Anal. Chem. **2005,** *77,* 6218−6224

Correlation of Relative Abundance Ratios Derived from Peptide Ion Chromatograms and Spectrum Counting for Quantitative Proteomic Analysis Using Stable Isotope Labeling

Boris Zybailov, Michael K. Coleman, Laurence Florens, and Michael P. Washburn*

Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, Missouri 64110

In this study, *S. cerevisiae* crude membrane fractions were prepared using the acid-labile detergent RapiGest from cells grown under rich and minimal media conditions using ^{14}N and ^{15}N ammonium sulfate as the sole nitrogen source. Four independent MudPIT analyses of 1:1 mixtures of sample were prepared and analyzed via quantitative multidimensional protein identification technology on a two-dimensional ion trap mass spectrometer. Using the method described in this study, low-abundance integral membrane proteins with up to 14 transmembrane domains were identified and their protein expression determined when sufficient spectrum counting and ion chromatogram information was generated. We demonstrate that spectrum counting and mass spectrometry derived ion chromatograms strongly correlate for determining quantitative changes in protein expression. Spectrum counting proved more reproducible and has a wider dynamic range contributing to the deviation of the two quantitative approaches from a perfect positive correlation.

Stable isotope labeling is widely used for quantitative proteomics analysis.[1] Organisms as diverse as *Saccharomyces cerevisiae*,[2,3] *Drosophila melanogaster*,[4] *Caenorhabditis elegans*,[4] and *Rattus norvegicus*[5] can be metabolically labeled using stable isotopes such as ^{15}N. In tissue culture, stable isotope labeling is achieved using ^{15}N or ^{13}C amino acids in an approach named stable isotope labeling by amino acids in cell culture (SILAC).[6,7] In these approaches, two cell cultures are grown under "heavy" and "light" media conditions and one condition is subjected to a stimulus. Protein extracts from both conditions are mixed 1:1 and analyzed with liquid chromatography and tandem mass spectrometry (MS/MS). In effect, each sample is an internal standard for the other, in a way that is most analogous to the approach used for oligonucleotide or cDNA array analyses.[8] Both full metabolic labeling and SILAC are reproducible methods for quantitative proteomic analysis[9,10] and, therefore, generate biologically relevant and informative data sets when properly applied.[11,12]

The interrogation of mass spectrometry peptide data sets for quantitative information takes many forms, ranging from the use of MS scans alone to MS/MS scans alone. At one end of the spectrum, MS scans alone are used in conjunction with liquid chromatography to determine changes in protein or peptide expression;[13−15] however, without the use of MS/MS scans, peptides and proteins in a complex mixture will not be identified. An intermediate, more commonly used approach is to carry out analyses with cycling MS and MS/MS events, to extract ion chromatograms from the full MS scans for relative abundance measurements, and to search MS/MS data sets for peptide and protein identification. This approach is commonly used with stable isotopes and metabolic labeling,[3,5−7,10,12,16] and has been proposed as a possible label-free approach where samples are run independently and compared.[17,18]

* To whom correspondence should be addressed. Phone: 816-926-4457. Fax: 816-926-4694. E-mail: mpw@stowers-institute.org.

(1) Julka, S.; Regnier, F. *J. Proteome Res.* **2004**, *3*, 350−363.
(2) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 6591−6596.
(3) Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R., 3rd. *Anal. Chem.* **2002**, *74*, 1650−1657.
(4) Krijgsveld, J.; Ketting, R. F.; Mahmoudi, T.; Johansen, J.; Artal-Sanz, M.; Verrijzer, C. P.; Plasterk, R. H.; Heck, A. J. *Nat. Biotechnol.* **2003**, *21*, 927−931.
(5) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Matthews, D. E.; Yates, J. R., 3rd. *Anal. Chem.* **2004**, *76*, 4951−4959.
(6) Blagoev, B.; Ong, S. E.; Kratchmarova, I.; Mann, M. *Nat. Biotechnol.* **2004**, *22*, 1139−1145.
(7) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol. Cell. Proteomics* **2002**, *1*, 376−386.
(8) Quackenbush, J. *Nat. Genet.* **2002**, *32* (Suppl.), 496−501.
(9) Washburn, M. P.; Ulaszek, R. R.; Yates, J. R., 3rd. *Anal. Chem.* **2003**, *75*, 5054−5061.
(10) Molina, H.; Parmigiani, G.; Pandey, A. *Anal. Chem.* **2005**, *77*, 2739−2744.
(11) Andersen, J. S.; Lam, Y. W.; Leung, A. K.; Ong, S. E.; Lyon, C. E.; Lamond, A. I.; Mann, M. *Nature* **2005**, *433*, 77−83.
(12) Washburn, M. P.; Koller, A.; Oshiro, G.; Ulaszek, R. R.; Plouffe, D.; Deciu, C.; Winzeler, E.; Yates, J. R., 3rd. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3107−3112.
(13) Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. *Anal. Chem.* **2003**, *75*, 4818−4826.
(14) Wiener, M. C.; Sachs, J. R.; Deyanova, E. G.; Yates, N. A. *Anal. Chem.* **2004**, *76*, 6085−6096.
(15) Silva, J. C.; Denny, R.; Dorschel, C. A.; Gorenstein, M.; Kass, I. J.; Li, G. Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P.; Geromanos, S. *Anal. Chem.* **2005**, *77*, 2187−2200.
(16) MacCoss, M. J.; Wu, C. C.; Liu, H.; Sadygov, R.; Yates, J. R., 3rd. *Anal. Chem.* **2003**, *75*, 6912−6921.
(17) Chelius, D.; Zhang, T.; Wang, G.; Shen, R. F. *Anal. Chem.* **2003**, *75*, 6658−6665.
(18) Radulovic, D.; Jelveh, S.; Ryu, S.; Hamilton, T. G.; Foss, E.; Mao, Y.; Emili, A. *Mol. Cell. Proteomics* **2004**, *3*, 984−997.

10.1021/ac050846r CCC: $30.25
Published on Web 08/31/2005

Recently, the sampling of tandem mass spectra from complex peptide mixtures has been identified as another source for quantitative information. It has long been recognized that, in a multidimensional protein identification technology (MudPIT) analysis of a complex protein mixture, proteins of higher abundance are identified by more peptides than proteins of low abundance.[19] To determine the potential of this insight as a quantitative proteomics approach, Liu et al. demonstrated that the total number of peptide identifications in a complex peptide mixture analyzed by MudPIT linearly correlated with the abundance of the proteins in a quantitative approach called spectral counting.[20] Similarly, Allet et al. proposed summing peptide identification scores to yield semiquantitative expression values[21] and, in a rigorous analysis, demonstrated the quantitative capabilities of peptide match score summation.[22] Venable et al. recently demonstrated that, in ion trap mass spectrometry, quantitative analyses can be carried out directly from tandem mass spectra in a data-independent all MS/MS approach[23] giving additional conformation to the quantitative nature of peptide identification.

If spectral counting[20] and peptide ion chromatogram interrogation[3,16,24] independently generate reliable relative abundance ratios for quantitative proteomics, then the ratios generated from either approach should positively correlate. The objective of the current study was to determine the correlation between spectral counting and ion chromatograms using stable isotope labeling and to determine sources of deviation from a perfect positive correlation. In this study, *S. cerevisiae* was grown in ^{14}N-rich media and ^{15}N minimal media four independent times and each sample was independently analyzed via quantitative MudPIT[3,9,12] on a linear ion trap mass spectrometer. Furthermore, we analyzed crude *S. cerevisiae* membrane samples in this study and describe a method for the quantitative analysis of integral membrane proteins. We demonstrate a strong positive correlation between relative abundance ratios derived from spectral counting and from reconstructed peptide ion chromatograms and define factors responsible for deviation from a perfect positive correlation.

MATERIALS AND METHODS

Materials. DIFCO bacto peptone, dextrose, yeast extract, and yeast nitrogen base without amino acids or ammonium sulfate were acquired from BD Diagnostics (Sparks, MD). Ammonium-^{15}N sulfate (99 atom %) and ammonium-^{14}N sulfate (99.99 atom %) were products of Aldrich (Milwaukee, WI). Standard chemical reagents and dibasic sodium phosphate, dibasic potassium phosphate, sodium fluoride, potassium fluoride, and tricholoroacetic acid (TCA) were obtained from Sigma (St. Louis, MO). RapiGest[25] was purchased from Waters Corp. (Milford, MA). Tris(2-carboxyethyl) phosphine hydrochloride (TCEP) was obtained from Pierce (Rockford, IL). Endoproteinase LysC and trypsin, modified sequencing grade, were products of Roche Diagnostics Corp. (Indianapolis, IN). HPLC grade water was from EMD Chemicals, Inc. (Gibbstown, NJ). HPLC grade formic acid and acetonitrile was purchased from Mallinckrodt Baker, Inc. (Phillipsburg, NJ).

Growth and *S. cerevisiae* Sample Preparation. Cultures of *S. cerevisiae* strain S288C were grown to midlog phase (OD_{600} ~0.6) in YEPD/rich media (10 g of Bacto yeast extract, 20 g of Bacto peptone, and 20 g of dextrose/L) 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/L) at 30 °C. Cells were pelleted at 1000*g*, 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) with pelleting by centrifugation. Cells were lysed by grinding in liquid nitrogen in lysis buffer (0.1 M Tris, 0.1 M NaCl, 50 mM DTT, 0.1 M NaF, and 0.1 M KF). Unbroken cell material was pelleted at 300*g* followed by washing twice in lysis buffer with subsequent centrifugation at 300*g*. Pooled supernatants were centrifuged at 20000*g* on a Beckman Coulter Avanti J-20 XPI centrifuge (Fullerton, CA). The pellet was washed with 0.1 M Na_2CO_3 and incubated for 1 h on ice followed by repeated centrifugation at 20000*g*. The pellet was solubilized by the addition of 100 mM Tris-HCl, pH 8.5, 8 M urea, and 5% RapiGest[25] incubated for 30 min at 37 °C with agitation, and centrifuged at 20000*g* for 10 min. The supernatants were removed and protein content was determined by the BCA protein assay (Pierce). A 1 to 1 mixture of 300 μg of ^{14}N proteins (from cells grown in YEPD) and 300 μg of ^{15}N proteins (from cells grown in minimal media) was prepared. The 1:1 mix of ^{14}N and ^{15}N proteins was precipitated by the addition of TCA to 20%, incubated overnight at 4 °C, pelleted at 21000*g*, and washed twice with 500 μL of acetone. The final pellet was dried via a SPD111V speed vacuum system with RVT400 refrigerated vapor trap (Thermo Electron, Milford, MA).

The precipitated proteins were resuspended by adding 100 mM Tris-HCl, pH 8.5, 8 M urea, reduced by adding TCEP to 5 mM, incubated at room temperature for 30 min, and carboxymethylated by adding IAM to 10 mM and incubating at room temperature for 30 min in the dark. Endoproteinase LysC was added to an enzyme/substrate ratio of 1:100 and incubated overnight at 37 °C. Samples were then diluted to 2 M urea with 100 mM Tris-HCl, pH 8.5, brought to 2 mM $CaCl_2$, further digested with the addition of trypsin to an enzyme/substrate ratio of 1:100, and incubated at 37 °C overnight while shaking. On the next day, the digestion reaction was quenched with the addition of formic acid to 5%.

Quantitative MudPIT on an LTQ. Each sample was subjected to MudPIT analyses largely as described previously[19,26] but on a Finnigan LTQ mass spectrometer (Thermo Electron) interfaced with a quaternary Agilent 1100 quaternary pump with autosampler (Agilent Technologies, Palo Alto, CA). Peptide

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

(20) Liu, H.; Sadygov, R. G.; Yates, J. R., 3rd. *Anal. Chem.* **2004**, *76*, 4193−4201.

(21) Allet, N.; Barrillat, N.; Baussant, T.; Boiteau, C.; Botti, P.; Bougueleret, L.; Budin, N.; Canet, D.; Carraud, S.; Chiappe, D.; Christmann, N.; Colinge, J.; Cusin, I.; Dafflon, N.; Depresle, B.; Fasso, I.; Frauchiger, P.; Gaertner, H.; Gleizes, A.; Gonzalez-Couto, E.; Jeandenans, C.; Karmime, A.; Kowall, T.; Lagache, S.; Mahe, E.; Masselot, A.; Mattou, H.; Moniatte, M.; Niknejad, A.; Paolini, M.; Perret, F.; Pinaud, N.; Ranno, F.; Raimondi, S.; Reffas, S.; Regamey, P. O.; Rey, P. A.; Rodriguez-Tome, P.; Rose, K.; Rossellat, G.; Saudrais, C.; Schmidt, C.; Villain, M.; Zwahlen, C. *Proteomics* **2004**, *4*, 2333−2351.

(22) Colinge, J.; Chiappe, D.; Lagache, S.; Moniatte, M.; Bougueleret, L. *Anal. Chem.* **2005**, *77*, 596−606.

(23) Venable, J. D.; Dong, M. Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R. *Nat. Methods* **2004**, *1*, 39−45.

(24) Li, X. J.; Zhang, H.; Ranish, J. A.; Aebersold, R. *Anal. Chem.* **2003**, *75*, 6648−6657.

(25) Yu, Y. Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S.; Gebler, J. C. *Anal. Chem.* **2003**, *75*, 6023−6028.

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

mixtures were pressure-loaded onto a 250-μm-inner diameter (i.d.) fused-silica capillary packed first with 4 cm of 5-μm strong cation exchange material (Partisphere SCX, Whatman), followed by 2 cm of 5-μm C18 reversed-phase (RP) particles (Aqua, Phenomenex). Loaded and washed microcapillaries were connected via a 2-μm filtered union (UpChurch Scientific) to a 100-μm-i.d. column, which had been pulled to a 5-μm-i.d. tip using a P-2000 CO_2 laser puller (Sutter Instruments), then packed with 10 cm of RP particles, and equilibrated in 5% acetonitrile, 0.1% formic acid (buffer A). The split three-phase column[27] was placed in line with the Agilent 1100 and Finnigan LTQ.

Each full MS scan (from 400 to 1600 m/z) was followed by five MS/MS events using data-dependent acquisition where the first most intense ion from a given MS scan was subjected to CID followed by the second to fifth most intense ions. A fully automated 12-cycle chromatography controlled by the Xcalibur software was carried out on each sample using buffer A, 80% acetonitrile, 0.1% formic acid (buffer B), and 500 mM ammonium acetate, 5% acetonitrile, 0.1% formic acid (buffer C). The first cycle is a 20-min linear reversed-phase gradient to 100% buffer B. Cycles 2−12 consist of 3 min of buffer A, followed by 2 min of X% buffer C, followed by 5 min of washing with buffer A, followed by a 15-min linear gradient to 15% buffer B, followed by a 92-min linear gradient to 45% buffer B. Cycles 2−12 used 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100% buffer C, respectively.

Data Analysis. Data sets generated by MudPIT runs on LTQ mass spectrometer were analyzed following the flowchart described in Figure 1. RAW files were extracted into ms2 file format[28] using RAW_Xtract (obtained from J. Venable, Scripps Research Institute). This extraction procedure does not filter out any spectrum based on intensity or quality and does not distinguish between +2 and +3 precursor ions. Because SEQUEST cannot search for both isotope versions of peptides at the same time, ms2 files were searched twice against a database of 6299 *S. cerevisiae* protein sequences combined with 172 sequences for the usual contaminants. In the ^{15}N search, +1, +2, +3, and +4 amu were statically added to the average masses of amino acids [GASPVTCLIDEMFY], [NQKW], H, and R, respectively. In both ^{14}N and ^{15}N searches, cysteine residues were considered to be fully carboxymethylated (+57 Da added statically). To reconcile the fact that two sets of SEQUEST output files (sqt format)[28] are generated from one set of ms2 files, we developed a suite of algorithms to be able to (i) tell apart peptides detected in the ^{15}N search from ^{14}N peptides (sqt-isotope), (ii) merge the two sets of sqt files into one (sqt-merge); (iii) and count how many spectra were matched to ^{15}N and ^{14}N peptides (isotope-stats). The sqt-isotope script reads in sqt files and adds "1", "2", "3", or "4" after the amino acids in the peptide sequences where appropriate. For example, the peptide "R.GNPTVEVELTTEK.G" is converted to "R.G1N2P1T1V1E1V1E1L1T1T1E1K2.G", mimicking the way differential modifications are reported by SEQUEST. To complete the switch from "statically searched" to "differentially searched" sqt files, sqt-isotope also modifies the sqt file header to allow for correct display of the calculated ion series by the spectrum viewer.

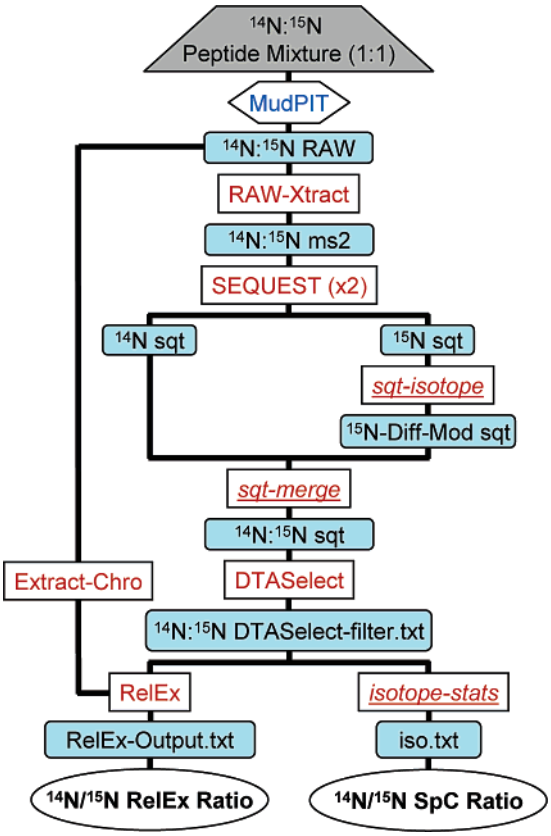

Figure 1. MS data analysis. This flowchart illustrates the processes and software involved in analyzing a complex peptide mixture to extract information on relative protein abundance using stable isotope labeling. Scripts developed for the current work are underlined.

The sqt-merge script reads in sqt files from normal (e.g., ^{14}N) and differentially modified (e.g., ^{15}N) searches generated from the same ms2 files and then merges and ranks the spectrum/peptide matches based on cross-correlation scores (XCorrs). DeltCn values are recalculated after this ranking event. This merging step is primordial as it allows only the best matches out of both (or n) differential SEQUEST queries to be ranked first ("sanity check"). The peptide information contained in the merged sqt files is converted into protein-level information using DTASelect.[29] The protein lists are established based on conservative filtering criteria as previously reported (DeltCn of at least 0.08, minimum XCorr of 1.8 for +1, 2.5 for +2, and 3.5 for +3 spectra and minimum peptide length of 7 amino acids). Protein lists from replicate MudPIT runs were compared and merged using Contrast.[29] The isotope-stats script reads in DTASelect-filter.txt (i.e., protein list with peptides) and reports for each protein the total number of spectra, the number of spectra matching ^{14}N and ^{15}N peptides, and the ratio of these (i.e., ^{14}N/^{15}N SpC), in a tab-delimited flat text file (iso.txt). The same DTASelect-filter.txt file was submitted to RelEx as described in MacCoss et al.[16] to generate the relative abundance ratios of proteins based on reconstructed ion chromatograms (i.e., ^{14}N/^{15}N RelEx Ratio).

RESULTS AND DISCUSSION

Signal-to-Noise Ratios. One critical choice when using quantitative proteomics software that measures peptide ion chro-

(27) McDonald, W. H.; Ohi, R.; Miyamoto, D. T.; Mitchison, T. J.; Yates, J. R. *Int. J. Mass Spectrom.* **2002**, *219*, 245−251.

(28) McDonald, W. H.; Tabb, D. L.; Sadygov, R. G.; MacCoss, M. J.; Venable, J.; Graumann, J.; Johnson, J. R.; Cociorva, D.; Yates, J. R., 3rd. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2162−2168.

(29) Tabb, D. L.; McDonald, W. H.; Yates, J. R., 3rd. *J. Proteome Res.* **2002**, *1*, 21−26.

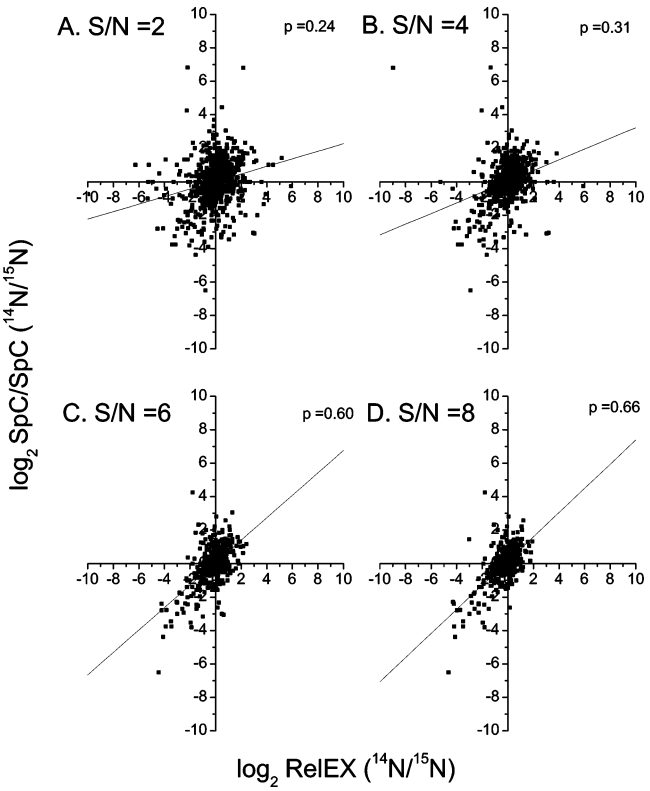

Figure 2. Signal-to-noise ratio impact on correlation. A crude membrane extract containing equal amounts of proteins from *S. cerevisiae* grown in ^{14}N-rich media and ^{15}N minimal media was analyzed by multidimensional protein identification technology on a linear ion trap mass spectrometer. After SEQUEST identification of peptides, DTASelect assembly of proteins, RelEx peptide quantification, and protein spectral count determination, the impact of the signal-to-noise ratio used for RelEx peptide quantification on the correlation of protein expression ratios from RelEx and spectral counts was compared by generating the $\log_2$ of the ratios for S/N cutoffs of 2 (A), 4 (B), 6 (C), and 8 (D). The Pearson product moment correlation coefficients between RelEx and spectral counting ratios are reported for each plot. The RelEx ratios were for proteins that had at least four quantifiable peptides using a given S/N.

matograms, such as RelEx,[16] is to determine what signal-to-noise ratio to use before generating a final list of proteins with corresponding protein expression ratios. We analyzed the correlation between relative abundance ratios derived from spectral counts (SpC) and ratios derived from reconstructed ion chromatograms (RelEx)[16] using signal-to-noise ratios of 2, 4, 6, and 8 for each of the four replicate MudPIT analyses carried out in this study. The scatterplots for one representative data set are shown in Figure 2 where the ratios are reported in $\log_2$ base, which is the standard approach for visualizing cDNA and oligonucleotide array data sets and makes the entire data set linear.[8] As the signal-to-noise ratio increased, fewer proteins were represented by 4 or more quantifiable peptides ranging from 804 using an S/N of 2 to 360 proteins using an S/N of 8 as cutoff. Clearly, the highest S/N cutoff gave the best Pearson product moment correlation (Figure 2D), and this trend was consistent for all four data sets (data not shown). RelEx[16] values reported in Supporting Information Table 1 were therefore filtered with a signal-to-noise ratio of 8.

Replicates. We then analyzed the correlation coefficients for each of the three additional experiments separately (data not shown) and all experiments combined (Figure 3). By merging four replicate experiments, we increased the number of total proteins that had at least four quantifiable peptides with an S/N of 8 to a total of 645 (Supporting Information Table 1). The total SpC for these 645 proteins ranged from 10 605 for glyceraldehyde-3-phosphate dehydrogenase 3 to 7 for Yhr032w, a poorly characterized locus that contains 10 predicted transmembrane domains. Both proteins had ^{14}N/^{15}N SpC ratios of 0.8 that were in close agreement to their RelEx ratios (Supporting Information Table 1). The correlation of the $\log_2$ RelEx ratios and spectral count ratios is shown in Figure 3 and had a strongly positive Pearson product moment correlation of 0.64. This correlation is even stronger when only the proteins with an arbitrary change of 1.5-fold or more are considered. A total of 319 proteins were 1.5-fold or greater overexpressed in minimal or rich media, and when only these proteins were considered for correlation analysis, a stronger positive Pearson correlation of 0.70 was obtained (data not shown).

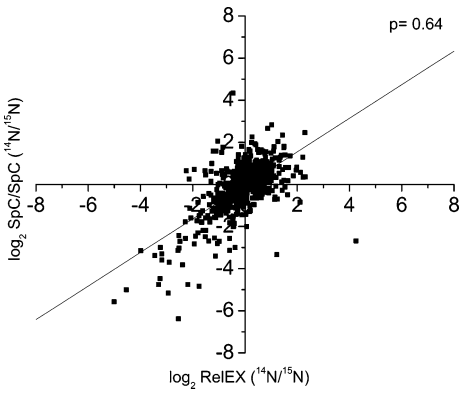

Figure 3. Correlation of log2 RelEx ratios to log2 spectrum counting ratios for replicate and merged data sets. Four crude membrane extracts containing equal amounts of proteins from four independent *S. cerevisiae* growths in ^{14}N-rich media and ^{15}N minimal media were analyzed by multidimensional protein identification technology on a linear ion trap mass spectrometer. After SEQUEST identification of peptides, DTASelect assembly of proteins, RelEx peptide quantification using a signal-to-noise ratio of 8, and protein SpC determination, the Pearson product moment correlation for the full data set of 645 proteins was determined to be 0.64. The RelEx protein ratios were for proteins with at least four quantifiable peptides with an S/N of 8.

An important question to consider is why does the Pearson product correlation between peptide ion chromatogram ratios and spectrum counting ratios deviate from a perfect positive correlation of 1.0? There are potential factors that contribute to this. First, RelEx[16] is a peptide or locally based quantitative approach whereas spectrum counting is a protein or global based quantitative approach. Next, large standard deviations can occur using ion chromatograms from statistical outliers or when a peptide is posttranslationally modified in one growth condition and not another. Last, ion chromatogram ratios and spectrum sampling have different dynamic ranges.

Outliers. Examples of the variation from ion chromatograms and its impact on the correlation of ratios is shown in Table 1 for Arg4p, Leu1p, and Phm2p. Arg4p, a protein involved in arginine biosynthesis and known to be overexpressed in cells grown in minimal media, had an average peptide ion chromatogram ratio of 2.31 ± 6.41 and a total ^{14}NSpC/^{15}NSpC of 0.099 (Supporting Information Table 1). The substantial error in the peptide ion chromatogram derived from the ratio of the peptide R.ETHHIS-

Table 1. Reproducibility of Spectral Counts and Peptide Ratios across Four Experimental Samples for Overexpressed Proteins

		experiment 1			experiment 2			experiment 3			experiment 4		
protein name	peptide	RelEx ratio	^{14}N SpC[a]	^{15}N SpC[a]	RelEx ratio	^{14}N SpC[a]	^{15}N SpC[a]	RelEx ratio	^{14}N SpC[a]	^{15}N SpC[a]	RelEx ratio	^{14}N SpC[a]	^{15}N SpC[a]
Arg4	total spectral counts[b]		2	27		1	10		3	19		2	25
	R.ETHHISGECVATAER.L			2	0.141		4	0.171		1	19.41		3
	R.HPNDEDIHTANERR.L	0.009		1						2			1
	R.RLGELIGR.D	0.127		1						3			
	L.TGFLMSLKGIPSTYDK.D	0.139		2									
	K.SGRVFGDLTGFLMSLK.G	0.103		2									
	K.SGRVFGDLTGFLMSLKGIPSTYDK.D	0.152	1										
	K.FVRHPNDEDIHTANER.R							0.402	2	2			
Leu1	total spectral counts[b]		2	22		0	45		0	28		2	18
	R.RVDCTLATVDHNIPTESR.K	0.108		2	0.113		5			5	0.065		3
	R.RVDCTLATVDHNIPTESR.K	0.06	2	1						2			
	R.HLVHEVTSPQAFEGLENAGR.K	0.081		9	1.051		6	0.078		2	0.108	1	2
	R.FRKDDQGKDQETDFVLNVEPWR.E			2	0.286		6						
	K.DDQGKDQETDFVLNVEPWR.E						3	0.013		1			
	K.HCLVNGLDDIGITLQKEEYISR.Y			2			3	0.379		4			
	G.GSKLLKFDNVPKRK.A										1.398		1
Phm2	total spectral counts[b]		64	26		59	14		85	22		75	19
	R.SSVDSWSERNESDFVEALDK.E	4.729	5	5		2	2		2			1	
	R.SSVDSWSERNESDFVEALDK.E	4.986	3										
	K.RTFTENTETGNSSFEEIR.L	4.033	5	2					5			3	2
	K.LSHFSNLEDASFK.S	2.326	2	2	3.028	2		2.375	2	2	2.142	5	2
	K.LSHFSNLEDASFK.S	3.237	2	1	3.353	2		2.131	2	2			
	K.LSHFSNLEDASFK.S	3.92	1										
	R.NPENWHRDDIDSNIPNPLR.F	1.801	2			1		3.817	2		3.458	2	4
	R.NPENWHRDDIDSNIPNPLR.F												
	K.FINNFIFKNDPSYK.N	3.132	5			4					4.009		1
	K.SIHDKLDNMR.R	2.961	4	2		1		2.446	3				
	K.SIHDKLDNMR.R	3.771		3		2							
	K.NYLINQLR.E		1		3.051	1					0.018	2	
	K.SNSLSSDGNSNQDVEIGK.S					2		0.02	5	5		4	
	R.KLDDLEENTK.S					2		0.121	6			2	1
	K.YVNILPFWLPDLETDIR.K		2	2		4	1	2.618	4	2		5	3
	R.ISYLYEFLR.S		2					3.662	4	2	3.428	4	1
	R.ISYLYEFLR.S							3.092	6	3		3	
	K.LHPNYPSVK.S							2.046	2			2	
	R.HLPALVYASVPNENDDFVDNLESDVR.V		2	3		6	5				3.732	7	

[a] ^{14}N SpC and ^{15}N SpC are the spectral counts in any given experiment for peptides from proteins in minimal media (^{15}N) or rich media (^{14}N). [b] The total spectral counts for each protein is greater than the sums of peptides in the table because only peptides with a S/N of 8 for each protein are listed and not all peptides identified were quantifiable by RelEX.

GECVATAER.L that had a value of 19.41 in the fourth experiment, but this same peptide had a ratio of 0.141 and 0.171 in experiments 2 and 3 (Table 1). Leu1p is involved in leucine biosynthesis, and it had an average peptide ion chromatogram ratio of 0.296 ± 0.43 and a total ^{14}NSpC/^{15}NSpC ratio of 0.035 (Supporting Information Table 1). The large standard deviation is a result of the ratios of one peptide from experiment 2 and one from experiment 4 (Table 1). For both Arg4p and Leu1p, 3 (for Arg4p in experiment 3) or fewer spectral counts were obtained in any of the four experiments from cells grown in rich media, but 10−45 spectral counts were obtained from cells grown in minimal media (Table 1). Furthermore, in every experiment, the ^{14}NSpC/^{15}NSpC ratios were 0.16 or smaller for Arg4p and Leu1p, indicating consistent overexpression of these proteins in cells grown in ^{15}N minimal media.

An example of a protein consistently overexpressed in rich media is the subunit of the vacuolar transporter chaperone complex, Phm2p. The averaged peptide ion chromatogram ratio for this protein from all four experiments was 2.813 ± 1.171 and a total ^{14}NSpC/^{15}NSpC ratio of 3.494 (Supporting Information Table 1). The ^{14}NSpC/^{15}NSpC ratio from each experiment was 2.46, 4.2, 3.86, and 3.95, demonstrating the consistency of overexpression of this locus in all four experiments. The peptide ion chromatogram ratio variation largely came from two peptides in experiment 3 with ratios of 0.02 and 0.121 and one peptide in experiment fo4ur with a ratio of 0.018 (Table 1). The overwhelming number of peptides correctly quantitatively analyzed via RelEx had ratios of 3−4-fold overexpression demonstrate that the peptides with ratios of 0.02 and 0.121 are likely statistical outliers. One cause for deviation in correlations between peptide ion chromatograms and SpC ratios is the variation in peptide ion chromatogram ratios that occurred in any given experiment or between experiments. We are currently investigating the use of an outliers test to alleviate this source of error, but a blanket use of an outliers test is biologically risky because a consistent outlier may indicate a posttranslational modification change in a protein from one growth condition to the next (data not shown).

Dynamic Range. Another primary source for the deviation of peptide ion chromatograms and spectral counting from a perfect positive correlation is the dynamic range of each approach. The large dynamic range of spectrum counting[20] that exceeds that of peptide ion chromatograms has been previously described using *S. cerevisiae* extracts from artificial ratios,[23] but not for a biological

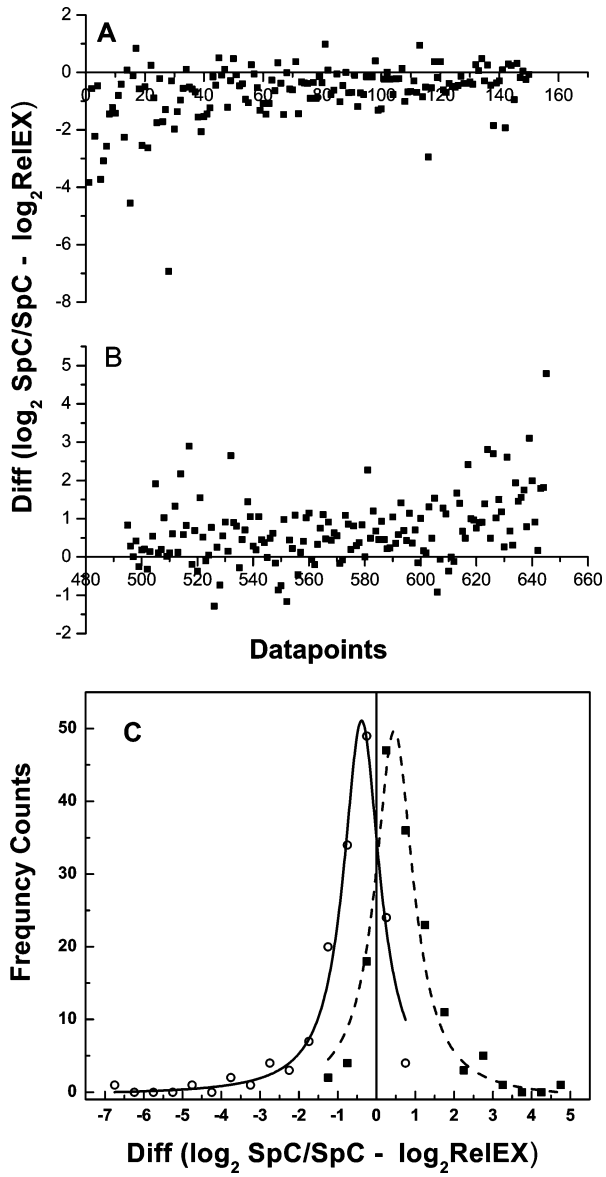

Figure 4. Difference in $\log_2$ RelEx ratios and $\log_2$ spectrum counting ratios for proteins overexpressed in either growth condition. The difference in $\log_2$ values were determined by subtracting the $\log_2$ RelEx ratios from the $\log_2$ SpC ratios. The 645 proteins with at least four peptides with a signal-to-noise ratio of 8 are sorted and ordered for greatest overexpression in ^{14}N minimal media to greatest overexpression in ^{15}N-rich media based on the SpC ratios. (A) The difference for the 150 proteins from greatest overexpression in ^{15}N minimal media to greatest overexpression in ^{14}N-rich media is shown. (B) The difference for the 150 proteins (645th data point down to 495th data point) from greatest overexpression in ^{14}N-rich media to greatest overexpression in ^{15}N minimal media is shown. (C) Frequency counts for data points used in (A), solid line and open circles, and (B), dashed line and closed squares, were plotted as a function of the difference in log2 ratios (bin size 0.5) and the distribution was fitted using a Lorentz function.

sample. In Figure 4, the difference in $\log_2$ ratios is shown for the 150 proteins with the greatest overexpression in minimal media (Figure 4A) and for the 150 proteins with the greatest overexpression in rich media (Figure 4B) when the data set is sorted from the most overexpressed in minimal media to the most overexpressed in rich media as determined by ^{14}NSpC/^{15}NSpC ratios. In Figure 4A, the majority of data points lie below the 0 line, indicating the $\log_2$ of SpC ratios was consistently more negative than the $\log_2$ of the peptide ion chromatogram ratios where a more negative value indicates a greater degree of overexpression in minimal media. For example, L-*myo*-inositol-1-phosphate synthase, Ino1p, was quantified by 13 total peptides with an S/N of 8 and had a peptide/ion chromatogram ratio of expression in rich to minimal of 0.171 ± 0.185 whereas Ino1p had a total SpC of 1 in rich media and 82 in minimal media for a ratio of 0.012 (Supporting Information Table 1). The difference in $\log_2$ values for these data points is -3.8 because SpC has a wider dynamic range. This same trend is seen in Figure 4B for the 150 proteins with the greatest overexpression in rich media where the SpC values were consistently more positive. In a combined approach, the frequency distribution of differences in $\log_2$ is shown in Figure 4C. The line defined by open circles represents the 150 proteins most overexpressed in minimal media, and the closed squares represent the 150 proteins most overexpressed in rich media. In both cases, the SpC ratios demonstrated larger differences in ratios with more negative diff log2 values for proteins in minimal media and more positive diff log2 values for proteins in rich media (Figure 4C.) Therefore, the larger dynamic range of spectral counting is a significant contributor to the deviation of the ^{14}N/^{15}N SpC and RelEx values from a perfect positive correlation with ratios derived from peptide ion chromatograms.

Membrane Proteins. A final important feature of the current research described is the ability to quantitatively analyze integral membrane proteins. As described in the Materials and Methods section, we used the acid-labile surfactant RapiGest[25] to extract proteins from a crude membrane preparation of *S.* cerevisiae grown on ^{14}N-rich media and ^{15}N minimal media. A total of 79 of the 645 proteins whose protein expression was determined in this body of work contained 3 or more predicted transmembrane domains, and 50 of these 79 contained 7 or more predicted transmembrane domains (Supporting Information Table 1). Included in this data set was the detection, identification, and protein expression determination of predicted low-abundance integral membrane proteins based on each protein's codon adaptation index value.[30] Shown in Table 2 are the protein expression ratios for 16 low-abundance integral membrane proteins with 5−14 predicted transmembrane domains. In this data set, the *myo*-inositol transporter, Itr1p, was greatly overexpressed in cells grown in minimal media compared to cells grown in rich media as shown by the combined peptide ion chromatogram ratio of 0.101 and the ^{14}NSpC/^{15}NSpC ratio of 0.037 (Table 2) where the ^{14}NSpC/^{15}NSpC for each of the four experiments was 1/15, 0/20, 1/23, and 1/23, respectively.

CONCLUSIONS

The acquisition scheme employed in the current body of work contained a single MS scan followed by five MS/MS scans. Effectively, 1/6th of the MudPIT analysis was used to acquire the MS scans needed for deriving peptide ion chromatograms for the powerful RelEx software platform [16] to determine quantitative changes in peptides. This likely negatively influenced the number of peptides from a given run with an S/N of 8 because better peptide ion chromatograms should be obtained when more

(30) Coghlan, A.; Wolfe, K. H. *Yeast* **2000**, *16*, 1131−1145.

Table 2. Quantitative Proteomic Analysis of Predicted Low-Abundance Integral Membrane Proteins[a]

description	CAI	TM	RelEx ratio (YPD/MIN)	RelEx Peps	unique Peps	total Sq Cv	SpC YPD	SpC MIN	SpC ratio (YPD/MIN)
metal resistance protein,Ycf1p	0.149	14	0.745 ± 0.388	10	36	16.2	51	67	0.761
bile acid transporter of ABC family; Ybt1p	0.187	14	1.611 ± 0.668	6	34	25.3	50	23	2.174
Ynl321wp	0.145	13	1.684 ± 0.929	10	18	20.2	35	26	1.346
probable multidrug resistance transporter; Pdr15p	0.183	12	1.802 ± 0.874	20	49	17.1	107	51	2.098
myo-inositol transporter; Itr1p	0.195	12	0.101 ± 0.089	10	25	39.2	3	81	0.037
ABC transporter; Snq2p	0.18	12	1.306 ± 1.327	9	36	26.2	42	30	1.4
Putative ABC transporter; Pdr10p	0.128	11	2.933 ± 2.555	6	22	10.5	42	24	1.75
endosomal membrane protein; Emp70p	0.189	9	1.213 ± 0.457	11	20	26.1	41	29	1.414
Phosphatidyl-ethanolamine *N*-methyltransferase; Cho2p	0.17	8	0.645 ± 0.372	11	36	29.9	66	116	0.569
Probable transmembrane protein; Ist2p	0.176	8	1.608 ± 0.546	17	32	27.5	47	39	1.205
HMG-CoA reductase isozyme; Hmg1p	0.195	7	1.86 ± 0.993	20	51	32.0	82	93	0.882
Active transport ATPase; Adp1p	0.14	7	1.022 ±0.382	6	23	26.8	33	29	1.138
plasma membrane protein; Syg1p	0.147	5	1.877 ± 0.523	4	15	16.9	19	25	0.76
Ynl115cp	0.147	5	0.664 ± 0.622	7	21	23.1	24	16	1.5
Ybr287wp	0.188	5	0.609 ± 0.546	5	14	39.8	17	19	0.895
involved in cobalt accumulation; Cot1p	0.138	5	0.803 ± 0.304	7	13	28.5	12	16	0.75

[a] CAI is Codon Adaptation Index value, TM is number of transmembrane domains, the number of peptides used to determine the RelEx ratio is listed in the RelEx peps (peptides) column, the number of unique peptides (peps) that identified a protein is in the Unique peps column, the corresponding total sequence coverage is in the Total Sq Cv column, and SpC stands for spectral count.

sampling of data points across an eluting peak are acquired. On the other hand, 5/6th of the MudPIT analysis was used to acquire MS/MS, which is used both for identification and for quantitative analysis of proteins using SpC. It is likely, based on this time commitment distribution, that the spectral count information is more reliable than the RelEx information.[16] However, as demonstrated throughout this body of work, both approaches validated each other even given the differences in dynamic range and variation of averaged peptide ion chromatogram information.

Given that RelEx[16] is a local approach determining peptide expression changes for a given protein and SpC is a global approach for changes in a given protein, they provide different values. The larger dynamic range for spectrum sampling is a clear advantage, and more data points are included since each identification is used to determine the expression changes. The advantage of RelEx[16] is that it detects changes at the peptide level, and it is possible and likely that many of the peptides with aberrant ratios when compared to other peptides from a protein are not always simple outliers. These peptides are clear candidates for posttranslational modification changes between cellular stimuli, which is a largely unexplored area in not only yeast biology but also higher organisms.

ACKNOWLEDGMENT

The authors thank Marjorie Fournier, Amber Mosley, Andrew Paoletti, and Selene Swanson for valuable discussions during the course of this analysis.

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review May 16, 2005. Accepted August 2, 2005.

AC050846R