ELSEVIER

Journal of Chromatography B, 782 (2002) 291–306

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Inverse labeling–mass spectrometry for the rapid identification of differentially expressed protein markers/targets

Y. Karen Wang*, Douglas F. Quinn, Zhixiang Ma, Emil W. Fu

Central Technologies, Drug Discovery Research, Novartis Pharmaceuticals Corporation, 556 Morris Avenue, Summit, NJ 07901, USA

Abstract

Comparative proteomic studies can lead to the identification of protein markers for disease diagnostics and protein targets for potential disease interventions. An inverse labeling strategy based on the principle of protein stable isotope labeling and mass spectrometric detection has been successfully applied to three general protein labeling methods. In contrast to the conventional single experiment approach, two labeling experiments are performed in which the initial labeling is reversed in the second experiment. Signals from differentially expressed proteins will distinguish themselves by exhibiting a characteristic pattern of isotope intensity profile reversal that will lead to the rapid identification of these proteins. Application of the inverse labeling method is demonstrated using model systems for protein chemical labeling, protein proteolytic labeling, and protein metabolic labeling. The methodology has clear advantages which are illustrated in the various studies. The inverse labeling strategy permits quick focus on signals from differentially expressed proteins (markers/targets) and eliminates ambiguities caused by the dynamic range of detection. In addition, the inverse labeling approach enables the unambiguous detection of covalent changes of proteins responding to a perturbation.

Keywords: Inverse labeling; Isotope labeling; Proteomics; Protein markers

1. Introduction

Complementing the field of genomics, proteomics is designed to elucidate protein level information of any cell type, tissue or whole organism. Comparative proteomics is a powerful tool to study changes in protein expression inherent to the developing pathophysiology of cells/tissues under different physiological conditions. It has been well established that almost all therapeutic intervention strategies involve targeting and modulating protein function and activity. Proteomics based approaches promise to revolutionize the study and treatment of individual disease processes by discovering new molecular markers for diagnostic profiling of the diseases, and by deciphering the biological pathways that lead to the diseases to identify potential points of intervention.

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is still the foundation of most proteomics studies. Major shortcomings of the technique include protein coverage, dynamic range, and automation capability. These drawbacks have prompted intensive recent effort in the development of alternative methodologies. One promising methodology to emerge from such effort is the application

*Corresponding author. Tel.: +1-908-277-7022; fax: +1-908-277-4910.

E-mail address: karen.wang@pharma.novartis.com (Y.K. Wang).

1570-0232/02/$ – see front matter
PII: S1570-0232(02)00561-5

of protein stable isotope labeling with mass spectrometric detection for gel-free proteomics. When using this approach for comparative proteomics of two protein pools (a disease vs. a normal or a drug treated vs. a control), the two protein pools are separately labeled, one with a heavy isotope label and the other with a light isotope label. The two pools are then mixed, processed, and analyzed by mass spectrometry. The signal intensity ratio of the light and heavy of an isotopic pair provides quantitative comparison of protein expression, whereas information obtained by mass spectrometry (MS) analysis (peptide finger print or sequence information) is used to search protein databases for protein identification. Application examples have been demonstrated based on this principle employing protein metabolic labeling [1,2], chemical tagging [3–8], and proteolytic labeling [9–11].

Correlation of isotopic pairs is a key step in the application of these approaches for differential comparison of protein expression. The conventional single-experiment approach relies on the use of a predefined mass difference. The intensive task of searching/fitting isotopic pairs of the predefined mass difference is accomplished with computer software. This approach has been successfully used with chemical tagging such as ICAT (Isotope Coded Affinity Tag), in which isotopic labeling occurs on cysteines with an 8-Da mass difference per modification. The same principle can be extended to protein proteolytic ^{18}O labeling, in which up to two ^{18}O atoms are incorporated into the C-terminus of peptides. The result is a 4-Da mass difference between isotopic counterparts. However, in the case of protein metabolic labeling (^{15}N, ^{13}C, or ^{2}H), correlating isotopic pairs via such an approach is not viable. Mass difference between isotopic pairs is sequence dependent and varies from peptide to peptide. The variable mass difference cannot be calculated based on molecular mass (M_r) (rather, it is determined by the total number of nitrogen, carbon, or hydrogen atoms in a sequence). Therefore, the conventional method based on the use of a predefined mass difference cannot be employed to obtain isotopic peak correlation.

High data complexity is another characteristic of gel-free proteomics. Efforts have been made to reduce mixture complexity and, as a result, analysis and data interpretation are simplified. Tags with a built-in affinity function offer the capability of isolating/enriching the tagged peptides and, consequently, lead to a dramatic reduction in peptide mixture complexity (e.g., ICAT). However, such approaches may yield a final mixture in which not all the initial proteins are represented. In the case of ICAT, information on proteins free of cysteine is completely lost. In addition, information associated with proteolytic peptides that are free of cysteine is lost. This may include important information such as protein processing or protein post-translational modifications that may be crucial in protein activity regulation, translocation, or signal transduction pathway regulation.

An alternative approach to simplify data interpretation is through data reduction. While the entire mixture is analyzed, data analysis is performed only on a small portion of the data after dramatic data reduction. Irrelevant signals are removed by data reduction. These include signals from proteins of no differential changes or from the regions of sequences in which neither qualitative nor quantitative changes have occurred. Such an approach is highly desired since it provides data complexity reduction while maintaining information integrity (assuming the dynamic range is accommodated through comprehensive fractionation).

Inverse labeling was introduced by Wang et al. in the context of applying protein proteolytic ^{18}O labeling to comparative proteomic studies [10]. When using this approach, signals from differentially expressed proteins can be rapidly identified. The approach offers dramatic reduction in the amount of work spent on data analysis and it eliminates ambiguity in data interpretation. Not only quantitative changes in expression, but also qualitative changes, such as protein modification, can be identified. In this report, we demonstrate that the approach can be applied to any protein labeling method. A predefined isotopic mass difference is not needed when using the method. The method works well for chemical tagging and proteolytic ^{18}O labeling where isotopic mass difference is well defined. It is essential in cases where isotopic mass deviations vary such as for protein metabolic labeling. Rapid and unambiguous identification of differentially expressed protein markers and targets can be achieved in all cases.

2. Experimental

2.1. Inverse chemical ICAT labeling

2.1.1. The six-protein model system

Commercial proteins of bovine serum albumin (BSA), aldolase, β-casein, apo-transferrin, β-lactoglobulin, and cytochrome *c* (Sigma) were used without further purification. The six proteins were mixed at a molar ratio of 1:1:1:1:1:1 for the ''control'' and 0.3:3:1:1:1:1 for the ''treated'' pool. The ICAT reagents were purchased from Applied Biosystems (Cambridge, MA, USA). The vendor protocol was followed. In brief, the control and treated protein mixtures were first reduced and denatured. ICAT tagging was carried out such that half of each mixture was reacted with D_0 and the remaining half with D_8. The inverse labeling was subsequently achieved by mixing the D_0-control with the D_8-treated, and the D_8-control with the D_0-treated. Trypsin digestion was then performed on both mixtures at 1:50 (w/w) trypsin-to-protein ratio for ~16 h at 37 °C (modified trypsin from Promega, sequencing grade). The resulting peptide mixtures first went through a cation-exchange column to remove the excess reagents. The labeled peptides were isolated using an avidin column via affinity isolation. Aliquots were taken from each pool and dried using a SpeedVac. They were reconstituted with mobile phase A [water with 0.1% formic acid+ 0.01% trifluoroacetic acid (TFA)] prior to liquid chromatography (LC)–MS and matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) MS analyses.

2.2. Inverse proteolytic ^{18}O labeling

2.2.1. The eight-protein model system

Commercial proteins of BSA, aldolase, carbonic anhydrase, β-casein, chicken albumin, apo-transferrin, β-lactoglobulin, and cytochrome *c* (Sigma) were used without further purification. The eight proteins were mixed at a molar ratio of 1:1:1:1:1:1:1:1 for the ''control'' and 0.3:3:1:1:1:1:1:1 for the ''treated'' pool. Two identical aliquots, containing 10 pmol each of the unchanged components, were taken from each pool and were dried using a Speedvac. The ^{18}O labeling was performed using two procedures, one during-proteolysis and the other post-proteolysis. In the during-proteolysis labeling, for both control and treated pool, one of the dried aliquots was reconstituted with 20 μl of regular water and the other with 20 μl of ^{18}O water, both containing 50 m*M* ammonium hydrogencarbonate. Trypsin was added at a 1:100 (w/w) trypsin-to-protein ratio to each solution and digestion was allowed to proceed at 37 °C for ~20 h. For the post-proteolysis labeling, trypsin digestion was performed in regular water–ammonium hydrogencarbonate buffer at the same trypsin to protein ratio for ~12 h for all aliquots. The resulting peptide mixtures were then taken to complete dryness using a Speedvac. To each aliquot of the dried peptide mixture, 10 μl of ^{18}O or regular water was added for the post-proteolysis ^{18}O labeling. The labeling process was allowed to proceed at room temperature for ~12 h (additional trypsin may be added, and a longer incubation at 37 °C may help to increase the labeling yield). For both during-proteolysis and post-proteolysis labeling, prior to analysis, the ^{16}O-control sample was mixed with the ^{18}O-treated sample and the ^{18}O-control sample was mixed with the ^{16}O-treated sample. MS analysis using both MALDI and electrospray LC–MS was performed on both peptide mixtures.

2.2.2. The whole cell lysate spiked with PTP protein

Approximately 5×10^7 harvested Chinese hamster ovary (CHO) cells were lysed mechanically (freeze–thaw) using a buffer containing 10 m*M* Tris, 1 m*M* EDTA, pH 7.4. The resulting cell lysate of 2.5 ml at 0.4 mg/ml protein concentration was divided into four aliquots. Two were spiked with 10 pmol of PTP-1B protein (internally expressed, residue 1–298) (PTP10) and the other two with 30 pmol of PTP-1B (PTP30). Trypsin was added to each solution at a 1:100 (w/w) trypsin-to-total protein ratio to initiate the digestion. Proteolysis was allowed to proceed at 37 °C for ~12 h. The resulting solutions were centrifuged and the solid discarded. The solutions were then taken to complete dryness with a Speedvac. For both PTP10 and PTP30, one of the two identical aliquots was reconstituted with 10 μl of ^{18}O water, the other with 10 μl of regular water. The post-proteolysis ^{18}O incorporation was allowed to proceed at room temperature for ~12 h. Prior to

analysis, the ^{16}O-PTP10 and ^{18}O-PTP30 samples were mixed, as were the ^{18}O-PTP10 and ^{16}O-PTP30 samples. Each mixture was diluted with 100 μl of mobile phase A and filtered through a 0.4-μm Microcon filter. The filtrate was injected into the LC–MS system for analysis.

2.3. Inverse metabolic ^{15}N labeling

2.3.1. The two-protein model system

Regular (^{14}N-) and ^{15}N-labeled PTP protein (1–298) and regular and ^{15}N-labeled HtrA protein (161–373) were prepared in the laboratory using standard culture conditions with the ^{15}N-labeled proteins being produced by fermentation in ^{15}N-enriched culture media. The authenticity of the proteins and the level of isotope incorporation were assessed by MS on the final protein products. The labeling yield was greater than 90% for both proteins according to MS results. The two-protein model system was made by mixing the two proteins PTP and HtrA. The ^{14}N-mixture consisted of the two ^{14}N-proteins, and the ^{15}N-mixture consisted of the two ^{15}N-proteins. The ''control'' was a mixture of the two proteins at a molar ratio of 1:1. The ''treated'' or ''altered state'' material was made to mimic ''protein differential expression'' for PTP protein while the level of ''expression'' of HtrA protein remained unchanged. The molar ratios of PTP:HtrA for the ''treated'' mixture was 3:1 mimicking a threefold up-regulation for PTP. To achieve inverse labeling, an aliquot of ^{14}N-control was mixed with an aliquot of ^{15}N-treated (each containing the same amount of HtrA protein) and, in the same fashion, a ^{15}N-control was combined with a ^{14}N-treated. Trypsin digestion was subsequently carried out on both mixtures at a 1:50 (w/w) trypsin-to-protein ratio at 37 °C for ~7 h in 50 m*M* ammonium hydrogencarbonate buffer (the two proteins are known to digest readily under this condition without prior reduction and alkylation). MS analysis was performed on the two peptide mixtures using both MALDI and electrospray LC–MS.

2.3.2. The algal cell lysate spiked with PTP protein

A crude ^{13}C-algal protein extract and ^{13}C-^{15}N-algal protein extract were purchased from Isotec (Miamisburg, OH, USA). A 1-ml volume of a solution containing 6 *M* Guanidine·HCl, 50 m*M* Tris, 50 m*M* NaCl, pH 7.4 was added to 10 mg each of the algal crude extracts. The mixtures were vortexed and sonicated for 40 min to solubilize the proteins. After centrifugation for 20-min at 20 000 rpm, the supernatants were removed for further use while the large amount of insoluble material was discarded. 10 m*M* dithiothreitol (DTT) was added to each solution and the reduction was allowed to proceed for 1 h at 50 °C. Cysteine alkylation was then initiated with the addition of 20 m*M* iodoacetic acid sodium salt followed by shaking at room temperature in the dark for 1 h. A Centricon filter of 10 000 molecular mass cutoff was used to remove the excess reagents and to exchange buffer to 50 m*M* ammonium hydrogencarbonate. Protein concentration of the extracts was measured using the standard Bradford method. 10 pmol of regular PTP protein was spiked into an aliquot of 50 μl ^{13}C-algal protein extract containing about 0.05 mg of total protein to form the ^{14}N-control, and 10 pmol of ^{15}N-PTP was spiked into an aliquot of 50 μl ^{13}C-^{15}N-algal protein extract containing about 0.05 mg of total protein to form the ^{15}N-control. For the ''treated'' samples, a threefold down-regulation of PTP was created by spiking 3 pmol of PTP into an identical aliquot of algal extract. The ^{14}N-material was the result of ^{14}N-PTP being spiked into an aliquot of ^{13}C-algal extract. The ^{15}N-material was produced by spiking ^{15}N-PTP into an aliquot of ^{13}C-^{15}N-algal extract. The inverse labeling experiments proceeded in the same manner by combining aliquots of ^{14}N-control with ^{15}N-''treated'' and ^{15}N-control with ^{14}N-''treated''. Trypsin digestion of both mixtures was performed at a 1:100 (w/w) trypsin-to-protein ratio at 37 °C for ~16 h in 50 m*M* ammonium hydrogencarbonate buffer. The resulting peptide mixtures were analyzed by electrospray LC–MS.

2.3.3. LC–MS and LC–MS–MS peptide analyses

The LC–MS analysis of peptide mixtures was carried out through LC–ESI-MS using a Finnigan LCQ ion trap mass spectrometer. A 150×1.0 mm Vydac C_{18} column was employed for on-line peptide separation with a gradient of 2–2–20–45–98–98% B at 0–2–10–65–66–70 min. Mobile phase A was water with 0.1% formic acid+0.01% TFA and

mobile phase B was acetonitrile with 0.1% formic acid+0.01% TFA. The flow-rate was 50 μl/min. After elution from the LC column, the flow was split 9:1 with about 5 μl/min going into the MS system and 45 μl/min being collected for later use. The LCQ ion trap mass spectrometer was operated in a data-dependent mode, automatically performing MS–MS on the most intense ion of each scan when the signal intensity exceeded a pre-set threshold. When needed, the collected samples were concentrated and re-analyzed to obtain MS–MS data that were not collected automatically in the first run for the peptides of interest. The relative collision energy was set at 45% at which, in our experience, most peptides fragment effectively. A 5–8-Da window for precursor ion selection was employed. In all cases, aliquots of samples containing 20 pmol each of the unchanged components (10 pmol light isotope labeled and 10 pmol heavy isotope labeled) were used for the LC–MS analysis.

2.3.4. MALDI-TOF-MS peptide analysis

Aliquots of the reconstituted, previously dried samples were diluted ~1:5 using the MALDI matrix solution (α-cyano-4-hydroxy cinnamic acid 25 mg/ml in water–acetonitrile 50:50 with 0.1% TFA). About 1 μl of the final solution (containing about 500 fmol each of the unchanged components) was loaded onto the MALDI target for analysis. The analysis was performed on a Bruker REFLEX III MALDI-TOF mass spectrometer operated in the reflectron mode with delayed ion extraction. When applicable, post source decay (PSD) was also performed on the peptide ions of interest.

2.3.5. Database searches

Search software PROWL (Proteometrics, New York, NY, USA) and MASCOT (Matrix Science, London, UK) were used to search the protein databases to identify proteins using peptide fingerprints, MS–MS fragments, and processed PSD spectra. For searches using peptide fingerprint information, peptide ions exhibiting the inverse labeling pattern were sorted out based on the direction of isotope pattern swap between the two inverse labeling experiments (from light isotope being stronger in the first experiment to heavy isotope being stronger in the second experiment for a down-regulation and a change in the opposite direction for an up-regulation). The mass difference between isotopic pairs may be multiples of 8-Da for ICAT, 2 or 4-Da for proteolytic ^{18}O (mostly 4-Da), and a variable in the range of 0.6–2.6% of the m/z at detection for metabolic ^{15}N. Each list of peptide masses, of up- or down-regulation, was then used for a database search to identify the proteins. For searches using peptide sequence information, the two MS–MS spectra from the two inverse labeling experiments were compared. Similarity in the fragmentation patterns provided confirmation of the isotopic correlation of the two precursor ions. MS–MS data of the light isotope labeled peptides were used to search a protein database for protein identification. An iterative search combining the data of ions with inverse labeling pattern from peptide map and MS–MS could also be performed. Any ions that demonstrated a clear inverse labeling pattern in the map and were further supported by similar fragmentation in MS–MS data were identified first using their MS–MS data. The peptides derived from the identified proteins were then removed from the list and a second round search was initiated using the masses of the remaining peptides of inverse labeling pattern. For ions in which no conclusion could be made, a second analysis was performed using the collected sample to obtain MS–MS data. The resulting data was used in the same manner to search the databases for protein identification.

3. Results and discussion

3.1. Inverse labeling and inverse labeling pattern

As depicted in Fig. 1, the rapid identification of differentially expressed proteins is achieved by quick identification of their peptides through the characteristic inverse labeling pattern that they exhibit between the two inverse labeling experiments. Protein expression of the vast majority of proteins is unaffected following a perturbation which is reflected by a similar abundance profile between pool 1 and pool 2. Therefore, for peptides from those proteins, there will be no significant difference in the labeling pattern between the two inverse labeling experiments (i.e., similar abundance of light isotope

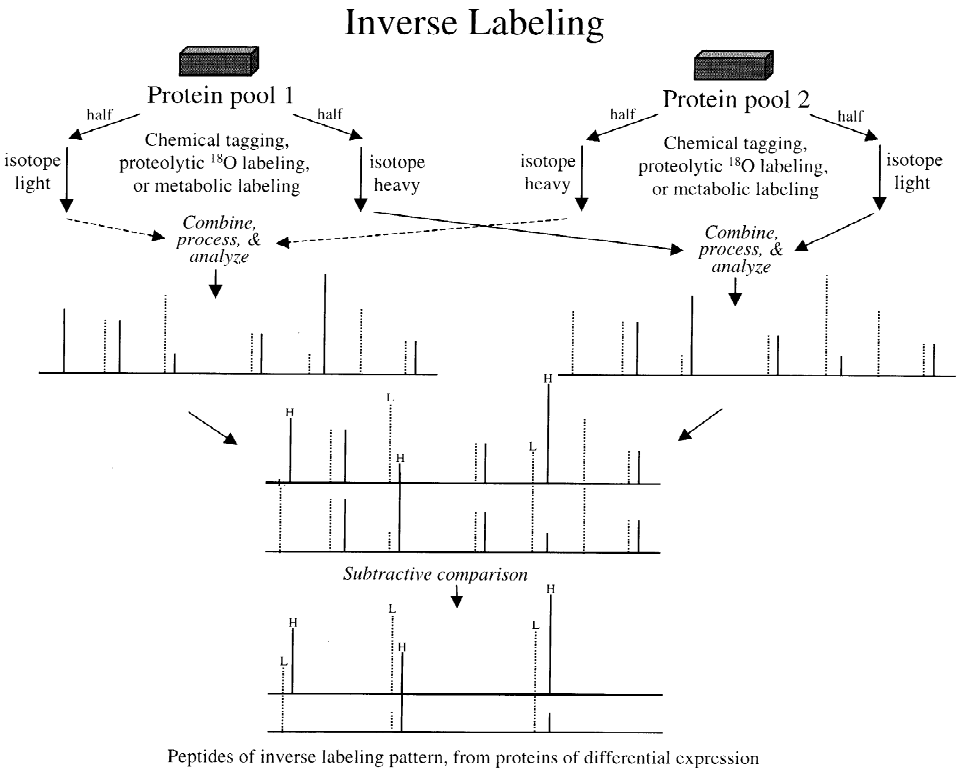

Fig. 1. The inverse labeling–mass spectrometry method for the rapid identification of marker/target proteins.

and heavy isotope signals in both experiments). And these signals can be subtracted out, in principle, by the comparative analysis of the two data sets. Chemical backgrounds and peptides without labeling are subtracted out as well. For a protein whose level of expression has been significantly up- or down-regulated by the perturbation, changes in the relative intensity of isotopically light and isotopically heavy signals will be observed. When the control pool is isotopically light labeled and the perturbed pool is isotopically heavy labeled, the heavy isotope signals of the resulting peptides will be of greater intensity than its light isotope signals if the protein is up-regulated as a consequence of perturbation. Conversely, the light isotope signals of the resulting peptides will be stronger if a down-regulation of protein has occurred upon perturbation. In our parallel or inverse analysis, the protein pool labeling is reversed and differentially expressed proteins will give rise to peptides that display an inverse isotopic intensity profile when contrasted with the first analysis. Thus, there is an apparent isotope swap for those peptides between the two inverse labeling experiments when the most intense isotopic ions are compared (i.e., from heavy isotope signal being stronger in the first experiment to light isotope signal being stronger in the second experiment or vice versa). A strength of the procedure lies in the fact that one does not need to look for the isotope pair of a predefined mass difference and to calculate quantitatively the ratio of light to heavy isotopic signals for every peptide in order to achieve differential comparison. One only needs to compare the two data sets and identify peptides of the characteristic isotopic intensity profile reversal (or the apparent isotope swap), which can be achieved rapidly and, potentially, automatically. All signals from differentially expressed proteins, either at quantitative level

or qualitative level, will give rise to such characteristic inverse labeling patterns. A predefined isotopic mass difference (or the apparent mass shift between the two experiments) is not a prerequisite in achieving the identification. Signals of any isotopic mass difference will be revealed after the subtractive cleanup of signals from proteins of no differential expression allowing variable mass differences to be identified all at once. Such examples include those due to variations in number of residues in a sequence susceptible to tagging. Inverse labeling offers an elegant way of overcoming the technical difficulties encountered by those labeling methods that result in variable isotopic mass differences, such as metabolic labeling. Any peptide from proteins of significant differential expression will display such characteristic inverse isotopic intensity profile, including those associated with extreme changes in expression or from proteins of covalent modifications, where unpaired isotope peaks are detected. Peptides derived from proteins with no expression deviations will not display a shift in isotope intensity pattern. The direction of the isotope intensity pattern reversal implicates the direction of differential expression of the protein (i.e., down-regulation or up-regulation).

3.2. Inverse chemical ICAT labeling

The inverse labeling and MS analysis were performed in the same manner as shown in Fig. 1 on the six-protein model system where BSA was ''down-regulated'' by threefold and aldolase ''up-regulated'' by threefold. MALDI-TOF-MS was performed directly on the mixture without any separation. Although the spectra was quite complex, it was clear that the inverse labeling strategy could help to quickly identify the peptide signals derived from proteins of differential expression. The alternative to inverse labeling would involve evaluation of a single spectrum (e.g., Fig. 2A) looking for the $\pm 8/16/24$-Da peak for each and every peptide to correlate isotopic pairs and, then, to calculate their intensity

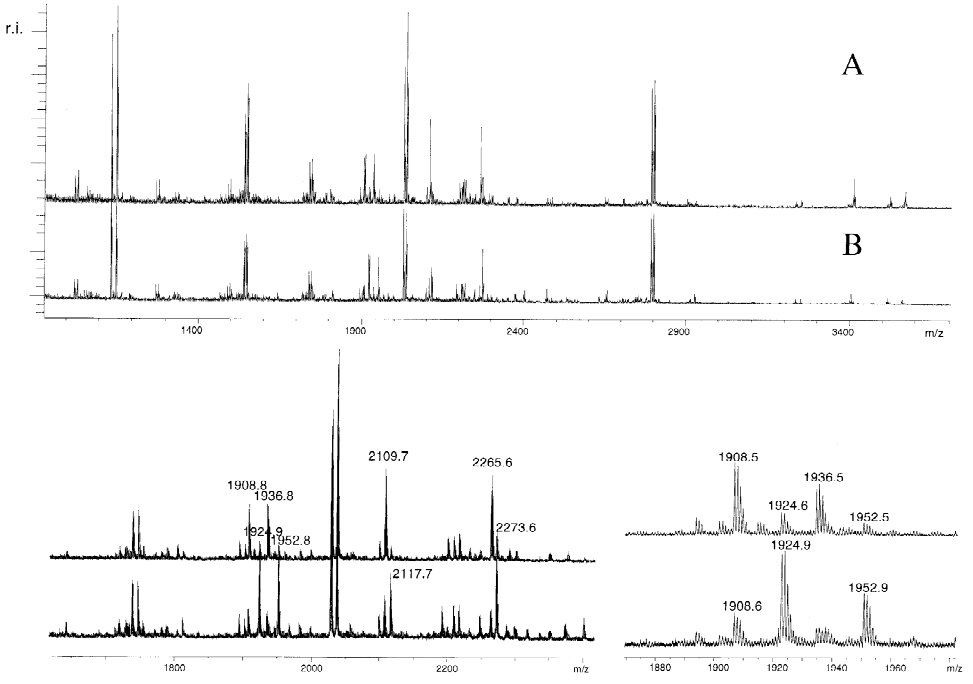

Fig. 2. MALDI-TOF detection of tryptic digests of the six-protein system. (A) D_0 control–D_8 ''treated'' sample; (B) D_8 control–D_0 ''treated'' sample, The lower panels are the selective zoomed-in m/z regions. The mass shifts or D_0/D_8 intensity ratio reversal indicates differential expression of proteins.

ratios for differential comparison. Utilizing the inverse labeling strategy one only needed to overlay the two spectra (Fig. 2A, B) and perform ''zoom and pick'' to identify the peaks that showed the characteristic inverse isotope pattern between the two spectra. Peaks of this characteristic inverse labeling pattern were thus quickly identified (e.g., the peaks with mass labels in Fig. 2), regardless of how many cysteines in a sequence or how much the mass difference would be between an isotopic pair (Fig. 2, 8-Da and 16-Da). Peptides from proteins of differential expression were quickly discovered by performing this simple qualitative comparison. Accurate quantitation on differential expression and MS–MS analysis for protein identification could then be performed on these peptides. When the same samples were analyzed using an LCQ with on-line reversed phase (RP) LC, this characteristic inverse labeling pattern was clearly detected on quite a number of peptides. A quick comparison of the two base-peak ion chromatograms (Fig. 3A, B) led to the rapid identification of the high intensity peptides of inverse labeling pattern (or of apparent mass shifts when the most intense isotopic ions are compared). Obviously, in order to identify less abundant peptides with inverse labeling patterns and those that co-elute with more abundant peptides, one would have to process all the MS data. The apparent mass changes due to the isotope swap will vary depending on the number of cysteines in the sequence and the charge state of the peptide being detected. Once such a change is detected, the mass shift value (i.e., isotope mass difference) can be used to elucidate the number of cysteine residues. This information can then be used in a protein database search to increase the search specificity and efficiency.

Following data analysis, two lists of peptide masses were quickly generated, segregated on the basis of direction of the isotope intensity swap. A light-to-heavy change is a shift from the light isotope signal being stronger in the first experiment to the heavy isotope signal being stronger in the inverse

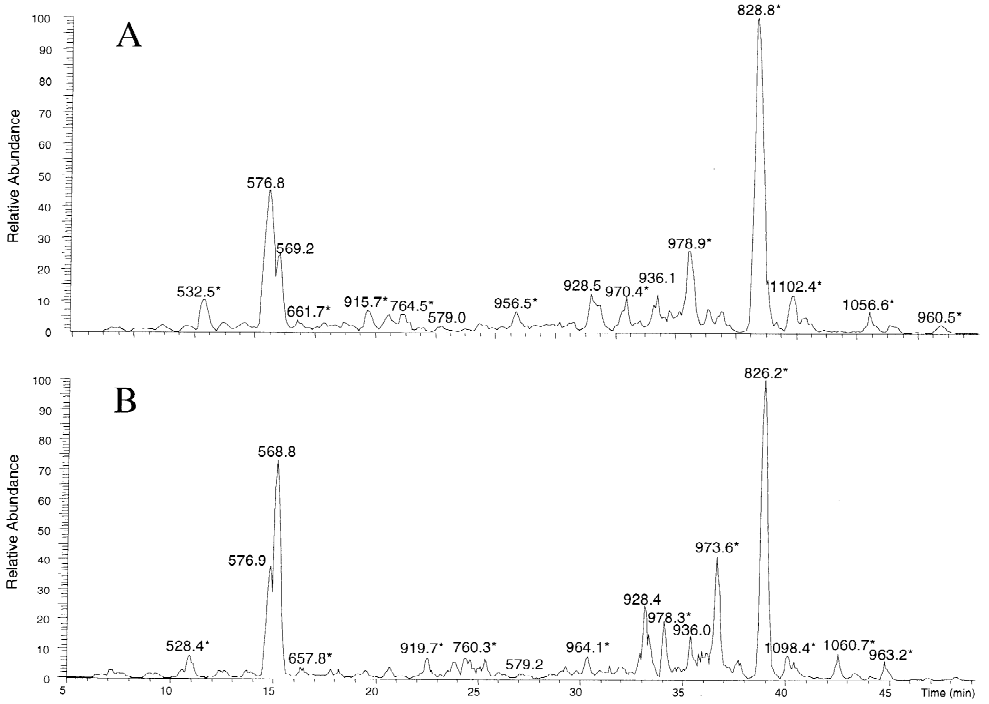

Fig. 3. LC–MS detection of tryptic digests of the six-protein system. (A) Base-peak ion chromatogram of the D_0 control–D_8 ''treated'' sample; (B) base-peak ion chromatogram of the D_8 control–D_0 ''treated'' sample, Signals of the characteristic inverse labeling pattern are clearly detected (marked with *). The differentially expressed proteins are quickly identified using the MS data.

experiment. A heavy-to-light change is just a shift in the opposite direction. These two lists were used to search the database. Aldolase was exclusively identified using the list of heavy-to-light shift, corresponding to an up-regulation of protein expression. BSA was identified using the list of light-to-heavy change, corresponding to a down-regulation in protein expression. MS–MS spectra were obtained automatically in data-dependent mode for a number of peptides. In order to emulate a broad-spectrum situation, where multiple proteins may be up- or down-regulated, an iterative search scheme was also applied. In this case we used the combined mass list of all peptides that showed an apparent mass shift, regardless of the direction of the shift. After a protein was identified with high confidence using either the mass list or an MS–MS spectrum (aldolase in our system), all peptides derived from this protein were removed from the mass list. The process was then repeated in order to identify the next protein displaying an inverse labeling pattern or an apparent mass shift (BSA in this case).

3.3. Inverse proteolytic ^{18}O labeling

3.3.1. The eight-protein model system

The inverse labeling and MS analysis were performed in the same manner as shown in Fig. 1 on the eight-protein model system where BSA was ''down-regulated'' by threefold and aldolase ''up-regulated'' by threefold. MALDI-TOF-MS performed directly on mixture without any separation resulted in a peptide-map spectrum that displayed a large degree of signal overlap. The complexity was more pronounced than that with ICAT due to the lack of fractionation/enrichment. This complexity made data interpretation somewhat difficult (Fig. 4A, B).

Ideally, one would employ off-line, multi-dimensional fractionation followed by MALDI analysis of the fractions. This would facilitate lower signal

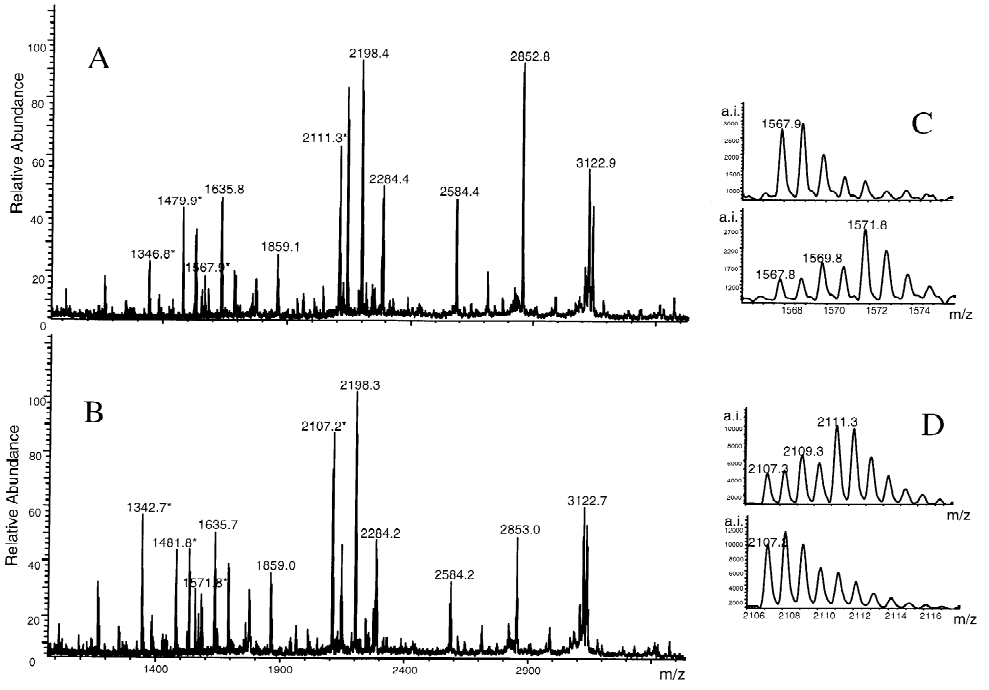

Fig. 4. MALDI-TOF detection of tryptic digests of the eight-protein mixtures. (A) ^{16}O control–^{18}O ''treated'' sample; (B) ^{18}O control–^{16}O ''treated'' sample; (C) monoisotopic patterns of a BSA peptide MH^+ 1567.9 in (A) (upper) and (B) (lower); and (D) monoisotopic patterns of an aldolase peptide MH^+ 2107.3 in (A) (upper) and (B) (lower). The inverse labeling pattern or $^{16}O/^{18}O$ intensity ratio reversal indicates differential expression of the proteins: ''down-regulation'' of BSA and ''up-regulation'' of aldolase. (* Ions showing the inverse labeling pattern).

interference and improve the detection dynamic range. Nonetheless, it was clearly demonstrated how the inverse labeling strategy could help to quickly identify the peptide signals from proteins of differential expression. We did not need to look for the ±2/4-Da pair for each peptide signal (would be challenging due to the complexity of the data) and to perform quantitation. A qualitative comparison of the two inverse labeling spectra (Fig. 4A, B) rapidly revealed signals of the characteristic isotope pattern reversal (2- or 4-Da mass difference). PSD was performed on a number of the peptides displaying the change and the corresponding proteins were identified using the PSD data (data not shown).

When the samples were analyzed using LC–MS, the characteristic inverse labeling pattern was clearly observed on a large number of peptides (data not shown). The MS–MS data were used to search protein databases and two differentially expressed proteins, aldolase and BSA, were readily identified.

3.3.2. Spiked cell lysate system

In an attempt to emulate a complex protein mixture PTP-1B protein was spiked at two different levels into two identical pools of whole cell lysate. The inverse labeling experiment was performed on the two pools followed by LC–MS analysis. As expected, single dimension LC was insufficient to adequately separate the tremendously large number of peptides present. Nonetheless, when the two sets of data from inverse labeling were compared, a number of ions possessing the characteristic inverse labeling pattern were extracted (Fig. 5A, B). The collected samples (via a splitter during the primary analysis) were subjected to a second analysis to obtain MS–MS on the ions that exhibited the inverse

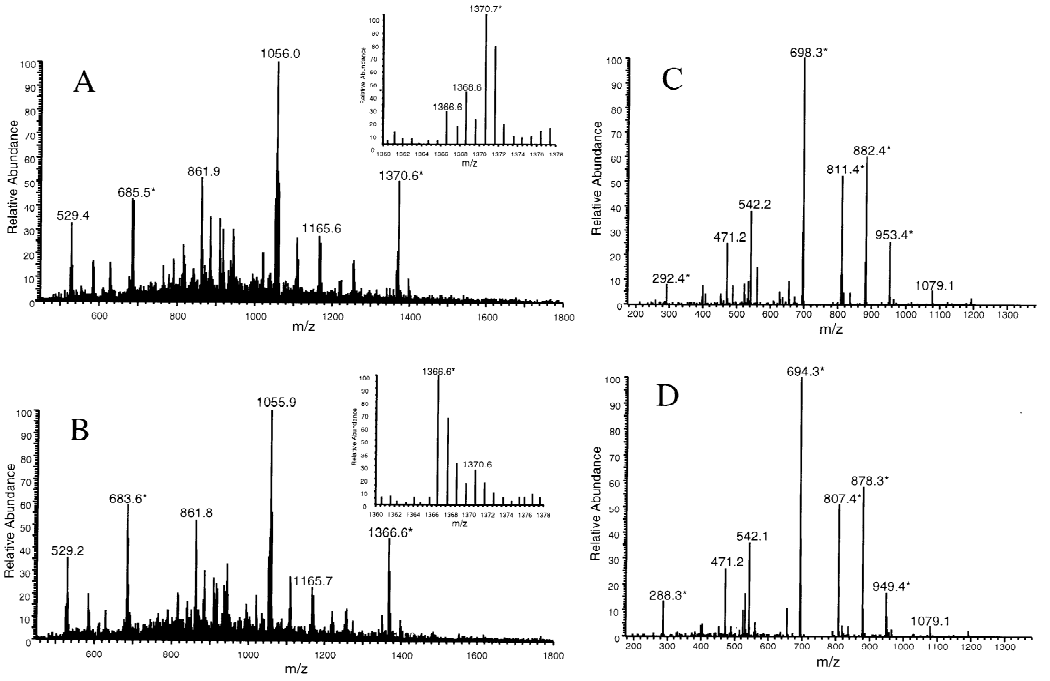

Fig. 5. LC–MS detection of a PTP tryptic peptide from a CHO cell lysate spiked with PTP-1B. (A) MS of the ^{16}O PTP10–^{18}O PTP30 sample; (B) MS of the ^{18}O PTP10–^{16}O PTP30 sample; (C) MS–MS of the peptide in (A) in-set; and (D) MS–MS of the peptide in (B) in-set, where PTP10 is a 0.25 mg CHO cell lysate spiked with 10 pmol of PTP-1B, PTP30 is a 0.25 mg CHO cell lysate spiked with 30 pmol of PTP-1B. A 4-Da isotope swap (or 2 Da on the doubly charged ion) between (A) and (B) (in-sets) on the most abundant isotopic ions indicates a significant ''differential expression'' of the protein. The inverse labeling pattern is further verified/confirmed in the MS–MS spectra by the 4-Da isotope swap of all Y ions, which also helps to identify Y ions and B ions and thus helps in the interpretation of the MS–MS spectra. PTP-1B protein is exclusively identified from database searching using the Y ions (those with a 4-Da shift). (* Ions showing the inverse labeling pattern).

labeling pattern or the apparent mass shifts. It is important to note that as a consequence of inverse labeling, MS–MS data are especially information rich. Since the ^{18}O label is incorporated at the C-terminus of each peptide, Y ions in an MS–MS spectrum carry the label and will exhibit the characteristic inverse labeling pattern for proteins that are differentially expressed. As shown in Fig. 5C, D, for proteins whose ''expression level'' has been significantly altered by ''perturbation'', the inverse labeling pattern of a 2/4-Da mass difference observed at the molecular ion level for the peptides is evident in the Y ions in the MS–MS spectra. Hence, MS–MS data provide further support on the inverse labeling pattern observed at the molecular ion level. Since most peptide fragments carry fewer charges than the precursor ion, the change is more prominent and easier to recognize when compared to that from the multiply charged precursor ion. In addition, the inverse labeling pattern reflected in Y ions offers facile assignment of Y ions and B ions for the interpretation of an MS–MS spectrum. The fragments displaying the inverse labeling pattern are Y and Y-related ions and those without are B or B-related ions. Although interpretation is not required to search the databases using MS–MS data, the added specificity helps to increase efficiency and accuracy of protein identification via database search. These are considerable advantages when dealing with novel proteins where de novo sequencing is required. The ability to assign Y and B ions greatly facilitates ''read out'' of the sequence from an MS–MS spectrum. In Fig. 5, a database search using the distinctive Y ions possessing the inverse labeling pattern led to the exclusive identification of the human PTP-1B protein.

3.4. Inverse metabolic ^{15}N labeling

3.4.1. The two-protein model system

Direct MALDI analysis was carried out successfully on the mixtures. This was largely due to the limited complexity of the two protein system. In reality, off-line coupling of separation (such as with two-dimensional chromatography) with MALDI-TOF-MS on a digest of a complex protein mixture (e.g., total cell lysate) can resemble, in each fraction, the complexity level demonstrated here. Metabolically labeled proteins possess variable isotopic mass differences. When a single-experiment approach is applied, correlation of isotopic pairs is difficult to achieve. This is true even in the case of a moderate change in expression at which both isotope counterparts are clearly detected such as that shown in the m/z range of 1550–1600 in Fig. 6A (a threefold change). However, by subtractive comparison of the two MALDI spectra from an inverse labeling experiment (Fig. 6A, B), signal pairs from proteins of no significant differential expression will be subtracted out (such as those marked with arrows along the horizontal axis). This subtraction results in significantly simplified spectra and easier peak correlation. The inverse labeling pattern is readily recognized after the subtractive cleanup of signals from proteins of no significant differential expression. The correlating, inversely labeled peak pairs have the characteristic of being signals at a similar level of relative intensity among the two inverse labeling analyses, since it is the detection of the same peptide population, but conversely labeled (in the two inverse labeling experiments). The mass difference between the correlating isotopic pairs should be in the range of 0.6–2.6% of the m/z detected. (The deviation in M_r due to ^{15}N normally ranges from 0.6 to 2.6% of the peptide M_r depending on sequence and averages about 1.2% of peptide M_r). As demonstrated, with inverse labeling, variable mass difference is no longer a problem. Signals of interest are all revealed by one simple qualitative comparative analysis.

3.4.2. Spiked algal cell lysate system

To demonstrate the application of the approach in a more complex mixture, the PTP-1B protein, both non-labeled and ^{15}N-labeled, was spiked into algal cell lysate-^{13}C and -$^{13}C/^{15}N$, respectively, to mimic protein differential expression (a threefold down-regulation). When the two sets of LC–MS data from inverse labeling experiment were compared, a number of ions possessing the characteristic inverse labeling pattern (apparent mass shifts) were extracted (Fig. 7A, B). The samples collected via the splitter were subjected to a second analysis to obtain MS–MS on the ions that exhibited the inverse labeling pattern. Their similar fragmentation patterns clearly validated the correlation of the isotopic peaks and the inverse labeling pattern on fragments confirmed that

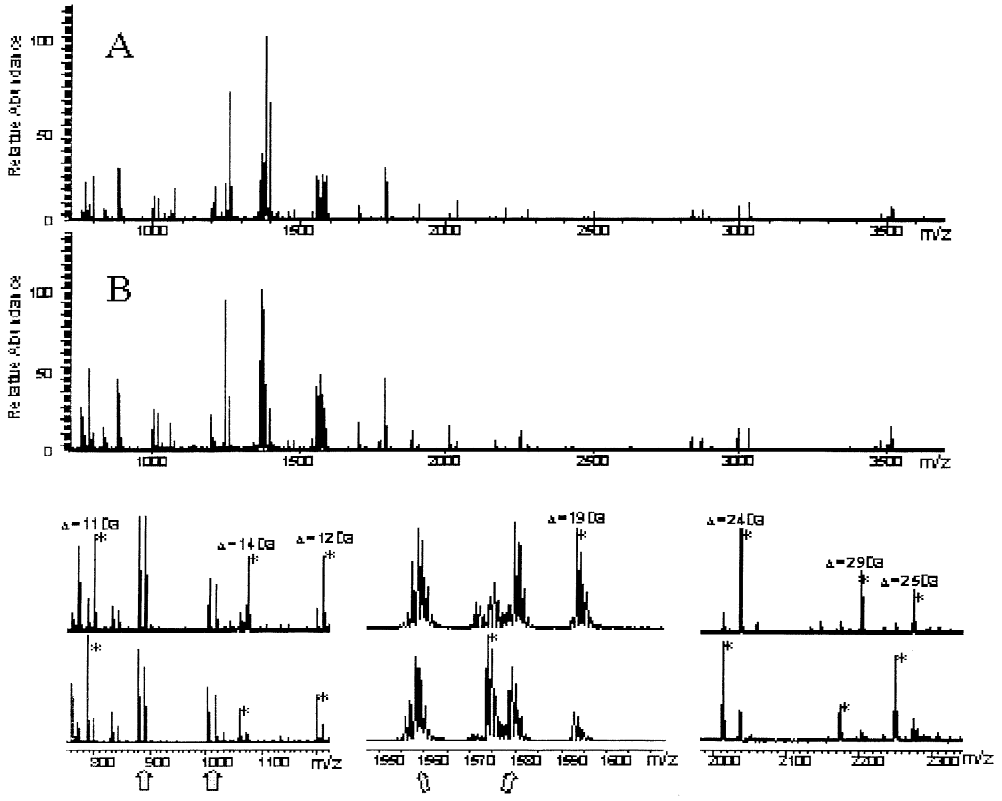

Fig. 6. MALDI-TOF detection of tryptic digests of a two-protein system with PTP protein threefold up-regulated in the ''treated''. (A) ^{14}N control–^{15}N ''treated'' sample; (B) ^{15}N control–^{14}N ''treated'' sample. The lower panels are the selective zoomed-in m/z regions.

observed on the precursor peptides and, thus, the differential expression of the parent protein (Fig. 7C, D). A database search using the MS–MS data of the ^{14}N-peptide led to the exclusive identification of the human PTP-1B protein.

3.4.3. Inverse labeling vs. conventional single experimental approach

By applying the inverse labeling strategy, the tremendous amount work spent on irrelevant proteins of no differential changes is eliminated (i.e., pair matching, intensity ratio calculation, etc., on hundreds of thousands of signals). The qualitative changes between the two experiments lead to the rapid identification of signals from proteins of interest (i.e., proteins of differential expression). In addition, the approach offers the advantage of unambiguous identification of signals from proteins of qualitative covalent changes. Although not demonstrated with examples in this study on covalent modifications of proteins, the advantages offered by inverse labeling are obvious. When responding to a perturbation, if a protein results in a covalent modification that does not exist in the control state, the modification-bearing peptide will have no isotopic counterpart in the analysis using the single-experiment approach. This signal without isotopic counterpart can be easily confused as a chemical background or as a peptide without isotope label (e.g., C-terminal peptides in proteolytic ^{18}O labeling) and be ignored. With inverse labeling, the peptide is still being detected as a signal without isotopic counterpart in either of the two inverse labeling experi-

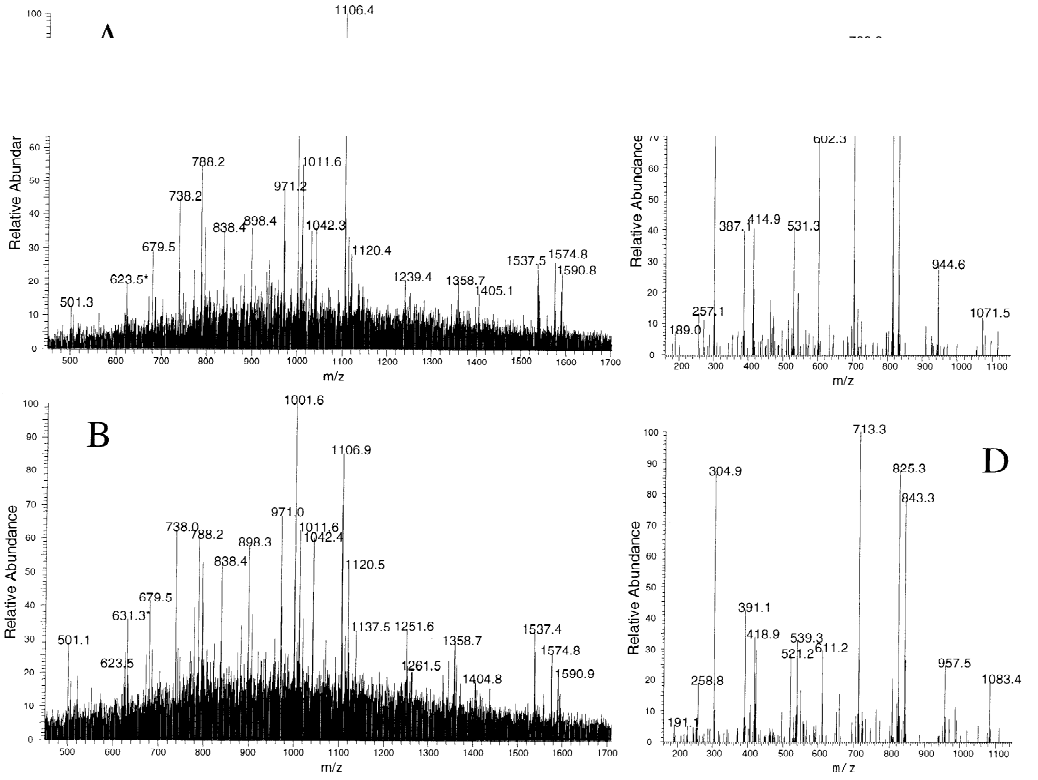

Fig. 7. LC–MS detection of tryptic digests of the algal cell lysate spiked with the PTP protein, with PTP threefold down-regulated in the ''treated''. (A) MS of the ^{14}N control–^{15}N ''treated'' sample, averaged spectrum over a 3-min range in the LC–MS analysis; (B) MS of the ^{15}N control–^{14}N ''treated'' sample, averaged spectrum over a 3-min range in the LC–MS analysis; (C) MS–MS of the peptide in (A) m/z 623.5; and (D) MS–MS of the peptide in (B) m/z 631.3, where ^{14}N control is a 0.05 mg ^{13}C-algal protein spiked with 10 pmol of PTP-1B, ^{15}N control is a 0.05 mg ^{13}C-^{15}N-algal protein spiked with 10 pmol of ^{15}N-PTP, ^{14}N ''treated'' is a 0.05 mg ^{13}C-algal protein spiked with 3 pmol of PTP-1B, and ^{15}N ''treated'' is a 0.05 mg ^{13}C-^{15}N-algal protein spiked with 3 pmol of ^{15}N-PTP. Isotope swap or inverse labeling pattern between (A) and (B) were observed on the marked ions (*). The inverse labeling or differential expression is further verified/confirmed in the MS–MS spectra by their similar fragmentation pattern. PTP-1B protein is exclusively identified from database searching using MS–MS data of the ^{14}N-peptide (C).

ments. However, when the two experiments are compared, the characteristic inverse labeling pattern will reveal the covalent change of the peptide without any ambiguity. Fundamentally different from the background signals is that such modification-bearing peptides do contain isotope label. When the labeling is reversed between the two inverse labeling experiments, the isotope swap will be clearly detected on these peptides. Thus, the inverse labeling pattern is unique to proteins of any real changes in expression, including changes at both quantitative and qualitative level. Although an additional experiment must be performed, in our opinion, the advantages of the method overwhelmingly outweigh any additional work or potential sensitivity drop due to any splitting of sample.

3.4.4. Data reduction and dynamic range of detection

By subtractively comparing data from the two inverse labeling experiments, signals from proteins of no differential expression can be subtracted out. As a result, important signals from differentially expressed proteins will be quickly discovered. In-

verse labeling, however, does not solve the problem of dynamic range of detection. Signal detection is determined by a number of factors, including detection sensitivity, dynamic range of detection, mixture complexity and the separation applied to reduce the complexity. In order to achieve detection of proteins at low abundance, comprehensive separation needs to be utilized to improve dynamic range of detection. Inverse labeling helps to improve data interpretation. It provides quick and unambiguous identification of protein signals of both qualitative and quantitative differential changes.

3.4.5. Different labeling methods

Applications of inverse labeling strategy to three general protein isotope labeling methods have been demonstrated in this study. With chemical labeling, a tag via a specific chemistry (for a particular group or amino acid residue) is used to isotopically label proteins. Mass difference between the two isotopically different tags is well defined by the stable isotopes incorporated in the tags. The isotopic mass difference for a peptide is determined by the number of reactive amino acid residues in the sequence. Applying inverse labeling to ICAT, signals from proteins of differential expression will all be revealed in one comparative analysis. The apparent mass difference may be 8-Da, 16-Da, or 24-Da depending on the number of cysteines in a sequence. ICAT labeling is performed at protein level. After labeling, the two protein pools are combined. As a result, processes such as fractionation, proteolysis, and peptide separation are carried out on the combined sample. Differences in sample handling and the effect they may have are no longer issues after the two pools are combined. With the built-in affinity functionality, the tag-containing peptides can be affinity isolated resulting in a dramatic reduction in peptide mixture complexity and simpler analysis. The other side of the token, however, is the incomplete coverage of proteins. About 20% of the protein population do not have cysteine in the sequences and they would not be detected using an ICAT based method. As with any chemical reaction, reaction yield and the post-reaction work-ups are always concerns. In addition, ICAT reagents are relatively expensive.

For proteolytic ^{18}O labeling, up to two ^{18}O atoms get incorporated into the C-terminus of peptides and a 4-Da mass difference results on comparison of a non-labeled ^{16}O peptide with its ^{18}O counterpart. A C-terminal peptide will not get labeled if it does not have a lysine or an arginine at the C-terminus. This unpaired, non-labeled C-terminal peptide signal can be confused with a dramatically down-regulated peptide signal for which the corresponding ^{18}O signal is so weak and is basically buried underneath the normal ^{13}C isotope peaks. Inverse labeling will eliminate this ambiguity since the C-terminal peptides, free of isotope labeling, will not present any inverse labeling pattern. Such signals without isotope labeling will be subtracted out together with signals of no differential changes. Any real changes, regardless of an up- or a down-regulation, a covalent change or a change in abundance, a moderate or a dramatic change in expression, will be detected without ambiguity by the characteristic inverse labeling pattern. Because the labeling occurs at the C-terminus, MS–MS data offers additional capability besides confirming protein differential expression. It offers the ability to differentiate Y and B fragment ions and thus the ability of de novo sequencing. Proteolytic ^{18}O labeling is probably the easiest to accomplish among three methods discussed. Since proteolysis is a natural step in standard sample preparation prior to analysis, no extra work is required to introduce the label. The supply of ^{18}O water may be limited, especially for large scale applications. To conserve ^{18}O water as well as to overcome technical difficulties in sample handling, post-proteolysis ^{18}O incorporation has been investigated. Kinetic differences in ^{18}O incorporation among peptides have been observed. Our results indicate that peptides with acidic residues in the immediate vicinity of the C-terminal Lys or Arg (for trypsin digest) exhibit a significantly slower rate of post-proteolysis ^{18}O incorporation. Longer incubation at 37 °C (vs. room temperature) will help to increase the labeling yield. With respect to label incorporation, ^{18}O labeling occurs at proteolysis, which is quite late in the sample preparation process. Since the two pools cannot be combined until after labeling, differences in sample handling between the two pools in any of the prior steps, such as protein level fractionation, may lead to discrepancies in protein recovery and faulty protein differential expression as a result.

With respect to metabolic ^{15}N labeling, the results of this study clearly demonstrate that inverse labeling offers an easy solution to the otherwise very challenging technical difficulty associated with this labeling method. In contrast to chemical ICAT labeling and proteolytic ^{18}O labeling, the conventional single experiment approach does not work for protein metabolic labeling. Special instrumentation (ultra-high resolution MS) or additional experiments (MS–MS on every signal) would have to be used or performed to overcome the hurdle. By performing inverse labeling, the qualitative isotope pattern change (regardless of what the isotopic mass differences may be) reveals signals of differentially expressed proteins. The massive data reduction dramatically reduces data complexity and helps to eliminate ambiguity in data interpretation. MS–MS fragmentation data provide further confirmation on protein differential expression and in addition, are utilized to search protein databases for protein identification. With protein metabolic labeling, isotope labels are incorporated during cell culture. The two pools are then readily combined. Since the combination of samples is prior to any major sample preparation procedure, discrepancies caused by sample handling are minimal. Proteins of low metabolic turnover may have difficulty getting labeled with this approach. At the present time, metabolic labeling of proteins seems economically feasible only for bacteria proteins, not for mammalian cells.

4. Conclusions

The inverse labeling procedure has been successfully applied to three general stable isotope labeling methods of proteins for comparative proteomics to rapidly identify protein markers/targets. Two parallel inverse labeling experiments are performed where the labeling is reversed in the second experiment, i.e., the heavy isotope labeled pool in the first experiment is light isotope labeled in the second experiment and vice versa. The separately labeled control and ''altered'' protein pools are then combined in each experiment. Both mixtures are subsequently processed (e.g., proteolysis if not yet at the end of labeling) and analyzed by MS coupled with on-line or off-line separation. The characteristic inverse labeling pattern (i.e., isotope intensity profile swap) presented by peptide signals between the two inverse labeling experiments indicate the differential expression of proteins from which the peptides are derived. MS–MS fragmentation data further support the conclusion (by similar fragmentation patterns and the inverse labeling pattern in fragments) and are used to search protein databases for protein identification. Predefined mass differences between isotopic pairs are not prerequisites using this method. The strategy works well with labeling methods that result in variable mass differences among peptides of different sequences. An MS instrument of high resolution is not required. The differentially expressed proteins and proteins of covalent changes resulting from an altered state are rapidly identified without ambiguity. In our opinion, the procedure presents a logical sequence in approaching the problem: data reduction of irrelevant signals, quick focus on signals of interest, detailed analysis on signals of interest only, and conclusions of minimum ambiguity. With the implementation of a comprehensive separation scheme, such as multi-dimensional LC for ESI-MS or MALDI-TOF-MS to achieve the necessary dynamic range of detection, and with the development of software for automated data interpretation, the inverse labeling strategy may be readily adapted into a high throughput proteomics platform. We believe it holds its own place in the process of shaping the future technology platform for high throughput proteomics.

Acknowledgements

The authors would like to thank Dr. Bryan Burkey and Ms. Mei Dong for providing the CHO cell lysate, Dr. James Koehn for providing the PTP and HtrA proteins used in the study, and Gary Trakshel for protein concentration measurements of the algal protein extracts. The authors would also like to thank Michael Sabio, Gary Trakshel, and Maria Cueto for critical reading of the manuscript.

References

[1] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Proc. Natl. Acad. Sci. USA 96 (1999) 6591.

[2] T.P. Conrads, K. Aiving, T.D. Veenstra, M.E. Belov, G.A. Anderson, D.J. Anderson, M.S. Lipton, L. Paša-Tolic, H.R. Udseth, W.B. Chrisler, B.D. Thrall, R.D. Smith, Anal. Chem. 73 (2001) 2132.
[3] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[4] D.K. Han, J. Eng, H. Zhou, R. Aebersold, Nat. Biotechnol. 19 (2001) 946.
[5] J. Ji, A. Chakraborty, M. Geng, X. Zhang, A. Amini, M. Bina, F. Regnier, J. Chromatogr. B 745 (2000) 197.
[6] D.R. Goodlett, A. Keller, J.D. Watts, R. Newitt, E.C. Yi, S. Purvine, J.K. Eng, P. von Haller, R. Aebersold, E. Kolker, Rapid Commun. Mass Spectrom. 15 (2001) 1214.
[7] M.B. Goshe, T.P. Conrads, E.A. Panisko, N.H. Angell, T.D. Veenstra, R.D. Smith, Anal. Chem. 73 (2001) 2578.
[8] E.C. Peters, D.M. Horn, D.C. Tully, A. Brock, Rapid Commun. Mass Spectrom. 15 (2001) 2387.
[9] X. Yao, A. Freas, J. Ramirez, P.A. Demirez, C. Fenselau, Anal. Chem. 73 (2001) 2836.
[10] Y.K. Wang, Z. Ma, D.F. Quinn, E.W. Fu, Anal. Chem. 73 (2001) 3742.
[11] I.I. Stewart, T. Thomson, D. Figeys, Rapid Commun. Mass Spectrom. 15 (2001) 2456.