Anal. Chem. **2002,** *74,* 1650−1657

Analysis of Quantitative Proteomic Data Generated via Multidimensional Protein Identification Technology

Michael P. Washburn,*,† Ryan Ulaszek,† Cosmin Deciu,† David M. Schieltz,† and John R. Yates, III†,‡

Proteomics, Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, California 92121, and Department of Cell Biology SR11, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

We describe the analysis of quantitative proteomic samples via multidimensional protein identification technology (MudPIT). Ratio amounts of the soluble portion of the *S. cerevisiae* proteome from cultures of *S. cerevisiae* strain S288C grown in either ^{14}N minimal media or ^{15}N-enriched minimal media were mixed and digested into a complex peptide mixture. A 1× ^{14}N/1× ^{15}N complex peptide mixture was analyzed by single-dimensional reversed-phase chromatography and electrospray ionization quadrapole time-of-flight mass spectrometry in order to demonstrate the replacement of ^{14}N by ^{15}N under the growth conditions used. After conformation of the incorporation of ^{15}N into the labeled sample, three separate samples consisting of a 1× ^{14}N/1× ^{15}N complex peptide mixture, a 5× ^{14}N/1× ^{15}N complex peptide mixture, and a 10× ^{14}N/1× ^{15}N complex peptide mixture were analyzed via MudPIT. We demonstrate the dynamic range of the system by analyzing a 1:1, 5:1, and 10:1 data set using the soluble portion from *S. cerevisiae* grown in either ^{14}N or ^{15}N-enriched minimal media. The method described provides an accurate way to undertake a large-scale quantitative proteomic study.

Multidimensional protein identification technology (MudPIT) is a method that directly couples a microscale two-dimensional chromatography system to a tandem mass spectrometer, allowing for the rapid analysis of a proteome.[1−3] In MudPIT, a complex peptide mixture is generated and loaded onto a biphasic microcapillary column consisting of strong cation exchange and reversed-phase material.[1−3] The biphasic microcapillary column is placed in-line with an HPLC and tandem mass spectrometer, and peptides are directly eluted off of the biphasic microcapillary column into the tandem mass spectrometer.[1−3] The data are searched using the SEQUEST algorithm, which interprets the tandem mass spectra (MS/MS) generated and identifies the peptide sequence from which it was generated, resulting in the determination of the protein content of the original sample.[4] First developed by Link et al.[1] as a rapid tool for analysis of the *Saccharomyces cerevisiae* ribosome, improvements in MudPIT resulted in the largest proteomic analysis to date with the detection and identification of 1484 proteins from the proteome of *S. cerevisiae*.[2] Included in the detection and identification of 1484 proteins were proteins traditionally difficult to detect and identify via 2D-PAGE such as transcription factors, protein kinases, and integral membrane proteins.[2]

While qualitative proteomic analyses play an important biological role in identifying proteins in complexes, for example, a variety of efforts are underway to establish the field of quantitative proteomics.[5−12] In all of these systems, relative and not absolute abundances are determined. In a quantitative proteomic analysis, cells are grown under at least two different conditions expected to alter protein expression levels. To determine the relative abundance of a protein from a sample by mass spectrometry, the masses of identical proteins from each sample need to be modified so peptides from these proteins have unique masses. The proteins from each cell growth condition are labeled with a "heavy" or "light" tag. The methods of quantitative proteomics can generally be placed in four classes, metabolic labeling,[5−7] postgrowth amino acid labeling,[8,9] digestion labeling,[11] and postdigestion labeling.[10,12] Each of these methods introduces "heavy" and "light" labels into a protein or peptide at distinct points in an experimental system

* To whom correspondence should be addressed: (e-mail) michael.washburn@syngenta.com; (phone) 858-812-1000; (fax) 858-526-1549.

† Torrey Mesa Research Institute.

‡ The Scripps Research Institute.

(1) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd. *Nat. Biotechnol.* **1999**, *17*, 676−682.

(2) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd. *Nat. Biotechnol.* **2001**, *19*, 242−247.

(3) Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd. *Anal. Chem.* **2001**, *73*, 5683−5690.

(4) Eng, J.; McCormack, A. L.; Yates, J. R., 3rd. *J. Am. Mass Spectrom.* **1994**, *5*, 976−989.

(5) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6591−6596.

(6) Pasa-Tolic, L.; Jensen, P. K.; Anderson, G. A.; Lipton, M. S.; Peden, K. K.; Martinovic, S.; Tolic, N.; Bruce, J. E.; Smith, R. D. *J. Am. Chem. Soc.* **1999**, *121*, 7949−7950.

(7) 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. *Anal. Chem.* **2001**, *73*, 2132−2139.

(8) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994−999.

(9) Griffin, T. J.; Gygi, S. P.; Rist, B.; Aebersold, R.; Loboda, A.; Jilkine, A.; Ens, W.; Standing, K. G. *Anal. Chem.* **2001**, *73*, 978−986.

(10) Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Anal. Chem.* **2000**, *72*, 4047−4057.

(11) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 2836−2842.

(12) Goodlett, D. R.; Keller, A.; Watts, J. D.; Newitt, R.; Yi, E. C.; Purvine, S.; Eng, J. K.; von Haller, P.; Aebersold, R.; Kolker, E. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1214−1221.

10.1021/ac015704l CCC: $22.00
Published on Web 03/05/2002

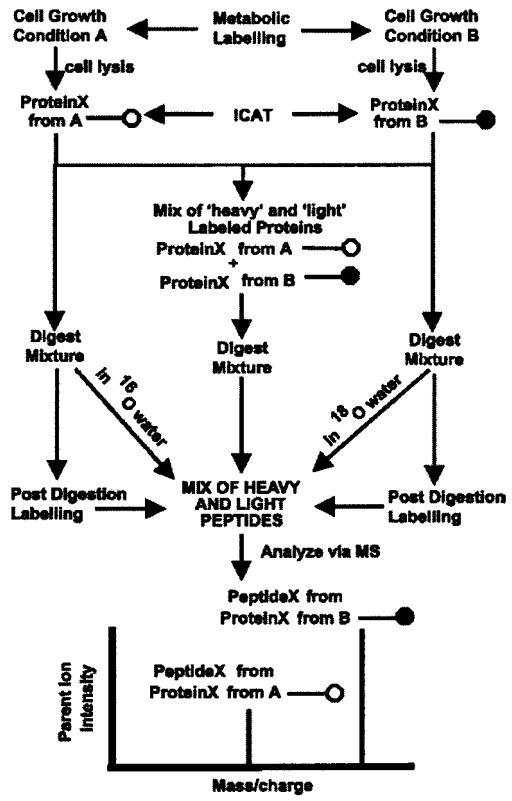

Figure 1. Method entry into a quantitative proteomic analysis scheme. When a quantitative proteomic analysis is carried out, the key is for the same peptide from two unique growth conditions to have unique masses when being analyzed by a mass spectrometer. "Heavy" and "light" peptides may be generated at many points in a sample preparation pathway. Metabolic labeling[5−7] introduces a label during the growth of the organism and is therefore the earliest point of introduction of heavy and light labels. Metabolic labeling is followed by ICAT,[8,9] digestion in ^{16}O and ^{18}O water,[11] and last postdigestion labeling.[10,12] Only after a label has been introduced can the samples be mixed and further processed. In this paper, metabolic labeling was used to label all of the proteins from *S. cerevisiae* strain S288C.

(Figure 1). However, introduction of heavy and light labels as early as possible minimizes differential sample loss.

The earliest introduction of a heavy label is possible when one can control the growth of an organism. This is particularly relevant in microorganisms such as *S. cerevisiae*,[5] *Escherichia coli*,[6] and *Deinococcus radiodurans*[7] but has also been demonstrated in mammalian tissue culture.[7] In metabolic labeling, the proteins in the cells grown under a particular condition can be uniformly labeled by adding a heavy isotope to the growth media.[5−7] The generalized flowchart of a metabolic labeling-based quantitative proteomic study is seen as part of Figure 1. An advantage of metabolic labeling is that the protein lysates are mixed together at an early stage in the experiment, decreasing the chances for an experimental error being inadvertently introduced to one sample and not the other (Figure 1). A disadvantage of this method is that it is limited to situations where the researcher can exquisitely control the growth media of a cell.

The goal of a quantitative proteomic study should be to provide as complete coverage of a proteome as possible. To date, the largest quantitative proteomic study detected, identified, and determined the relative abundance of 491 unique proteins from the microsome of human myeloid leukemia cells.[13] In genomics, a cDNA or high-density oligonucleotide array can detect, identify, and determine the relative abundance of thousands of unique mRNAs.[14] While mRNA and protein provide unique challenges for quantitative analysis, quantitative proteomics currently lacks the coverage of a proteome that array analysis provides of a transciptome.

To increase the potential coverage of a proteome in a quantitative proteomic method, we coupled the detection and identification power of MudPIT to the quantitative proteomic analysis method of metabolic labeling in *S. cerevisiae*. In the current paper, *S. cerevisiae* was grown in either minimal media with ^{14}N ammonium sulfate (defined as ^{14}N minimal media) or minimal media with ^{15}N ammonium sulfate (defined as ^{15}N minimal media). In this way, the source of nitrogen in the media was controlled, allowing for the uniform labeling of all the peptides in *S. cerevisiae* grown in either media. Using three separate complex peptide mixtures, we determined that the dynamic range of the system is at least 10:1, and we were able to detect and identify more than 800 unique proteins from each sample, demonstrating the potential of metabolic labeling coupled to MudPIT to be an excellent tool for quantitative proteomic analysis.

MATERIALS AND METHODS

Materials. Urea, ammonium acetate, ammonium bicarbonate (AmBic), dithiothreitol (DTT), iodoacetamide (IAM), EDTA, dibasic sodium phosphate, dibasic potassium phosphate, and sodium carbonate were obtained from Sigma (St. Louis, MO). Ammonium-^{15}N sulfate (99 atom %), Ammonium-^{14}N sulfate (99.99 atom %), sodium vanadate, sodium fluoride, and sodium pyrophosphate were products of Aldrich (Milwaukee, WI). Difco dextrose and yeast nitrogen base without amino acids or ammonium sulfate were products of Becton Dickinson Microbiology Systems (Sparks, MD). Poroszyme bulk immobilized trypsin was a product of Applied Biosystems (Framingham, MA). HPLC grade acetonitrile (ACN) and HPLC grade methanol was purchased from Fischer Scientific (Fair Lawn, NJ). Endoproteinase Lys-C was purchased from Roche Diagnostics (Indianapolis, IN). Heptafluorobutyric acid (HFBA) was obtained from Pierce (Rockford, IL). SPEC Plus PT C18 solid-phase extraction (SPE) pipet tips were purchased from Ansys Diagnostics (Lake Forest, CA). Glacial acetic acid was purchased from Mallinckrodt Baker Inc. (Paris, KY).

Growth and Lysis of *S. cerevisiae*. The *S. cerevisiae* strain S288C was chosen since it is a parental yeast strain that is able to grow in minimal media without supplementation of amino acids whose metabolic pathways have been genetically disrupted. After preparing overnight cultures in identical media *S. cerevisiae* strain S288C was grown to mid log phase (O.D. 0.6) in ^{14}N or ^{15}N minimal media (1.7 g of yeast nitrogen base without amino acids and ammonium sulfate, 20 g of dextrose, and 5 g of either ammonium sulfate per liter) at 30 °C followed by centrifugation at 1000*g*. The pellets were washed three times with 1× phosphate-buffered saline (1.4 mM NaCl, 0.27 mM KCl, 1 mM Na_2HPO_4, 0.18 mM KH_2PO_4, pH 7.4). Lysis buffer (0.1 M Na_2CO_3, 310 mM NaF, 3.45 mM $NaVO_3$, 12 mM EDTA, 250 mM NaCl) was added to the cells, and cells were lysed in a mortar and pestle frozen in

(13) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. *Nat. Biotechnol.* **2001**, *19*, 946−951.
(14) Lockhart, D. J.; Winzeler, E. A. *Nature* **2000**, *405*, 827−836.

liquid nitrogen. After lysis, the lysed cells and solution were thawed and placed on ice for 30 min. After incubation, the lysed cells were spun at 14 000 rpm for 10 min at 4 °C in an Eppendorf microfuge and the supernatant was removed.

Preparation and Analysis of Ratio Mixtures. A Bradford assay (Bio-Rad, Hercules, CA) was carried out on the soluble portion of the proteome from each of the samples. Three unique samples were then generated consisting of protein from each sample combined into ratio mixtures of a 1× ^{14}N proteins to 1× ^{15}N proteins sample (250 μg of total protein, to be named the 1× ^{14}N/1× ^{15}N sample), a 5× ^{14}N proteins to 1× ^{15}N proteins sample (360 μg of total protein, to be named the 5× ^{14}N/1x ^{15}N sample), and a 10× ^{14}N proteins and 1× ^{15}N proteins sample (350 μg of total protein, to be named the 10× ^{14}N/1× ^{15}N sample). Each of the three samples were then prepared for MudPIT analysis by denaturing in urea, reduction with DTT, modification with IAM, sequential digestion with endoproteinase Lys-C and trypsin beads, and buffer exchange/concentration with SPEC Plus PTC18 SPE tips as described previously.[2]

Single Dimension Reversed-Phase Mixture Analysis via the Q-TOFII. The 1× ^{14}N/1× ^{15}N sample was analyzed on a Q-TOFII (Micromass Inc., Beverly, MA), a quadrapole time-of-flight instrument coupled to a quaternary HP 1100 series HPLC (Agilent Technologies, Palo Alto, CA). The Q-TOFII was equipped with a Z-spray nanoelectrospray source fitted with the glass capillary probe option. A fused-silica microcapillary column (100 μm i.d. × 365 μm o.d.) was pulled with a model P-2000 laser puller (Sutter Instrument Co., Novato, CA) as described.[15] The pulled microcapillary column was packed with 10 cm of 5-μm Zorbax Eclipse XDB-C_{18} (Agilent Technologies) and affixed to a PEEK finger-tight microcross electrospray interface described previously.[15] The capillary/corona voltage line was attached directly to the PEEK finger-tight microcross electrospray interface through a gold rod. The microcapillary column packed with RP material was threaded through the opening to the glass capillary option stage of the Q-TOFII tail end first after which it was connected to the PEEK finger tight microcross.

The Q-TOFII was run according to the manufacturer's instructions for usage of the instrument with a HPLC pump and the glass capillary option of the Z-spray nanoelectrospray source. Internal parameters were set as follows. The electrospray capillary voltage was set to 2.2 kV, the cone voltage set to 40 V, and the source temperature set to 80 °C. The MS survey scan was m/z 400−1400 with a scan time of 0.42 s and an interscan time of 0.08 s. When the intensity of a peak rose above a threshold of 8 counts, tandem mass spectra were acquired. Normalized collision energies for peptide fragmentation was set using the charge-state recognition files for +1, +2, and +3 peptide ions. The scan range for MS/MS acquisition was from m/z 50 to 1900 with a scan time of 1.92 s and an interscan time of 0.08 s. After fragmentation, an m/z value was dynamically excluded for 4 min. Upon the completion of a run, the MS/MS data were extracted to *.RAW files according to the manufacturer's instructions and analyzed via the SEQUEST[4] algorithm. The SEQUEST[4] algorithm was run as described below for analysis of both the ^{14}N and ^{15}N peptides with two separate sequest.params files. The only specialized alteration to the sequest.params file for analysis of data from a Q-TOFII was the setting of the parent and fragment mass types to monoisotopic. Xcorr and ΔCn criteria were applied to the results as described below for data from MudPIT analyses.

Multidimensional Protein Identification Technology Analysis of Complex Ratio Peptide Mixtures on the Finnigan LCQ Ion Trap Mass Spectrometer. The MudPIT system was used as described previously[2,3] Briefly, a quaternary HP 1100 HPLC pump was interfaced with a Finnigan LCQ ion trap mass spectrometer (Finnigan MAT, San Jose, CA). A 100 × 365 μm fused-silica capillary microcolumn (Agilent Technologies) was prepared with a P-2000 laser puller (Sutter Instruments Co.) as described previously.[15] The fritless microcapillary column was first packed with 10 cm of 5-μm Zorbax Eclipse XDB-C_{18} and then with 4 cm of 5-μm Partisphere strong cation exchange (SCX) resin (Whatman, Clifton, NJ). The column was connected to a PEEK microcross as described elsewhere,[15] in order to split the flow of the HPLC pump to an effective flow rate of 0.15−0.25 μL/min and supply a spray voltage of 1.8 kV. Operation of the LCQ was executed as described previously.[3]

A fully automated 13-cycle chromatographic run was carried out on each sample. The four buffer solutions used for the chromatography were 5% ACN/0.012% HFBA/0.5% acetic acid (buffer A), 80% ACN/0.012% HFBA/0.5% acetic acid (buffer B), 250 mM ammonium acetate/5% ACN/0.012% HFBA/0.5% acetic acid (buffer C), and 500 mM ammonium acetate/5% ACN/0.012% HFBA/0.5% acetic acid (buffer D). Each of the first 10 cycles were 110 min with the following profile: 3 min of 100% buffer A, 2 min of *X*% buffer C, 5 min of 100% buffer A, a 10-min gradient from 0 to 10% buffer B, and a 90-min gradient from 10 to 45% buffer B. The 2-min buffer C profiles in the first 12 cycles were as follows: cycle 1 10%, cycle 2 20%, cycle 3 30%, cycle 4 40%, cycle 5 50%, cycle 6 60%, cycle 7 70%, cycle 8 80%, cycle 9 90%, and cycle 10 100%. Cycle 11 consisted of a 5-min 100% buffer A wash followed by a 20-min 100% buffer C wash, a 5-min 100% buffer A wash, a 20-min gradient from 0 to 10% buffer B, and a 130-min gradient from 10 to 60% buffer B. Cycle 12 was identical to cycle 11 except that the 20-min salt wash was with 100% buffer D, and cycle 13 was identical to cycle 12 except that the last 130-min gradient was from 10 to 100% buffer B.

SEQUEST and Quantification Analysis. The SEQUEST algorithm was used to interpret MS/MS as described previously.[1,2,4,16,17] The SEQUEST[4] algorithm was run two separate times on each of the three data sets against the yeast_orfs.fasta database from the National Center for Biotechnology Information. Each sample had to be run twice to separately detect and identify peptides from the ^{14}N minimal media sample and from the ^{15}N minimal media sample by using two separate SEQUEST[4] parameters files where the masses of each amino acid was set to the corresponding growth conditions and nitrogen content therein. Specifically, to search the data set for peptides where all of the amino acids were fully ^{15}N labeled, the mass of each amino acid needed to be statically modified to an increased mass depending on the number of nitrogen atoms in that amino acid. In both data sets, all accepted results had a ΔCn of 0.1 or greater and specific

(15) Gatlin, C. L.; Kleemann, G. R.; Hays, L. G.; Link, A. J.; Yates, J. R., 3rd. *Anal. Biochem.* **1998**, *263*, 93−101.

(16) Yates, J. R., 3rd.; Morgan, S. F.; Gatlin, C. L.; Griffin, P. R.; Eng, J. K. *Anal. Chem.* **1998**, *70*, 3557−3565.

(17) Yates, J. R., 3rd.; Carmack, E.; Hays, L.; Link, A. J.; Eng, J. K. *Methods Mol. Biol.* **1999**, *112*, 553−569.

cross-correlation score criteria were applied depending on the tryptic nature and charge state of a given peptide.[2,3] Briefly, fully tryptic singly charged peptides with cross-correlation scores (Xcorrs) of 1.9 or higher were accepted. Fully or partially tryptic doubly charged peptides with Xcorrs at least 2.2 and all doubly charged peptides with Xcorrs above 3 were accepted. Last fully or partially tryptic triply charged peptides were accepted if their XCorrs were above 3.75. Visual assessment[1] of matches was carried out when a protein was identified by three or fewer unique peptides possessing SEQUEST[4] scores that passed the above criteria.

Next, a single separate file was generated for each of the three ratio samples. Each file contained both the ^{14}N and ^{15}N peptides corresponding to particular proteins from each of the three samples that passed the SEQUEST[4] criteria described in the previous paragraph. Included in these files for each of the peptides in the file is the parent ion mass, the charge state of the identified peptides, and the MudPIT step information. Each file was then submitted to a software program that determined the relative abundance of each peptide in the chromatographic runs of MudPIT. When only one of the peptides from a peak pair generated a tandem mass spectrum of sufficient quality to be included in the list, the software calculated the expected m/z of the missing partner by assuming 100% replacement of ^{14}N in every amino acid in the identified peptide with ^{15}N or vice versa. Next, the MS scans used are searched to confirm that no charge state of any other peptide, either ^{14}N or ^{15}N, in the same scan range has an m/z within 2.0 amu of the predicted pair. Then, the chromatographic elution peak area for the missing partner was determined. Relative abundance ratios for individual peptides were generated, and when multiple hits to a protein existed, the software determined the average ratio using all of the successful ratio determinations. The software then reported back a new data file containing the peptides detected and identified from each protein and the relative abundance of each of these peptides.

RESULTS AND DISCUSSION

Analysis of ^{15}N Enrichment of Amino Acids in *S. cerevisiae*. In order for accurate analysis of relative abundances using metabolic labeling, isotopic peak distribution shape of the same peptide from a sample grown in ^{14}N minimal media and ^{15}N minimal media must be nearly identical (data not shown). During the sample preparation, colonies of the *S. cerevisiae* strain S288C plated on normal YPD were first transferred and grown in 5-mL overnight cultures in either ^{14}N minimal media or ^{15}N minimal media. The next day one 5-mL culture was expanded into 125 mL of the same media it was grown in the night before. To ensure replacement of ^{14}N with ^{15}N in *S. cerevisiae* cultures grown in ^{15}N minimal media, the ^{15}N enrichment of peptides of *S. cerevisiae* cultures grown in ^{15}N minimal media was analyzed via a MicroMass Q-TOFII. The Q-TOFII is a high-resolution quadrapole time-of-flight tandem mass spectrometer that allows one to resolve the isotope distribution of any peptide. A 40-μg aliquot of the 1× ^{14}N/1× ^{15}N sample (see Materials and Methods section for definition) was applied to a 15-cm RP column connected to the electrospray source of the Q-TOFII. A 2-h reversed-phase gradient was run on this sample. Using the SEQUEST[4] scoring criteria described in Materials and Methods, 67 unique proteins and 102 unique peptides were detected and identified using the ^{14}N sequest.

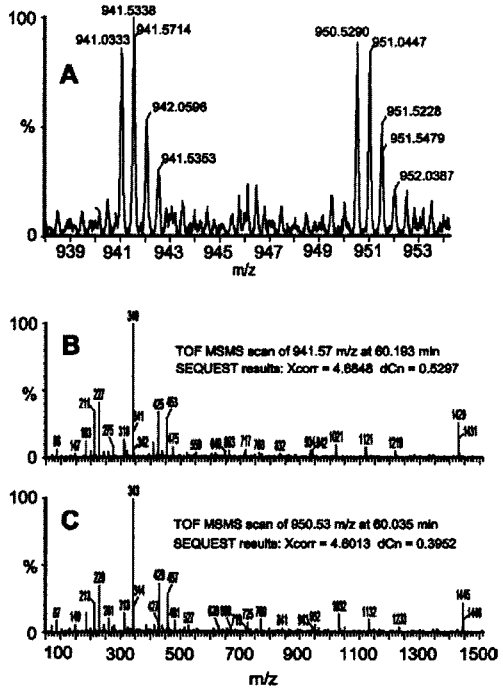

Figure 2. Analysis of ^{15}N enrichment via a high-resolution mass spectrometer. A 1× ^{14}N/1× ^{15}N complex peptide mixture generated from *S. cerevisiae* grown in normal minimal media and ^{15}N-enriched minimal media was loaded onto a 10-cm reversed-phase microcapillary column. The loaded microcapillary column was attached to a PEEK microcross incorporating a gold rod, HPLC flow line, and a flow−split line in the other three ports. MS scans and MS/MS scans were collected over the course of a 2-h reversed-phase gradient. After the run, the QTOFII data were converted to SEQUEST[4] compatible files and the SEQUEST algorithm was run on the data set using the yeast.fasta file from the National Center for Biotechnology Information. (A) The MS scans from 59.567 to 60.713 min were combined, and the region from *m*/*z* 938 to 957 was expanded to show the isotopic distribution of the *m*/*z* 941/950 peak pair. (B) The MS/MS taken at 60.193 min was identified by the SEQUEST algorithm as the ^{14}N version of the peptide IPEIPLVVSTDLESIQK from RPL4A/4B. The MS/MS taken was of the 1-^{13}C version of the peptide. (C) The MS/MS taken at 60.035 min was identified by the SEQUEST algorithm as the ^{15}N version of the peptide IPEIPLVVSTDLESIQK from RPL4A/4B. The MS/MS taken was of the all ^{15}N and ^{12}C version of the peptide.

params files, and 64 unique proteins and 84 unique peptides were detected and identified using the ^{15}N sequest.params files for a total of 103 unique proteins and 160 unique peptides (data not shown).

Figure 2A contains the summed MS scans of the elution of the ^{14}N and ^{15}N versions of the peptide IPEIPLVVSTDLESIQK from the ribosomal protein RPL4A/4B. The monoisotopic peak for the $(M + 2H)^{2+}$ for this peptide with all of the nitrogen as ^{14}N is m/z 941.0333. The monoisotopic peak for the $(M + 2H)^{2+}$ for this peptide with all of the nitrogen as ^{15}N is m/z 950.5333. In addition, the isotopic pattern of each peptide is nearly identical with only the 3-^{13}C peak of the ^{15}N peptide being difficult to separate from the background at m/z 952.0387 (Figure 2A). The tandem mass spectrum of the ^{14}N peptide is shown in Figure 2B and the tandem mass spectrum of the ^{15}N peptide is shown in Figure 2C. The tandem mass spectra were acquired within 1.5 s of each other, and they yielded highly similar SEQUEST results (Figure 2B and C). On the Q-TOFII, similar patterns were

observed throughout the 2-h run (data not shown). That is, $^{14}N/^{15}N$ peptide peak pairs eluted at the same time, and when tandem mass spectra were generated for both of the peptides, the SEQUEST scores were nearly identical. Furthermore, the analysis of the $1\times$ $^{14}N/1\times$ ^{15}N sample mixture via the Q-TOFII demonstrated that the growth of *S. cerevisiae* under the conditions described replaced ^{14}N in all the amino acids in a protein with ^{15}N. As a result, when a ^{14}N and ^{15}N peptide peak pair elutes into a mass spectrometer off of a one-dimensional or two-dimensional column into a mass spectrometer, the chromatographic peak shape will only be influenced by peptide concentration and the isotopic distribution will not skew the quantitative information in the peak.

Analysis of the Dynamic Range of Quantitative MudPIT. MudPIT on the Finnigan LCQ ion trap mass spectrometer has been demonstrated to be a rapid and comprehensive method for proteome analysis.[2,3] After demonstrating the replacement of ^{14}N with ^{15}N on a high-resolution instrument, three separate complex peptide mixtures were designed to determine the dynamic range and potential of coupling metabolic labeling[5−7] to MudPIT on a lower resolution system as a comprehensive quantitative proteomic tool. Each of the three separate mixtures generated (the $1\times$ $^{14}N/1\times$ ^{15}N sample, the $5\times$ $^{14}N/1\times$ ^{15}N sample, and the $10\times$ $^{14}N/1\times$ ^{15}N sample) was analyzed via MudPIT. In representative analyses, MudPIT identified 638 unique proteins from 1460 unique ^{14}N peptides and 664 unique proteins from 1642 unique ^{15}N peptides for a combined total of 842 unique proteins and 1836 unique peptides from the $1\times$ $^{14}N/1\times$ ^{15}N sample. MudPIT identified 774 unique proteins from 2254 unique ^{14}N peptides and 288 unique proteins from 443 unique ^{15}N peptides for a combined total of 895 unique proteins and 2400 unique peptides from the $5\times$ $^{14}N/1\times$ ^{15}N sample. Last, MudPIT identified 763 unique proteins from 2119 unique ^{14}N peptides and 222 unique proteins from 283 unique ^{15}N peptides for a combined total of 872 unique proteins and 2264 unique peptides from the $10\times$ $^{14}N1\times$ ^{15}N sample.

We analyzed the three data sets and compiled relative abundance ratios for the majority of the proteins identified. A sample data point from the $1\times$ $^{14}N/1\times$ ^{15}N sample is shown in Figure 3. Figure 3A contains the elution profile of the peptide LSVISIDHLPSLLPR from the protein LYS1 (YIR034C). The average mass for the ^{14}N $(M + 2H)^{2+}$ version of this peptide is m/z 831.00 and the average mass for the ^{15}N $(M + 2H)^{2+}$ version of this peptide is m/z 841.00. Both the ^{14}N version (m/z 831.0) and ^{15}N version (m/z 840.9) of this peptide eluted from 52.77 to 54.50 min in a 120-min MudPIT step for an elution time of 1.73 min. Taking into account a 1 amu m/z window on either side of the peak m/z, the ratio of ^{14}N peptide to ^{15}N peptide was 1.03. MS/MS of the ^{14}N peptide was taken at 53.48 min and of the ^{15}N peptide at 53.26 min. The MS/MS of the ^{14}N peptide had an Xcorr of 3.88 and a ΔCn of 0.43 (Figure 3B) and the MS/MS of the ^{15}N peptide had an Xcorr of 3.75 and a ΔCn of 0.36 (Figure 3C). At the level of the chromatogram, the total ion currents of the ^{14}N and ^{15}N version of the peptide LSVISIDHLPSLLPR yielded a 1:1 ratio, resulting in nearly identical SEQEUST score results. Figure 3A also contains portions of the elution profiles of additional ^{14}N and ^{15}N peptide pairs seen at m/z 545.3/551.7 (the +2 charge state ^{14}N and ^{15}N versions of the peptide WPFWLSPR from YIL078w (THS1)), at m/z 890.8/902.0 (the +3 charge state ^{14}N and ^{15}N versions of the peptide RYTNPVVIQGAGAGAAVTAAGVLGDVIK from YJR139c (HOM6)), and at m/z 1335.2/1352.7 (the +2 charge state ^{14}N and ^{15}N versions of the peptide RYTNPVVIQGAGAGAAVTAAGVLGDVIK from YJR139c (HOM6)). In these cases, the majority of the peak pairs also indicate a 1:1 ratio even though the complete elution profile is not shown except for the peptide LSVISIDHLPSLLPR.

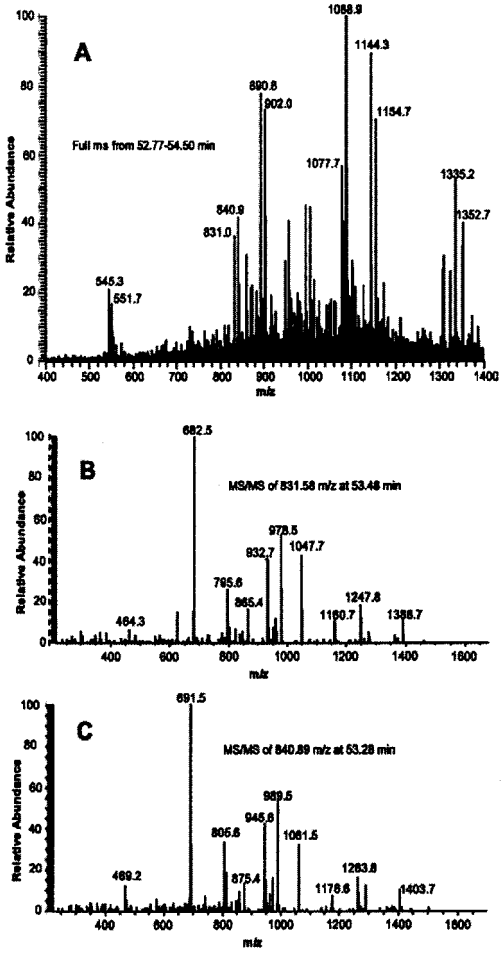

Figure 3. Partial LCQ chromatogram of the 1:1 ^{14}N to ^{15}N samples of *S. cerevisiae* grown in minimal media. A $1\times$ $^{14}N/1\times$ ^{15}N complex peptide mixture generated from *S. cerevisiae* grown in normal minimal media and ^{15}N-enriched minimal media was loaded onto a two-dimensional microcapillary column. The biphasic loaded microcapillary column was attached to a PEEK microcross and analyzed via a 13-step MudPIT analysis on a Finnigan LCQ ion trap mass spectrometer as described in Materials and Methods. (A) The normalized relative abundance of the combined MS spectra from 52.77 to 54.50 min from one of the MudPIT steps is shown to detail the elution profile of the peptide peak pair at m/z 831.0 and 840.9. (B) The SEQUEST[4] algorithm identified the MS/MS taken at 53.48 min as the ^{14}N version of the peptide LSVISIDHLPSLLPR from the protein LYS1 (YIR034C). (C) The SEQUEST algorithm identified the MS/MS taken at 53.28 min as the ^{15}N version of the peptide LSVISIDHLPSLLPR from the protein LYS1 (YIR034C).

In Figure 4, a sample data point for the peptide LSVISIDHLPSLLPR from the protein LYS1 (YIR034C) is shown from the $5\times$ $^{14}N/1\times$ ^{15}N sample data set. This is the same peptide whose data from the $1\times$ $^{14}N/1\times$ ^{15}N sample data set is shown in Figure 3. In the $5\times$ $^{14}N/1\times$ ^{15}N sample data set, the peptides eluted over a 1.54-min time window defined by the ^{14}N peptide. The ^{14}N peptide

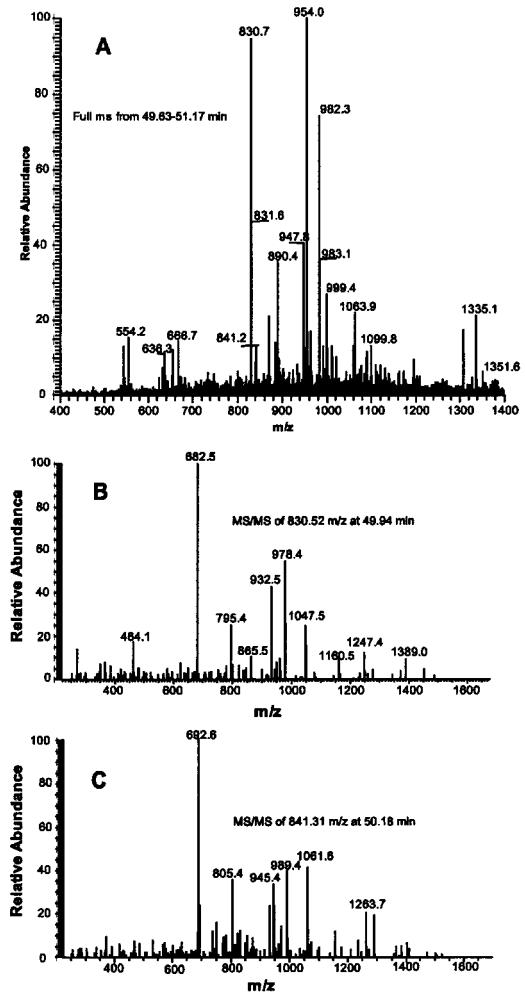

Figure 4. LCQ chromatogram of the 5:1 ^{14}N to ^{15}N samples of *S. cerevisiae* grown in minimal media. A 5× ^{14}N/1× ^{15}N complex peptide mixture generated from *S. cerevisiae* grown in normal minimal media and ^{15}N-enriched minimal media was loaded onto a two-dimensional microcapillary column. The biphasic loaded microcapillary column was attached to a PEEK microcross and analyzed via a 13-step MudPIT analysis on a Finnigan LCQ ion trap mass spectrometer as described in Materials and Methods. (A) The normalized relative abundance of the combined MS spectra from 49.63 to 51.17 min from one of the MudPIT steps is shown to detail the elution profile of the peptide peak pair at *m*/*z* 830.7 and 841.2. (B) The SEQUEST[4] algorithm identified the MS/MS taken at 49.94 min as the ^{14}N version of the peptide LSVISIDHLPSLLPR from the protein LYS1 (YIR034C). (C) The SEQUEST algorithm identified the MS/MS taken at 50.18 min as the ^{15}N version of the peptide LSVISIDHLPSLLPR from the protein LYS1 (YIR034C).

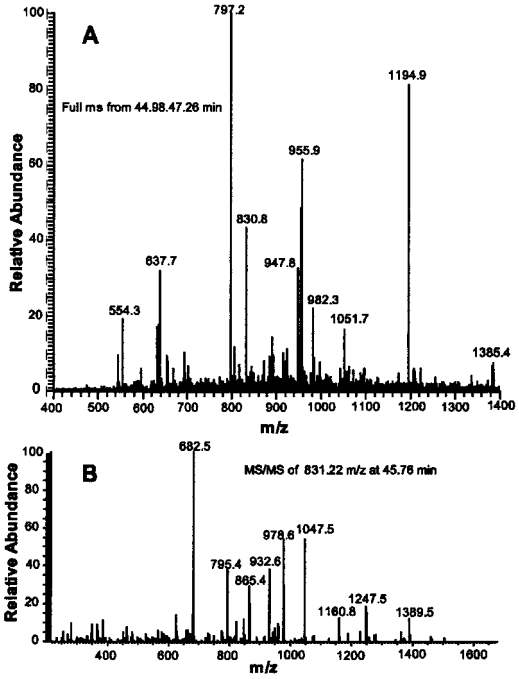

Figure 5. LCQ chromatogram of the 10:1 ^{14}N to ^{15}N samples of *S. cerevisiae* grown in minimal media. A 10× ^{14}N/1× ^{15}N complex peptide mixture generated from *S. cerevisiae* grown in normal minimal media and ^{15}N-enriched minimal media was loaded onto a two-dimensional microcapillary column. The biphasic loaded microcapillary column was attached to a PEEK microcross and analyzed via a 13-step MudPIT analysis on a Finnigan LCQ ion trap mass spectrometer as described in Materials and Methods. (A) The normalized relative abundance of the combined MS spectra from 44.98 to 47.26 min from a single step in MudPIT is shown to detail the elution profile of the peptide peak at *m*/*z* 830.8. (B) The SEQUEST[4] algorithm identified the MS/MS taken at 45.76 min as the ^{14}N version of the peptide LSVISIDHLPSLLPR from the protein LYS1 (YIR034C). The ^{15}N version of the peptide LSVISIDHLPSLLPR from the protein LYS1 (YIR034C) was of insufficient quantity to generate a tandem mass spectrum of sufficient quality to be accurately identified by SEQUEST.

is seen at m/z 830.7 and the ^{15}N peptide seen at m/z 841.2 in Figure 4A. The ratio of the $^{14}N/^{15}N$ peptide was 5.35:1. Even with a 5× decrease in the ^{15}N peptide versus the ^{14}N peptide, MS/MS were generated for the peptide within 14 s of each other (Figure 4B and C). In addition, even though there was a 5× decrease in the ^{15}N peptide concentration from the ^{14}N peptide concentration, both of the peptides had highly similar SEQUEST results with the MS/MS of the ^{14}N peptide having an Xcorr of 3.14 and a ΔCn of 0.40 and the ^{15}N peptide having an Xcorr of 3.26 and a ΔCn of 0.31 (data not shown).

In the 10× $^{14}N1/1\times$ ^{15}N sample data set, we again detected and identified the peptide LSVISIDHLPSLLPR from the protein LYS1 (YIR034C). However, in the 10× $^{14}N/1\times$ ^{15}N sample data set, only the ^{14}N peptide was detected and identified. Figure 5A shows the elution of the ^{14}N peptide over a 2.28-min time period. The MS/MS of the ^{14}N peptide shown in Figure 5B had an Xcorr of 4.08 and a ΔCn of 0.43. Even though the ^{15}N peptide was not detected and identified using the SEQUEST results criteria described previously, we were able to determine the relative abundance of the ^{14}N peptide was 9.1-fold greater than that of the ^{15}N peptide. As described in Materials and Methods, the m/z of the ^{15}N peptide was calculated by assuming the complete replacement of ^{14}N by ^{15}N in the peptide LSVISIDHLPSLLPR. Using this m/z as a pointer, the abundance of the ^{15}N peptide was calculated and the ratio of the ^{14}N to ^{15}N peptide was then determined. Especially in the 10:1 data set, confident SEQUEST results were typically obtained for only one peptide in each peptide pair, in this case the ^{14}N peptide, but accurate relative abundances were determined (see discussion below).

While Figures 3−5 demonstrate the ability of the system to detect, identify, and quantify a peptide, LSVISIDHLPSLLPR, from one loci, LYS1, our data sets contained substantially more data points. A portion of this data set is shown in Table 1, where the average ratios and their respective standard deviations from the

Table 1. Ratios of Peptide Abundance from Proteins from 1:1, 5:1, and 10:1 ^{14}N to ^{15}N Samples of *S. cerevisiae* Grown in Minimal Media

	data set					
	1× ^{14}N/1× ^{15}N[a]		5× ^{14}N/1× ^{15}N[a]		10× ^{14}N/1× ^{15}N[a]	
loci (name)	*N*[b]	ratio[c]	*N*[b]	ratio[c]	*N*[b]	ratio[c]
YAL003w (EFB1)	2	0.99 ± 0.32	3	4.76 ± 0.71	4	9.20 ±3.25
YAL012w (CYS3)	7	1.45 ± 0.73	7	5.57 ± 2.36	10	9.25 ± 2.76
YAL016w (TPD3)	2	1.42 ± 0.12	3	5.41 ± 1.26	1	8.34
YCL018w (LEU2)	4	1.26 ± 0.25	5	4.00 ± 1.25	6	10.42 ± 2.55
YDL229w (SSB1)	11	1.26 ± 0.31	5	6.07 ± 1.98	6	10.69 ± 1.9
YDR037w (KRS1)	3	1.14 ± 0.17	3	4.38 ± 1.27	3	9.16 ± 2.58
YDR155c (CPH1)	5	1.18 ± 0.46	4	4.64 ± 2.31	3	8.99 ± 3.29
YEL046c (GLY1)	1	0.95	1	4.72	1	9.22
YGL202w (ARO8)	3	0.93 ± 0.02	6	4.47 ± 2.47	5	7.42 ± 3.13
YGR061c (ADE6)	4	1.57 ± 0.64	6	4.60 ± 1.85	3	9.61 ± 3.81
YHR027c (RPN1)	3	1.06 ± 0.66	1	5.11	1	9.29
YIR034c (LYS1)	5	1.07 ± 0.07	5	4.86 ± 0.67	3	10.54 ± 1.36
YJR105w (ADO1)	4	1.01 ± 0.10	4	4.81 ± 0.50	4	10.07 ± 1.19
YKL060c (FBA1)	19	1.24 ± 0.55	19	5.34 ± 2.71	17	9.42 ± 2.55
YMR105c (PGM2)	1	1.4	1	6.3	1	10.02
YNL216w (RAP1)	1	1.11	2	5.20 ± 0.005	1	8.69
YOL059w (GPD2)	1	1.16	2	5.83 ± 1.30	2	10.05 ± 3.50
YOL086c (ADH1)	7	1.5 ± 0.68	8	4.91 ± 1.45	8	10.39 ± 3.04
YOL140w (ARG8)	1	0.98	2	4.98 ± 0.31	3	10.42 ± 6.09
YPR074c (TKL1)	7	1.01 ± 0.22	6	5.77 ± 1.83	6	9.36 ± 2.67

[a] The data points shown were determined by MudPIT analysis followed by relative peptide abundance determination of each peptide from each of the loci shown in the table. Each of the three samples was run separately via MudPIT and analyzed separately for relative abundance determination. [b] *N* represents the number of peptides from a particular loci for which the relative abundance was determined. Each peptide used was detected and identified in each of the three different samples and possessed tandem mass spectra of sufficient quality to pass the SEQUEST criteria described previously.[2] [c] The ratio shown is the average and standard deviation of the ratios determined for each peptide (*N*) for each loci shown.

1:1, 5:1, and 10:1 peptide mixture data sets for unique proteins found in each mixture is shown. For instance, additional peptides were detected, identified, and quantified from the protein LYS1, and these additional peptides also had accurate ratios as demonstrated by the averages and standard deviations shown (Table 1). Furthermore, similar to the ratio of the peptide LSVISIDHLPSLLPR from the protein LYS1 in the 10× ^{14}N/1× ^{15}N sample data set, confident SEQUEST identifications were only needed for either the ^{14}N or ^{15}N version of the same peptide for accurate relative abundance determination (data not shown). While this occurred in all three data sets, this was particularly the case in the 10×^{14}N/1× ^{15}N data set (data not shown), and yet accurate ratios were determined (Table 1).

Of the 20 loci shown in Table 1, the 1× ^{14}N/1× ^{15}N sample had an average ratio of 1.18, the 5× ^{14}N/1× ^{15}N sample had and average ratio of 4.89, and the 10× ^{14}N/1× ^{15}N sample had an average ratio of 9.53. To determine whether the analysis of all three data sets was carried out in a uniform fashion, the standard deviation as a percentage of the average was determined for each protein in Table 1. Then the average and standard deviation of this percentage was determined. In the 1:1 data set the average standard deviation was 29 ± 17% of the average, 28 ± 16% in the 5:1 data set, and 30 ± 12% in the 10:1 data set. Therefore, the variation in each of the three data sets was consistent regardless of the computational challenges presented in each dataset.

There are several potential sources of this variation. One potential source of the variation is the lower resolution of the ion trap mass spectrometer. Yao et al. analyzed ^{18}O-labeled peptides on a high-resolution MALDI-Fourier transform time-of-flight mass spectrometer.[11] After calculating the results from Table 1 in Yao et al., the average standard deviation as a percentage of the average ratio using ^{18}O and a MALDI-TOF is 16 ± 11%. Although only ratios between 2.6 and 4.7 were reported, if the value of 16 ± 11% persists throughout at least a 10× dynamic range, the use of higher resolution instruments for quantitative proteomics may result in less variation in ratios. A second source of potential error is that of biological variation where even though both *S. cerevisiae* cultures were grown under identical conditions (excepting the isotopic alteration of nitrogen), the protein expression of each protein in each culture is not perfectly identical. In mRNA array analysis, only genes with postnormalization ratios of greater than 2× up- or downregulation are considered significant.[18] Effectively this corresponds to a ratio of 1 ± 1 for upregulated genes and 1 ± 0.5 for downregulated genes using the relative abundance reporting method used in this paper. Using these values, our 30% average standard deviation is a dramatic improvement.

While multiple peptide hits increases one's confidence in the relative abundance calculation, many single hits resulted in accurate relative abundance calculations. For example, YEL046c (GLY1) was detected and identified by only one peptide in all three mixtures, and the relative abundance ratio for each peptide hit resulted in accurate ratios of 0.95, 4.72, and 9.22, respectively (Table 1). In general, if a tandem mass spectrum is of sufficient quality to generate a SEQUEST result whose Xcorr and ΔCn passed the criteria described in Materials and Methods, then accurate relative abundances were easily determined. However, there are exceptions to this rule. For example, if the elution of a peptide comes at the very beginning or very end of a chromatographic step, inaccurate relative abundances are obtained because of partial chromatographic peak shapes (data not shown). In

(18) Quackenbush, J. *Nat. Rev. Genet.* **2001**, *2*, 418−427.

general, accurate relative abundance ratios were obtained regardless of the number of peptide hits obtained from a protein, but proteins identified by single hits need to be regarded with particular scrutiny since they are most susceptible to an error like the one described above.

CONCLUSIONS

We have described a system by which quantitative proteomics may be carried out via metabolic labeling [5−7] and MudPIT.[1−3] Using this system, a highly complex peptide mixture generated from *S. cerevisiae* was grown in either ^{14}N or ^{15}N minimal media was analyzed. Artificial 1× ^{14}N proteins, 5× ^{14}N proteins, and 10× ^{14}N proteins mixtures were generated and analyzed. Accurate quantitation was achieved to a dynamic range of 10× ^{14}N peptides to 1× ^{15}N peptides and did not require SEQUEST [4] identification of the lower abundant peptide to determine the relative abundance of the $^{14}N/^{15}N$ peptide pair. When tandem mass spectra were generated of both a ^{14}N and ^{15}N peptide pair, the SEQUEST[4] scores were similar even when the ^{15}N peptide was 5× less abundant than the ^{14}N peptide. The system described will be useful for detailed quantitative proteomic analysis of a variety of biological systems. Essentially, any organism that can be grown under conditions where the nitrogen content of a cell can be controlled can be studied via metabolic labeling and MudPIT. This includes not only microorganisms such as *S. cerevisiae* [5]and *E. coli*[6] but also mammalian tissue culture systems.[7]

ACKNOWLEDGMENT

The authors thank Dr. Guy Oshiro of the Genomics Institute of the Novartis Research Foundation for assisting in growing *S. cerevisiae* strain S288C in the media described.

Received for review November 27, 2001. Accepted January 30, 2002.

AC015704L