were flagged. If the m/z space between two flagged peaks matches the m/z space of z_m/n ($n = 2, 3, \dots, z_m$) and this happens to all the

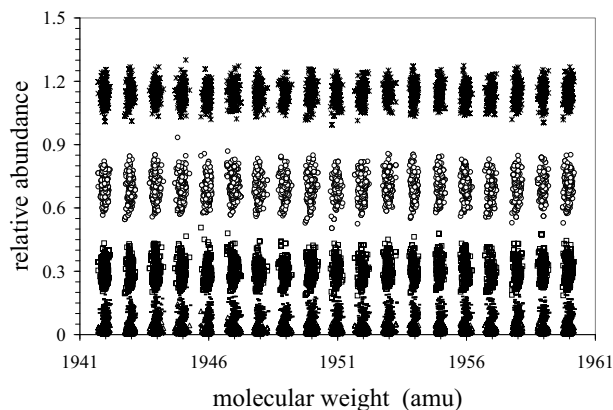


Figure 3. In silico generated isotopic peak profiles of tryptic peptides derived from the Swiss-Prot database. The abundance of each isotopic peak is normalized to the monoisotopic peak M0 by setting the abundance of the monoisotopic peak to 1. Asterisk represents M1/M0, open circle represents M2/M0, open square represents M3/M0, minus sign represents M4/M0, open triangle represents M5/M0, and plus sign represents M6/M0.

flagged peaks, the flagged peaks are temporally assigned the charge z_m . The cluster is then reevaluated with this alternate charge according to algorithm described from Step 2 to Step 4. The most probable charge state of the processed peaks is the charge or mix of charges that give the smallest r^2 .

- The residual of each processed peak is put back into the spectrum as an unprocessed peak. The program then goes back to Step 2 if there are unprocessed peaks in the selected peak group.
- Steps 2 through 6 were repeated until all peaks in the selected peak group have been processed. The program then goes back to Step 1.

Modeling Peptide Isotope Profiles

Natural isotopic peak profiles are a function of the amino acid composition of a peptide. Several algorithms have been developed to calculate isotopic peak profiles for a given peptide [20, 21], or to resolve overlapped isotopic peaks in a known peptide mixture [22]. However, peptide sequence is unknown during peak detection in the case being examined here. To generate a generic peptide peak profile, isotopic peak profiles were calculated for a collection of peptides and recorded within banded mass ranges. This method is intended to produce a model that describes the naturally occurring range of isotope ratios that are likely to be encountered for a given peptide mass range rather than modeling a fixed set of isotope ratios for that same mass range.

All proteins in the Swiss-Prot database [23] were cleaved in silico with trypsin specificity. For each in silico peptide, the isotopic peak distribution was calculated and normalized to a monoisotopic peak intensity of 1. Figure 3 plots the normalized isotopic peak profiles of all in silico tryptic peptides with molecular weight

between 194 and 196 u. It can be seen that the normalized isotopic peaks form various clusters based on peptide molecular weight and type of isotopic peaks (e.g., M1, M2, ...). Cluster distribution is relatively tight, which means that predicting isotopic peak profile for each specific peptide using the in silico profile information will not introduce large variation.

For the purpose of computational efficiency, peptides of molecular weight in the region of $20^*i - 20^*(i + 1)$ were sorted into the same group, where $i = (0, 1, 2, \dots)$. In the mass region of i , where peptide molecular weight is larger than 20^*i and smaller than $20^*(i + 1)$, abundance of the isotopic peak j of the model peptide ($M_{i,j}$) was defined as a median abundance of isotopic peak j with all in silico peptides in the mass region i . High and low extremes for relative isotopic intensity were calculated from the means of the top and bottom 5% of calculated relative isotope ratios. Extremes of the distribution calculated for peptides in this mass range and recorded as $M_{i,j,t}$ and $M_{i,j,b}$, respectively. This boundary information was recorded as in silico boundary $T_{i,j}$.

$$T_{i,j} \in (M_{i,j,b}, M_{i,j,t}) \quad (1)$$

Measurement error estimates are used to calculate a range for the observed experimental distribution. The experimental boundary of a measured isotopic peak ($E_{i,j}$) in an isotopic peak cluster, whose molecular weight is within mass region i , was calculated with the equations

$$I_{i,j,b} = (I_j - \varepsilon_j) / (I_0 + \varepsilon_0) \quad (2)$$

$$I_{i,j,t} = (I_j + \varepsilon_j) / (I_0 - \varepsilon_0) \quad (3)$$

$$E_{i,j} \in (I_{i,j,b}, I_{i,j,t}) \quad (4)$$

where, I_j is the intensity of isotopic peak j , I_0 is the intensity of the monoisotopic peak, and ε_j is the uncertainty of I_j , which is the sum of chemical noise at isotopic peak j and statistical uncertainty of I_j . $I_{i,j,b}$ is the minimum intensity of the isotopic peak j , and $I_{i,j,t}$ is the maximum intensity of peak j .

The peptide molecular weight of a certain group of experimental isotopic peaks is calculated from the m/z and charge state. A modeled peptide distribution for the matching molecular weight band is extracted from an isotopic peak distribution array and used in a challenge to the experimentally measured isotope profile. An overlap between the modeled boundary $T_{i,j}$ and the experimental boundary $E_{i,j}$ indicates that the measured isotopic peak height is allowable, e.g., the measured height of the isotopic peak j is reasonably derived from a single peptide. In the case where there is no overlap between these two boundaries, likely a second nonisobaric peptide contributed to the measured peak height. The contribution of the first peptide to the measured peak height is estimated as $I_0 * M_{i,j}$ and the residual $(I_j - I_0 * M_{i,j})$ is considered as the contribution of the second peptide. The residuals were placed back in the spectrum as unprocessed peaks.

After deisotoping, the monoisotopic peak of a peptide isotope peak cluster is preserved as the sole recorder for subsequent analysis while the other isotopic peaks were disregarded (though a verbose component peak table that includes the complete profile is optionally produced). To determine whether these deisotoped peaks are true peptide peaks at the chromatographic level, peaks with the same m/z in adjacent scans were selected to construct a chromatographic profile. The width of the chromatographic peak was compared with a user defined minimum chromatographic peak width. The chromatographic peak was removed if its peak width was less than the minimum value.

It is common in liquid chromatography that some organic compounds may contaminate the sample, solvent, or system. With high mass accuracy, peptide fractional mass provides analytical information. For modest sized peptides, fractional mass is simply the difference between the accurate monoisotopic mass and its nearest smaller integer. It has been pointed out that peptide molecular weights cover only limited “allowed” regions of fractional mass on the mass scale interrupted by “forbidden” zones [24, 25]. With increasing mass, the “allowed” regions of fractional mass increase at the cost of “forbidden” zones up to about 4000 u, where the “forbidden” zones vanish. To use fractional mass to differentiate peptides from other chemical components, fractional masses of all *in silico* peptides generated from the Swiss Prot protein database were calculated (Figure 4). The fractional mass of each experimental peak is evaluated with the predicted fractional mass in Figure 4. If the majority of the data match the expected fractional mass for peptides, experimental peaks with fractional mass in the “forbidden” zone were considered as nonpeptide peaks and were removed from the peak list. Otherwise, the results were flagged for further evaluation because the majority of the experimental peaks were not peptide-like peaks. This occurs when there is some experimental issue, such as a shift in mass calibration.

Peak Quantification

Stable isotope quantification is based on the fact that paired samples can be differentially coded with derivatizing agents that are isotopically distinct. Doublet ion clusters of peptide isoforms resulting from derivatization with these isotopically unique coding agents characterize spectra derived from differentially coded mixtures of these peptides. Based on the difference in mass of the labeling agent isoforms, a doublet cluster can be recognized by its m/z spacing, retention time, and charge state. The light and heavy isoforms of a peptide can also differ slightly in retention time, especially with deuterium labeling [26, 27]. The shift of retention time between hydrogen and deuterium labeled peptides depends on the number of deuterium atoms present and also the size of labeled peptides. The greater the number of deuterium atoms in the peptide, the greater the retention time shifts. In contrast, the

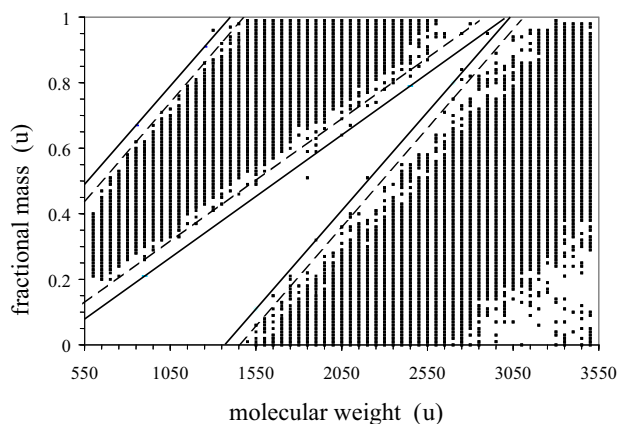


Figure 4. Fractional mass filter. The dotted line is the boundary of the peptide fractional mass that covers 99.9% tryptic peptides generated from the Swiss-Prot protein database. The solid line is the boundary with mass variants. The “forbidden” zone is among solid lines and axes.

larger the peptide, the smaller the chromatographic isotope effect will be.

Two methods were developed to quantify doublet ion clusters. The first calculates peak ratios in each scan. The second calculates the peak ratio across scans. The first method works well for labeling agents that don’t cause chromatographic isotope effects. The peak ratio calculated in each scan should reflect the actual peptide ratio and this ratio should be preserved across multiple scans corresponding to the chromatographic width of the labeled peptide peak. After calculating ratios of doublets in each scan, peaks in adjacent scans having matching m/z , charge, mass difference, and similar ratios are selected. Peak ratios that vary significantly from the norm are rejected as outliers. Outliers were detected by using the equation [28].

$$|X_i - M|/M_{ad} > M_{ax} \quad (5)$$

where, X_i is any of the ratios being evaluated with respect to whether it is an outlier and M is the median of the ratios in the sample. M_{ad} is the median absolute deviation, and M_{ax} is the critical value the inequality must exceed to conclude the value X_i is an outlier. The value M_{ax} is set at 5.0, which identifies ratios deviating from the mean by more than three standard deviations. After removal of outlier ratios, the peptide ratio is assigned as the median value of the ratios calculated from the remaining scans.

The second quantification method calculates a single doublet ratio from the heights of separately processed peaks. Thus, peaks with the same m/z and charge in continuous scans are selected and considered to be a single chromatographic peak. Savitzky-Golay smoothing filters [16] are used to smooth the selected spectral peaks. After smoothing, the center of a chromatographic peak is determined by the first derivatives of peak intensities, where the sign of the first derivative changes from positive to negative and further tested for symmetry about this

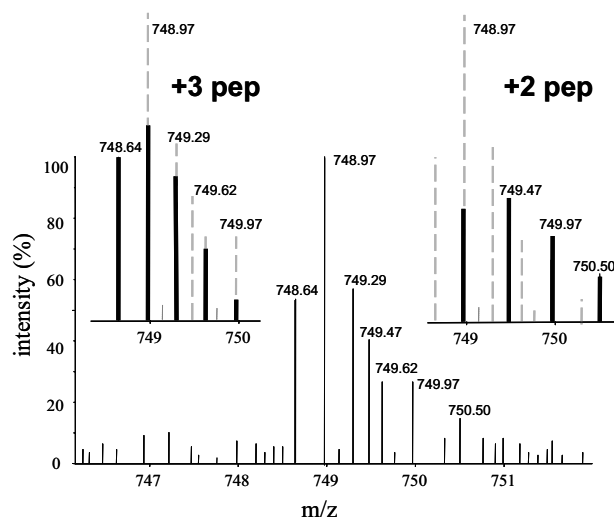


Figure 5. Charge fitting and deisotoping. The illustration shows isotopic peaks of two peptides that overlap along with background noise. The peak at m/z 748.97 is larger than can be matched to the triply charged peptide with M_0 748.67. The left inset marks the isotope distribution fitted portions for this peptide. The right inset shows the peak portions that were then fit to the doubly charged peptide of M_0 748.97

point. Peaks that don't match acceptable width and shape tests are removed. Doublet ion clusters are recognized by comparing the mass difference and retention time shift between two peptide peaks.

Changes in peptide concentration are calculated using eq 6, where κ is change, I_L is peak intensity of the light isoform of the peptide, and I_H is peak intensity of the heavy isoform of the peptide. Eq 6 is designed to provide a continuous variable where positive values report up-regulation of the light form and negative values the down-regulation.

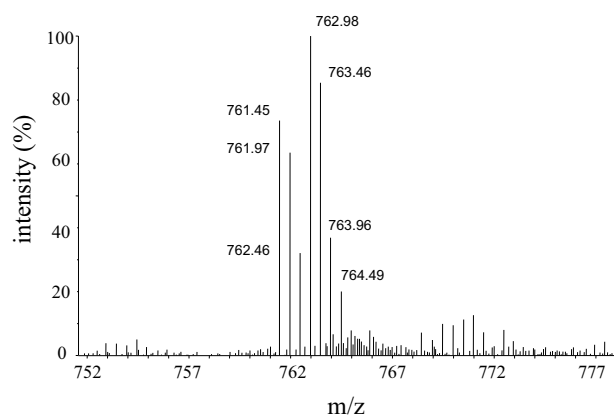


Figure 6. Deisotoping and peak ratio calculation with a GIST labeled peptide doublet. The monoisotopic peak M_0 of ^{12}C labeled peptide is $m/z = 761.45$ while the monoisotopic peak M_0 of ^{13}C labeled peptide is $m/z = 762.98$. The ^{12}C and ^{13}C labeled peptides were mixed in 1:1 ratio. The peak ratio calculated after deisotoping was 0.91, as opposed to a value of 0.74 by direct calculation from peak height.

$$\kappa = \begin{cases} I_L/I_H - 1.0 & (I_L \geq I_H) \\ -I_L/I_H + 1.0 & (I_L < I_H) \end{cases} \quad (6)$$

Ranking Doublets

All doublets are ranked according to peak quality and complexity of the doublets using a 1 through 5 scale. Doublets that do not overlap with other doublets are ranked as 1 through 3 according to signal-to-noise ratio with best S/N pairs ranking 1. A doublet is ranked as 4 for the ambiguous case where more than one doublet pairing is found. Peptides that don't have a pairing cluster, i.e., a singlet peak cluster, are ranked as 5.

Two methods have been developed to rescue some of the rank 4 doublets. Peptides often occur in multiple charge states in electrospray ionization (ESI) mass spectrometry. If a labeled peptide is present as multiple charge forms, all forms should yield the same doublets ratio. A doublet with a ranking of 4 at one charge state could have

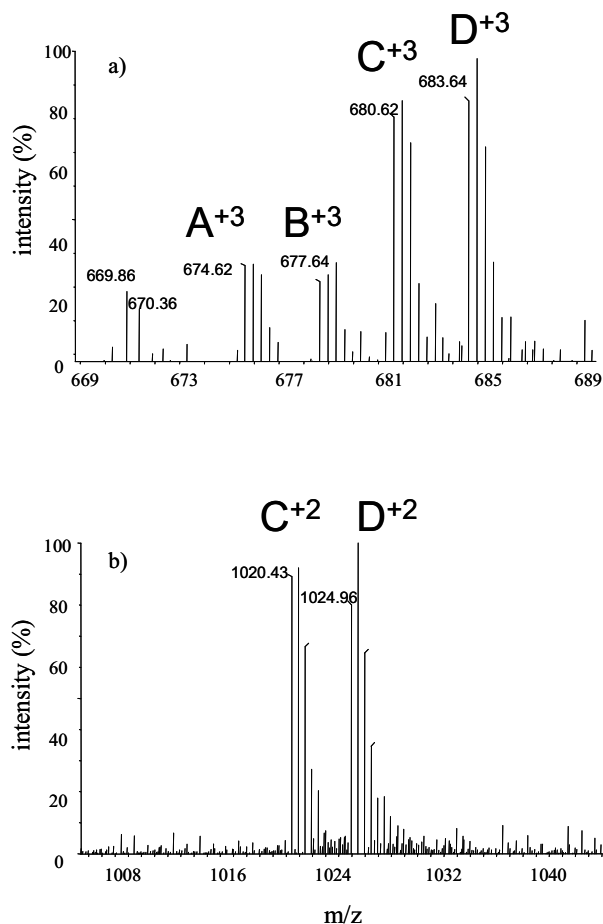


Figure 7. An example of rescuing doublet ions from a doublet complex by verifying molecular weight, ratio, and mass difference of another doublet in the same scan. (a) four GIST labeled peptides were detected by MS. Six possible pairings can be made out of these four peptides, according to the mass difference in the labeling agents. (b) alternate charge state forms of two of these four peptides are observed resolving the most plausible pairings as A with B and C with D.

Table 1. Partial results from an acetate labeled BSA digest. Peptide doublet ratios were calculated from a MS-only analysis. The peptides were identified by SEQUEST from data dependent MS/MS analysis

Peptide sequence	Results of SEQUEST		Results of GISTool		
	Xcorr	Charge	Charge	No. of lysine	Ratio
CCTESLVNR	2.75	2	2(0.94) ^a	0K	0.87
DAFLGSFLYEYSR	2.91	2	2(0.98)	0K	0.71
FKDLGEEHFK	3.32	2	2(0.92)	2K	0.96
HLVDEPQNLIK	3.48	2	2(1.12)	1K	0.96
KPDPNTLCDEFK	2.43	2	2(2.00)	2K	0.96
KQTALVELLK	3.22	2	2(1.07)	2K	0.98
KVPQVSTPTLVEVSR	2.84	2	2(1.87)	1K	1.00
LCVLHEK	2.77	2	2(1.03)	1K	0.95
LGEYGFQNALIVR	4.03	2	2(1.09)	0K	0.91
LVNELTEFAK	2.54	2	2(1.02)	1K	0.92
RPCFSALTPDETYVPK	2.93	2	2(1.78)	1K	0.91
SHCIAEVEK	2.54	2	2(1.38)	1K	0.95
SLHTLFGDELCK	3.54	2	2(1.02)	1K	0.90
TCVADESHAGCEK	4.17	3	3(1.74)	1K	0.81

^aValues in parentheses are scores of charge status assignment.

a ranking of 1–3 at another. When this occurs, the 4 ranked doublet is upgraded to have the same rank as its sister doublet at another charge state. Any other doublets composed of one of the peak cluster of the upgraded doublet were deleted, and the rest of the doublets were reranked accordingly.

Further simplification of a 4 ranked doublet family is achieved by calculating the product of peak intensities of the light and heavy components for each doublet and sorting them in decreasing order ($a_0, a_1, \dots, a_i, \dots, a_n$), where a_i is the product of the peak intensities of doublet i . The standard deviation δ_i and mean μ_i of (a_{i1}, \dots, a_{in}) are also calculated, where $i = (1, \dots, n)$. The doublets are then separated into two groups by comparing a_i with $\mu_{i+1} + 2*\delta_{i+1}$. All doublets wherein $a_i > \mu_{i+1} + 2*\delta_{i+1}$ are ranked as 4 while the rest of doublets are ranked as 4.5. This algorithm biases the preferred pairing toward the most intense doublets in keeping with the principle of mediocrity.

Testing the Algorithm

Charge fitting and deisotoping. Figure 5 is an example of charge state fitting, deisotoping and deconvolution of overlapping peptide peaks. The peptide peaks were split into two groups by initial charge assignment. The first group ($m/z = 748.64, 748.97, 749.29, 749.62, 749.97$) had an initial charge assignment of +3, while the second group ($m/z = 749.47, 750.50$), which remained in the unprocessed peak set, had not received a charge assignment. Peaks in the first group were fitted with a theoretical isotopic peak profile by assuming that height of the peak with $m/z = 748.64$ was solely from peptide one and that this peak was a monoisotopic peak in the peak cluster. The isotopic peaks with $m/z 748.97$ and 749.97 were flagged because their peak heights were significantly higher than expected, indicating the contribution of another peptide to these two peaks. The unmatched residual intensity from the isotopic

peaks 748.97 and 749.97 was passed to the second group of peaks for further analysis. The first group of peaks was recorded as processed peaks with a charge assignment of +3, of which only the monoisotopic peak $m/z = 748.64$ was used to represent the peptide and the other isotopic peaks were deleted. The second group of peaks and the residual of the flagged peaks of the first group were then processed together. The charge state of these peaks is assigned as +2 and the peak profile is then fitted with the theoretical isotopic profile.

Figure 6 is an example from the spectrum of a GIST labeled BSA peptide, where the light labeled peptide and the heavy labeled peptide were experimentally mixed in 1:1 ratio. A peptide peak ratio of 0.74 (light:heavy) was calculated directly from peak heights. This value is lower than the actual ratio because the isotopic peak M_3 ($m/z = 762.98$) of the light isoform contributes to the monoisotopic peak M_0 of the heavy isoform of the peptide. Using the deisotoping algorithm described above, the contribution of the light component was estimated and deducted from peak heights of the heavy component. The peak ratio using this approach was calculated to be 0.91 in comparison to the mixing ratio of 1. The deisotoping algorithm did not detect any other peaks overlapping with these peptide isotope peaks. The doublet was therefore ranked as 1. Charge state assignment, deisotoping, doublet recognition, and peak quantitation worked as designed in this case. Moreover, the observed error in estimating the actual isotope ratio is no greater than has been observed with other methods of isotope ratio quantification in proteomics [7–14].

Rescuing Doublets from Doublet Complex

Even though protein digests can be extensively fractionated during MDLC-MS analysis, often multiple peptides coelute from reversed-phase columns and enter the mass spectrometer together. Stable isotope coding adds to this

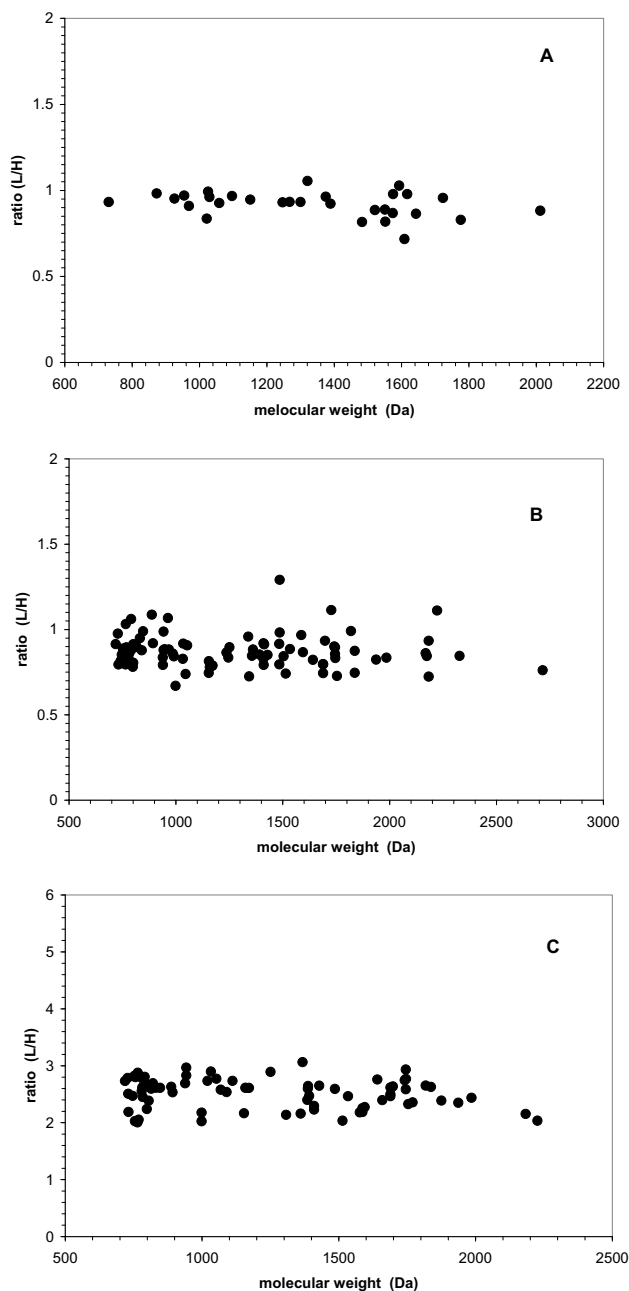


Figure 8. Confirmation of quantification using a single protein digestion. (a) deuterium (^2H)-based coding with 1:1 (light:heavy) mixture of BSA peptides. (b) ^{13}C -based coding in 1:1 (light:heavy) mixture of human transferrin. (c) ^{13}C -based coding with a 3:1 (light:heavy) mixture of human transferrin.

problem by doubling the number of peptides being analyzed. This complicates doublet recognition by increasing the potential for coincidental doublet overlaps with other peptides. A method for rescuing some doublets from a doublet complex by using other charge states of the peptide has been described above.

Figure 7 shows the effectiveness of rescuing doublets from a GIST propionate coded doublet complex. There is a doublet complex in this spectrum from which 6 possible doublet combinations are theoretically feasible

based on the mass difference between light and heavy labeled peptides ($\Delta m = 3.0, 6.0, 9.0$ u). Each of these doublets was initially ranked 4. By checking through the entire spectrum, a doublet at m/z 1020.43 with charge +2 was ranked as 2. This doublet and another doublet ($m/z = 680.62$ and $m/z = 683.64$) in the doublet complex have the same molecular weight ($MW = 203.84$ u), the same isotope ratio ($R = 1.03$), and the same mass difference ($\Delta m = 6.0$ u) between the light and heavy components. Based on these similarities it was assumed that these two doublets came from the same peptide pair. The doublet ($m/z = 680.62$ and $m/z = 683.64$) in the doublet complex was upgraded to a ranking of 2. Fixing the association between 680.62 and 683.64 eliminates other four false pairs and leaves a single pairing option for 674.62 and 677.64. This doublet was then reranked accordingly.

Analysis of Isotope Labeled Protein Digest

An *in silico* analysis indicates that a tryptic digestion of BSA will contain about 140 peptides including those with up to one missed cleavage. However, only $\sim 70\%$ of these peptides were seen in LC/MS because of a lack of retention of some peptides on the reversed-phase chromatography column, suppression of ionization, and difficulties with sequencing very small or large peptides by MS/MS. Table 1 lists some of the peptides that were identified by SEQUEST [29] in a data dependent LC-MS/MS analysis. Database matches that passed initial cross-correlation (Xcorr) score cut-offs of 2.00, 2.70, and 3.30 for 2+, 3+, respectively, and a ΔC_n value of 0.08 were then manually inspected to assure that all of the major b- and y-fragment ions correlate with the sequence from the database entry. The sample was also analyzed in MS-only mode and the data were processed through GISTool. Information corresponding to these sequenced peptides from the GISTool analysis is also listed in the table. The number of lysine residues in a peptide was calculated according to the labeling agent and mass difference between light and heavy labeled peptides. It was observed that peptide charge determined by the instrument software is identical to the charge calculated by the processing algorithm developed here. Isotope ratios were consistent between peptides derived from the same protein and close to the expected value.

Figure 8 shows ratios of all BSA and human transferrin tryptic peptides detected in other three different experiments. A 1:1 mixture (light:heavy) of peptides from a BSA digest differentially coded with ^1H and ^2H labeled acetate was examined by LC/MS and found to have an isotope ratio of 0.94 with a standard deviation 0.07. The isotope ratio of a 1:1 ^{12}C and ^{13}C propionate coded mixture of peptides derived from a trypsin digest of human transferrin was found experimentally to be 0.91 with a standard deviation 0.09. Using the same coding strategy with a 3:1 (light:heavy) mixture of tryptic peptides from human transferrin the experimen-

Table 2. Propionate labeled tryptic digest of a seven-protein mixture. Peptides were identified from a data dependent MS/MS analyses using SEQUEST. Measured protein ratios were the average ratio of peptides identified

Protein	Protein concentration (μM)		Protein ratio ($^{12}\text{C}/^{13}\text{C}$)		Number of identified peptides (2+, 3+)
	^{12}C labeled	^{13}C labeled	Expected ratio	Measured ratio	
Bovine serum albumin	2.0	0.4	5.0	4.5 (1.7) ^a	17
Horse heart myoglobin	1.0	1.0	1.0	1.3 n/a	1
Chicken lysozyme	1.0	1.0	1.0	1.1 (0.14)	4
Bovine β -lactoglobulin	2.3	0.23	10	singlets	7
Bovine carbonic anhydrase II	1.5	0.22	6.8	5.0 (1.2)	4
Glucose oxidase	0.7	2.1	0.33	0.48 (0.17)	13
Human transferrin	1.0	0.38	2.6	3.3 (0.57)	10

^aValues in parentheses are standard deviation.

tally determined isotope ratio was 2.70 with a standard deviation 0.26. It should be noted that mean ratio values are close to the expected values and peak ratio distribution based on standard deviation is relatively tight in both cases. This indicates the effectiveness of the data analysis algorithm.

In yet another experiment, seven proteins were individually digested with trypsin and two distinct samples were created. These two samples were then differentially labeled with nonenriched and ^{13}C enriched propionate. Following mixing of these coded samples, the labeled digests were then analyzed by LC/MS in the MS-only mode. The concentration of proteins in the two samples and the ratios of detected peptides are presented in Table 2. A total of 252 doublets were detected in the MS-only spectra. The percentage of propionate labeled peptides that were observed with single charge was higher than nonlabeled peptides because acylation of all primary amines in peptides reduces their overall charge [30]. If needed, this problem can be eliminated by labeling with trimethylammonium butyrate that carries a quaternary amine [31].

The correlation of LC/MS and LC/MS/MS results were performed based on comparing the similarity of peptide charge, m/z , sequence tag, and retention time measured in LC/MS experiment and the same information of parent ion measured in LC/MS/MS experiment. Sequence tag is the number of stable isotope labeled functional sites in a peptide. For example, sequence tag in a peptide labeled by propionate will be $N + 1$ where N is number of lysine in that peptide and 1 refers to the N-terminal site if free amine group is available. This mechanism works very well for simple protein mixture. It may cause false-positive results in case of complex mixture such as human serum, where a further analysis of peptide ratios of all peptides identified from the same protein is necessary. The false-positive correlation can also be leveraged by statistical outlier analysis on the peptide ratio [28].

It can be seen from Table 2 that the measured protein ratios were close to the expected values in each case. However, the variation between measured protein ratio and calculated protein ratio is higher than the variation obtained for pure protein (see Figure 8). In general, the

variation is about 30% of calculated protein ratio. Seven peptides were identified for bovine β -lactoglobulin from the MS/MS data. Expected ratio for peptides from bovine β -lactoglobulin is 10:1. However, all of these seven peptides were reported as singlets in the MS-only data analysis, which means that the ^{13}C labeled peptides were not detected by the algorithm.

To evaluate the effectiveness of the algorithm, all spectra of the seven-protein mixture were manually inspected as well. The manually interpreted results confirmed the results automatically analyzed by the program. The singlets of bovine β -lactoglobulin are because the peak height of the ^{13}C labeled peptides was not higher than noise level. It should be noted that the maximum ratio of doublet peptides that the program can measure depends on multiple variables, including peptide concentration and noise. As the less intense peak clusters in a doublet approaches the noise level, isotope ratio measurements will decrease in accuracy. At the point where only the intense peak can be seen, the problem becomes how to determine both the isotope ratio and sample origin of the peptide.

It should be noted that the sequence coverage of the seven proteins is significantly low even though 252 doublets were detected in LC/MS experiment. The reasons are (1) each peptide pair is usually carrying two different charges, (2) only the top three most abundant peaks were sequenced during data dependent analysis, (3) no doublet peaks were put into the inclusion list for sequence analysis, (4) data dependent LC/MS/MS analyses were executed by selecting only 2+, 3+, and 4+ ions for MS/MS, and (5) the percentage of propionate labeled peptides that were observed with single charge was higher than nonlabeled peptides because acylation of all primary amines in peptides reduces their overall charge.

Conclusions

An algorithm for analyzing stable isotope coded peptide mixtures has been proposed. Based on data examined in these studies it is concluded that in spectra with ion clusters arising from multiple substances this algorithm is capable of (1) recognizing peptide peak clusters based on isotope profile, (2) recognizing and deleting nonpeptide

peaks, (3) determining the relative peak height of isotopic ions in a cluster, (4) deconvoluting overlapping peptide signals, (5) combining the detected ions in a cluster, i.e., “deisotope” the cluster, (6) recognizing and grouping isotopically coded isoforms of a peptide, and (7) quantifying the relative abundance of peptide isoforms. It is further concluded that in the measurement of isotope ratio values, the deviation from known values of the isotope ratio of peptide isoforms is in the range of $\pm 15\%$ for single protein and about $\pm 30\%$ for protein mixture. Finally, it is concluded that the flexibility and versatility of this model will allow it to be used with a wide variety of noncoded and isotopically coded peptides, including those arising from the ICAT and ^{18}O labeling methods.

Acknowledgments

The authors thank Dr. Paul Stroobant, Dr. Robert McBurney, and other R and D committee members of Beyond Genomics Inc., for reviewing the manuscript. The authors also highly appreciate the discussion with Dr. Ted Marple during the preparation of this manuscript.

References

- Julka, S.; Regnier, F. E. Quantification in Proteomics Through Stable Isotope Coding: A Review. *J. Proteom. Res.* **2004**, *3*(3), 350–363.
- Sechi, S.; Oda, Y. Quantitative Proteomics Using Mass Spectrometry. *Curr. Opin. Chem. Biol.* **2003**, *7*(1), 70–77.
- Geng, M.; Ji, J.; Regnier, F. E. Signature-Peptide Approach to Detecting Proteins in Complex Mixtures. *J. Chromatogr. A* **2000**, *870*, 295–313.
- Wang, S.; Regnier, F. E. Proteomics Based on Selecting and Quantifying Cysteine Containing Peptides by Covalent Chromatography. *J. Chromatogr. A* **2001**, *924*, 345–357.
- Wang, S.; Zhang, X.; Regnier, F. E. Quantitative Proteomics Strategy Involving the Selection of Peptides Containing Both Cysteine and Histidine from Tryptic Digests of Cell Lysates. *J. Chromatogr. A* **2002**, *949*, 153–162.
- Charkraborty, A.; Regnier, F. E. Global Internal Standard Technology for Comparative Proteomics. *J. Chromatogr. A* **2002**, *949*, 173–184.
- Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative Analysis of Complex Mixtures using Isotope-Coded Affinity Tags. *Nat. Biotechnol.* **1999**, *17*(10), 994–999.
- 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.
- Brown, K. J.; Fenselau, C. Investigation of Doxorubicin Resistance in MCF-7 Breast Cancer Cells Using Shot-Gun Comparative Proteomics with Proteolytic ^{18}O Labeling. *J. Proteom. Res.* **2004**, *3*(3), 455–462.
- Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic ^{18}O Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus. *Anal. Chem.* **2001**, *73*(13), 2836–2842.
- Andersen, J. S.; Mann, M. Functional Genomics by Mass Spectrometry. *FEBS Lett.* **2000**, *480*(1), 25–31.
- Jensen, P. K.; Pasa-Tolic, L.; Anderson, G. A.; Horner, J. A.; Lipton, M. S.; Bruce, J. E.; Smith, R. D. Probing Proteomes Using Capillary Isoelectric Focusing-Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **1999**, *71*(11), 2076–2084.
- Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. Accurate Quantitation of Protein Expression and Site-Specific Phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6591–6596.
- Conrads, T. P.; Alving, K.; Veenstra, T. D.; Belov, M. E.; Anderson, G. A.; Anderson, D. J.; Lipton, M. S.; Pasa-Tolic, L.; Udseth, H. R.; Chrisler, W. B.; Thrall, B. D.; Smith, R. D. Quantitative Analysis of Bacterial and Mammalian Proteomes Using a Combination of Cysteine Affinity Tags and ^{15}N -Metabolic Labeling. *Anal. Chem.* **2001**, *73*(9), 2132–2139.
- <http://msquant.sourceforge.net/>
- MacCoss, M. J.; Wu, C. C.; Liu, H.; Sadygov, R.; Yates, J. R. III. A Correlation Algorithm for the Automated Quantitative Analysis of Shotgun Proteomics Data. *Anal. Chem.* **2003**, *75*, 6912–6921.
- Li, X.; Zhang, H.; Ranish, J. A.; Aebersold, R. Automated Statistical Analysis of Protein Abundance Ratios from Data Generated by Stable-Isotope Dilution and Tandem Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 6648–6657.
- Press, H. W.; Teukolsky, A. S.; Vetterling, T. W.; Flannery, P. B. *Numerical Recipes in C++*; Cambridge University Press: Cambridge, United Kingdom 2002, Chap. XIV.
- Zhang, Z.; Marshall, A. G. A Universal Algorithm for Fast and Automated Charge State Deconvolution of Electrospray Mass-to-Charge Ratio Spectra. *J. Am. Soc. Mass Spectrom.* **1998**, *9*(3), 225–233.
- Yergey, J. A. A General Approach to Calculating Isotopic Distributions for Mass Spectrometry. *Int. J. Mass Spectrom. Ion Phys.* **1983**, *52*, 337–349.
- Hsu, C. S. Diophantine Approach to Isotopic Abundance Calculations. *Anal. Chem.* **1984**, *56*(8), 1356–1361.
- Mirgorodskaya, O. A.; Kozmin, Y. P.; Titov, M. I.; Korner, R.; Sonksen, C. P.; Roepstorff, P. Quantitation of Peptides and Proteins by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Using ^{18}O -Labeled Internal Standards. *Rapid. Commun. Mass Spectrom.* **2000**, *14*, 1226–1232.
- <http://www.expasy.org>
- Zubarev, R. A.; Hakansson, P.; Sundqvist, B. Accuracy Requirements for Peptide Characterization by Monoisotopic Molecular Mass Measurements. *Anal. Chem.* **1996**, *68*, 4060–4063.
- Juhasz, P.; Martin, S. A. The Utility of Nonspecific Proteases in the Characterization of Glycoproteins by High-Resolution Time-of-Flight Mass Spectrometry. *Int. J. Mass Spectrom. Ion Processes* **1997**, *169/170*, 217–230.
- Zhang, R.; Regnier, F. E. Minimizing Resolution of Isotopically Coded Peptides in Comparative Proteomics. *J. Proteom. Res.* **2002**, *1*(2), 139–147.
- Zhang, R.; Sioma, C. S.; Wang, S.; Regnier, F. E. Fractionation of Isotopically Labeled Peptides in Quantitative Proteomics. *Anal. Chem.* **2001**, *73*(21), 5142–5149.
- Sheskin, D. J. *Handbook of Parametric and Nonparametric Statistical Procedures, 2nd ed.*; Chapman and Hall/CRC: Boca Raton, FL 2000; Test 11.
- Eng, J. K.; McCormack, A. L.; Yates, J. R., III. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
- Kindy, J. M.; Taraszka, J. A.; Regnier, F. E.; Clemmer, D. E. Quantifying Peptides in Isotopically Labeled Protease Digests by Ion Mobility/Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2002**, *74*(5), 950–958.
- Zhang, R.; Sioma, C. S.; Thompson, R. A.; Xiong, L.; Regnier, F. E. Controlling Deuterium Isotope Effects in Comparative Proteomics. *Anal. Chem.* **2002**, *74*(15), 3662–3669.