

## APPLICATION NOTE

# ZoomQuant: An Application for the Quantitation of Stable Isotope Labeled Peptides

Brian D. Halligan, Ronit Y. Slyper, and Simon N. Twigger

Bioinformatics, Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, Wisconsin, USA

Wayne Hicks, Michael Olivier, and Andrew S. Greene

Biotechnology and Bioengineering Center, Medical College of Wisconsin, Milwaukee, Wisconsin, USA

The main goal of comparative proteomics is the quantitation of the differences in abundance of many proteins between two different biological samples in a single experiment. By differentially labeling the peptides from the two samples and combining them in a single analysis, relative ratios of protein abundance can be accurately determined. Protease catalyzed  $^{18}\text{O}$  exchange is a simple method to differentially label peptides, but the lack of robust software tools to analyze the data from mass spectra of  $^{18}\text{O}$  labeled peptides generated by common ion trap mass spectrometers has been a limitation. ZoomQuant is a stand-alone computational tool that analyzes the mass spectra of  $^{18}\text{O}$  labeled peptides from ion trap instruments and determines relative abundance ratios between two samples. Starting with a filtered list of candidate peptides that have been successfully identified by Sequest, ZoomQuant analyzes the isotopic forms of the peptides using high-resolution zoom scan spectrum data. The theoretical isotope distribution is determined from the peptide sequence and is used to deconvolute the peak areas associated with the unlabeled, partially labeled, and fully labeled species. The ratio between the labeled and unlabeled peptides is then calculated using several different methods. ZoomQuant's graphical user interface allows the user to view and adjust the parameters for peak calling and quantitation and select which peptides should contribute to the overall abundance ratio calculation. Finally, ZoomQuant generates a summary report of the relative abundance of the peptides identified in the two samples. (J Am Soc Mass Spectrom 2005, 16, 302–306) © 2004 American Society for Mass Spectrometry

One of the main goals of many proteomics experiments is to determine the difference in abundance of proteins between samples representing different biological conditions. Analyses of the peptides from the two samples as independent LC/MS runs can have difficulties related to issues of non-linearity of chromatography, reproducibility of peptide coverage due to timing of data dependent sampling by the MS, and ion suppression in complex samples. To overcome these limitations, a number differential stable isotope labeling methods have been developed so that labeled peptides from one sample can be combined with unlabeled or differently labeled peptides from a second sample and both samples analyzed in a single analytical run [1–9]. This approach overcomes the problems of timing and reproducibility between the analyt-

ical runs and provides an accurate determination of the ratio of abundance of proteins between the two samples.

Differential peptide labeling for analysis by mass spectroscopy relies on introducing a mass difference using stable isotopes and it is assumed that this isotopic difference does not alter the peptides' chromatographic or fragmentation properties. Labeling can take place biosynthetically, using media containing  $^{15}\text{N}$  amino acids for example [1], after sample isolation by introducing a chemical modification using the ICAT reagent [2–4], or by exchanging atoms for a heavier isotopic form, as in  $^{18}\text{O}$  exchange [5–9]. Since the biosynthetic labeling method is restricted to systems that can be grown with labeled media, it is not practical for many biological or clinical samples.

The two methods that have been most widely used for post-translational labeling of peptides have been differential labeling of cysteine residues with a heavy and light form of the ICAT reagent [2, 4] or labeling the C-terminus of the peptide by endoprotease catalyzed

Published online January 13, 2005

Address reprint requests to Dr. B. D. Halligan, Bioinformatics Research Center-HMGC, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53213, USA. E-mail: halligan@mcw.edu

$^{18}\text{O}$  exchange [5–9]. The ICAT method has been well developed and widely used, but has some limitations. It can only be used with peptides that contain a cysteine residue and therefore, the majority of peptides are not targeted by this method. Many of the low abundance proteins, which only have one or several peptides identified in each run, may therefore be very difficult to quantitate using the ICAT method. Differential  $^{18}\text{O}$  labeling of samples for relative peptide quantitation overcomes many of these problems. The chemistry of the labeling reaction is simple and can be done as part of the initial digestion with trypsin or other serine proteases [10]. Alternatively, incubation in the presence of isotopically labeled water and protease can be used to exchange the carboxyl oxygens post-proteolytically. Since labeling takes place in a sequence independent manner, this allows for all peptides to be potentially labeled with  $^{18}\text{O}$  except for the C-terminal peptide of the protein.

The analysis of  $^{18}\text{O}$  labeled comparative proteomics experiments has mostly been done by hand calculations, spreadsheets [11], or with programs designed for ICAT [3, 12], or metabolic labeling experiments [13]. Unlike these other methods that rely on obtaining data on each of the isotopes from extracted ion chromatograms of the different ion species, ZoomQuant uses the high resolution zoom scan spectra to quantitate the ionic species. This is an advantage in two respects. First, extracted ion chromatograms are constructed by summing the intensity for a fixed region over the duration of a chromatographic peak. If the centers of the isotope peaks vary during the time window of the analysis, then intensity can spill from one isotope to the adjacent one which can lead to inaccuracies if the peaks are not baseline separated. Second, inaccuracies can occur if there are multiple species eluting in the same time window with masses within the isotope distribution. By using the zoom scan spectra, we have a high resolution snapshot of the isotope distribution at a single time point that we can use to accurately quantitate the isotope species present.

The ZoomQuant program has been written in Perl, using the Tk module to provide a graphical user interface and has been tested on the Windows, Unix, and Macintosh platforms. Information about obtaining the ZoomQuant, Epitomize, and rawbitz programs can be found on the MCW NHLBI Proteomics Center web site (<http://proteomics.mcw.edu>).

## Overview and Examples

The ZoomQuant program takes two data files as its input. The first input is a filtered list of Sequest identified peptides produced by the Epitomize program (manuscript in preparation). Epitomize takes the .out files generated by Sequest [14] and filters the files based on user defined levels of Xcorr, intensity, number of matched ions, and other values. The results are summarized by protein identification and can be output in

a variety of formats. For use with ZoomQuant, a colon-delimited text file format has been developed. The file contains the list of proteins that have been identified, the sequence and atomic composition of peptides that have been identified, the scan number, charge state, Xcorr, and mass data for each of the scans involved in an identification. This data file is combined with a second file containing the information from the zoom scans extracted from the .raw binary raw data file. The zoom scans are extracted using a Visual Basic script, rawbitz, which utilizes the ThermoFinnigan (San Jose, CA) Xcalibur Developer's Kit (XDK). The rawbitz program selects all of the zoom scans in the .raw binary file and exports them to a single ASCII format text file.

ZoomQuant reads in both files and matches the zoom scan information to the Sequest identified peptides. In the main window's first column, the user is presented a menu of the proteins that have been identified by Sequest (Figure 1). The user selects a protein to examine and the peptides from that protein are displayed in the second column. In a third column, the corresponding peptide and the scan information for each of these peptides is displayed. When the user selects one of the peptides and clicks on the "Peaks" button in the main window, the spectra window opens and displays the zoom scan spectra (Figure 2). The ZoomQuant program uses a simple threshold based peak finding algorithm to identify and integrate the peaks. If the program finds a set of seven or more nearly equally spaced peaks that are expected to result from the unlabeled, partially labeled, and fully  $^{18}\text{O}$  labeled peptide species, it marks each of the peaks with an oval to indicate that they are to be used in the ratio calculations. If the user is satisfied with the peak definitions and assignments, clicking the "Get Ratios" button in the main window initiates the calculation of the ratios of labeled versus unlabeled peptides and the efficiency of labeling. If the user desires, the peak definitions and assignments can be adjusted in the spectra window. Two tools (split and combine) allow the ends of the peaks to be changed so that large peaks can be manually split or smaller peaks combined and an additional tool (delete) can remove peaks or area from further consideration.

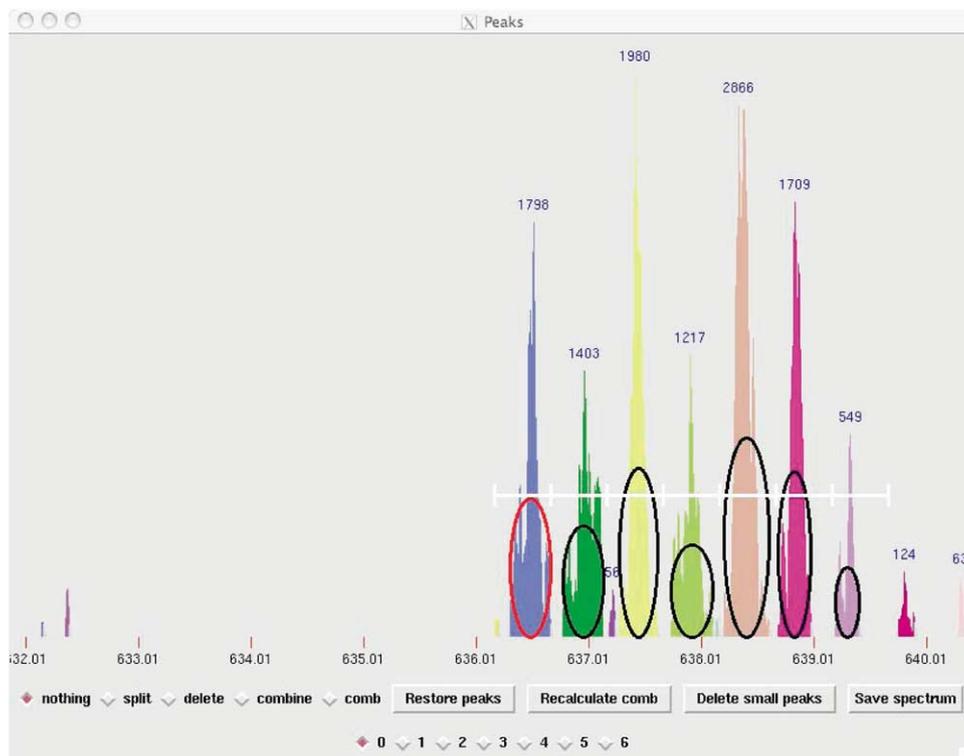
The ratios of  $^{18}\text{O}$  labeled to  $^{16}\text{O}$  unlabeled peptides are calculated using three different methods. The first ratio method uses the peptide's mass, charge, and atomic composition determined from the Sequest identification to calculate the theoretical isotope distribution using the binomial expansion method of Yergey et al. [15]. The first of the series of peaks is assumed to be the unlabeled monoisotopic peak,  $M$ . The second peak,  $M+1$ , represents the naturally occurring +1 isotope variant of the unlabeled peak. The distance between the  $M$  and  $M+1$  peaks and ratio of the areas of the  $M$  and  $M+1$  peaks are checked for consistency with the charge and calculated isotope distribution. The contribution of the unlabeled species to the area of peaks  $M+2$ ,  $M+3$ , and  $M+4$  is calculated using the area of peak  $M$  and the



**Figure 1.** ZoomQuant main window displays the proteins and peptides identified by the peptide identification program and the data for individual scans. The user can choose individual proteins and scans to analyze. The user also has the option to perform ratio calculations and generate a report for the experiment.

theoretical isotope distribution. To determine the area of the unlabeled species,  $A_0$ , the areas of peak  $M$  and  $M+1$  are added to the calculated area contributed by the unlabeled species to peaks  $M+2$ ,  $M+3$ , and  $M+4$ .

The contribution of the unlabeled species to peaks  $M+2$ ,  $M+3$ , and  $M+4$  are subtracted to give new area values for the peaks  $M+2'$ ,  $M+3'$ , and  $M+4'$ . To determine the area of the partially labeled species in



**Figure 2.** The peaks window displays the spectra and the peaks identifications and quantitation. The user can use the controls on the bottom of the window to split, combine, or delete peaks. Additionally, all small peaks can be removed or the group of peaks (comb) to be used in the ratio calculations can be changed or refreshed. Ovals indicate the peaks selected for ratio calculations. The optional ruler indicates the expected peak widths based on the ion charge.

which only one of the carboxyl oxygens have been exchanged,  $A_1$ , the process used for the unlabeled species is repeated using peak  $M+2$  as the starting point. This is followed by the subtraction of the contribution of the partially labeled species to peak  $M+4$ ,  $M+5$ , and peak  $M+6$  to give peaks  $M+4'$ ,  $M+5$ , and  $M+6'$ . The area of the fully labeled species in which both of the carboxyl oxygens have been exchanged,  $A_2$ , is then calculated by summing the areas of peaks  $M+4''$ ,  $M+5'$ , and  $M+6'$  plus the values for peaks  $M+7$  and  $M+8$  calculated using the theoretical isotope distribution and peak  $M+4''$ . The ratio of labeled to unlabeled peptide is calculated as:

$$R = \frac{A_1 + A_2}{A_0} \quad (1)$$

and is expressed as a percentage. The efficiency of labeling is calculated as:

$$E = \frac{A_2}{A_1 + A_2} \quad (2)$$

and is also expressed as a percentage. In all of these calculations adjustments are made for abundance of  $^{18}\text{O}$  (95 APE) in the labeled water used in the oxygen exchange labeling reaction using the method described by Johnson and Muddiman [8].

The second method used to calculate the ratio of labeled to unlabeled species was described by Johnson and Muddiman [8]. In their approach, instead of using the exact atomic formula to calculate the isotope distribution,

an approximation is made by using the mass and determining to how many average amino acid or “averagine” units the peptide corresponds. It allows an approximation of the average atomic formula for a given peptide mass. This has the advantage of allowing quantitation of a peptide in the absence of an accurate identification and it simplifies calculations. Since the atomic ratios of the peptide remain constant, the isotope distribution does not have to be individually calculated and can be incorporated into the ratio formula. This method can be inaccurate if the peptide actually has a significantly higher than average number of amino acids that contain sulfur [8].

The third ratio calculation method offered is a modification of the method of Yao et al. as corrected [5]. The ratio calculation they describe uses the intensity of the  $M$ ,  $M+2$ , and  $M+4$  peaks and the theoretical isotope distribution to calculate the  $^{16}\text{O}/^{18}\text{O}$  ratio. We have modified it to use the areas of the  $M$ ,  $M+2$ , and  $M+4$  peaks instead of intensities. The ZoomQuant program offers all three methods of ratio calculations and also has a checkbox that allows ratio calculations using the averagine approximation for the first ratio method instead of the exact atomic formula deduced from the peptide sequence identification.

An additional checkbox in the main window allows selection of which peptide scans are to be included or excluded in the calculation of the overall abundance ratio for the protein. This allows peptides that appear to be atypical or labeled poorly by the exchange reaction to be eliminated from the overall ratio calculation. When the inspection of the peptides is complete, the user can use the “Save Report” button in the main window to produce a tab delimited text file with the peak information and

**Table 1.** Report format

PROTEIN: sp P02188 MYG HORSE Myoglobin - Equus caballus (Horse), and Equus burchell							
PEPTIDE: LFTGHPETLEK[							
Scan	Mass	Charge	Xcorr	Ratio 1	Ratio 2	Ratio 3	Efficiency
785	1275.371	2	2.8768	0.6904	0.7258	0.7592	0.9944
PEPTIDE: LFTGHPETLEK							
Scan	Mass	Charge	Xcorr	Ratio 1	Ratio 2	Ratio 3	Efficiency
800	1271.378	2	3.2132	0.9686	0.9578	0.9843	0.6437
782	1271.403	2	3.0675	1.2818	1.1681	1.1953	0.7516
Mean				1.1252	1.0629	1.0898	0.6977
SD				0.2215	0.1487	0.1492	0.0763
PEPTIDE: VEADIAGHGQEVLR							
Scan	Mass	Charge	Xcorr	Ratio 1	Ratio 2	Ratio 3	Efficiency
1119	1606.427	2	4.3203	0.9887	0.9888	1.1786	0.6861
1110	1606.546	2	4.2017	1.1267	0.8020	0.9854	0.8907
Mean				1.0577	0.8954	1.0820	0.7884
SD				0.0975	0.1321	0.1366	0.1446
PEPTIDE: LFTGHPETLEKFDK							
Scan	Mass	Charge	Xcorr	Ratio 1	Ratio 2	Ratio 3	Efficiency
1131	1661.477	3	2.5041	0.8220	0.4379	0.6143	0.7635
PEPTIDE: HGTVVLTALGGILK							
Scan	Mass	Charge	Xcorr	Ratio 1	Ratio 2	Ratio 3	Efficiency
2061	1378.420	2	4.6019	0.8025	0.6054	0.6666	0.7811
PEPTIDE: HGTVVLTALGGILK[							
Scan	Mass	Charge	Xcorr	Ratio 1	Ratio 2	Ratio 3	Efficiency
2067	1382.507	2	5.3122	0.8255	0.6916	0.7570	0.8645
Protein Mean				0.9383	0.7972	0.8926	0.7969
Protein SD				0.1939	0.2336	0.2253	0.1143

Sample report of the analysis of horse myoglobin from a mixture of four proteins labeled with  $^{18}\text{O}$  using trypsin and mixed at a 1:1 ratio with unlabeled material.

results of the ratio calculations, suitable for import into Excel or other programs (Table 1).

## Discussion

ZoomQuant is a software tool that aids the process of comparative proteomics by enabling the analysis of differentially isotopically labeled peptide samples. It provides an integrated method for identifying the peaks from differentially labeled samples, integrating the peaks, and calculating the ratios between the labeled and unlabeled samples using three different methods, as well as calculating the efficiency of labeling. Although this version of the program is designed to use  $^{18}\text{O}$  labeled samples, it is possible to adapt it to other stable isotope labeling methods. Additionally, since we use a modular design that separates the extraction of the spectral data and processing of peptide identification data, we can easily adapt the system to different instruments and peptide identification programs. The rawbitz program is dependent on the ThermoFinnigan XDK and dll files and therefore can only be run on a computer on which BioWorks has been installed. By creating other small programs that can read files created by instruments from other manufacturers or other data types, such as extracted ion chromatographs for example, ZoomQuant could be used to analyze data from these other sources. Similarly, the peptide identification filter program, Epitomize, could also work with peptide identification results from Mascot, thus allowing alternative peptide identification methods as well.

The ZoomQuant program relies on the use of the higher resolution zoom scans that are a part of the standard "triple play" method of proteomic analysis (ThermoFinnigan Application Note 326). Using the zoom scan option is a means of collecting relatively higher resolution MS data with a common lower resolution instrument that is now being exploited by a commercial kit for quantitative proteomics using  $^{18}\text{O}$  labeling (Stratagene). One disadvantage of performing zoom scans on instruments with a slower data acquisition time is that this causes the loss of instrument duty cycles that could have been used for additional MS/MS scans, thereby leading to lower peptide coverage. For this reason, many researchers currently omit performing the zoom scans when analyzing complex peptide mixtures. A benefit of including the zoom scans is that they can reduce the computational time and increase the accuracy of peptide identification. With an accurate value for the peptide charge, only a single charge state possibility is used when searching the peptide database, thereby reducing the number of searches and decreasing false positive matches to the incorrect charge state of the peptide. Alternatively, ZoomQuant can be used with any instrument that can produce a high-resolution MS scan and peptide identification derived from a coupled MS/MS scan. It is expected that a linear ion trap FT-ICR instrument would produce much higher resolution data for the  $^{18}\text{O}/^{16}\text{O}$  ratio calculation and its higher scan rate would also allow for better peptide coverage.

## Acknowledgments

This work was supported in part by NHLBI Proteomics Center Grant (NIH-BAA-HL-02-04).

## References

1. Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R. III. Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal. Chem.* **2002**, *74*, 1650–1657.
2. Gygi, S. P.; Rist, B.; Gerber, 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. 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.
4. Smolka, M. B.; Zhou, H.; Purkayastha, S.; Aebersold, R. Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Anal. Biochem.* **2001**, *297*, 25–31.
5. Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic  $^{18}\text{O}$  labeling for comparative proteomics: Model studies with two serotypes of adenovirus. *Anal. Chem.* **2001**, *73*, 2836–2842.
6. Stewart, I. I.; Thomson, T.; Figeys, D.  $^{18}\text{O}$  labeling: A tool for proteomics. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2456–2465.
7. Heller, M.; Mattou, H.; Menzel, C.; Yao, X. Trypsin catalyzed  $^{16}\text{O}$ -to- $^{18}\text{O}$  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.
8. Johnson, K. L.; Muddiman, D. C. A method for calculating  $^{16}\text{O}/^{18}\text{O}$  peptide ion ratios for the relative quantification of proteomes. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 437–445.
9. Brown, K. J.; Fenselau, C. Investigation of doxorubicin resistance in MCF-7 breast cancer cells using shot-gun comparative proteomics with proteolytic  $^{18}\text{O}$  labeling. *J. Proteome Res.* **2004**, *3*, 455–462.
10. Reynolds, K. J.; Yao, X.; Fenselau, C. Proteolytic  $^{18}\text{O}$  labeling for comparative proteomics: Evaluation of endoprotease Glu-C as the catalytic agent. *J. Proteome Res.* **2002**, *1*, 27–33.
11. Bantscheff, M.; Dumpelfeld, B.; Kuster, B. Femtomol sensitivity post-digest  $^{18}\text{O}$  labeling for relative quantification of differential protein complex composition. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 869–876.
12. Li, X. J.; 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.
13. Schulze, W. X.; Mann, M. A novel proteomic screen for peptide-protein interactions. *J. Biol. Chem.* **2004**, *279*, 10756–10764.
14. Eng, J. K.; McCormack, A. L.; Yates, J. R. 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.
15. Yergey, J.; Heller, D.; Hansen, G.; Cotter, R. J.; Fenselau, C. Isotopic distributions in mass spectra of large molecules. *Anal. Chem.* **1983**, *55*, 353–356.