Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra

John D Venable[1], Meng-Qiu Dong[1], James Wohlschlegel[1], Andrew Dillin[2] & John R Yates III[1]

To take advantage of the potential quantitative benefits offered by tandem mass spectrometry, we have modified the method in which tandem mass spectrum data are acquired in 'shotgun' proteomic analyses. The proposed method is not data dependent and is based on the sequential isolation and fragmentation of precursor windows (of 10 *m/z*) within the ion trap until a desired mass range has been covered. We compared the quantitative figures of merit for this method to those for existing strategies by performing an analysis of the soluble fraction of whole-cell lysates from yeast metabolically labeled *in vivo* with ^{15}N. To automate this analysis, we modified software (RelEx) previously written in the Yates lab to generate chromatograms directly from tandem mass spectra. These chromatograms showed improvements in signal-to-noise ratio of approximately three- to fivefold over corresponding chromatograms generated from mass spectrometry scans. In addition, to demonstrate the utility of the data-independent acquisition strategy coupled with chromatogram reconstruction from tandem mass spectra, we measured protein expression levels in two developmental stages of *Caenorhabditis elegans*.

A major hurdle in comparative proteomics is the identification and subsequent quantification of proteins and their expression levels in complex biological systems. Because of the inherent need for speed and sensitivity in these analyses, mass spectrometry (MS)-based methods have excelled. In general, quantification using MS is achieved by comparing an unlabeled or 'light' peptide (comprised of naturally abundant stable isotopes) to an internal standard that is chemically identical with the exception of atoms that are enriched with a stable 'heavy' isotope. Typically, the relative ion-intensity abundances of the light and heavy molecules are compared; however, absolute quantification can be obtained through the introduction of a known amount of internal standard (**Fig. 1a**).

Most published accounts of quantitative proteomic studies have involved data-dependent acquisition of tandem mass spectra for peptide identification and MS scans for the subsequent quantification of light and heavy peptides[1–6]. However, for several reasons, it is desirable to perform quantification from tandem mass spectra rather than MS scans. First, most mass spectrometers (especially ion traps) operating in conjunction with liquid chromatography produce MS scans permeated by chemical noise. This can be particularly problematic at low *m/z* values, where solvent-derived adducts are plentiful and can significantly limit the signal-to-noise ratio of resulting chromatograms, which effectively decreases the dynamic range available for comparison. It is well established that tandem mass spectrometry (MS/MS) experiments offer extraordinarily high sensitivity because of the ability to accumulate and trap precursor ions and that tandem mass spectra, in general, are afflicted with less chemical noise than MS scans[7–9]. In fact, it is not uncommon for peptides to be easily detected in the MS/MS mode but below detection in MS mode. Second, because of the inherent complexity of shotgun proteomic analyses (that is, the large numbers of peptide ions and charge states), it is desirable to obtain the increased specificity associated with tandem mass spectrum transitions instead of relying on the MS-mode resolution to effectively remove noise components (such as isobaric peptide signals)[9]. Finally, the presence of multiple fragment ions has potential benefits for quantitative analysis. For example, the ion intensities from several transitions can be summed to produce signal-to-noise enhancements[10] or averaged to obtain more accurate measurements[7].

Because of the limitations in sensitivity and dynamic range of existing procedures, we have explored an alternative data acquisition strategy that is not data dependent and permits quantitative analysis directly from tandem mass spectra. This method is based on the sequential isolation and fragmentation of relatively large (10-*m/z*) precursor windows within the ion trap until a desired mass range has been covered. Subsequent identification using SEQUEST and automated ion chromatogram reconstruction from fragment-ion intensities in tandem mass spectra (RelEx) provided several quantitative benefits over quantification directly from MS scans, including increased signal-to-noise ratio, sensitivity, selectivity and dynamic range.

RESULTS

Data-independent scan sequence

The proposed data-independent scan sequence (**Fig. 1b**) is based on the isolation and subsequent fragmentation of successive windows of 10 *m/z* throughout the mass range. Temporal characteristics of a theoretical chromatogram obtained from fragment-ion intensities are also shown (**Fig. 1c**).

[1]Department of Cell Biology, The Scripps Research Institute, La Jolla, California 92014, USA. [2]The Salk Institute for Biological Studies, P.O. Box 85800, San Diego, California 92186-5800, USA. Correspondence should be addressed to J.R.Y. (jyates@scripps.edu).

PUBLISHED ONLINE 29 SEPTEMBER 2004; DOI:10.1038/NMETH705

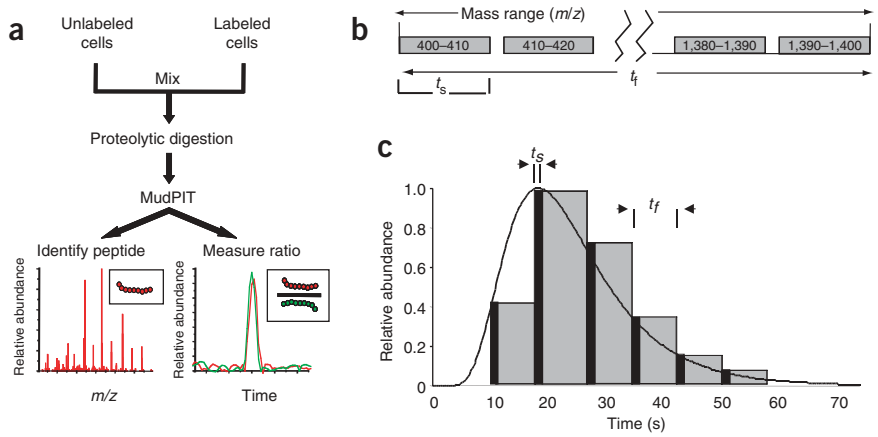

Figure 1 | Quantification using MS/MS. (**a**) Overview of the experimental approach for quantitative MS using MudPIT and isotopic labeling. (**b**) Data-independent scan sequence that is based on the isolation and subsequent fragmentation of successive windows throughout the mass range. (**c**) Theoretical chromatogram constructed from fragment-ion intensity, where t_s is the time required to complete one tandem mass spectrum and t_f is the time required to complete one cycle of the entire scan sequence.

Considerations for data-independent acquisition

Sampling rate and size of mass range to be interrogated. With data-independent acquisition, the scan rate of the instrument used coupled with the width of the isolation window effectively determines the maximum mass range that can be interrogated while maintaining a scan rate (t_f) that allows for repetitive sampling across the chromatographic peak. For example, a ThermoElectron LTQ in normal scan-speed mode can acquire a single tandem mass spectrum (1 'microscan') approximately every 0.25–0.3 s, which can be extrapolated to ~40 scans/10 s. If each tandem mass spectrum were acquired from the isolation of a 15-*m*/*z* window, a mass range of ~600 *m*/*z* could be covered every 10 s. Also, increased scan speeds ('turbo scans') can be used to increase the sampling rate, although at the expense of some degree of spectral quality.

Impact of isolation width on spectral quality. One of the primary concerns in using a relatively large isolation window is the effect on the overall quality of tandem mass spectra. Spectra derived from larger isolation widths typically have higher background owing to chemical noise (see **Supplementary Fig. 1** online). However, the overall impact of this increased noise level on the peptide identification process does not seem to be very significant. A further concern about acquiring tandem mass spectra from a relatively large isolation window is that fragmentation will vary depending on the size of the window or the relative position of the precursor ion within the window. To measure the impact of isolation window width and the precursor-ion position on fragmentation, we collected 200 replicate tandem mass spectra for a peptide standard (angiotensin I) over a range of precursor-ion positions relative to the center of an isolation window of 25 *m*/*z*. The peptides were infused through an infusion pump at a concentration of 1 pmol/μl and a flow rate of 1 μl/min. The resulting tandem mass spectra were searched using SEQUEST, and the average XCorr scores were plotted as a function of position relative to the center of the mass windows (**Fig. 2**). Average XCorr measurements for replicate tandem mass spectra of angiotensin I were relatively unaffected by the size of the isolation windows studied and the positions of precursor ions within these windows.

Validation of approach: qualitative aspects

Because there are significant differences in both the number and the appearance of tandem mass spectra collected using data-dependent and data-independent acquisition, we explored the effects on the overall peptide identification process. Four replicate experiments analyzing ~10 μg of the soluble fraction from the tryptic digest of a yeast whole-cell lysate were carried out using each approach. For these experiments, data-dependent and data-independent acquisition were used as described in the **Supplementary Methods** online.

We obtained ~20% more spectra with the data-independent approach than with the data-dependent approach because of the lack of MS scans. The increase in number of spectra, however, did not translate into a significant increase in the number of peptide or protein identifications. On average, for the data-dependent acquisition strategy, 798 ± 200 peptides were identified, which corresponded to 214 ± 42 nonredundant protein identifications. The data-independent strategy identified an average of 773 ± 45 peptides that corresponded to 240 ± 22 proteins within the same mass range. A summary of peptide and corresponding protein identifications for these analyses is provided (**Table 1**). In total, 491 different proteins were identified when the identifications from the replicate experiments were merged. Of these, ~50% were identified by both strategies, ~28% were identified by the data-dependent strategy exclusively, and the remaining ~22% were identified only using the data-independent strategy.

All in all, the number of peptide identifications and subsequent protein identifications seemed to be comparable between the two

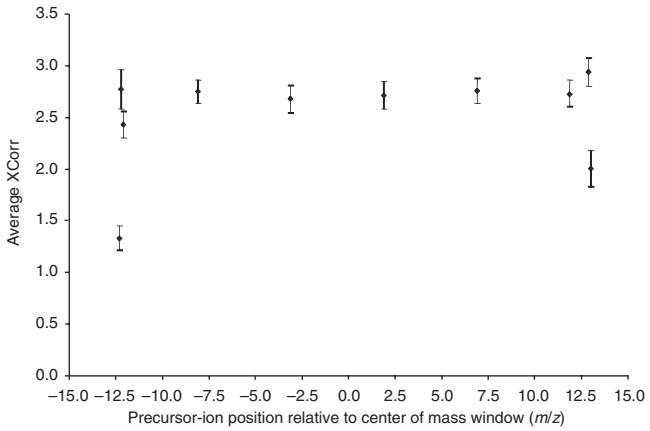

Figure 2 | The effect of precursor-ion position (relative to the center of the isolation window) on the average XCorr and standard deviation for a peptide standard, angiotensin I.

Table 1 | Qualitative summary of four replicate analyses of yeast whole-cell lysates using either data-dependent or data-independent MS/MS acquisition

Data-dependent acquisition

Rep.	MS/MS scans	Peptide identifications	Protein identifications (P2)
1	86,243	676	193
2	94,831	1,093	275
3	88,619	745	201
4	86,573	677	185

Data-independent acquisition

Rep.	MS/MS scans	Peptide identifications	Protein identifications (P2)
1	112,612	784	246
2	98,264	739	213
3	114,140	830	264
4	108,873	738	232

P2, proteins identified having ≥2 peptides per locus.

acquisition strategies. This result could be viewed as surprising, however, because the data-independent approach could potentially acquire tandem mass spectra for every peptide eluting in the mass window. Numerous possibilities exist that could account for the minimal difference in the number of identifications between acquisition strategies, but one of the most likely explanations is that the data-independent approach produces a significant number of multiplexed spectra (spectra derived from multiple precursors). For each of these spectra, we are only reporting one match, which results in fewer identified peptides than could be potentially be found. These tandem mass spectra require more refined approaches to database searching to extract the information contained, and we currently do not have the tools to extract this information reliably. We are developing methods to improve the analysis of multiplexed tandem mass spectra to extract more identification data. Nonetheless, it is clear that the larger isolation window used, and the corresponding greater background noise and potential for spectral convolution, did not reduce the information content of the qualitative analysis. A more thorough analysis of the qualitative differences between the two approaches would be interesting and is planned for future studies.

Validation of approach: quantitative aspects

To study the benefits offered by quantitative analysis from tandem mass spectra (in signal to noise, selectivity, dynamic range, accuracy, and precision of peptide ratio measurements), we carried out five replicate six-step multidimensional liquid chromatography–tandem mass spectrometry of peptide mixtures (MudPIT) experiments for two defined-ratio mixtures (1:1 and 10:1) of trypsin-digested unlabeled and ^{15}N-labeled yeast whole-cell lysates. For each experiment, ~10 μg of one of the mixtures was loaded onto a triphasic column and analyzed by MudPIT on a ThermoElectron LTQ mass spectrometer using data-independent acquisition. For each identified peptide, chromatograms were generated (by RelEx, as described in the **Supplementary Methods**) from both MS scans and tandem mass spectra.

Signal-to-noise enhancements. One of the most noticeable distinctions between extraction strategies was the difference in signal-to-noise ratio of reconstructed chromatograms. For example, we identified a +2 peptide ion (YGYQLYTSNPSGNYTGWK) with an XCorr of 3.6, and a precursor-ion average mass of 1,050.6. In the full scan data, the signal-to-noise ratio of this ion is ~2 (**Fig. 3**). However, in the tandem mass spectra generated from the isolation window (1,050–1,060 *m/z*), the signal-to-noise ratios of fragment ions Y12 (1,311.8) and Y13 (1,474.7) are significantly higher (>10).

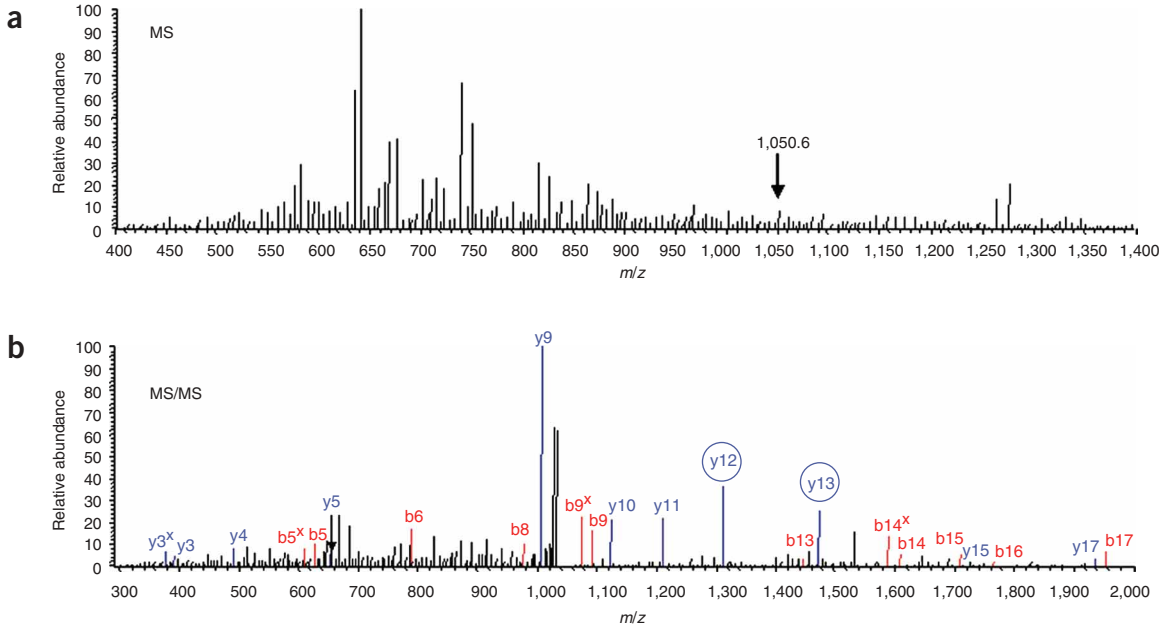

Figure 3 | Sample MS and MS/MS scans. (**a**,**b**) Screen snapshot from Xcalibur (ThermoElectron) showing an MS scan (**a**) taken immediately before a tandem mass spectrum (**b**) that was later identified, using SEQUEST, as a +2 peptide ion with sequence YGYQLYTSNPSGNYTGWK. This precursor ion has an $[M+H]^+$ of 1,050.6.

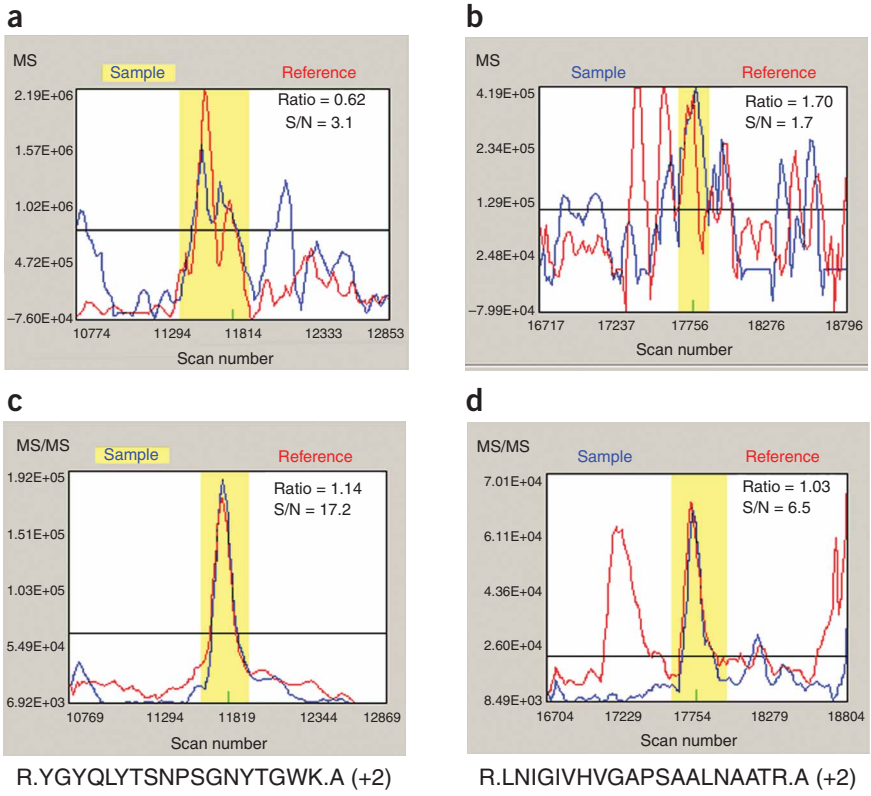

Figure 4 | RelEx chromatograms from MS and MS/MS scans. (**a**–**d**) RelEx screenshots of reconstructed chromatograms for two different peptides generated from both (**a**,**b**) MS and (**c**,**d**) tandem mass spectra.

Ultimately, the increased signal-to-noise ratio and selectivity offered by tandem mass spectra translates to chromatograms with greater signal to noise, as can be seen in the chromatograms generated by RelEx for two identified peptides (**Fig. 4**). In addition, signal-to-noise ratio enhancements can also be observed in the chromatograms used to quantify RPL19A in the 1:1 yeast whole-cell lysate standard mixture (**Supplementary Figs. 2–5**). **Supplementary Table 1** shows a more comprehensive view of the signal to noise enhancements we observed for chromatograms extracted from both MS scans and tandem mass spectra for a collection of peptides identified in the 1:1 mixture of yeast whole-cell lysate. The average increase in signal-to-noise ratio was approximately 350%.

Effect of signal-to-noise ratio on dynamic range. To illustrate the dependency of ratio measurements on the signal-to-noise values of corresponding mass spectra, we created scatterplots of the measured ratio versus the signal to noise of corresponding chromatograms (**Fig. 5a–d**) to evaluate systematic bias[11,12]. We also generated frequency distributions for the corresponding ratio measurements (**Fig. 5e**,**f**). For the 1:1 ratio measurements (**Fig. 5a**,**c**), there does not seem to be a systematic deviation from the expected ratio; this can also be seen in **Figure 5e**. The enhancement in signal to noise previously detailed is clearly visible, as the mass of the scatterplot in **Figure 5a** shifts significantly to the right in **Figure 5c**. However, for the 10:1 ratio measurements (**Fig. 5b**,**d**), there seems to be a systematic deviation from the expected ratio at lower signal-to-noise values, which is clearly visible in **Figure 5f**. This deviation is much less pronounced in the ratios calculated from tandem mass spectra, most likely because of the shift to higher signal-to-noise ratio. As has been previously documented[8,13], these results show that there can be systematic errors in ion intensity ratios when they are calculated from MS scans with low signal to noise. However, by performing quantification directly from tandem mass spectra, the signal-to-noise ratio is increased, which results in increased accuracy and a larger quantitative range.

Accuracy and precision of ratio measurements. Even after filtering (see **Supplementary Methods**), the limitations in accuracy, and subsequent effect on the dynamic range, that result collectively from the aforementioned sources of error are apparent. For ~30 proteins that were identified and quantified using each of the extraction strategies (**Supplementary Table 2**), the accuracies of measured ratios were comparable between the two strategies, with average ratios typically within 10–20% of the known value for the

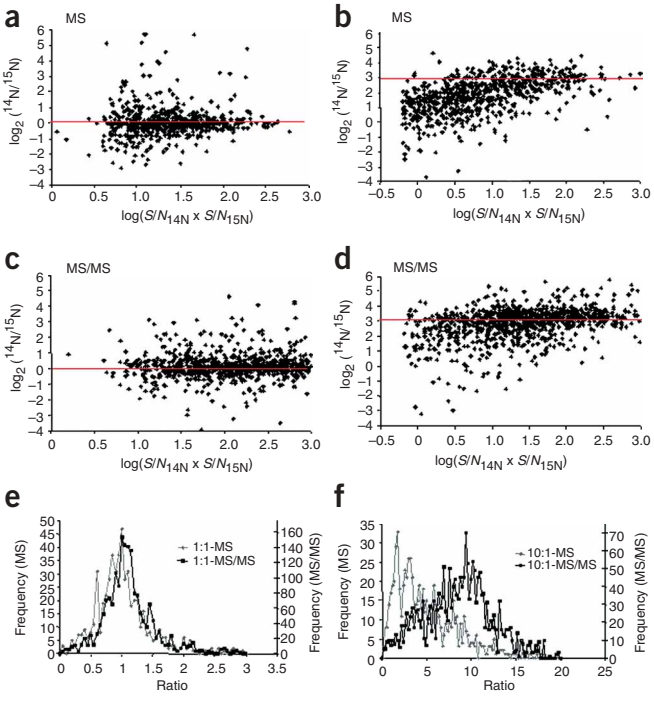

Figure 5 | Effect of signal-to-noise ratio on dynamic range using MS and MS/MS. (**a**–**d**) Plots of $\log_2(^{14}N/^{15}N)$ versus $\log_{10}(S/N_{14N} \times S/N_{15N})$ for peptide ratio measurements obtained from unfiltered chromatograms for two different standard mixtures, 1:1 and 10:1. *S*, signal; *N*, noise. For each standard mixture analyzed, ratios were calculated from MS- (**a**,**b**) and MS/MS-derived (**c**,**d**) reconstructed chromatograms. Expected ratios are highlighted by a red line. (**e**,**f**) Frequency distributions for the two data sets.

Table 2 | Summary information from yeast analysis

	MS		MS/MS	
	1:1	**10:1**	**1:1**	**10:1**
Proteins identified (SEQUEST)	446	599	446	599
Peptides identified	1,314	1,933	1,314	1,933
Chromatograms after filtering	569	621	1,109	1,319
Proteins quantified having:				
No. peptides ≥1	161	196	217	290
No. peptides ≥2	97	110	165	206

1:1 ratio standard mixture. However, the data for the 10:1 standard showed significant differences between extraction strategies. One of the most notable observations is the underestimation by ~40%, on average, of the true ratio by the MS chromatogram reconstruction approach. In contrast, average protein ratios obtained from MS/MS chromatograms were within 20% of the true value. The percent relative standard deviations for the calculated protein ratios ranged from ~2% to 118% and averaged ~20% for the 1:1 mixture and ~35% for the 10:1 mixture; they seemed to be independent of the extraction strategy used.

In summary, the increased signal to noise and selectivity of the MS/MS approach extends the quantitative capabilities of isotopic labeling strategies. The total number of peptides identified was 1,314 and 1,933 for the 1:1 and 10:1 ratio mixtures, respectively (**Table 2**). The number of chromatograms extracted from MS scans that subsequently passed the filtering criteria was 569 and 621 for the 1:1 and 10:1 mixtures. Of the chromatograms extracted from tandem mass spectra, 1,109 and 1,319 passed the same filtering criteria, representing increases of 94% and 112%, respectively. This increase in the number of quantifiable chromatograms led to an 87% increase in the number of proteins we were able to compare (number of peptides ≥2). In addition, the range for accurate quantification in the analysis of complex mixtures was at least twofold larger using the tandem mass spectra strategy, whereas the precisions of both extraction methods were comparable.

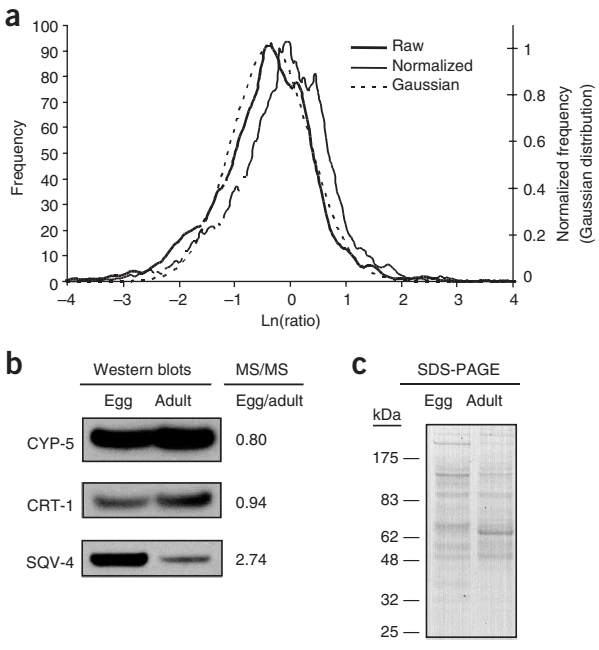

Figure 6 | Quantification of *C. elegans* proteins. (**a**) Normalization of measured peptide-ion current ratios obtained from *C. elegans* embryos and young adults. The gray line represents the distribution of ratios before normalization ('raw') and the black lines show the distribution after normalization. A Gaussian distribution with the same average and standard deviation as the original distribution is shown as a dashed line. (**b**) Relative abundances of *C. elegans* CYP-5, CRT-1 and SQV-4 in the soluble fractions of ^{14}N-labeled eggs and ^{15}N-labeled young adults were determined by western blotting and compared to the ratios determined by MS/MS analysis, shown at right. (**c**) A replicate gel was stained with Coomassie blue dye to show the total amount of protein loaded (10 μg/lane).

Validation of approach using a real-world sample

To demonstrate the utility of our approach for quantitative analysis from tandem mass spectra in a real-world sample, we analyzed protein expression levels in *C. elegans* at two developmental stages (eggs and young adults). Three replicate 12-step MudPIT experiments were performed on a trypsin-digested 1:1 mixture of unlabeled eggs and ^{15}N-labeled young adults. For each replicate experiment, ~10 μg of the mixture was loaded onto a triphasic column and then analyzed by MudPIT on a ThermoElectron LTQ mass spectrometer using data-independent acquisition (see **Supplementary Methods**). For each identified peptide, chromatograms were generated from tandem mass spectra and the resulting chromatograms were filtered by RelEx (as discussed in **Supplementary Methods**). From the database search, we identified 10,062 spectra that translated into 767 different proteins. Of the 10,062 chromatograms generated (one for each identified spectrum), 4,566 passed the required filtering criteria. After redundant measurements and outliers were removed, however, 3,109 peptides ratio measurements that corresponded to 573 different proteins remained. Of these, 333 proteins were quantified by at least three measurements (**Supplementary Table 3**). Measured peptide-ion current ratios were then normalized using a global normalization factor (**Fig. 6a**) to adjust the median peptide-ion current ratio to reflect a 1:1 mixture[11,12,14].

Our results showed that of the 333 proteins quantified, ~16% were enriched in the embryonic stage and ~24% were enriched in the young adult stage by a factor of 2 or greater. The remaining proteins were not appreciably enriched in either stage. For many of the functionally characterized proteins in this list, our measurements correlate well with the putative protein function (data not shown). To corroborate our quantitative MS/MS analysis, we determined the expression levels of several proteins of interest (CYP-5, CRT-1 and SQV-4) using an unrelated approach, western blotting (**Fig. 6b**). Equal amounts of protein were analyzed for the western analysis, as can be seen by Coomassie staining in **Figure 6c**. It should be pointed out that the western blotting samples were prepared independently from those for MS analysis (see **Supplementary Methods**), so sample variation could have contributed to any differences seen. Overall, the correlation between western blots and MS measurements seems reasonable for the three proteins for which we were able to get antibodies; suggesting that our approach is a reliable way to determine relative protein levels.

DISCUSSION

Because of the limitations in existing procedures for shotgun proteomic quantitative analysis, we have developed a method

that uses data-independent tandem mass spectrum acquisition (that is, eliminating MS scans as part of the data acquisition process). As well as its benefits, this method has some potential limitations that bear discussion. One of the primary considerations is the effect that the reduced mass accuracy of the precursor has on the database search results and the time needed to perform the search. Because peptide precursor ions could be present anywhere within the width of the isolation window, the accuracy with which the mass of the precursor ion is known is dependent on the size of the isolation window. Search algorithms typically use the precursor mass to narrow the list of possible matches in the database, and SEQUEST and most other search engines can be tailored to consider a range of precursor masses to allow for the decreased mass accuracy. However, the number of peptide interpretations considered increases markedly when the mass accuracy is decreased, which potentially increases the number of false positives as well as the time required to perform the search.

There is evidence that the identification process is not significantly affected by the reduced mass accuracy even though a larger number of peptides are considered in the search[15,16]. However, other authors[17,18] have recently shown that the decreased mass accuracy can potentially result in higher search-algorithm false-positive rates. Therefore, to offset the potential for increased false positives, we used strict spectral filtering criteria (including enzyme specificity, if applicable, and a requirement for $\geq$2 peptides per locus, among others). Alternatively, preliminary work in this area shows that it is possible to identify the molecular weight of a peptide ion from a tandem mass spectrum by a cross-correlation algorithm (J.D. Venable, unpublished observations; see **Supplementary Methods**).

Another consideration relates to the extent of convolution in tandem mass spectra caused by the simultaneous isolation and fragmentation of peptides of similar *m*/*z* ratios. Researchers have previously explored this territory[7,19]; however, the impact of this phenomenon on a search algorithm's ability to discern the correct interpretation has not been extensively studied. We have found specific cases where this phenomenon has negatively affected our identification step, but on the whole, the impact seems to be minimal, as is evidenced by the qualitative study discussed earlier. This is most likely due to the large dynamic range inherent to complex peptide mixtures, which tends to produce spectra with a dominant ion series that can be successfully identified and quantified. Moreover, unless pseudo-isobaric peptides perfectly coelute, is usually possible to obtain tandem mass spectra primarily composed of one peptide precursor as long as the precursor region is sampled sufficiently often. Unfortunately, less abundant components in convoluted spectra are not usually identified, and so we are currently working on methods to identify and separate convoluted spectra from one another.

In summary, our approach for quantitative analysis of complex mixtures from tandem mass spectra offers several quantitative benefits and was validated using complex peptide mixtures obtained from yeast whole-cell lysates and two developmental stages of *C. elegans*. Average signal-to-noise improvements of ~350% were obtained as compared to analysis from MS spectra, and the selectivity afforded by tandem mass spectrum transitions led to simpler chromatograms with less background noise. Furthermore, the effective dynamic range for quantitative analysis from MS/MS scans seems to be larger, by at least a factor of 2, than that from MS scans. The most notable potential limitations of the technique include an increased chance of false-positive search results owing to decreased mass accuracy, and spectral convolution caused by the isolation and fragmentation of pseudo-isobaric peptides. However, in most cases the quantitative benefits outweigh these potential limitations.

METHODS

Materials. Angiotensin I (*Homo sapiens*) and [Glu^1]fibrinopeptide B (*H. sapiens*) were obtained from Sigma Chemical Co. Peptide stock solutions were prepared by dilution of standards with 5% formic acid (J.T. Baker) to a final concentration of 10 pmol/μl.

Metabolic ^{15}N labeling and preparation of *Saccharomyces cerevisiae* samples. Yeast were grown and labeled using a previously published protocol[6,14]. Unlabeled and ^{15}N-labeled yeast were mixed in known ratios (1:1 and 10:1) as determined by OD_{600}/ml. Mixtures of yeast cells were collected by centrifugation at 1,000*g*, 4 °C, after which yeast were lysed. Lysates were subjected to methanol/chloroform precipitation followed by digestion with two different proteases (see **Supplementary Methods**).

Metabolic ^{15}N labeling of *C. elegans* samples. Worm plates (12 g agarose, 3 g NaCl and 972 ml H_2O) were autoclaved, cooled to 60 °C, mixed with 1 ml 5 mg/ml cholesterol, 1 ml 1 M $CaCl_2$, 1 ml 1 M $MgCl_2$ and 25 ml 1 M K_2HPO_4/KH_2PO_4, pH 6.0, and poured into 100-mm Petri dishes. OP50 bacteria grown in either ^{14}N- or ^{15}N-labeled medium (Celtone-U or Celtone-N, Spectra Stable Isotopes) were concentrated 50-fold and added to the plates (1 ml/plate). Unlabeled eggs and ^{15}N-enriched young adults were prepared (see **Supplementary Methods**) at the third generation of labeling at 20 °C. Synchronized young adults staged before embryogenesis were harvested at hour 67 after L1 larvae were seeded on the OP50 plates.

MudPIT and software. A detailed description of MudPIT and the software used for this study can be found in the **Supplementary Methods**. This software is available from the authors for individual use and evaluation through an Institutional Software Transfer Agreement (see http://fields.scripps.edu/relex for details).

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

The authors would especially like to thank M. MacCoss for his insightful comments about the manuscript and detailed knowledge of RelEx. They thank H. McDonald and C. Wu for discussion and critical reading of the manuscript, and H.R. Horvitz, A.P. Page and J. Ahnn for the SQV-4, CYP-5, and CRT-1 antibodies. J.D.V. and J.W. are supported by National Research Service Award (NIH) fellowships. J.R.Y. is supported by National Institutes of Health grants RR11823, R01MH067880 and R21067598-01.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 21 June; accepted 24 August 2004
Published online at http://www.nature.com/naturemethods/

1. Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
2. Griffin, T.J. *et al.* Abundance ratio-dependent proteomic analysis by mass spectrometry. *Anal. Chem.* **75**, 867–874 (2003).

3. Wu, C., MacCoss, M. & Yates, J.R. III. Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal. Chem.* in the press (2004).
4. Han, D.K., Eng, J., Zhou, H. & Aebersold, R. Quantitative profiling of differentation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19**, 946–951 (2001).
5. Gygi, S.P., Rist, B., Griffin, T.J., Eng, J. & Aebersold, R. Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *J. Proteome Res.* **1**, 47–54 (2002).
6. 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.* **74**, 1650–1657 (2002).
7. Arnott, D. *et al.* Selective detection of membrane proteins without antibodies. *Mol. Cell. Proteomics* **1**, 148–156 (2002).
8. Ong, S.-E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
9. Berger, S.J. *et al.* High-throughput global peptide proteomic analysis by combining stable isotope amino acid labeling and data-dependent multiplexed-MS/MS. *Anal. Chem.* **74**, 4994–5000 (2002).
10. Owens, K.G. Application of correlation analysis techniques to mass spectral data. *Appl. Spectrosc. Rev.* **27**, 1–49 (1992).
11. Quackenbush, J. Microarray data normalization and transformation. *Nat. Genet.* **32**, 496–501 (2002).
12. Yang, I.V. *et al.* Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome Biol.* **3**, 0062 (2002).
13. Ong, S.-E., Kratchmarova, I. & Mann, M. Properties of ^{13}C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.* **2**, 173–181 (2003).
14. MacCoss, M.J., Wu, C.C. & Yates, J.R. III. A correlation algorithm for the automated analysis of quantitative 'shotgun' proteomics data. *Anal. Chem.* **75**, 6912–6921 (2003).
15. Fenyö, D. & Beavis, R.C. A method for assessing the statistical significance of mass spectrometry–based protein identifications using general scoring schemes. *Anal. Chem.* **75**, 768–774 (2003).
16. Anderson, D.C., Li, W., Payan, D.G. & Noble, W.S. A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores. *J. Proteome Res.* **2**, 137–146 (2003).
17. Olsen, J.V., Ong, S.-E. & Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteom.* (2004).
18. Clauser, K.R., Baker, P. & Burlingame, A.L. Role of accurate mass measurement (10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* **71**, 2871–2882 (1999).
19. Purvine, S., Eppel, J.-T., Yi, E.C. & Goodlett, D.R. Shotgun collision–induced dissociation of peptides using a time of flight mass analyzer. *Proteomics* **3**, 847–850 (2003).