한국환경농학회지 제25권 제4호 (2006)
Korean Journal of Environmental Agriculture Vol. 25, No. 4, pp. 371-381

연구보문

Quantitative Proteomics Towards Understanding Life and Environment

Jong-Soon Choi*, Keun-Yook Chung[1], and Sun-Hee Woo[2]

Proteomics Team, Korea Basic Science Institute, Daejeon, Korea

[1]Department of Agricultural Chemistry, Chungbuk National University, Cheongju, Korea

[2]Department of Crop Science, Chungbuk National University, Cheongju, Korea

(Received Novemver 27, 2006, Accepted December 11, 2006)

ABSTRACT: New proteomic techniques have been pioneered extensively in recent years, enabling the high-throughput and systematic analyses of cellular proteins in combination with bioinformatic tools. Furthermore, the development of such novel proteomic techniques facilitates the elucidation of the functions of proteins under stress or disease conditions, resulting in the discovery of biomarkers for responses to environmental stimuli. The ultimate objective of proteomics is targeted toward the entire proteome of life, subcellular localization, biochemical activities, and the regulation thereof. Comprehensive analysis strategies of proteomics can be classified into three categories: (i) protein separation via 2-dimensional gel electrophoresis (2-DE) or liquid chromatography (LC), (ii) protein identification via either Edman sequencing or mass spectrometry (MS), and (iii) proteome quantitation. Currently, MS-based proteomics techniques have shifted from qualitative proteome analysis via 2-DE or 2D-LC coupled with off-line matrix assisted laser desorption ionization (MALDI) and on-line electrospray ionization (ESI) MS, respectively, toward quantitative proteome analysis. *In vitro* quantitative proteomic techniques include differential gel electrophoresis with fluorescence dyes, protein-labeling tagging with isotope-coded affinity tags, and peptide-labeling tagging with isobaric tags for relative and absolute quantitation. In addition, stable isotope-labeled amino acids can be *in vivo* labeled into live culture cells via metabolic incorporation. MS-based proteomics techniques extend to the detection of the phosphopeptide mapping of biologically crucial proteins, which are associated with post-translational modification. These complementary proteomic techniques contribute to our current understanding of the manner in which life responds to differing environment.

Key Words: 2-dimensional gel electrophoresis, Edman sequencing, mass spectrometry, 2-dimensional liquid chromatography, matrix assisted laser desorption ionization, electrospray ionization, differential gel electrophoresis, isotope-coded affinity tag, isobaric tags for relative and absolute quantitation, stable isotope labeled amino acid, phosphopeptide mapping

INTRODUCTION

The advent of genomic technology has facilitated our general knowledge regarding genetic information, and enables us to design tailored transformed organisms, via the use of genetic tools. The genome itself gives us a sort of guide map, allowing us to determine where we depart and where we arrive. The actual life events occur during the coordination and cross-talk occurring between cellular proteins. However, there remains a technical gap to bridge, between our understanding of the genomic sequence and cellular behaviors. Thus, the determination of what proteins are transiently and spatially expressed can be extended from a snap picture showing where proteins are, to a relatively complete movie showing the manner in which cells function internally. The terms "proteome" and "proteomics" were initially coined by Mark Wilkins in the Siena Two-dimensional Electrophoresis Meeting, as analogues to the terms "gene" and

*Corresponding author:
Tel: +82-42-865-3428 Fax: +82-42-865-3419
E-mail: jschoi@kbsi.re.kr

"genomics", respectively[1]. Three technical breakthroughs underlying the birth of proteomics include (i) the full collection of genes and protein-sequence databases, (ii) the development of user-friendly bioinformatic tools to handle the vast array of biological data, and (iii) the development of a microchip array platform technology for the interpretation of genome-wide expression. Currently, a host of useful tools have been placed within our hands, allowing us to view from a high vantage the entire system of life, allowing biologists to begin to think in a more expansive manner (Figure 1). In the present article, these novel proteomic techniques will be reviewed and considered, in an effort to describe, in part, disciplines which may someday be applied in industry.

CONSIDERATIONS INHERENT TO BIOLOGICAL SYSTEMS

Proteomics is often considered more of a technology than a science. The techniques of proteomics constitute a powerful prism by which the multiproteins in a cell can be screened, the purpose of which is to characterize not the behavior of the single protein molecule, but the entire network inherent to a biological system. In this context, the ultimate objective for proteomists is totally different from those who apply the methods of protein chemistry. Protein chemistry principally concerns the complete sequences of target proteins, in order to determine cellular functions, which can be deduced from determinations of three-dimensional structures. In contrast to classical protein chemistry, proteomics involves the high-throughput identification of expressed proteins via partial sequence analysis and, ultimately, the quantitation of proteins for the systematic interpretation of cellular functions. Thus, proteomics is targeted toward an investigation of the cellular dynamics of proteomes in biological environments, the characteristics of which are completely different from what can be seen in the static and fixed picture allowed for by genomic techniques. Regardless of cell type, the observed proteins are constitutively expressed proteins in common for biologically essential functions, including DNA replication, protein synthesis, and energy metabolism. Proteins with unique cell-specific functions like rhodopsin in the retinal epithelium can be considered biomarker target proteins responsible for specific functions in specialized biosources. Specifically, proteins associated with signal transduction, transcription factors, and cell-cycle control turn over rapidly, thus performing key roles in specific functions in specialized cells. Although the settlement of oligoucleotide chip analysis allows for the interpretation of genome-wide expression patterns in a cell[2], we are unable, at this point, to determine directly the corresponding proteins. There are three possible reasons for the differences between transcriptomes and proteomes. mRNA stability and translational efficiency subsequently affect protein expression, both qualitatively and/or quantitatively. Furthermore, mRNA levels alone are insufficient to determine in detail the manner in which a given cell is regulated. In the previous literature, a great deal of evidence has been uncovered to demonstrate the poor correlation between mRNA copies and coding protein quantities[3]. Proteomics explains what is happening within a cell, whereas transcriptomics and genomics provide us with information as to what may *actually be happening* and what *could be happening*, respectively. In the extraordinary example of human α-antitrypsin, for instance, 22 different types of isoforms have been detected in human plasma[4]. Thus, the true complexity of a protein can be determined only via proteomic study, and remains undeterminable via transcriptomic and genomic studies. After the completion of the sequencing of the entire human genome[5],

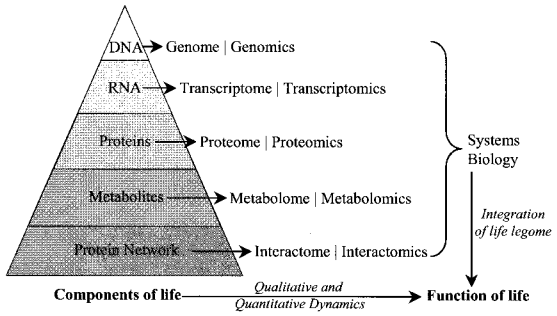

Fig. 1. Hierarchical structure of biological components. There are different approach strategies for organisms grouped into genome (the complete set of genes), transcriptome (the complete set of mRNA molecules), proteome (the complete set of proteins), metabolome (the complete set of metabolites), and interactome (all the protein-protein interaction nodes). Systems biology integrates the entire set of these biological components to examine dynamics systematically in terms of qualitative and quantitative values.

many reductionists were surprised to discover that the numbers of human genes, comprising 10^{13} cells with ~26,000 genes, have fewer than twice as many as the 959-cell nematode, *Caenorhabditis elegans*, which is comprised of ~18,000 genes (OnLine Databases at http://wit.integratedgenomics.com/ GOLD)[6]. Thus, scientists have suggested that the higher order of complexity seen in Figure 2 may be caused, not by the gene number alone, but from a variety of post-translational modifications of proteins and/or a scale- free network of the proteomic society[7]. In summary, proteomics seeks to identify and quantify proteins initially, and subsequently to determine their localization, modification, interaction, activity, and, ultimately, their cellular function[8].

PROTEIN SEPARATION BASED ON 2-DIMENSIONAL GEL

Analytical proteomics generally begins with the separation of proteins or peptides. As is shown in Figure 3, the simple strategies for MS proteomic analysis can be divided into (i) the separation of the protein mixture followed by tryptic digestion, and conversely, (ii) the entire digestion of the protein mixture followed by separation for the identification of proteins via on-line MS. The former procedure can be exerted simply via two-dimensional gel electrophoresis (2-DE), whereas the latter one, referred to as shot-gun proteomics, can be conducted via tandemly-linked two-dimensional liquid chromatography (2D-LC) coupled with on-line MS[9]. As is shown in Figure 4, 2-DE is a well-known protein separation method, which is conducted in accordance with the isoelectric points (p*I*) and molecular weights (M*r*) of proteins[10]. The commercial immobilized pH gradient strip gels provide stable pH gradients for 2-DE experimenters[11]. The advantages of 2-DE include the ability to display simply the protein profiles visually on 2-DE gels, and to detect some isoforms of post-translational modifications. Typically, both phosphorylated and non-phosphorylated proteins can be resolved horizontally via 2-DE, due to the alterations of p*I* values as the result of phosphorylation[12]. Recently, an improved 2-DE method has been introduced, and is referred to as differential gel electrophoresis (DIGE). DIGE utilizes two different fluorescence dyes (Cy5 and Cy3) for the two states of the protein samples, which are subsequently run on the same single 2-D gel[13]. The fluorophore dyes are amine-specific to the N-terminal proteins or ε-amine side chains of lysine residues, of which the fluorescence intensities of single protein spots in a combination of the two protein groups can be compared quantitatively at specific excitation and emission Cy5 wavelengths (Ex λ_{max}=633 nm, Em λ_{max} =670 nm) and Cy3 dyes (Ex λ_{max}=530 nm, Em λ_{max} =580 nm). The DIGE procedure is schematized in Figure 5.

Although 2-DE-based proteomics are relatively simple and low-cost, they remain labor-intensive and time-consuming, in particular due to the drawbacks inherent to automation. In addition, proteins with extreme pH values, such as strong acidic or basic p*I*s

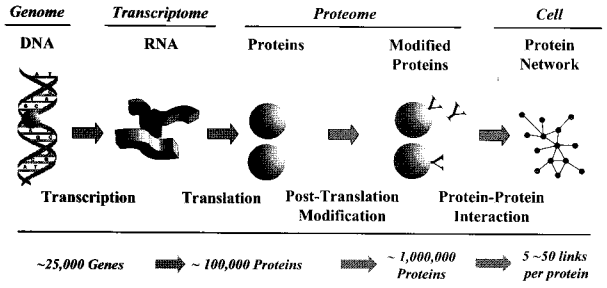

Fig. 2. Origins of cellular complexicity. During the process of DNA doctrine (DNA makes RNA makes protein), multiple proteins can be generated via RNA splicing and editing to various mature mRNAs, in which mature RNA, in turn, translates proteins and further experiences protein isoforms via post-translational modification. In the cell, more than one million proteins engage in cross-talk in the formation of the cellular network.

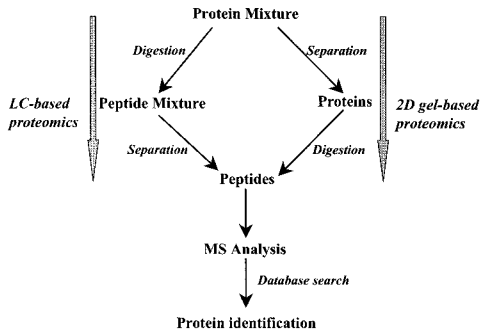

Fig. 3. Workflow of two proteomic strategies. One general method is that the protein mixture is separated on 2-dimensional gels, followed by MS analysis. Another shot-gun proteomic method predicated on LC-based proteomics is the direct tryptic digestion of a protein mixture to a peptide mixture, followed by MS analysis.

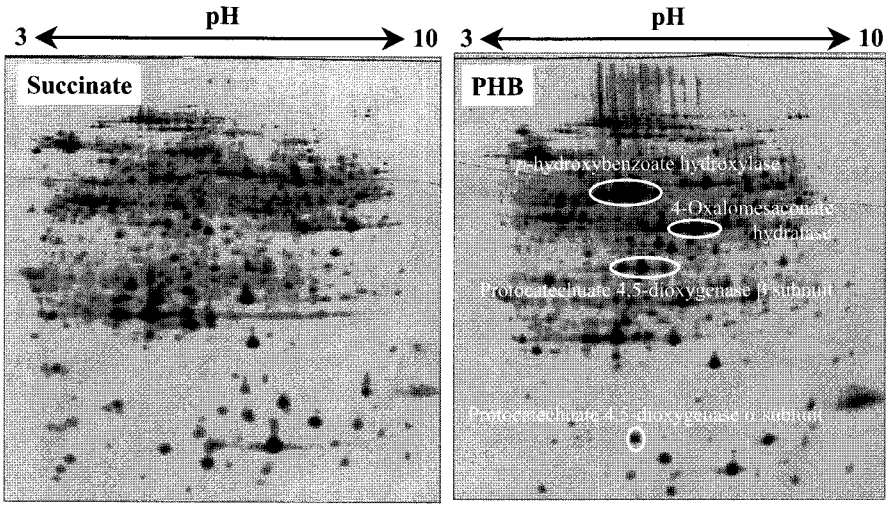

Fig. 4. Differential 2-D gel images of soluble fractions from *P.putida* KT2440 induced by monocyclic aromatic carbons: (A) succinate-cultured, (B) polyhydroxy benzoate-culture cells. IPG gel in a pH range of 3-10 was used. The circles of spots corresponded to the polyhydroxybenzoate-induced proteins identified via PMF or MS/ MS fragmentation analysis[10].

and hydrophobic proteins, are notoriously difficult to separate and visualize on conventional 2-D gels. In order to overcome this limitation, a shot-gun proteomic method that bypasses 2-DE has been developed. This new technique involves the conversion of the entire peptide mixture after the whole digestion of proteins with trypsin in combination with liquid chromatographic techniques[14]. The multidimensional protein identification (MudPIT) technique is a recently developed 2-D LC separation method involving the tandem linkage of a cation-exchange column and a reverse-phase column, and MudPIT has been applied to the analysis of several whole proteomes with enhanced computing power (Figure 6).

PROTEIN IDENTIFICATION METHODS

One of the oldest methods for the identification of the unknown protein is N-terminal sequencing via Edman degradation chemistry[15]. The Edman reagent, phenylisothiocyanate, chemically modifies the free amino termini of polypeptides to form the phenylthiocarbamyl (PTC) polypeptide. The addition of anhydrous acid to PTC amino-terminal residues resulted in rapid cleavage from the polypeptide chain, yielded anilinothiazolinone (ATZ) amino acid and *n-1* polypeptide. The unstable ATZ derivative amino acid converted to a more stable phenylthiohydantoin (PTH)-

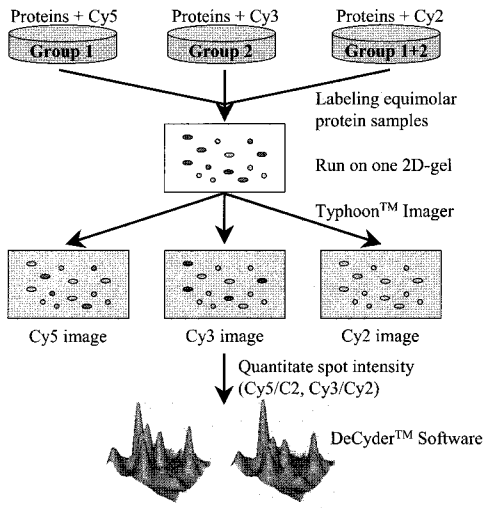

Fig. 5. Schematic representation of differential gel electrophoresis. Two protein samples were labeled separately with Cy5 for control (Group 1), Cy3 for experiment (Group 2), and Cy2 as internal standard dye (half of Group 1 and 2). The samples were then mixed and run on the same 2D gels. The intensities of each protein spot (Cy5 and Cy3) were divided by the intensity of Cy2 for normalization, thereby allowing the determination of the differential quantity of each protein spot using DEcyder™ image analysis software.

amino acid, in which this derivative can be run in a reverse-phase column to identify the amino acid via comparison with the standard PTH amino acid (Figure

7). Although MS has principally replaced Edman sequencing in modern proteomics, it remains useful in the differentiation between isoleucine and leucine with the same molecular weight of 113 Da, the acquisition of N-terminal processed information following translation, and the identification of proteins lacking genomic information via homology searches[16].

Besides direct protein sequencing, information gained from partially digested peptide fragments from original proteins can provide some clues, via comparisons of collections of experimentally acquired molecular weight data with a genomic database. Theoretically, each protein can generate a peptide mass fingerprint (PMF), or a specific pattern of a given protein generated by a specific protease[17]. Using the example of trypsin, the post-tryptic digestion products can be expected to represent the peptide fragments cleaved after the carboxy-termini of arginine or lysine. Thus, the protein source from either the 2-DE-separated protein or the eluted solution from liquid chromatography can be utilized as a substrate for trypsin. As is shown in Figure 8, the purified tryptic digests dissolved with 50% *(v/v)* aqueous acetonitrile and 0.1% *(v/v)* formic acid were applied to a matrix solution of α-cyano-4-hydroxy cinnamic acid in 50% acetonitrile. The mixture is then applied to a MALDI plate, allowing the crystals to be ionized via nanosecond laser pulses, in which the ionized peptides, in turn, are evaporated and mobilized into a vacuum tube, allowing them to reach the detector. The molecular weights can be calculated as a form of mass-to-charge *(m/z)* via time-of-flight (TOF) analysis from the applied kinetic energy, $E=\frac{1}{2}mv^2$. The collected *m/z* data are then applied to a search engine, i.e., MASCOT (http://www.matrixscience.com) in order to construct reports regarding the statistical summary of PMF in the order of best-fitting scores. Identification methods, such as Edman sequencing and PMF, can provide complementary protein information under a given physiological condition[18].

When the above-described protein identification

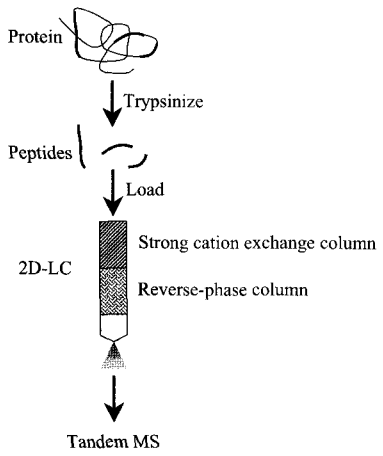

Fig. 6. The MudPIT scheme. Denatured protein complex is digested into peptides, in which they are separated in tandemly connected columns, i.e., strong cation exchange and reverse-phase column. The subsequently MSMS fragmented peptides are searched against SEQUEST.

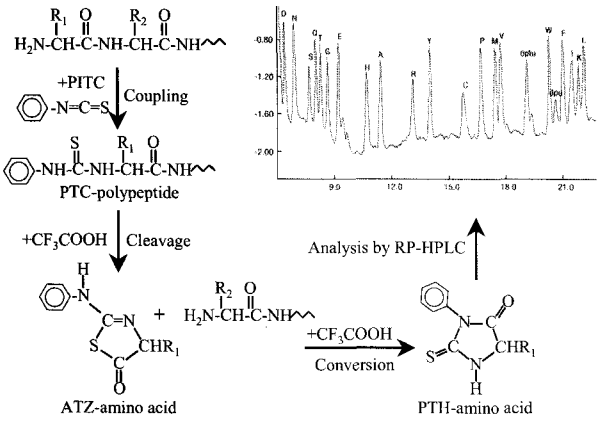

Fig. 7. The schematic procedure of Edman sequencing. Amino-terminal residues of protein couples with PITC to form a PTC amino acid. Acid cleavage removes the amino-terminal residue as an unstable ATZ amino acid. Subsequently, the unstable ATZ amino acids are then converted to stable PTH amino acids. Amino acids are sequentially identified via reverse phase-HPLC.

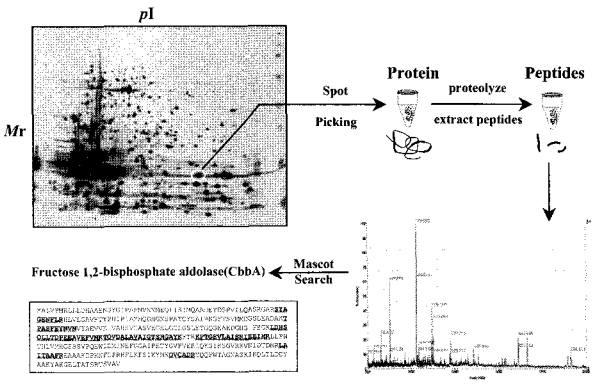

Fig. 8. Peptide massfingerprinting flowchart. The protein spot to be analyzed was picked and trypsinized. The tryptic digests were then ionized on MALDI plates and scanned by MS spectra. The 5~10 prevalent peaks were entered into the Mascot database in order to identify fructose 1,6-bisphophate aldolase against the CyanoBase[18].

methods are not satisfied, a much higher degree of protein information can be acquired via the further fragmentation of the prevalent parent ion molecules, resulting in the production of daughter ions. During this MS/MS analysis, the bidirectional peptide sequences of a single peptide can be determined via the analysis of the *b*- and *y*-ion series from the database (Figure 9). The *de novo* peptide sequence data generated by MS/MS is quite useful for high-confidence protein identification via peptide sequence information searches across both nucleotide and protein databases. Protein identification via MS/MS analysis, such as with Edman sequencing, has been determined to be the most appropriate method for organisms with less fully annotated databases due to incomplete genome sequencing. Automated protein identification via large-scale MS/MS can be conducted via on-line nanoscale-LC based electrospray ionization quadrupole TOF MS or off-line LC-connected to a MALDI TOF-TOF MS system. These tandem MS/MS analyzers are consistent with shot-gun proteomics techniques, including MudPIT.

The proteomic approach to protein phosphorylation is one of the most challenging new technologies thus far developed. Protein phosphorylation is one of the most extensively utilized and important post-translational modifications, and is known to regulate biological functions[19]. Phosphorylation is inherent to a variety of cellular dynamic processes including cell division, metabolic regulation, and signal transduction. Regarding the signaling pathway process, phosphorylated proteins have been shown to occupy 1~2% of the entire proteome, and are also encountered in conjunction with the dynamic characteristics of kinases and phosphatases[20]. A common technique for the detection of phosphoproteins involves radioactive labelling of the phosphate groups with inorganic ^{32}P and subsequent autoradiography. Another method for the detection of phosphoproteins involves Western-blotting using phosphospecific antibodies, in which the anti-phosphoserine, -phosphothreonine, and -phosphotyrosine antibodies are obtained commercially[21]. However, for the direct mapping of phosphorylated sites via MS, methods for the enrichment of phosphoproteins and phosphopeptides via affinity chromatography are required. One of the most frequently employed methods for phosphoprotein and phosphopeptide enrichment is immobilized metal-ion affinity chromatography, which is utilized in the purification of His-tagged

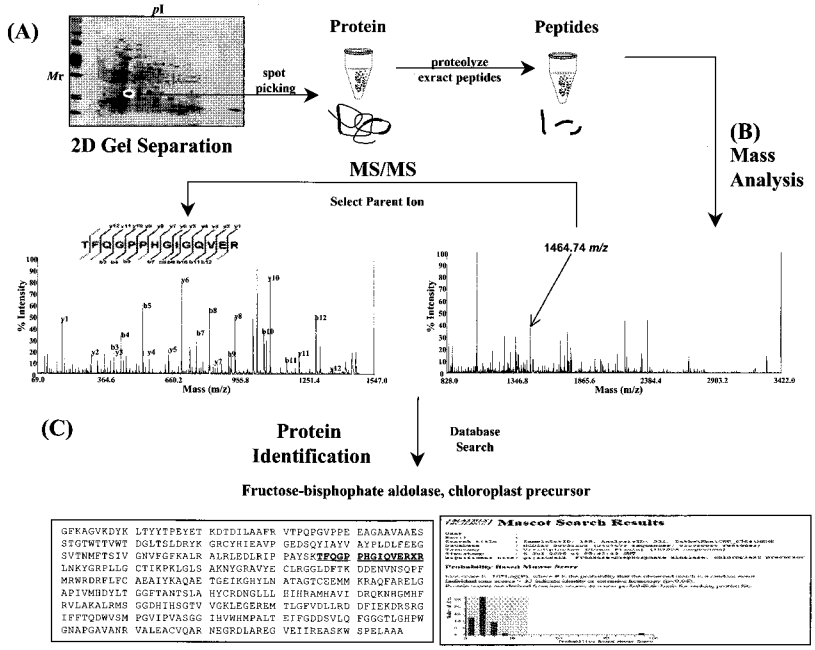

Fig. 9. Flowchart of protein identification by MS/ MS fragmentation. The protein spot to be analyzed was picked and trypsinized. The tryptic digests were ionized on MALDI plates and scanned by MS spectra. One of the parent peaks with an *m/z* value 1464.74 was further allowed to fragment, in order to determine the MS/ MS spectra. The partial sequence of TFQGPPHGIGQVERXR was identified as ribulose-1,5-bisphophate carboxylase oxygenase with statistical significance ($p<0.05$) via Mascot search (unpublished data).

proteins[22]. Metal-ions such as Fe^{3+}, Ga^{3+}, Al^{3+} or Zr^{3+} have been extensively employed in order to achieve higher specificity and enrichment of phosphopeptides. Another possible enrichment technique involves the use of titanium oxide nanoparticles, in which impurities such as acidic peptides can be reduced for superior phosphopeptide mapping[23]. As is shown in Figure 10, enrichment strategies for phosphoproteins and phosphopeptides are schematized. Among the methods for phosphoprotein analysis, techniques for the selective purification and determination of phosphopeptide sites have been established successfully in our laboratory, employing titanium oxide particle absorption onto standard phosphoproteins (Figure 11).

SYSTEMATIC QUANTITATIVE PROTEOMICS TECHNIQUES

The systematic quantitation of proteomes can be categorized into gel-based and LC-based techniques, as is described in Figure 3. Gel-based comparative proteomics techniques, including DIGE, have some limited drawbacks, including the disadvantages associated with full-automation, as well as the underrepresentation of low-abundance proteins, and poor resolution of highly acidic/basic proteins and proteins with extreme molecular weights and hydrophobicity. Contrary to gel-based proteomics techniques, LC-based quantitative techniques allow for a variety of approaches to high-throughput and automation of the proteome, via the downsizing of proteome complexing to optimal proteins or peptides. The multi-dimensional LC approach, which is abbreviated as MDLC, constitutes the basic analysis template, in which MDLC can be applied in turn to the systematic quantitative analysis of the proteome in two cellular states. The chemical tagging of proteins and/or peptides in the two proteomes under study allows for the direct relative quantitation of proteins via LC-MS analyses (Figure 12).

The prototype of chemical tagging is the isotope-

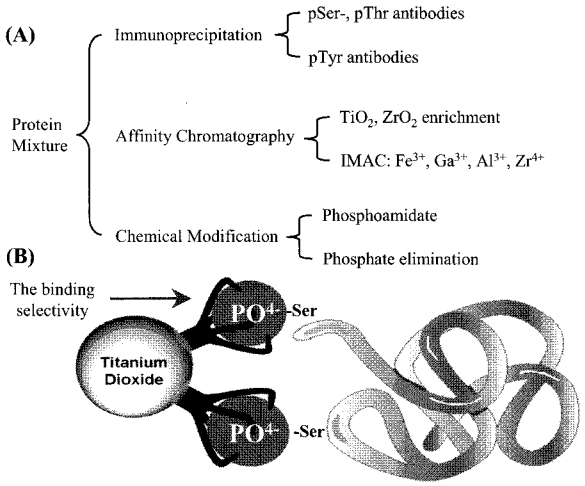

Fig. 10. Schematic representation of the enrichment of phosphoproteins and phosphopeptides. (A) Classification of phosphopeptide determination. Protein mixtures can be selectively enriched and detected via immunoprecipitation, affinitychromatography, and chemical modification. (B) Picture depicting the capture of 2 moles of phosphates in Serine residues by 1 mole of titanium oxide particles.

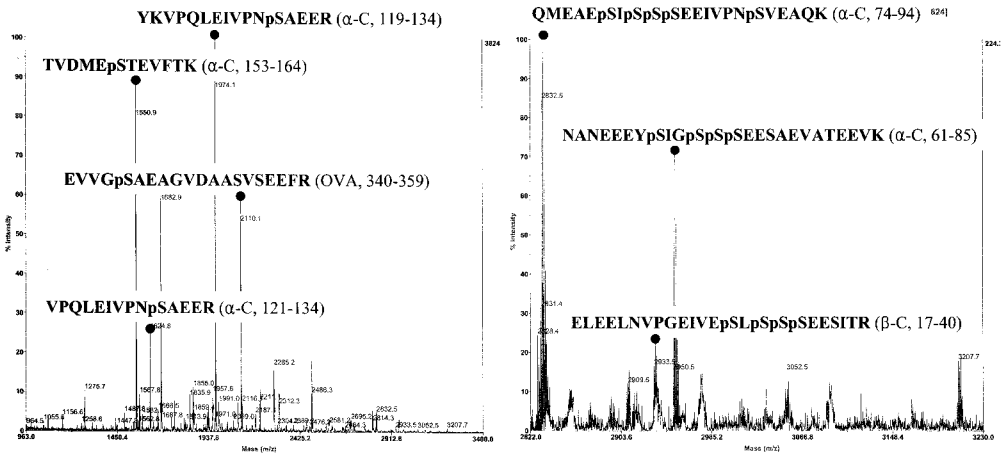

Fig. 11. Enrichment and determination of phosphorylated peptides. Standard mixture including α-casein (α-C), β-casein (β-C), and ovalbumin (OVA) was undergone to fragmentation by MALDI MS MS obtained from the TiO_2-enriched peptide mixtures, using acetic acid as the loading buffer. The phosphorylated serine sites are marked as pS (unpublished data).

coded affinity tag (ICAT) technique, developed by Aebersold *et al.*[24]. The ICAT reagent consists of the biotin group which binds to the avidin column, the polyether linker with eight hydrogens (D_0) or deuteriums (D_8), and the reactive group which is used to label cysteine residues specifically (Figure 13). The combined proteomes labeled with each light and heavy ICAT reagent allow for trypsin digestion followed by the capture of biotin-containing peptides to the avidin column. After the washing of the unbound peptides with buffer, the labeled cysteine-containing peptides are eluted and quantified via MS. Doublet peaks with 8 Da difference are targets for the quantification of peak intensity implementing the relative abundance of two states of cell populations. The identification of the target peaks can be conducted via the bidirectional readout of amino acid sequences by MS/MS analysis. Recently, a new cleavable ICAT reagent with the ^{13}C isotope linker and acid-cleavable biotin group has been developed, allowing for us overcome the limitations of ICAT, which include the co-elution of D_0 and D_8-labeled peaks and the suppressed peptide peaks caused by the masking effect of biotin group[25].

A novel method for quantitation of proteins, *i*sobaric *t*ags for *r*elative and *a*bsolute *q*uantitation (iTRAQ), was recently introduced to employ four-plex labeling of amine-reactive isobaric tags (*m/z* 145), thus, iTRAQ is theoretically able to label the whole proteome (Figure 14). The iTRAQ labeling is based on the labeling of peptide at amine groups of N-termini and side chain of lysine, thereby labeling all the proteins after trypsinization. The peptides labeled with four-plex isobaric tags cannot be directly used to identify each state of the proteome in MS. However, when the collected four-plex tagging peptides were fragmented via MS/MS analysis, signature ion peaks from *m/z* 114 to 117 are generated to obtain the relative quantitative information. The MS/MS spectra are concomitantly provided in order to identify the protein to be targeted[26]. Chemical tagging methods, including ICAT, cICAT, and iTRAQ allow for the labeling of the proteins or peptides in vitro, thus allowing for labeling, enrichment, and purification, followed by MS and MS/MS analyses. The serial chemical reaction processes include poor reproducibility and low confidence. Thus, labeling stable isotope labeling by essential amino acid culture (SILAC) is another innovative approach that has been developed which can bypass the unnecessary chemical reactions[27]. Moreover, SILAC involves *in vivo* labeling into live cells via metabolic incorporation, in which the proteomes in two states of cultured cell populations are capable of being utilized for direct quantitative analysis from purified proteins in the cultured cells being studied (Figure 15).

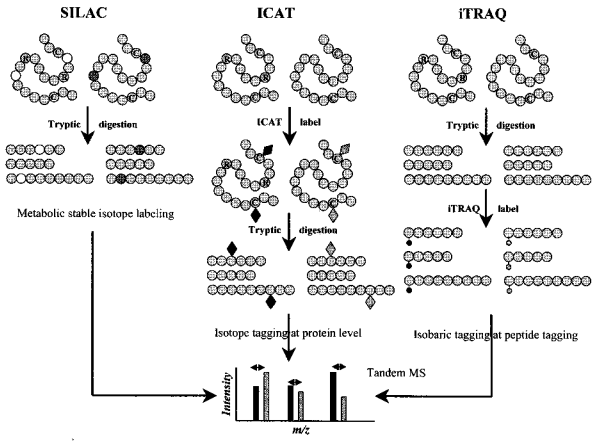

Fig. 12. Schematic representation of MS-based quantitative analysis. Proteins are metabolically incorporated into media with stable isotopes containing ^{15}N-amino acids (SILAC). Proteins and post-tryptic peptides are specifically labeled at the cysteine and amine groups of lysine with ICAT and iTRAQ reagent dyes, respectively. In each case, the labeled proteins or peptides are combined, separated, and analyzed via tandem MS. During analysis, the relative abundance and protein identification were determined by MS and MS/MS analyses, respectively.

TOWARDS INTEGRATED BIOLOGY

Due to the rapid development of proteomic technologies, in particular, mass spectrometric methods, biologists have begun to build "protein legos", which

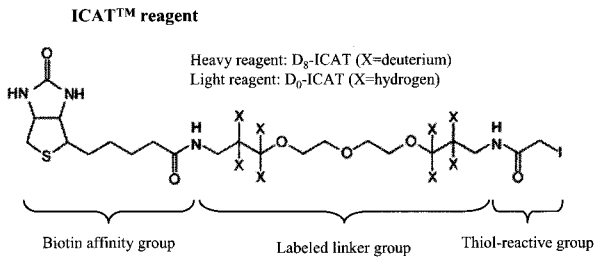

Fig. 13. Chemical structure of ICAT™ reagent. A biotin affinity tag was used to selectively isolate the tagged peptides. A mass-encoded polyether harbors 8 hydrogens (D_0) or 8 deuteriums (D_8) for the quantitation of relative abundance in two cell or tissue states. The thiol-reactive group is tagged to the cysteine residues.

allow for the construction of dynamic motion pictures of internal cellular functions. This is an important advance from previous techniques, as the cell itself is not a static construct. Current techniques tend to integrate a variety of data forms concerned with quantitative information (proteome dynamics), proteome localization (subproteome atlas), and protein-protein interaction (proteome society), enabling researchers to analyze and integrate systematically a variety of intracellular occurrences, including biomolecular networking, using enhanced computing power coupled with mathematical and bioinformatic techniques[28]. Combined with a host of breakthroughs that have obviated previous technical barriers, current proteomics techniques are rapidly becoming a quantitative technical platform by which biological functions can be expressed in numerical language. Thus, we hope to deal with previously intractable cellular conundra using mathematics, a pursuit which has been referred to as either integrated biology or systems biology.

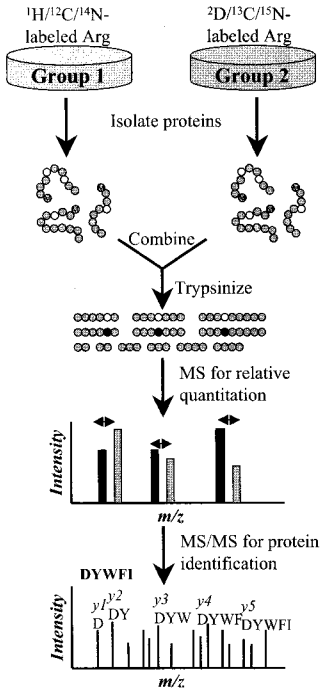

Fig. 15. SILAC *in vivo* labeling method scheme. An equivalent number of two sets of cells were grown on minimal media with either normal amino acid (generally used in $^{1}H/^{12}C/^{14}N$-arginine) or stable isotope amino acid ($^{2}D/^{13}C/^{15}N$-labeled arginine). These mass tags have been subsequently combined and allowed for the quantitation of the relative abundance and identification of proteins via MS and MS/ MS analysis, respectively.

ACKNOWLEDGEMENTS

I thank all the members of proteomics team in Korea Basic Science Institute for all of their valuable comments and reference data. This research was

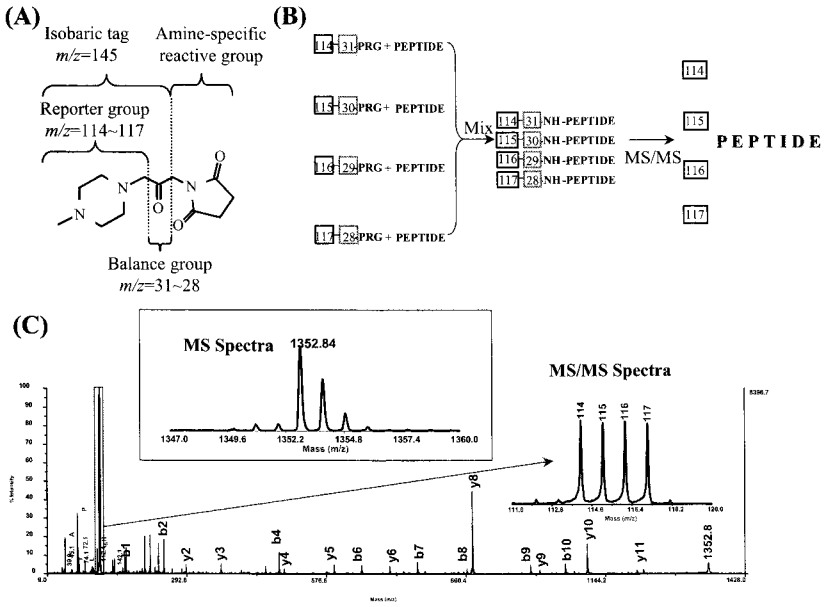

Fig. 14. Chemical structure of four-plex iTRAQ^{TM} reagent. (A) The overall chemical structure includes the reporter group (*m/z* 114~117), isobaric tag (total mass=145), and the amine-specific peptide reactive group (NHS). (B, C) Four-plex isobaric tagged peptides were combined and subjected to MSMS analysis, resulting in a determination of relative abundance by different reporter mass intensities and the identification of internal sequences via peptide fragmentation.

financially supported by a grant from the Korea Basic Science Institute, in fiscal year 2006.

REFERENCES

1. Wilkins, M. R. Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., and Williams, K. L. (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* 13, 19-50.
2. DeRisi, J. L., Iyer, V. R. and Brown, P. O. (1997) Exploring the metabolic and genetic control of gene expression in a genomic scale. *Science* 278, 680-686.
3. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19, 1720-1730.
4. Hoogland, C., Sanchez, J. C., Walther, D., Baujard, V., Baujard, O., Tonella, L., Hochstrasser, D. F., and Appel, R .D. (1999) Two-dimensional electrophoresis resources available from ExPASy. *Electrophoresis* 20, 3568-3571.
5. 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., (2001) The sequence of the human genome. *Science* 291, 1304-1351.
6. The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *Caenorhabditis elegans*. A platform for investigating biology. *Science* 282, 2012-2018.
7. Davidson, E.H., Rast, J. P., Oliveru, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., et al., (2002) A genomic regulatory network for development. *Science* 295, 1669-1678.
8. Fields, S. (2001) Proteomics. Proteomics in genomeland. *Science* 291, 1221-1224.
9. MacCoss, M. J., McDonald, W.H., Saraf, A., Sadygov, R., Clark, J. M., Tasto, J. J., Gould, K. L., Wolters, D., Washburn, M., Weiss, A., Clark, J. I., and Yates, J. R.3rd. (2002) Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Nat'l. Acad. Sci.* USA 99, 7900-7905.
10. Kim, Y. H., Cho, K., Yun, S. H., Kim, J. Y., Kwon, K.H., Yoo, J. S., and Kim, S. I. (2006) Analysis of aromatic catabolic pathways in *Pseudomonas putida* KT2440 using a combined proteomic approach: 2-DE/MS and cleavable isotope-coded affinity tag analysis. *Proteomics* 6, 1301-318.
11. Görg, A., Postel, W., Gunther, S., Weser, J., Strahler, J.R., Hanash, S. M., Somerlot, L., and Kuick, R. (1988) Approach to staionary two-dimensional pattern: influence of focusing time and immobiline/carrier ampholytes concentrations. Electrophoresis 9, 37-46.
12. Lewis, T. S., Hunt, J. B., Aveline, L. D., Jonscher, K. R., Louie, D. F., Yeh, J. M., Nahreini, T. S., Resing, K. A., and Ahn, N. G. (2000) Identification of novel MAP kinase pathway signaling targets by functional proteomics and mass spectrometry. Mol. Cell 6, 1343-1354.
13. Unlu, M., Morgan, M. E., and Minden, J. S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis 18, 2071-2077.
14. Washburn, M. P. Wolters, D., and Yates, J. R. 3rd. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242-247.
15. Edman, P. (1949) A method for the determination of the amino acid sequence in peptides. Arch Biochem 22, 475-476.
16. Gooley, A. A., Ou, K., Russell, J., Wilkins, M. R., Sanchez, J. C., Hochstrasser, D. F., and Williams, K. L. (1997) A role of Edman degradation in proteome studies. Electrophoresis 18, 1068-1072.
17. Pappin, D. J., Hojrup, P., and Bleasby, A. J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. Curr. Biol. 3, 327-332.
18. Choi, J.S., Kim, D. S., Lee, J., Kim, S. J., Kim, S. I., Kim, Y. H., Jong, J., Yoo, J. S., Suh, K. H., and Park, Y. M., (2000) Proteome analysis of light-induced proteins in Synechocystis sp. PCC 6803: Identification of proteins separated by 2D-PAGE using N-terminal sequencing and MALADI-OF MS. Mols. Cells 10, 705-711.
19. Hunter, T. (1995) Protein kinases and posphatases: the yin and yang of protein phosphorylation and signaling. Cell 80, 225-236.
20. Schlessinger, J. (1993) Cellular signaling by receptor tyrosine kinases. Harvey Lect. 89, 105-123.

21. Kaufmann, H., Bailey, J. E., and Fussenegger, M. (2001) Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis Proteomics 1, 194-1999.
22. Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. Nature 258, 598-599.
23. Sano, A. and Nakamura, H. (2004) Titania as a chemo-affinity support for the column-switching HPLC analysis of phosphopeptides: application to the characterization of phosphorylation sites in proteins by combination with protease digestion and electrospray ionization mass spectrometry. Anal Sci 20, 861-864.
24. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature Biotech. 17, 994-999.
25. Hansen, K. C., Schmitt-Uhlms, G., Chalkley, R. J., Hirsch, J., Baldwin, M. A., and Burlingame, A. L. (2003) Mass spectrometric analysis of protein mixtures at low levels using cleavable 13C-isotope-coded affinity tag and multidimensional chromatography. Mol. Cell. Proteomics 2, 299-314.
26. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Barlet-Jones, M., He, F., Jacobson, and A., Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cervisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics 3, 1154-1169.
27. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell. Proteomics 1, 376-386.
28. Kisinger, T., Cox, B., Kannan, A., Chung, C., Hu, P., Ignatchenko, A., Scott, M. S., Gramolinoi, A. O., Morris, Q., Hallet, M. T., Rossant, J., Highes, T. R., Frey, B., and Emili, A. (2006) Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. Cell 125, 173-186.