Anal. Chem. **2002,** *74,* 4994−5000

High-Throughput Global Peptide Proteomic Analysis by Combining Stable Isotope Amino Acid Labeling and Data-Dependent Multiplexed-MS/MS

Scott J. Berger,[†] Sang-Won Lee,[‡] Gordon A. Anderson, Liljana Paša-Tolić, Nikola Tolić, Yufeng Shen, Rui Zhao, and Richard D. Smith*

Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Mail Stop K8-98, Richland, Washington 99352

In this work, we describe the application of a stable isotope amino acid (lysine) labeling in conjunction with data-dependent multiplexed tandem mass spectrometry (MS/MS) to facilitate the characterization and identification of peptides from proteomic (global protein) digests. Lysine auxotrophic yeast was grown in the presence of ^{13}C-labeled or unlabeled lysine and combined after harvesting in equal proportions. Endoproteinase LysC digestion of the cytosolic fraction produced a global proteomic sample, consisting of heavy/light labeled peptide pairs. Then data-dependent multiplexed-MS/MS was applied to simultaneously select and dissociate only labeled peptide ion pairs. The approach allows differentiation between N-terminal (e.g., b-type ions) and C-terminal fragment ions (e.g., y-type ions) in resulting tandem mass spectra, as well as the capability of differentiation between near-isobaric glutamine and lysine residues. We also describe the utility of peptide composition and fragment information to support peptide identifications and examine the potential application of lysine labeling for differential quantitative protein analysis.

Global peptide analysis is a rapidly developing approach for proteomic studies, wherein the protein complement of whole organisms, or significant subcellular fractions, is proteolytically digested and analyzed via separations coupled with mass spectrometric analysis. This approach has been significantly improved with the development of high-resolution single-dimension[1,2] and multidimensional peptide separations[3] and advances in mass spectrometric techniques that provide higher mass measurement accuracy, higher mass resolution, a larger dynamic range, and data-dependent MS/MS capabilities.[4]

Peptide identifications by tandem (MS/MS) mass spectrometry involve correlation of parent peptide mass and masses of generated fragment ions to the amino acid sequence of possible parent peptides by utilizing computerized database search algorithms. However, the complexity of proteomic samples and the large number of possible peptides from an organism due to modifications during or after transcription (such as frame shifts, alternate RNA splicing, and posttranslational modifications) significantly limits confident and high-throughput peptide identification. The ability to introduce additional constraints in the identity of possible parents and the interpretation of fragment ion spectra can potentially lead to faster and more confident identifications of the peptide and the parent protein.

The introduction of peptide compositional constraints has been achieved by using several in vivo stable isotopic labeling strategies. We recently described the use of PEO−biotin technology in conjunction with the whole-cell ^{14}N/^{15}N labeling of both microbial and mammalian cells to simultaneously introduce nitrogen and cysteine content constraints into the peptide identification process.[5] At the protein level, stable isotope amino acid labeling of auxotrophic *Escherichia coli* has permitted identification of proteins using accurate mass measurement of intact proteins and compositional constraints on as few as three amino acids.[6] Equivalent experiments at the peptide level have been also reported.[7,8]

Sechi and Chait reported the first in vitro isotope labeling technique wherein cysteine residues were modified with unlabeled acrylamide and deuterium-labeled acrylamide to obtain cysteine content information from the mass spectrometric analysis by MALDI-TOF MS.[9] Gygi et al. reported an approach for quantitative proteomics that introduced a cysteine constraint as well as the capability of probing differential expression and purifying cysteine-containing peptides via an affinity tag (i.e., *i*sotope *c*oded *a*ffinity *t*ag or ICAT).[10] Subsequently, additional in vitro approaches for stable isotope labeling were reported for de novo peptide sequenc-

[†] Current address: Waters Corp., 34 Maple St., Milford, MA 01581.

[‡] Current address: Department of Chemistry, Korea University, 1 Anam, Seongbuk, Seoul 136-701, South Korea.

(1) Shen, Y.; Zhao, R.; Belov, M. E.; Conrads, T. P.; Anderson, G. A.; Tang, K.; Paša-Tolić, L.; Veenstra, T. D.; Lipton, M. S.; Udseth, H. R.; Smith, R. D. *Anal. Chem*. **2001**, *73*, 1766−1775.

(2) Shen, Y.; Tolić, N.; Zhao, R.; Paša-Tolić, L.; Berger, S. J.; Harkewicz, R.; Anderson, G. A.; Belov, M. E.; Smith, R. D. *Anal. Chem*. **2001**, *73*, 3011−3021.

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

(4) Li, L.; Masselon, C. D.; Anderson, G. A.; Paša-Tolić, L.; Lee, S.-W.; Shen, Y.; Zhao, R.; Lipton, M. S.; Conrads, T. P.; Tolic, N.; Smith, R. D. *Anal. Chem*. **2001**, *73*, 3312−3322.

(5) Conrads, T. P.; Alving, K.; Veenstra, T. D.; Belov, M. E.; Anderson, G. A.; Anderson, D. J.; Lipton, M. S.; Paša-Tolić, L.; Udseth, H. R.; Chrisler, W. B.; Thrall, B. D.; Smith, R. D. *Anal. Chem*. **2001**, *73*, 2132−2139.

(6) Veenstra, T. D.; Martinović, S.; Anderson, G. A.; Paša-Tolić, L.; Smith, R. D. *J. Am. Soc. Mass Spectrom*. **2000**, *11*, 78−82.

(7) Chen, X.; Fei, Z.; Smith, L. M.; Bradbury, E. M.; Majidi, V. *Anal. Chem*. **1999**, *71*, 3118−3125.

(8) Hunter, T. C.; Yang, L.; Zhu, H.; Majidi, V.; Bradbury, E. M.; Chen, X. *Anal. Chem*. **2001**, *73*, 4891−4902.

(9) Sechi, S.; Chait, B. T. *Anal. Chem*. **1998**, *70*, 5150−5158.

10.1021/ac020105f CCC: $22.00
Published on Web 09/06/2002

ing and peptide/protein identification.[11,12] One of the approaches involves proteolytic digestion (a hydrolytic reaction) in the presence of a mixture of $^{16}O/^{18}O$ water, which produces a population of digested peptides containing peptide pairs differing only by the one or two $^{18}O/^{16}O$ on their respective carboxyl termini.[13] This asymmetry of addition is also exploited during peptide MS/MS analysis, as N-terminal-derived fragments are uniform between peptide pairs, while two sets of C-terminus-derived fragments are generated (i.e., the labeled and unlabeled carboxyl terminus).[14,15] In this approach, simultaneous CID was achieved using low-resolution ion selection with an expanded mass selection window containing both isotope pairs. Similar applications have been developed employing differential CH_3/CD_3 modification of peptide N-termini, using *N*-hydroxysuccinimide or its trideuterioacetylated analogue and determining peptide isotope ratio by MALDI mass spectrometry.[16] More recently, Goodlett et al. reported a differential stable isotope labeling technique using CH_3OH/CD_3OH for quantitation and a correlative database searching algorithm for confident identification.[17]

A disadvantage of in vitro labeling approaches is the extra sample processing steps required and the effects of modification and side reactions on chromatographic behavior or sample complexity. For example, during reversed-phase separation, it is observed that the ICAT-d_0-labeled peptides can elute significantly offset from the ICAT-d_8-labeled peptides, leading to difficulties in quantitation.[18]

In the present work, we demonstrate the utility of coupling in vivo stable isotope lysine labeling with data-dependent tandem mass spectrometry to simplify data interpretation and achieve more confident peptide identifications. Complex peptide mixtures were produced from global LysC digestion of soluble proteins obtained from yeast combined after growth in synthetic media containing ^{13}C-labeled lysine or normal lysine. Data-dependent MS/MS experiments were employed to select only fully digested peptides for subsequent MS/MS analysis, which permitted N-/C-terminal fragment ion differentiation in the resultant tandem spectra and reduced overhead in the downstream sequence database searching for identification. This approach provides an additional advantage of differentiation between two nearly isobaric residues, lysine and glutamine. From these data, we were able to characterize the overall proteolytic processing of the global digest and also assess the potential of lysine labeling for differential quantitative protein analysis.

EXPERIMENTAL SECTION

Cell Growth and Stable Isotope Labeling. *Saccharomyces cerevisiae* strain BY4702 (ATCC designation 200867), a S288C-derived lysine auxotroph, was grown to log phase in a lysine-free synthetic defined (SD-LYS) liquid yeast media (Q-Biogene/Bio101) supplemented with 50 mg/mL lysine (Sigma Chemical Co.). Two cultures (each 600 mL in 2-L polycarbonate baffled culture flasks) were inoculated with 50 μL (1:12000 overall dilution) of the source culture. One flask contained SD-LYS supplemented with lysine; the other was supplemented with 98%+ stable isotope-labeled [$^{13}C_6$]lysine (Cambridge Isotope Labs). Cultures were grown in parallel using a temperature-controlled shaker incubator (30 °C, 250 rpm) until achieving midlog phase ($A_{600} = 6.0$). A roughly equal mixture of the labeled and unlabeled cells was harvested by centrifugation (JLA 10.5 rotor, 5000*g*, 15 min). The resultant pellet was washed once with cold distilled water, repelleted, and stored at −30 °C.

Cell Lysis and Cytosol Preparation. The pellet of mixed labeled/unlabeled cells was resuspended in 3 volumes of 1× TBS/g of wet pellet, and cells were lysed by three rounds of French press (16 000 psi) in a chilled Micro French Press Cell (SLM-Amino/Spectronic model FA-003). The resultant cell homogenate was centrifuged (17000*g*, 4 °C, 15 min) in an Eppendorf microcentrifuge (model 5814), and the cleared homogenate was stored at −80 °C. A cytosolic fraction was obtained by ultacentrifugation (TLA100.1, 240000*g*, 4 °C, 10 min). The protein content of the labeled cytosol was found to be ∼36 mg/mL using a modified Bradford protein assay (Bio-Rad) with bovine IgG as the standard.

Global Endoproteinase LysC Digestion. Cytosol (5 mg/mL) supplemented with freshly prepared urea (2 M), Tris (pH 8.0 at room temperature, 25 mM), and EDTA (1 mM) was heated to 90 °C for 5 min and then quickly cooled to 37 °C, where it was subjected to a multistage protease digestion procedure. Sequencing grade endoproteinase LysC (Roche) was added at a 1:100 ratio (w/w cytosol) and incubated at 37 °C overnight, and a second batch of LysC was added (1:50 w/w final) for an additional 8-h incubation. The digest was centrifuged (17000*g*, room temperature, 5 min) and the supernatant mixed with a third aliquot of LysC (1:25 w/w final) for a second overnight digestion. Aliquots were used directly or stored at −80 °C.

Capillary LC. The high-pressure capillary reversed-phase LC system has been described in detail elsewhere.[2] Briefly, mobile phases (solvent A, 0.1% trifluoroacetic acid (TFA) in water; solvent B, 90% acetonitrile, 0.1% TFA in water) were delivered by two Isco pumps (model 100DM, Isco) at 10 000 psi. Liquid flow control was provided using four Valco ultrahigh pressure positive-feedback valves (Valco, Houston, TX). The solvents were mixed in a stainless steel mixer (volume 2.5 mL) with a magnetic stirrer before the flow splitter and a packed capillary column (150 μm i.d. × 360 μm o.d., 85 cm long). The column was prepared using C-18 bonded particles (3-μm particle with 300-Å pore size, Phenomenex, Torrance, CA), as described previously.[1] An exponential gradient was generated over 200 min and encompassed a range of 0−82% solvent B. HPLC grade solvents were purchased from Aldrich (Milwaukee, WI) and used without further purification.

(10) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994−999.
(11) Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Anal. Chem.* **2000**, *72*, 4047−4057.
(12) Gao, H.; Shen, Y.; Veenstra, T. D.; Harkewicz, R.; Anderson, G. A.; Bruce, J. E.; Paša-Tolić, L.; Smith, R. D. *J. Microcolumn Sep.* **2000**, *12*, 383−390.
(13) Reviewed in: Stewart, II.; Thomson, T.; Figeys, D. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2456−2465.
(14) Qin, J.; Herring, C. J.; Zhang, X. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 209−216.
(15) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 2836−2842.
(16) Geng, M.; Ji, J.; Regnier, F. E. *J. Chromatogr., A* **2000**, *870*, 295−313.
(17) 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.
(18) Zhang, R.; Sioma, C. S.; Wang, S.; Regnier, F. E. *Anal. Chem.* **2001**, *73*, 5142−5149.

Mass Spectrometry. All experiments were performed using a previously described external ion source 7-T FTICR mass spectrometer.[4,19] Briefly, the instrument is equipped with three serial quadrupoles, which are separated by two conductance limits and are operated synchronously, to guide ions produced by the ESI source to a rectangular closed cell through four stages of differential pumping. The ICR cell region is efficiently pumped by a custom-made cryopump with two concentric cryopannels extending to the ICR cell, providing a base pressure below 10^{-9} Torr. The electrospray interface was equipped with an electrodynamic ion funnel to achieve high ion transmission efficiency from the atmospheric ESI source.[20] A home-built piezoelectric pulsed valve driven by a pulsed-valve driver (Lasertechniques, Albuquerque, NM) was used to inject N_2 gas (to $\sim 10^{-5}$ Torr) for ion trapping and for SORI-CID experiments.

Data-Dependent Multiplexed-MS/MS for Labeled Peptide Pair Selection. Data-dependent control of dynamic ion selection and fragmentation was accomplished by running the Odyssey data station in parallel with an ancillary PC running the ICR-2LS application developed in our laboratory.[4] While the Odyssey data station provides the overall experimental sequence, which consists of repeated iterations of a MS sequence script followed by MS/MS and data storage at the end of each sequence, the ancillary PC acquires the same time domain signals from the ICR cell using an ADC card (National Instruments 6070E Analog IO card) and processes the acquired data to allow dynamic ion selection. ICR-2LS, on the ancillary PC, then generates the appropriate SWIFT waveform for isolation and SORI waveform for ion fragmentation in response to the dynamic fluctuation of ion populations eluting from on-line capillary LC. These waveforms are subsequently downloaded to an arbitrary waveform generator (AWG, National Instrument 5411), installed in the PC. The data station then sends trigger signals to the arbitrary waveform generator to send the already generated SWIFT and SORI waveforms to an amplifier and subsequently to the ICR cell while it also energizes a relay that direct the excite signals from the AWG to the amplifier.

Utilization of the ancillary PC for "on-the-fly" or "in-parallel" data processing provides considerable flexibility in ion selection during on-line experiments where ion populations change dynamically. In this work, fully LysC-digested lysine labeled/unlabeled peptide pairs were identified by searching the spectra for pairs of isotopic distributions with characteristic m/z differences of 6.02 (singly charged), 3.01 (doubly charged), or 2.0 (triply charged) within a tolerance of 0.1. After a pair of interest is found, ICR-2LS generates the corresponding SWIFT waveform for ion selection and SORI waveform for fragmentation. The SORI waveform consists of two excitation spikes; one 3 m/z lower than the m/z of the light isotope peptide and the other 4 m/z higher than that of the heavy isotope peptide. These frequency differences were experimentally chosen to induce approximately similar extents of fragmentation for both peptides across their isotope distributions.

Data Processing. Each capillary RPLC/multiplexed-MS/MS experiment produces data sets containing ~800 each MS and MS/MS spectrum (1 MB each). ICR-2LS processed the data set to generate MS and MS/MS files containing the neutral masses of all ions observed, using the THRASH algorithm developed by Horn et al.[21] The data can be visualized in a 2D display of neutral mass versus scan number, where a "spot" size represents component intensity, using software developed at our laboratory (Figure 2). The neutral masses of the parent ions and of the fragments were then used for database searches against a whole yeast protein database[22] using criteria of 100 ppm mass accuracy and using LysC peptide digestion rules.

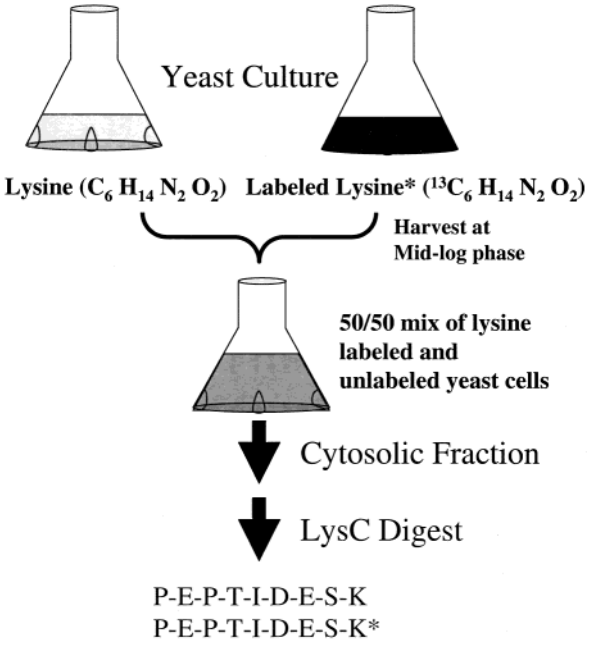

Figure 1. Sample preparation scheme of lysine-labeled yeast. Lysine auxotrophic yeast were grown to midlog phase on defined media containing either unlabeled or $^{13}C_6$-labeled lysine(*), as described in the Experimental Section. Cells from both cultures were combined in equal proportion and processed to produce a LysC-digested cytosolic fraction for global LC/FTICR-MS analysis. The procedure generates peptide pairs differing only by the 6 Da mass for the labeled/unlabeled C-terminal lysines.

RESULTS AND DISCUSSION

Global Labeling of Yeast with Isotopically Labeled Lysine. Global isotopic labeling of lysine residues in proteins was accomplished by addition of lysine auxotropic yeast cells to a synthetic defined media containing $^{13}C_6$-enriched lysine in parallel with growth of the auxotroph in a culture containing lysine with a natural abundance of carbon isotopes (Figure 1). Yeast cultures containing labeled and unlabeled yeast were mixed in equal proportion, based on optical density measurements, after growth to midlog phase. The generation of a global soluble peptide digest from the mixture of labeled and unlabeled yeast was accomplished by production of a cytosolic fraction and proteolytic digestion with endoproteinase LysC (LysC), which cleaves C-terminal to lysine

(19) Winger, B. E., Hofstadler, S. A., Bruce, J. E., Udseth, H. R., Smith, R. D. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 566.
(20) Kim, T., Tolmachev, A. V., Harkewicz, R., Prior, D. C., Anderson, G., Udseth, H. R., Smith, R. D., Bailey, T. H., Rakov, S., Futrell, J. H. *Anal. Chem.* **2000**, *72*, 2247.
(21) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320−332.
(22) Costanzo, M. C.; Hogan, J. D.; Cusick, M. E.; Davis, B. P.; Fancher, A. M.; Hodges, P. E.; Kondu, P.; Lengieza, C.; Lew-Smith, J. E.; Lingner, C.; Roberg-Perez, K. J.; Tillberg, M.; Brooks, J. E.; Garrels, J. I. *Nucleic Acids Res.* **2000**, *28*, 73−76.

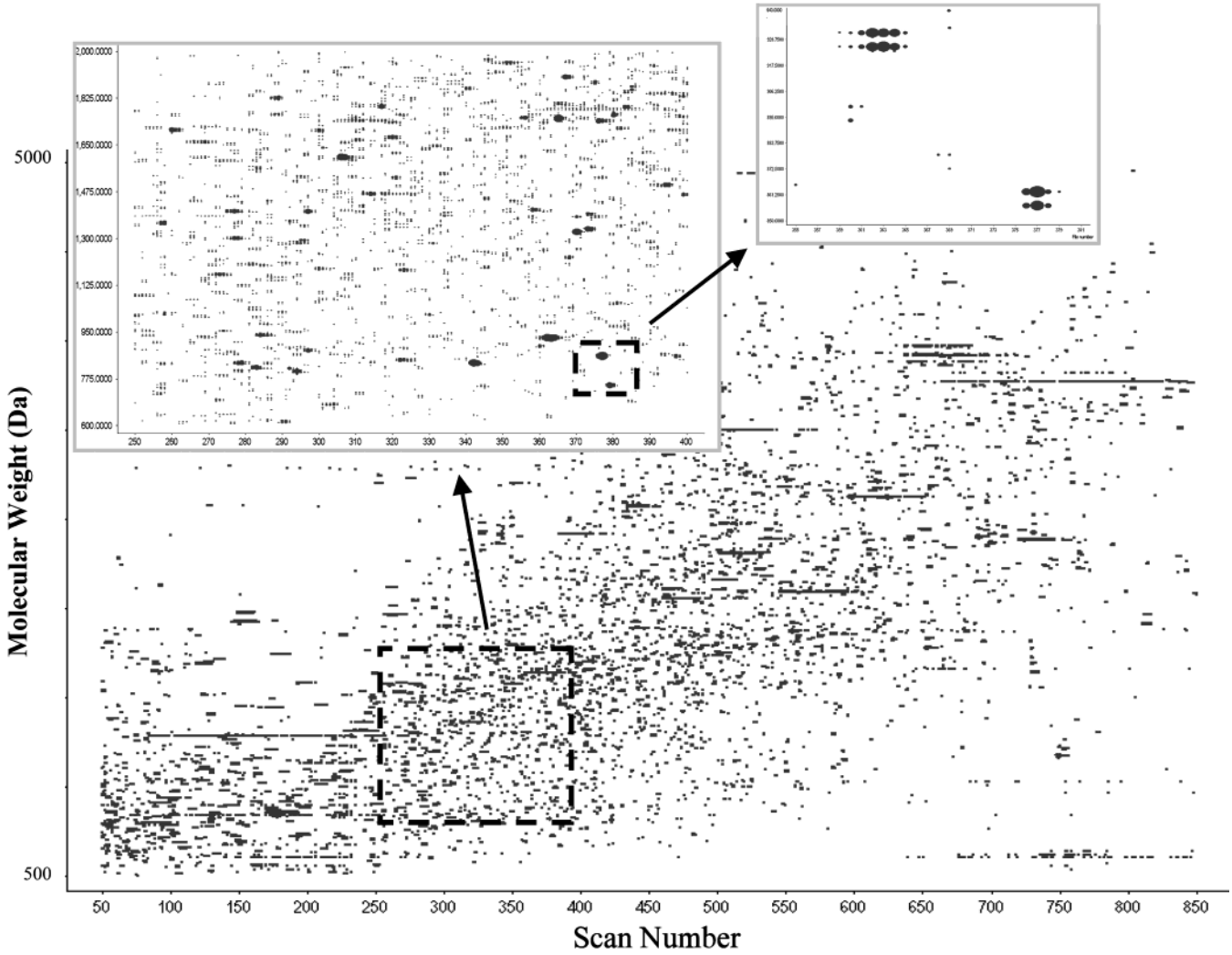

Figure 2. Two-dimensional display (neutral mass vs scan number or elution time) reconstructed from FTICR-MS data obtained during capillary LC separation of LysC-digested labeled yeast cytosol. Each dot corresponds to a component with isotopic distribution, and the dot size is scaled to the intensity of the ion. It shows ~25 000 isotopic distributions, two-thirds of which are observed as components of lysine-labeled peptide pairs. Peptide pairs with one missed cleavage containing two lysines ($\Delta m = 12$ Da) were also observed but with significantly lower frequency and lower intensities (data not shown). A closer examination of the data (upper right) reveals coelution of peptide pairs during the high-resolution reversed-phase separation.

residues (Figure 1). In this process, pairs of peptides were generated differing by $\Delta m = n6.02$ Da, where n is the number of lysine residues (e.g., $n = 1$ for fully digested products and $n = 2$, 3, ..., for partially digested products). One caveat to the labeling is that the C-terminal peptides of proteins will not be observed as pairs unless lysine is present at the C-terminus. The distinctive feature of an unlabeled C-terminal peptide is also observed with $H_2{}^{16}O/H_2{}^{18}O$ labeling; in the latter approach, this distinction has been exploited to permit identification of C-terminal peptides for rapid and simplified identification of proteins from in-gel digests of 2D-gel spots.[23,24] Equally important, however, is that the occurrence of peptide pairs distinguishes peptides from electronic noise, non-protein sample/solvent components, and autolyzed LysC products. The use of a 1:1 mixture permits not only the mass difference to be utilized in identifying pairs but also the relative intensity of those pairs to be used is distinguishing peptide from non-peptide signals.

Production of lysine-labeled peptide pairs could conceivably be accomplished via a single culture supplemented with a 50/50 mixture of labeled and unlabeled lysine. This experiment is far less appealing, however, in that only fully digested peptides will exhibit a 1:1 labeling ratio (LR) of labeled to unlabeled peptide. For example, in a 50% labeling experiment, peptides containing two lysines (one missed cleavage) would appear as a distribution containing 1:2:1 unlabeled/labeled/double labeled peptide, while equivalent distributions with a neutral mass difference of 12 Da (Δ12) would be observed with peptides obtained from a mixture of labeled and unlabeled cells. Also it would be complicated due to proteolysis effects and other processes during cell culture.[25]

Several benefits arise from use of LysC rather than trypsin proteolytic cleavage. First, global LysC digestion results in fewer protein fragments in global digests, simplifying separation requirements and the potential number of MS/MS experiments. Second, LysC peptides will be larger and more likely to contain sites for multiple charging (internal arginines), resulting in peptide ions

(23) Kosaka, T.; Takazawa, T.; Nakamura, T. *Anal. Chem.* **2000**, *72*, 1179−1185.
(24) Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D. *Electrophoresis* **1996**, *17*, 945−953.
(25) Martinovic, S.; Veenstra, T. D.; Anderson, G. A.; Paša-Tolić, L.; Smith, R. D. *J. Mass Spectrom.* **2002**, *37*, 99−107.

more amiable for MS/MS experiments and unique identifications in database searches.[26] Third, LysC is a robust enzyme that maintains activity in higher concentrations of denaturants than trypsin, producing more efficient, and potentially more comprehensive global digests. Most importantly, LysC generates pairs of fully digested peptides with lysine located solely at the C-terminus, permitting constraints to be introduced for interpretation of peptide MS/MS data, as will be described later in this work.

LC/FTICR-MS Analysis of Global LysC Digests of Soluble Yeast Proteins. The complexity encountered during the analysis of global peptide digests has driven improvements in peptide separations. Two approaches have found particular success in addressing these challenges, multidimensional chromatography and high-resolution single-dimension chromatography. Multidimensional separations such as the MUDPIT approach utilized by Yates and co-workers[3] have combined lower resolution orthogonal separations to generate high overall peak capacities. The alternative approach adopted by our laboratory[1,2] has been the development of higher resolution single-dimension, reversed-phase, capillary-scale separations, utilizing longer capillaries packed with small particles and operating at ultrahigh-pressure regimes to increase peak capacity. The advantage of this approach is that less sample is required and specific losses, which increase with every additional mode of separation, can be minimized. We previously described the benefits of combining high-resolution separations with the high sensitivity, dynamic range, and mass measurement accuracy of FTICR mass spectrometry for analysis of global protein digests[1] and demonstrated the capability of analyzing in excess of 100 000 components in a single analysis.

The 2D LC/MS display of a global soluble lysine-labeled yeast digest (neutral monoisotopic mass versus spectrum number) reconstructed from FTICR mass spectra reveals 17 857 different peptides observed in a 3.5-h capillary LC/MS/MS experiment (Figure 2). Following each MS acquisition, the most abundant $\Delta m = 6$ peptide pair was selected in a subsequent spectrum for MS/MS analysis.

Analysis of the LC/MS data set using the ICR-2LS software indicated that overall 48% of the components were observed to be paired, while 59% of components in the top 10% of signal intensity were observed as pairs. Of the pairs, 89.4% corresponded to fully digested peptides ($\Delta m = 6$), 10.1% to peptides containing one missed LysC cleavage ($\Delta m = 12$), and ~0.5% to peptides with multiple missed LysC cleavages. This capability to determine proteolytic digest efficiency should prove useful as a benchmarking statistic for comparing independent proteomic analyses of global protein digests. In fact, other work from our laboratory has demonstrated lysine labeling as a useful indicator of trypsin digest efficiency.

The knowledge of digestion state for an individual peptide by means of the number of lysines presents opportunities for introduction of constraints or a validation criterion in database searching from MS and MS/MS data. For example, stable isotope amino acid labeling in *S. cerevisiae* using multiple amino acids has demonstrated that constraints can be introduced based on amino acid composition that permit increased rates of peptide identification from peptide mass information.[7,8] Knowledge of digestion processing efficiency can reduce the possible number of possible global LysC digestion products (500−5000 MW) for the 6231 predicted yeast ORFs, from 410 717 to only 144 704 possible peptides if complete digestion is obtained.

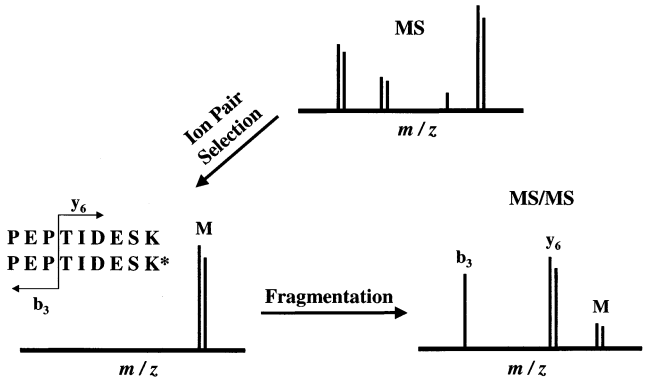

Figure 3. Stable isotope labeling in conjunction with multiplexed MS/MS simplified identification of parent peptide ions. Selection of a hypothetical peptide pair of PEPTIDESK and PEPTIDESK* for simultaneous MS/MS analysis generates spectra with characteristic 6-Da fragment ion pairs for C-terminal fragments (e.g., sets of y and y* ions), while N-terminal fragments (e.g., b ions) are observed as unpaired peaks.

A closer examination of the LC/MS 2D plot (Figure 2) demonstrates two features of in vivo global lysine labeling: the coelution and equiabundant intensities of labeled and unlabeled peptide pairs. The observations by others that ICAT-d_0 and ICAT-d_8 peptide typically elute with variable relative retention times[18] has limited their usefulness as tools for quantitative proteomic analysis. Results from our own group have shown that N^{14}/N^{15} peptide pairs also have small differences in elution profile.[5] However, LC/MS analysis of the [$^{12}C_6/^{13}C_6$]lysine-labeled digest reveals coelution of labeled and unlabeled peptides over the entire course of the experiment, not only improving quantification capabilities but also simplifying multiplexed ion pair selection.

Examination of the LC/MS data set also reveals that peptide pairs elute with roughly equal intensities. Analysis of $\Delta m = 6$ pairs reveals a symmetrical distribution of LRs centered at a LR = 1.0 with a standard deviation of 0.2. We hypothesize that this error may be due to the natural variation in the level of individual proteins, even though the labeled and unlabeled cultures were grown under identical conditions and seeded from the same starter culture. For this calculation, the intensities of the observed peptides were obtained by summing up the three most abundant ^{13}C isotope peaks of an individual isotope envelope instead of choosing the most abundant peak to accurately describe the intensity ratio between the pair peptides. Further elucidating these effects may prove useful in the development of in vivo labeling approaches for quantitative proteomics.

Simplified and Confident Peptide Identification. Figure 3 illustrates the schemes for coupling multiplexed-MS/MS with stable isotope labeling for highly informative peptide analysis. The sample described above predominantly produces peptide pairs, almost all of which appear to be fully LysC digested. The most abundant isotope pair is then selected by the data-dependent ion selection, after only searching for the pairs that are separated by mass difference of 6 Da. The selected $\Delta m = 6$ peptides are subsequently subjected to simultaneous fragmentation by SORI.

(26) Conrads, T. P.; Anderson, G. A.; Veenstra, T. D.; Paša-Tolić, L.; Smith, R. D. *Anal. Chem.* **2000**, *72*, 3349−3354.

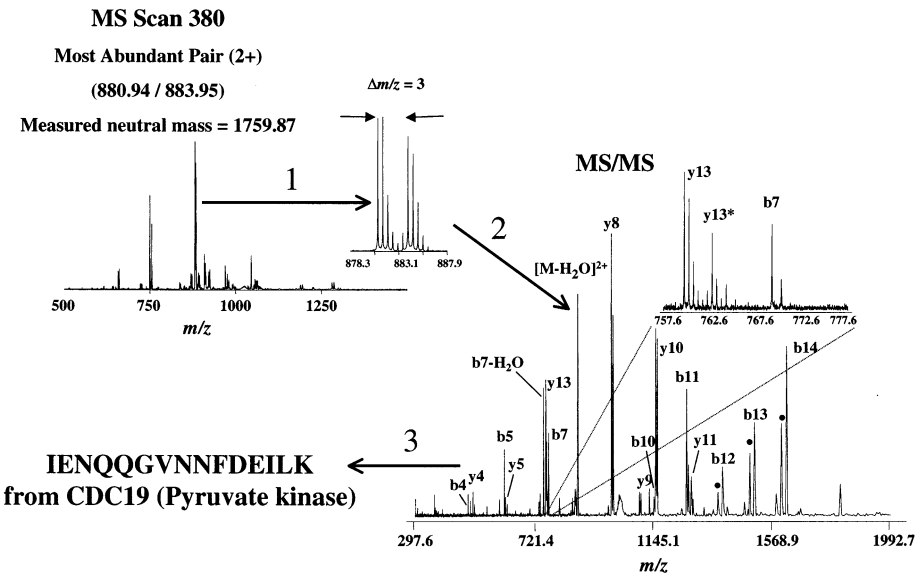

Figure 4. Data-dependent MS/MS analysis of a peptide pair. The most abundant peptide ion pair in an LC/MS spectrum 380 (upper left), was isolated by a SWIFT waveform in the subsequent spectrum (arrow 1), as described in the Experimental Section. The MS/MS spectrum (lower right) resulting from the isolated pair (arrow 2) was sufficient for an unambiguous identification (arrow 3) of the fully digested peptide arising from yeast pyruvate kinase from a database search of a predicted yeast ORF. Closer examination (inset) of the MS/MS spectra shows two fragments.

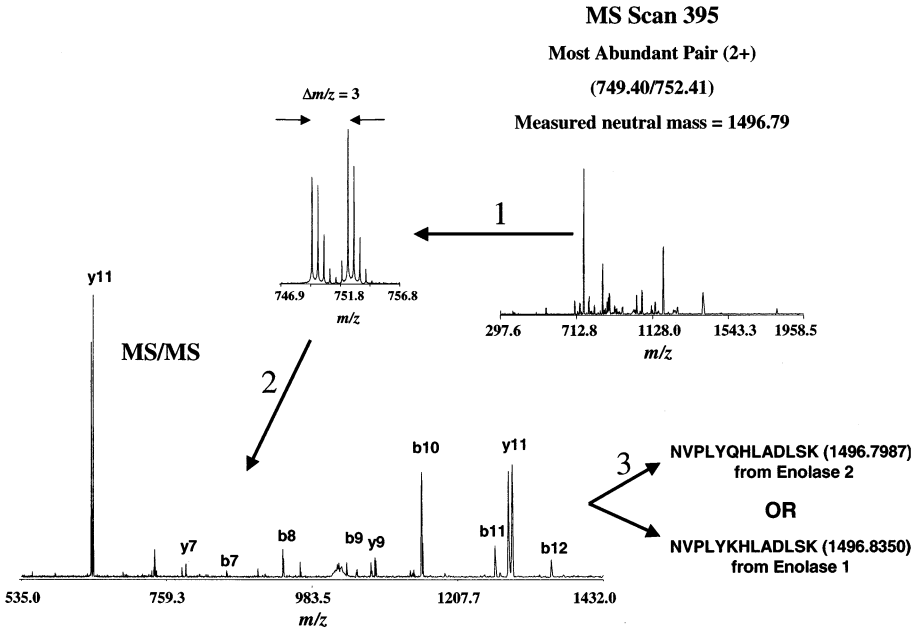

Figure 5. Two candidate peptides, NVPLYQHLADLSK and NVPLYKHLADLSK. in spectrum 395, identified by the database search. The two peptides differ by an amino acid (Q/K substitution) and originate from two closely related proteins, enolase 1 and enolase 2. The mass differences between the two peptides is 0.036. However, the parent peptide pair had a single labeled lysine (indicated by $\Delta m = 6$ Da). allowing its correct assignment.

The resultant MS/MS spectra contain fragment peaks from both N-terminus and C-terminus. However, as shown in Figure 3, only C-terminal fragments (typically y-type ions) contain lysine isotope information and will show up as doublets in MS/MS spectra, while N-terminal fragments (e.g., a- or b-type ions) and internal fragments appear as singlet peaks. The ability to distinguish N-terminal fragments from C-terminal fragments increases confidence of peptide identification.

Figure 4 shows a representative example of high-confidence peptide identification using the aforementioned scheme. In MS spectrum 380, the most abundant ion pair (m/z 880.94/883.95, doubly charged) was selected for SORI-CID. From the spacing between ^{13}C isotope peaks of the individual peptide, the charge state is determined to be +2 and thus $\Delta m/z = 3$ between the isotope peptide pair corresponds to a neutral mass difference of 6 Da, indicating the peptide of interest contains a single lysine. The data-dependent MS/MS scheme as described in the Experimental Section selects the peptide pair using a double-notched SWIFT waveform. Multiplexed-MS/MS allowed simultaneous fragmentation of the pair peptides, and as shown in Figure 4, the resultant MS/MS spectrum is complex, containing fragments from the both peptides. As expected, some fragments appeared as

doublets while others were unpaired singlets, indicating C-terminal and N-terminal fragments, respectively. At this point we have not implemented the information of C-/N-terminal fragments into the database search algorithm, but rather we use the information to verify search results. The database search resulted in several candidate peptides, and the MS/MS spectrum was best described by a peptide (IENQQGVNNFDEILK) from CDC19, a pyruvate kinase. As shown in Figure 4, all the fragment ions were consistent with the singlet/doublet rule of N-terminal/C-terminal fragments, confirming correct identification. The average mass measurement accuracy for this peptide and its fragment ions was 7.3 ppm.

Lysine labeling offers an additional advantage over other stable isotope labeling techniques: differentiation between lysine and glutamine. In MS spectrum 395, the multiplexed-MS/MS experiment and subsequent database search resulted in two candidate peptides, NVPLYQHLADLSK and NVPLYKHLADLSK (Figure 5). The two candidate peptides originate from two closely related yeast proteins, enolase 1 and enolase 2, and differ by a single amino acid substitution, one containing glutamine (Q) and the other containing lysine (K). The calculated mass difference between the two peptides is 0.036 Da, corresponding to 24.1 ppm, and cannot be readily distinguished by most mass spectrometers. However, as shown in the expanded mass spectrum of the parent peptide pair in Figure 5, the parent contained only a single lysine that is evidenced by the mass difference of 6 Da. Thus, we were able to confidently assign the parent peptide to NVPLYQHLADLSK from enolase 2 since the NVPLYKHLADLSK peptide from enolase 1 contains two lysines and would have resulted in a mass difference of 12 Da.

CONCLUSIONS

We have demonstrated the combination of stable isotope lysine labeling with multiplexed-MS/MS experiments for high-throughput global proteome analysis. In vivo [^{13}C]lysine labeling offers several advantages including the following: high-efficiency isotope incorporation, the coelution of peptide pairs, enabling discrimination between peptides of interest and contaminants, providing evaluation of global digestion efficiency, and introduction of constraints that aid peptide identifications. Data-dependent ion selection proved a highly efficient process allowing only fully digested yeast peptide pairs to be selected for simultaneous MS/MS analysis, while peaks from trypsin or other contaminants could be ignored regardless of signal intensity. Multiplexed-MS/MS of isotopically labeled peptide pairs allows differentiation between N-terminal- and C-terminal-derived fragments, providing greatly improved confidence in peptide identification. Finally, the approach allows distinguishing glutamine from lysine, which is not straightforward using conventional methods.

To achieve more efficient and high-throughput analysis, we are currently expanding this approach by including C-/N-terminal fragments into database search algorithms and implementing even higher order data-dependent multiplexed-MS/MS (as shown earlier[4]) analyses where multiple peptide pairs are simultaneously analyzed. Furthermore, lysine being an essential amino acid in mammals should permit us to apply lysine labeling approaches for proteomic analysis of mammalian cell culture systems.

ACKNOWLEDGMENT

Portions of this research were supported by the office of Biological and Environmental Research, U.S. Department of Energy, and the National Cancer Institute under Grant CA 93306. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC06-76RLO 1830.

Received for review February 15, 2002. Accepted August 2, 2002.

AC020105F