Abundance Ratio-Dependent Proteomic Analysis by Mass Spectrometry

Timothy J. Griffin,[†] Chris M. Lock,[‡] Xiao-jun Li,[†] Alpesh Patel,[§] Iryna Chervetsova,[‡] Hookeun Lee,[†] Michael E. Wright,[†] Jeffrey A. Ranish,[†] Sharon S. Chen,[†,||] and Ruedi Aebersold*[,†]

Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103, MDS-Sciex, 71 Four Valley Drive, Concord, Ontario, L4K 4V8, Canada, Applied Biosystems, Foster City, California 94404, and Department of Bioengineering, University of Washington, Seattle, Washington 98199

The goal of quantitative proteomics is to determine the identity and relative quantity of each protein present in two or more complex protein samples. Here we describe a novel approach to quantitative proteomics. It is based on a highly accurate algorithm for the automated quantification of chromatographically fractionated, isotope-coded affinity-tagged peptides and MALDI quadrupole time-of-flight tandem mass spectrometry for their identification. The method is capable of detecting and selectively identifying those proteins within a complex mixture that show a difference in relative abundance. We demonstrate the effectiveness and the versatility of this approach in the analysis of a standard protein mixture, protein expression profiling in a human prostate cancer cell line model, and identification of the specific components of the multiprotein transcriptional machinery in *S. cerevisiae*.

The sequencing of the human genome[1,2] has ushered in the era of systems biology and with it the goal of studying the function and control of biological systems via the systematic and quantitative analysis of the system's components. Central to systems biology is the quantitative analysis of gene expression changes due to external (pharmacological, environmental) or internal (genetic, pathological) perturbations, measured globally at both the mRNA and protein levels,[3−8] and the identification of the components of the protein complexes that constitute the molecular machinery of cell physiology.[9−13] The integrated information gathered from these analyses constitutes the basis for the formulation of mathematical models that simulate the system and predict its biological responses to perturbative effects.[6,8,14]

The capability to undertake quantitative proteome profiling experiments has been significantly advanced by the combination of stable-isotope protein labeling, multidimensional chromatography, tandem mass spectrometry (MS/MS), and automated sequence database searching.[15] Stable isotope labeling in quantitative proteomics is most commonly achieved via isotope-coded affinity tag (ICAT) reagent labeling.[5] In the initial implementation of the ICAT reagent method, proteins were first identified by a "shotgun" approach using on-line reversed-phase microcapillary liquid chromatography (μLC) electrospray ionization (ESI) MS/MS and sequence database searching. In this method, the mass spectrometer first acquires a mass spectral (MS) survey scan of all peptides eluting at a specific point in time from the μLC column. The instrument control software then sequentially selects peptide ions of a specific mass-to-charge (m/z) ratio for collision-induced dissociation (CID), usually in order of descending intensity of the precursor ion signal. The resulting CID spectra are then searched against sequence databases to identify the protein from which the sequenced peptide originated. Subsequently, the abundance ratio for each protein is calculated based on the signal intensity ratios of isotopic pairs of peptide ions detected in the MS scan of the automated MS/MS routine. Therefore, in this method peptide identification and quantification are decoupled and identification precedes quantification. This has significant implications for

* Corresponding author. Tel: 206-732-1204. Fax: 206-732-1255. E-mail: raebersold@systemsbiology.org.
[†] Institute for Systems Biology.
[‡] MDS-Sciex.
[§] Applied Biosystems.
[||] University of Washington.

(1) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; et al. *Nature* **2001**, *409*, 860−921.
(2) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; et al. *Science* **2001**, *291*, 1304−1351.
(3) Schena, M.; Shalon, D.; Davis, R. W.; Brown, P. O. *Science* **1995**, *270*, 467−470.
(4) Lockhart, D. J.; Winzeler, E. A. *Nature* **2000**, *405*, 827−836.
(5) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994−999.
(6) Ideker, T.; Thorsson, V.; Ranish, J. A.; Christmas, R.; Buhler, J.; Eng, J. K.; Bumgarner, R.; Goodlett, D. R.; Aebersold, R.; Hood, L. *Science* **2001**, *292*, 929−934.
(7) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. *Nat. Biotechnol.* **2001**, *19*, 946−951.
(8) Griffin, T. J.; Gygi, S. P.; Ideker, T.; Rist, B.; Eng, J.; Hood, L.; Aebersold, R. *Mol. Cell. Proteomics* **2002**, *1*, 323−333.
(9) Alberts, B. *Cell* **1998**, *92*, 291−294.
(10) Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T. A.; Judson, R. S.; Knight, J. R.; Lockshon, D.; Narayan, V.; Srinivasan, M.; Pochart, P.; et al. *Nature* **2000**, *403*, 623−627.
(11) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd. *Nat. Biotechnol.* **1999**, *17*, 676−682.
(12) Gavin, A. C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; et al. *Nature* **2002**, *415*, 141−147.
(13) Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; et al. *Nature* **2002**, *415*, 180−183.
(14) Aebersold, R.; Hood, L. E.; Watts, J. D. *Nat. Biotechnol.* **2000**, *18*, 359.
(15) Gygi, S. P.; Aebersold, R. In *Proteomics: A Trends Guide*; Mann, M., Blackstock, W., Eds.; Elsevier: London, 2000; pp 32−37.

10.1021/ac026127j CCC: $25.00
Published on Web 00/00/0000

quantitative proteomic analysis using stable isotope dilution. The main consequence is that numerous peptides are identified that are derived from proteins of unchanged abundance between the samples. As these constitutively expressed proteins usually make up the large majority (up to 90%)[7] of proteins that are detected, instrument, data analysis, and operator time are consumed analyzing peptides that are many times of little biological significance to the system being studied. Additionally, these constitutively represented peptides are repeatedly selected for MS/MS analysis, thus preventing the analysis of more informative peptides that are derived from proteins showing changes in abundance.

We previously described an approach in which peptides were first quantified in a high-resolution ESI time-of-flight (TOF) mass spectrometer and peptides with particular abundance ratios were then sequenced using an ESI quadrupole mass spectrometer.[16] This method was limited by the need for two different ESI mass spectrometers and the off-line sample-handling steps that decreased sensitivity and throughput. Here we describe a novel approach to quantitative proteome analysis, based on the automated, accurate quantification of isotope-labeled, chromatographically separated peptide mixtures, followed by the automated selection for CID of those peptides that show a significant change in abundance using a matrix-assisted laser desorption/ionization quadupole TOF (MALDI QqTOF) mass spectrometer.[17−19] This approach represents a significant advance in quantitative proteomics technology, enabling more intelligent, selective identification of biologically interesting proteins from within a background of constitutively represented proteins. We demonstrate the effectiveness and versatility of this method in the quantitative profiling of androgen-regulated prostate cancer proteins and also the identification of the specific components of an immune precipitated multiprotein complex, the transcriptional machinery of *Saccharomyces cerevisiae.*

EXPERIMENTAL SECTION

All ICAT reagents used were purchased from Applied Biosystems. α-Cyano-4-hydroxycinnapinic acid matrix was purchased from Agilent Technologies (Palo Alto, CA) and diluted 3-fold from its original concentration in 50% methanol, 49.9% water, and 0.1% TFA and used for all analyses.

Preparation of a Standard Protein Mixture. A standard protein mixture containing the proteins *Escherichia coli β*-galactosidase (BGAL_ECOLI), rabbit phosphorylase B (PH2_RABIT), chicken albumin (OVAL_CHICK), and bovine α-lactalbumin (LCA_BOVIN) were prepared at varying d_0/d_8 ratios as described.[5] A mixture of proteins from a total yeast lysate at a 1:1 d_0/d_8 ratio was prepared, digested with trypsin, fractionated by strong cation exchange (SCX) HPLC as described,[19] and avidin cartridge purified. Approximately 25% of a single SCX fraction of yeast peptides was combined with ∼0.5 μg of total proteins from the standard protein mixture. This peptide mixture was dried by vacuum centrifugation and redissolved in 8 μL of 18% ammonium hydroxide. A 1.5-μL aliquot of an 8 M solution of *O*-methylisourea hemisulfate (Acros, Geel, Belgium) in water was added to this and the resultant mixture incubated for 15 min at 60 °C to convert lysine side chains to homoarginine.[20] The peptides were separated by μLC using a gradient profile starting at 10% solvent B (80% acetonitrile, 0.1% TFA) and increasing to 35% buffer B over 45 min using an HP 1100 HPLC system (Agilent). The eluting peptides were fractionated manually to the MALDI plate in 1-min time intervals, where 0.6 μL of α-cyano matrix was immediately added to each spot and allowed to air-dry.

Protein Expression Profiling from Human Prostate Cells. The androgen-dependent human prostate cell line LnCAP FGC was grown separately in the presence or absence of the synthetic androgen hormone R1881; the nuclear protein fractions were isolated as described[21] and labeled using the d_8 and d_0 versions of the ICAT reagent label, respectively. These proteins were digested, fractionated by SCX HPLC, and purified by avidin affinity chromatography as described.[19] A single, purified SCX fraction was selected, and ∼50% of this sample was dried by vacuum centrifugation and brought up in 18% ammonium hydroxide containing 0.5 M 2-methoxy-4,5-dihydro-1*H*-imidazole (E. Peters, Genomic Institute of the Novartis Research Foundation).[22] This mixture was incubated at 55 °C for 1 h. The sample was dried by vacuum centrifugation and separated by μLC as described[19] using a 2-h gradient profile with automatic spotting to the MALDI plate using a Probot microfraction collector (LC Packings). Both the precolumn (250 μm i.d. × 7 mm length) and analytical column (150 μm × 100 mm) used were packed in-house with 5-μm, 200-Å, Magic C18AQ (Michrom BioResources, Inc., Auburn, CA). α-Cyano matrix was automatically added using a syringe pump and introduction into the eluent flow via a postanalytical column T-adapter.

Analysis of Proteins Associated with Srb4p from Yeast. A PCR epitope tagging method[23] was used to C-terminally tag the yeast *SRB4* open-reading frame (ORF) in *S. cerevisiae* strain BWG 1-7a[24] with three copies of the FLAG epitope. Correct integration was confirmed by western blotting with an antibody to the FLAG epitope and a polyclonal antibody to Srb4p (S. Hahn, Fred Hutchinson Cancer Research Center). Nuclear extract was prepared from strain BWG 1-7a (negative control) and JRY1 (FLAG tagged Srb4p) as described.[25] A 250-μL aliquot of a 50% anti-flag M2 agarose bead slurry (Sigma) were aliquoted into separate tubes and washed twice with 10 mL of wash buffer 1 (20 mM HEPES, pH 7.9, 100 mM potassium acetate, 5 mM magnesium acetate, 1 mM EDTA, 0.01% NP40), containing protease inhibitors PMSF, benzamidine, chymostatin, leupeptin, and pepstatin A. The beads were then blocked by incubation with 1 mg of BSA in 1 mL of wash buffer for 15 min at room temperature and washed twice in wash buffer. A 1. 4-mg sample of protein each in 2 mL of wash buffer 1 from the control lysate and positive lysate were

(16) Griffin, T. J.; Han, D. K.; Gygi, S. P.; Rist, B.; Lee, H.; Aebersold, R.; Parker, K. C. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1238−1246.
(17) Loboda, A. V.; Krutchinsky, A. N.; Bromirski, M.; Ens, W.; Standing, K. G. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1047−1057.
(18) Shevchenko, A.; Loboda, A.; Shevchenko, A.; Ens, W.; Standing, K. G. *Anal. Chem.* **2000**, *72*, 2132−2141.
(19) Griffin, T. J.; Loboda, A.; Gygi, S. P.; Rist, B.; Jilkine, A.; Ens, W.; Standing, K. G.; Aebersold, R. *Anal. Chem.* **2001**, *73*, 978−986.
(20) Beardsley, R. L.; Reilly, J. P. *Anal. Chem.* **2002**, *74*, 1884−1890.
(21) Dignam, J. D.; Lebovitz, R. M.; Roeder, R. G. *Nucleic Acids Res.* **1983**, *11*, 1475−1489.
(22) Peters, E. C.; Horn, D. M.; Tully, D. C.; Brock, A. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2387−2392.
(23) Knop, M.; Siegers, K.; Pereira, G.; Zachariae, W.; Winsor, B.; Nasmyth, K.; Schiebel, E. *Yeast* **1999**, *15*, 963−972.
(24) Pfeifer, K.; Prezant, T.; Guarente, L. *Cell* **1987**, *49*, 19−27.
(25) Kang, J. J.; Auble, D. T.; Ranish, J. A.; Hahn, S. *Mol. Cell. Biol.* **1995**, *15*, 1234−1243.

incubated separately with the beads for 2 h at room temperature with gentle shaking. The beads were then washed twice with 10 mL of wash buffer 1, followed by two washes with 10 mL of wash buffer 2 (20 mM HEPES, pH 7.9, 100 mM potassium acetate, 5 mM magnesium acetate, 1 mM EDTA, 0.0025% NP40), eluted by two 30-min incubations in 250 μL of wash buffer 2 containing 6 M urea, and rinsed once with an additional 250 μL of this solution. The separate eluent mixes were first concentrated to ~50 μL using Centricon 10 spin columns (Millipore, Bedford, MA), then diluted into 500 μL of 10 mM Tris, 6 M urea, pH 8.3, and concentrated to ~75 μL. A 3-μL alquot of 1% SDS was added to each mixture. The negative control and positive samples were then labeled with the d_0 and d_8 forms of the ICAT reagent, respectively, as described,[26] combined, trypsin digested, and purified by cartridge SCX followed by cartridge avidin purification (Applied Biosystems) as per manufacturer's instructions. Approximately 30% of the total purified peptide mixture was reacted with *O*-methylisourea hemisulfate as described above and then fractionated by μLC as described above for the standard protein mixture, using a 2-h gradient.

Analysis by MALDI QqTOF MS. Mass spectrometric analysis was performed using a prototype MALDI source (Applied Biosystems/MDS-Sciex) mounted on a Qq-oaTOF mass spectrometer (Applied Biosystems/MDS-Sciex) as described.[17] The design was modified from that previously reported by the addition of a flow-restricting aperture between the sample plate and the Q0 rods, effectively raising the pressure in the ionization region to ~1 Torr, resulting in a rapid collisional cooling effect on the MALDI-generated plume of ions and, as a consequence, a significant decrease in metastable fragmentation. The source design utilizes a 96-well plate format, with *xy* motion controlled through a pair of high-precision stepper motors. All aspects of the source were computer controlled via a dedicated NT-based software package (M. Yang, MDS-Sciex, Concord, ON, Canada). Laser pulses were generated via a nitrogen laser (Laser Science Inc., VSL-337ND) operating at 337.1 nm and a frequency of 29 Hz. The laser power was approximately 30−50mJ/cm^2 over a target area of ~0.1 mm^2. In a typical analysis, the single-stage MS spectra were acquired for 30 s, and each subsequent CID spectrum was acquired for 1 min. Data were acquired in an automated fashion using a Visual BASIC-based acquisition script.

Automated Quantification Software. The automated software routine works as follows: (i) Text files of the initial MS survey scan spectra are put in order of elution from the μLC column, and for each spot, scan intensities in the previous, current, and succeeding spectra are added together to generate a combined spectrum. This step accounts for the mobility shift between d_0- and d_8-labeled peptides,[27] which can introduce errors if the abundance ratio is calculated based upon the signal detected in a single spot only for eluting peptides that are split across adjacent spots. Peaks in the original spectra and the combined spectra are then clustered into isotope series, and these are scanned for pairs of isotope series with a mass difference (Δm) of 8.05 ± 0.05 or 16.1 ± 0.05 Da (for double-cysteine containing peptides). (ii) Relative abundances of the outputted mass list of ICAT reagent-labeled peptides are then calculated by combining the signal intensities of each isotope series from adjacent sample spots, then summing up the area of the first three consecutive peaks of the isotope series for each of the peaks in a given pair, and dividing the summed areas to produce a d_0/d_8 value. (iii) ICAT reagent-labeled singlet peaks are identified by scanning for remaining peaks in the original spectra above a local S/N threshold of 2.0. (iv) Masses of peptides having d_0/d_8 values that pass the user-defined d_0/d_8 threshold criteria along with masses of singlet peptides are outputted to a text-formatted inclusion list, along with the sample spot on the MALDI plate containing the most intense signal for these peptides, for subsequent automated MS/MS analysis.

Database Searching of MS/MS Data. Prior to the database search, prominent fragment peaks at m/z values known to be derived from the ICAT reagent linker were automatically removed from the MS/MS data in order to minimize false matches to these peaks. In all cases, the data were searched twice using the search program SEQUEST,[28] in one case with no requirement that the peptides be tryptic, the other being constrained to only tryptic peptides. In both cases, the mass tolerance of the precursor peptide was set at ±0.1 Da, and the database was constrained to consider only peptides containing cysteine. MS/MS data were differentially searched for both d_0-modified cysteines (545.23 Da) and d_8-modified cysteines (553.28 Da). Peptides matched with SEQUEST correlation scores greater than 1.5 or delta correlation scores greater than 0.1 were considered as potential matches. All of these MS/MS spectra were manually checked to verify the validity of the results as described.[7] The standard mixture of proteins was searched against a database containing proteins derived from all of known *S. cerevisiae* ORFs, with the four standard proteins also added to this database. The human prostate samples were searched against an annotated database containing ~75 000 human proteins compiled at the Frederick Biomedical Supercomputing Center, Frederick, MD. Data from the IP study were searched against the *S. cerevisiae* ORF database.

RESULTS AND DISCUSSION

Description of the System. A flowchart of the abundance ratio-dependent protein identification method is detailed in Figure 1, which consists of the following steps: (1) Affinity-purified ICAT reagent-labeled peptides are separated by μLC, collected onto a MALDI sample plate, and inserted into the mass spectrometer. (2) The peptides deposited in each position of the sample plate are surveyed by MALDI-TOF MS and the data transferred to the quantification software. (3) Abundance ratios are determined for ICAT reagent-labeled normal (d_0) and heavy (d_8) peptides detected as peak pairs with a Δm of 8.05 Da per cysteine. Singlet peptide peaks (i.e., peptides derived from proteins that are detected in only one of the samples being compared) are identified by virtue of no accompanying peak detected with the expected Δm. These singlets are identified using signal-to-noise ratios and isotopic signature considerations, to eliminate singlet peaks caused by chemical noise. Potential singlet peaks due to non-ICAT reagent-labeled peptides are minimized using the optimized affinity selection protocol that provides greater than 90% enrichment for cysteine-containing peptides. (4) The software outputs a list

(26) Smolka, M. B.; Zhou, H.; Purkayastha, S.; Aebersold, R. *Anal. Biochem.* **2001**, *297*, 25−31.
(27) Zhang, R.; Sioma, C. S.; Wang, S.; Regnier, F. E. *Anal. Chem.* **2001**, *73*, 5142−5149.
(28) Eng, J.; McCormack, A. L.; Yates, J. R., 3rd. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.

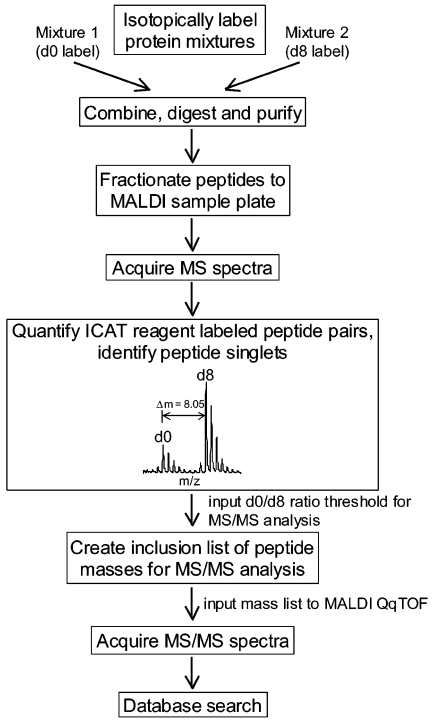

Figure 1. Flowchart of the automated abundance ratio-dependent protein identification method.

containing the mass and sample plate position of those peptides that show an abundance ratio that exceeds a user-defined threshold. (5) The precursor ion masses identified in step 4 are automatically selected for CID, and the resulting spectra are recorded. (6) Last, sequence database searching of the acquired tandem mass spectra is initiated to identify the peptide sequences and the proteins from which they are derived.

Validation of the Quantification Software. To demonstrate the effectiveness of the system for quantifying μLC-separated ICAT reagent-labeled peptides within a complex sample, we used a mixture of standard peptides at predetermined, varying abundance ratios within a matrix of constitutively represented peptides from yeast. Peptides showing abundance ratios close to 1 and also those showing abundance changes were identified by MS/MS analysis. Table 1 shows the quantification results for the identified peptides obtained from this experiment. The average absolute error of 3.1% and the average coefficient of variation of 15.4% for these data demonstrate the accuracy and the precision of the method in quantifying the ICAT reagent-labeled peptides. Two main factors contributed to the excellent performance of the system. First, the mass accuracy and the prevalence of singly charged peptide ions that are characteristic of MALDI-TOF mass measurement facilitated the detection of pairs of isotopically labeled peptides even within complex mass spectra. Second, the automated data acquisition of the MALDI QqTOF mass spectrometer reproduced the acquisition parameters for each sample spot including sample acquisition time, laser intensity, and spatial sampling.

One complicating factor in the quantification of peptides labeled with deuterated isotope tags such as the ICAT reagents used in this study is the tendency of the deuterated isotopically heavy (d_8) labeled peptide to elute several seconds earlier than the isotopically normal (d_0) labeled form of the same peptide[7,27] during the μLC separation. Consequently, eluting peptide pairs that are

Table 1. Automated Quantification of a Standard Protein Mixture

	d_0/d_8	
protein[a]	expected[b]	average
OVAL_CHICK (4)	2.7	2.9 ± 0.5
BGAL_ECOLI (5)	0.28	0.28 ± 0.07
LCA_BOVIN (5)	1.9	1.9 ± 0.2
PHS2_RABIT (2)	0.73	0.67 ± 0.09
yeast proteins (11)	1.0	1.0 ± 0.1

a Numbers in parentheses indicate the number of identified peptides from each protein. In the case of the yeast proteins, 11 unique peptides were identified and quantified. *b* For the standard proteins, the d_0/d_8 values were determined from the unfractionated mixture by MALDI-TOF measurement of relative intensities of tryptic peptides at specific M + H values known to be derived from each protein. The yeast proteins were isolated from a yeast lysate that was divided in half and labeled with the d_0 or d_8 ICAT reagent; thus the expected abundance ratio for these peptides was 1.0.

split into adjacent fractions on the MALDI plate may show dramatic differences in d_0/d_8 ratios between the two adjacent spots, potentially introducing large errors into the measurement of relative abundance ratios.[27] We accounted for this difference in retention time by scanning for peptides that had been split across sample spots and summing their signal intensities from each adjacent spot to determine an accurate abundance ratio over the entire peptide elution profile. It should be noted that this difference in retention time between deuterated and nondeuterated peptides significantly complicates precursor ion selection based upon relative abundance using "on-the-fly" LC−ESI MS/MS methods. The ability of a MALDI-based analysis platform to resample peptides contained within sample spots after their accurate quantification, without the time constraints inherent to automated LC-ESI-based methods, makes this platform ideally suited for abundance ratio-dependent analysis. Furthermore, the predominance of single-charged peptide ions generated by MALDI simplifies the quantitative analysis, compared to the multiple charge states of peptide ions produced by ESI.

Application to Protein Expression Profiling. To demonstrate the effectiveness of the method to quantitative protein profiling, we analyzed changes in nuclear protein abundance caused by androgen stimulation in androgen-dependent, cancerous human prostate cells. For the work described here, a single fraction from the cation-exchange HPLC separation of the starting nuclear protein mixture was further affinity purified to select ICAT reagent-labeled peptides. This sample served as a representative example of a complex mixture of mammalian peptides to test the utility of the system for abundance ratio-dependent protein identification in protein profiling studies. The experiment resulted in the automated detection of 306 putative ICAT reagent-labeled peptides of which 79 (26% of the total) showed abundance differences considered to be significant (Figure 2) by virtue of either their abundance ratio (48 peptides) or their identification as a peptide singlet (31 peptides). A theshold d_0/d_8 ratio of less than 0.67 or greater than 1.5 was considered significant, based on the demonstrated accuracy of the method using the standard peptide mixture. The amino acid sequences of 38 peptides were successfully identified by MS/MS analysis and sequence database searching, resulting in the identification of 36 unique proteins. For comparison, an aliquot of the same sample was also analyzed

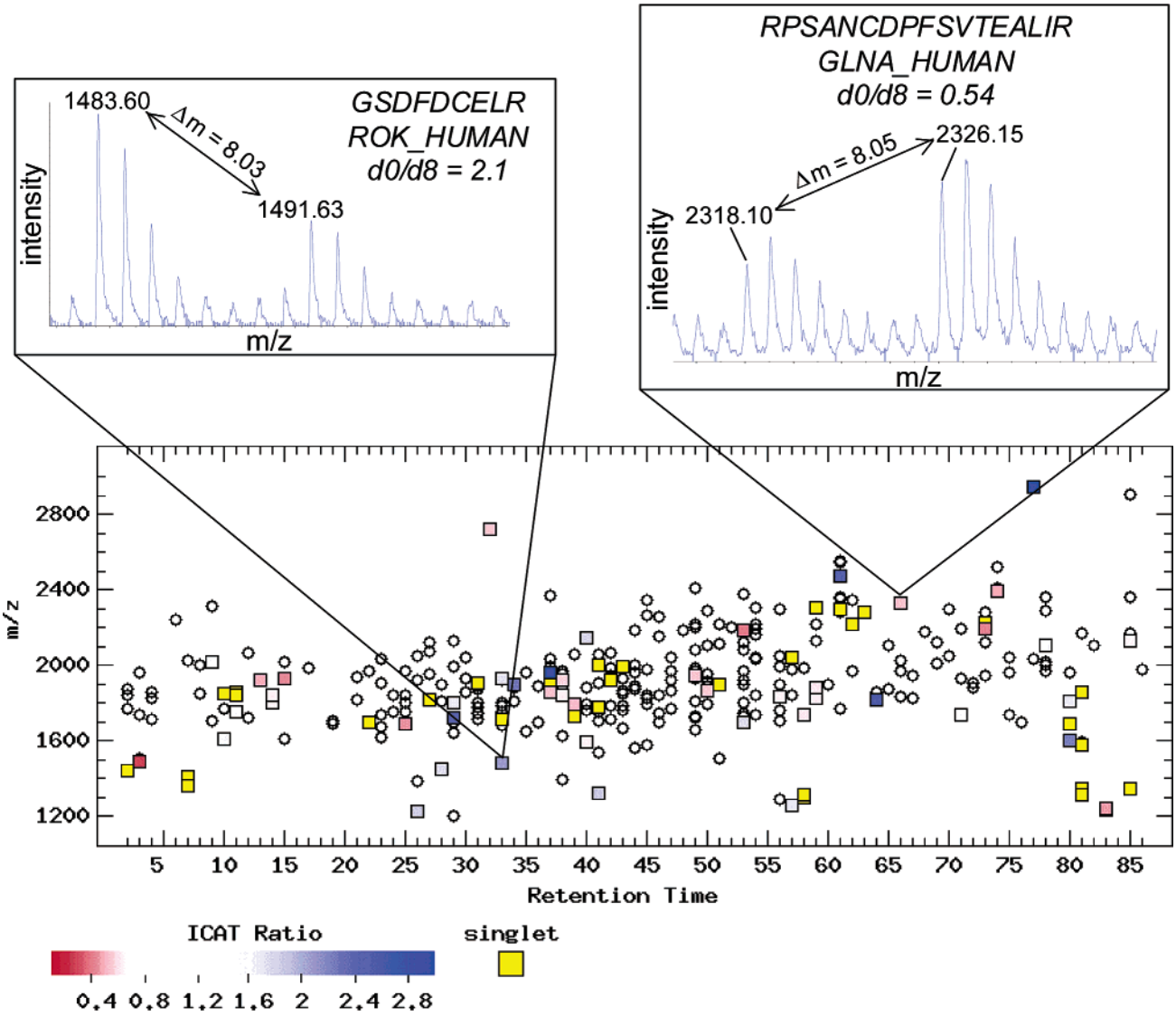

Figure 2. Protonated peptide masses (M + H) of automatically determined, putative ICAT reagent-labeled peptides derived from human prostate cells plotted against chromatographic retention time. Circles indicate constitutively represented peptides, while colored squares indicate peptides showing significant abundance changes. Representative results from two identified, differentially expressed peptides are shown along with the proteins from which they are derived.

by μLC ESI quadrupole ion trap MS/MS using data-dependent MS/MS.[5,7] The results from these two analyses are compared in Table 2. These results demonstrate several points. (i) The abundance ratio-dependent approach is dramatically more efficient than the data-dependent approach in detecting and analyzing differentially expressed peptides. This is evidenced by the difference in the ratio of the number of identified peptides to the number of acquired CID spectra (identification/CID ratio), which was 50% for the abundance ratio-dependent method compared to 16% for the data-dependent method. The identification/CID ratio for peptides showing significant abundance change was only 2.6% (28/1200) using the data-dependent method. Furthermore, despite the large number of acquired CID, there were still significantly more peptides present in the sample than the data-dependent method was able to identify, evidenced by comparison of the 195 identified peptides and the 309 signals detected using the MALDI QqTOF mass spectrometer as putative peptide pairs. The main consequence of this difference in efficiency is an increase in the number of peptides (and hence proteins) significant for the system being studied that are identified using the abundance-dependent method (Table 2). Comparison of other samples analyzed by both

Table 2. Comparison of Protein Profiling Results Obtained from a Model Prostate Sample Analyzed Using Data-Dependent and Abundance Ratio-Dependent Methods

	data dependent[a]	abundance ratio dependent[b]
total CID acquired	1200	79
peptides identified	195	38
differentially expressed peptides identified[†]	28 (26 proteins)	35[c] (33 proteins)
identification/CID ratio	16% (195/1200)	50% (38/79)
identification/CID ratio (differentially expressed)	2.6% (28/1200)	44% (35/79)

[a] The methods used are described in the text. [b] Identified peptides with d_0/d_8 ratios greater than 1.5 or less than 0.67 were considered to be differentially expressed. [c] Three of the 38 peptides identified were quantified incorrectly due to overlap with peptides of similar mass.

the abundance-dependent method and the data-dependent method have shown similar results (data not shown). (ii) Analysis of stable isotope-tagged peptides by MALDI TOF mass spectrometry provides increased quantitative accuracy. The accuracy of quantification using data generated by the ion trap mass spectrometer can be affected by the lower mass resolution afforded by that instrument and the data-dependent mode of operation in which the instrument alternates between MS and MS/MS scans.[7] These factors can result in decreased signal-to-noise (S/N) levels and erratic ion chromatograms. For the prostate sample analyzed here, 25 identified peptides gave ambiguous quantitative results. In contrast, the mass resolution of the TOF detector (>7000 $m/\Delta m$ at fwhm[17]) provided for improved S/N, increasing the accuracy of relative abundance measurements. Manual inspection of the spectra for the prostate peptide sequences identified using the MALDI QqTOF mass spectrometer revealed that all but three of the peptides showed reliable quantitative measurements. Measurement inaccuracies for the three peptides were caused by interfering peptides of similar mass that were not filtered out by the software algorithm. (iii) The approach provides potentially dramatically increased throughput as a consequence of both the increased efficiency and the highly accurate automatic quantitative analysis. Taken together, these factors can effectively decrease instrument and computer analysis time, as well as the amount of time consumed for the manual interpretation of the data.

Application to the Characterization of Protein Complexes. A common method for the identification of the components of protein complexes is the isolation of the whole complex by immunoprecipitation (IP), followed by its separation into individual components, typically by gel electrophoresis. Protein bands detected at a significantly higher intensity in the positive IP sample versus a suitable control IP are excised, digested, and identified by MS or MS/MS analysis.[12,13,29] To test whether the specific components of protein complexes could be identified by ICAT reagent labeling of the complex and a suitable control followed by abundance ratio-dependent MS/MS analysis, we immunoprecipitated epitope-tagged SRB4p, a subunit of the RNA polymerase II (PolII) holoenzyme in *S. cerevisiae*[30] and its associated proteins. A control sample was prepared by following the same isolation protocol using a cell lysate of a yeast strain that did not express epitope-tagged SRB4p. The samples were labeled with the d_8 and d_0 forms of the ICAT reagent, respectively. The d_8/d_0 ratio was used to guide the MS/MS analysis, in which peptides showing a d_8/d_0 ratio of >1.5 and singlet peaks were selected for CID, under the assumption that specifically associated d_8-labeled proteins would show increased abundance relative to the control. A portion of each protein sample was separated and visualized by gel electrophoresis. Inspection of the electropherogram shown in Figure 3A indicates the presence of many prominent, constitutively represented proteins present in the two samples, along with some bands at apparent higher staining intensities in the positive sample, mostly in the range above 100 kDa in size.

Table 3 summarizes the proteins identified that were determined to be associated with the SRB4 protein based on measured

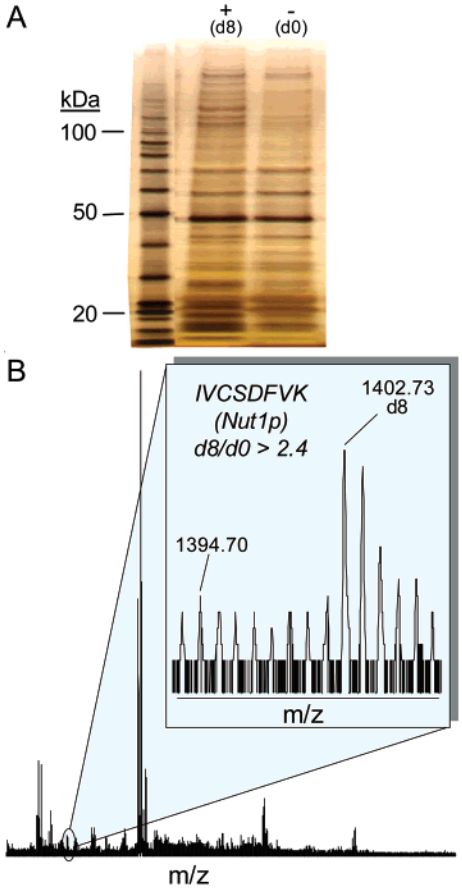

Figure 3. (A) SDS−PAGE gel showing the positive (+) and control (−) IP experiments for the isolation of SRB4p-associated proteins, which were labeled with the d_8 and d_0 forms of the ICAT reagent, respectively. (B) Representative result showing the MS scan from one μLC elution time point. The inset shows a peak that was automatically identified as an ICAT reagent-labeled singlet and identified by MS/MS analysis as a peptide derived from the NUT1 protein, a known component of the SRB mediator complex.[31]

abundance ratios. In all, 17 unique proteins were identified, 10 of which are either known components of the PolII holoenzyme or known to have a functional role in PolII transcription. Some of the identified proteins shown in Table 3 would not be expected to be associated with SRB4p, most notably the highly abundant translation factor TEF2, its associated proteins (Tef4p, Ded1p),[31] and the two ribosomal proteins identified. Further experimental evidence is necessary to test the biological relevance of these observed interactions. Many of the proteins presented in Table 3 have molecular masses consistent with bands of higher staining intensity in the positive IP lane of the gel (Figure 3A), most notably those above 100 kDa in size. However, there are proteins, such as Med4p and Rpb3p, migrating between 20 and 70 kDa on the gel, for which it would be difficult to establish specific complex association based on the gel band pattern alone. We also selected a portion of the peaks that were showing d_8/d_0 values close to unity for MS/MS analysis. These peptides were identified as being derived from either the conserved portion of the immunoglobulin antibody or nonspecifically bound, high-abundance proteins from yeast (bottom portion of Table 3). The results clearly indicate the

(29) Vorm, O.; King, A.; Bennett, K. L.; Leber, T.; Mann, M. In *Proteomics: A Trends Guide*; Mann, M., Blackstock, W., Eds.; Elsevier: London, 2000; pp 43−47.

(30) Thompson, C. M.; Koleske, A. J.; Chao, D. M.; Young, R. A. *Cell* **1993**, *73*, 1361−1375.

(31) Hodges, P. E.; Payne, W. E.; Garrels, J. I. *Nucleic Acids Res.* **1998**, *26*, 68−72.

Table 3. Srb4p Co-Immunoprecipitated Proteins

protein[a]	protein MW	d_8/d_0[b]	protein description[c]
Sin4p (2)	111 305	>2.4	component of RNA polymerase II holoenzyme and Kornberg's mediator subcomplex*[a]
Rif1p (1)	217 806	>2.9	protein involved in telomere length regulation and transcriptional silencing*[a]
Rsc3p (1)	101 722	>5.9	component of the abundant RSC chromatin remodeling complex*[a]
Nut1p (2)	128 750	>2.4	histone acetyltransferase (HAT), component of RNA polymerase II mediator (SRB) subcomplex*[a]
Rpb3p (1)	35 143	>2.9	RNA polymerase II, third-largest subunit*[a]
Rgr1p (1)	123 366	>5.0	component of RNA polymerase II holoenzyme and Kornberg's mediator (SRB) complex*[a]
Rpb2p (1)	138 600	>2.5	RNA polymerase II, second-largest subunit*[a]
Med4p (2)	32 048	5.0	component of RNA polymerase II holoenzyme and mediator subcomplex*[a]
Rpb10p (1)	8 147	>2.0	shared subunit of RNA polymerases I, II, and III*[a]
Rpn13p (1)	17 758	>5.0	component of the 19S regulatory cap of the 26S proteasome complex, also functions in RNA polymerase II*[a] transcription elongation
Ded1p (1)	65 422	>2.9	ATP-dependent RNA helicase of DEAD box family involved in protein synthesis, associated with Tef1p
Rpl4ap (1)	39 022	3.2	ribosomal protein L4
Tef2p (7)	49 912	2.4	translation elongation factor EF-1α
Rpl23A/Bp (1)	14 485	1.8	peptide derived from identical ribosomal proteins L23A and L23b
Tef4p (1)	46 389	2.5	translation elongation factor EF-1γ
Glt1p (1)	234 874	1.8	glutamate synthase
Ydjp (1)	44 379	1.5	protein involved with protein import into mitochondria and ER
Ig (8)[d]		0.83	immunoglobulin, conserved regions
Cdc19p (2)[d]	54 412	0.91	pyruvate kinase
Tdh3p (1)[d]	35 612	0.91	glyceraldehyde-3-phosphate dehydrogenase 3
YAR010C (1)[d]		0.83	abundant transposable element, common to multiple proteins

[a] The names of each identified protein are given, with the number of peptides identified from each protein shown in parentheses. [b] The automatically determined abundance ratio is given for each protein. The "greater than" sign (>) in front of a value indicates ICAT reagent singlets, where the value shown was determined by measuring the S/N of the peak singlet. Average d_8/d_0 values are given for proteins from which multiple peptides were identified. [c] Protein descriptions are from the Yeast Protein Database (www.proteome.com).[31] Proteins previously described as having a role in transcription are indicated with an asterisk. [d] Identified proteins from which peptides were selected for MS/MS analysis that showed d_8/d_0 values close to 1.

effectiveness of this approach to automatically identify specifically associated components of protein complexes isolated by IP from within a prominent background of nonspecific contaminants and demonstrate its potential for the rapid characterization of affinity-isolated protein complexes.

One limitation of the method in its current form is the quality of CID spectra of single-charged peptide ions generated by MALDI,[32] which was the main factor in the inability to identify some of the MALDI MS/MS spectra via sequence database searching. Chemical modifications of peptides[22,33] to increase the extent and uniformity of fragmentation along the peptide backbone has great potential to dramatically improve the CID quality of MALDI-generated peptides and increase the effectiveness of the database searches. Derivatization of the prostate cell peptides using a novel imidazole reagent[22] helped to improve the MS/MS spectral quality of lysine-terminated peptides relative to that observed from guanidinated or underivatized peptides. The use of alternative isotopic-labeling procedures[34] that replace the existing, relatively large ICAT reagent label also may provide improved performance. Furthermore, alternate protein labeling chemistries that target other amino acid residues and therefore increase peptide coverage of identified proteins would be beneficial to the method. The two expected cysteine-containing, tryptic peptides from SRB4p were not identified in the IP experiment. Reasons for failing to detect those peptides include incomplete labeling with the ICAT reagent, miscleavage of the peptide by trypsin resulting in a peptide having a mass outside the range of the quadrupole TOF mass spectrometer used, or lack of ionization and detection in the MALDI mass spectrometer. Regardless, labeling chemistries targeting other amino acid residues would increase the probability of detecting more proteins.

These limitations notwithstanding, the results presented here demonstrate the effectiveness of abundance ratio-dependent protein identification as a general approach to the characterization of functionally interesting proteins. This should be highly useful for the rapid and accurate identification of clinically important diagnostic disease markers and potential therapeutic targets and also for measuring the changes in the composition of protein complexes induced by perturbations of the cells from which they were isolated. The increased efficiency realized by abundance ratio-dependent identification also has great potential for the investigation of the dynamics of protein expression, making more highly resolved time course or dose response studies a possibility.

(32) Cramer, R.; Corless, S. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2058−2066.
(33) Keough, T.; Lacey, M. P.; Strife, R. J. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2227−2239.
(34) Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. *Nat. Biotechnol.* **2002**, *20*, 512−515.

Additionally, the approach is amenable to any of the hybrid MALDI-TOF MS/MS instruments that continue to be introduced[35] and should find utility with any of these high-speed analysis platforms. As such, the software routine described here for precursor ion quantification has been developed as a stand-alone, generic tool amenable to data from any MALDI instrument and is freely available by request. Finally, the mass accuracy afforded by the TOF detection may also facilitate the identification of proteins by measurement of intact peptide mass in conjunction with other constraints,[36] obviating the need for MS/MS analysis and improving the sensitivity and throughput of proteomic analyses.

(35) Griffin, T. J.; Goodlett, D. R.; Aebersold, R. *Curr. Opin. Biotechnol.* **2001**, *12*, 607−612.
(36) Goodlett, D. R.; Bruce, J. E.; Anderson, G. A.; Rist, B.; Pasa-Tolic, L.; Fiehn, O.; Smith, R. D.; Aebersold, R. *Anal. Chem.* **2000**, *72*, 1112−1118.

ACKNOWLEDGMENT

The authors thank Lyle Burton and Jimmy Eng for their assistance in this work, and Eric Peters for generously donating the imidazole lysine-modifying reagent. T.J.G. was supported in part by a postdoctoral Genome Training Fellowship from the NIH. J.A.R. was supported by an NIH postdoctoral fellowship, GM19884. The work was supported in part by Grants 1R33CA84698 and 1R33CA93302 from the National Cancer Institute, Grant RR11823 from the National Center for Research Resources, and a funding partnership with Oxford Glycosciences.

Received for review September 11, 2002. Accepted December 13, 2002.

AC026127J