Original

Development of Large−Scale Protein Identification Technology Using High Resolution LC−MS

Masato Taoka

Laboratory of Biochemistry, Department of Chemistry, Graduate School of Science,

Tokyo Metropolitan University, 1−1 Minami−osawa, Hachioji−shi, Tokyo, Japan

Received for review January 19, 2005. Accepted June 28, 2005.

Abstract

The author demonstrated a newly developed fully integrated online two−dimensional (2 D) LC−MS/MS system that allows non−gel based mass identification of components in a complex protein mixture by analyzing its proteolytic digests. The system was composed of a 2 DLC, which consists of two independent HPLC one with an ion exchange and the other with a reversed phase capillary columns, a hybrid type MS spectrometer with an electrospray ionization source, and a software program for retrieval of mass spectral information. The system successfully showed that proteins present in standard mixtures could be identified at fmol levels over a 300−fold difference in molar quantity with significant reproducibly. Approximately 4,000 peptides derived from 1,000 proteins, which include the proteins rarely seen in 2 D−PAGE coupled mass spectrometric analysis, were identified in a single analytical run from *E. coli* K 12 lysate by using the system. This 2 DLC−MS/MS system can be used for large−scale protein and peptide analysis in strategy of proteomics, in particular, for mass−spectrometric identification of protein components in crude biological sources.

Keywords: proteomics/twodimensional/liquid chromatography/tandem mass spectrometry/*Escherichia coli*

1. Introduction

Postgenomic era requires the challenge of analyzing proteome expressed by an organism, tissue, cell, etc. to aid in the understanding of the operation of complex cellular pathways, networks, and modules under various physiological conditions. The proteome is not fixed, but changes with the state of the development and the environmental conditions. A typical proteomic study includes: determination of quantitative changes in the expression levels of proteins to assess the effects of a wide variety of perturbations to cells, and analysis of protein interactions by mass identification of protein components in the functional protein complexes and cellular organelle [1]. Thus, the technologies of proteomics depend largely on the methods of protein separation and identification.

Among the current technologies two−dimensional polyacrylamide gel electrophoresis (2 D−PAGE) followed by mass spectrometry (MS) is most widely used for protein separation and identification [2,3]. 2 D−PAGE allows resolving thousands of proteins in a single analysis. However, the protein identification is essentially of spot−basis, and requires time−consuming multi−step manual handling for digestion, extraction, and pre−treatments for MS analysis [4,5]. In addition, the experimental 2 D−PAGE procedure is yet to be automated, and portions of cellular proteins such as those with extremes in pI and Mr and membrane−associated proteins are rarely found in 2 D−PAGE studies [6]. These technical limitations prevent high−throughput large−scale protein analysis, which is a key issue in proteomics.

Liquid chromatography (LC)−based methods for the analysis of components of protein mixtures have provided an alternative approach for proteomic analyses [7]. The LC−based technology, in combination with the electrospray ionization (ESI)−tandem mass spectrometry (MS/MS), allows online automation of protein analysis. The identification is based on the method called “peptide se-

Correspondence: Masato Taoka, Ph. D., Laboratory of Biochemistry, Department of Chemistry, Graduate School of Science, Tokyo Metropolitan University, Minami−osawa 1−1, Hachioji, Tokyo 192−0397, Japan.
Phone & Fax: +81−426−77−2525
E−mail: mango@comp.metro−u.ac.jp

quence tag", in which the internal amino acid sequence is considered for reliable peptide assignment and protein identification in a high throughput manner [8,9]. However, the complex mixture often overwhelms the resolution capability of any single dimensional LC since protease digestion has increased the diversity of the mixture. Two−dimensional methods combining orthogonal modes for separation could improve the resolution of highly complex mixtures.

In almost two decades ago, our group presented an automated two−dimensional high performance liquid chromatography (2 DLC) technique for systematic separation of very complex protein and peptide mixtures [10−12] and applied the system for sequence analysis of very large proteins and protein genetic variants [13−15], and for profiling proteins expressed in developing rat cerebella [16 −18]. This technique has now been refined for more sophisticated approach by incorporating new MS technology, e.g., replacing a conventional UV detector by MS with electrospray ionization source. For example, Yates *et al.* reported the micro−scale 2 DLC −MS/MS system for comprehensive analysis of functional protein complex and crude biological materials fractionated [19]. The multidimensional system employs different solvent for each chromatographic step and control salt by using volatile one to avoid detrimental effect to a mass spectrometer [20−22]. For multidimensional−LC system, however, several types of chromatography with variety of buffer solutions were necessary and complete replacement of chromatographic solvents with MS compatible ones is required to combine the multidimensional system with a mass spectrometer. Therefore, it is necessary to develop a new analytical platform for proteomics, i.e., online 2 DLC−MS/MS, which provides a fully integrated system equipped with solvent replacing device, "trap column" [23]. In this report, the author demonstrates the characteristics and advantages of this system for non−gel based large−scale protein identification by analyzing standard protein mixtures and a crude biological source.

2. Materials and methods

Materials

Chromatography grade of water, acetonitrile, and formic acid were obtained from Wako Pure Chemical Industries, Ltd. (Tokyo, Japan). Human serum albumin (A 9511) and a protein mixture for SDS−PAGE marker (M 4038) containing myosin (MYOH), beta−galactosidase (BGAL), phosphorylase b (PHS), fructose−6−phosphate kinase (K 6 PH), albumin (ALB), glutamate dehydrogenase (DHE), ovalbumin (OVAL), glyceraldehyde−3−phosphate dehydrogenase (G 3 P), carbonic anhydrase (CAH), trypsinogen (TRY), trypsin inhibitor (ITRA), alpha−lactoalbumin (LCA) and aprotinin (IBPS) were purchased from Sigma−Aldrich, Inc. (St. Louis, MO, USA). Lysate of *E. coli* in the late logarithmic phase was prepared as described in our previous report [24]. All other reagents were of analytical−reagent grade from Wako Pure Chemicals, unless otherwise mentioned. Empty column was purchased from LC Assist (Tokyo, Japan) and Sugiyama (Yokohama, Japan). Column−filter (2 SR 1) was from Valco Instruments Co. Inc. (TX, USA). Fused silica capillary and PEEK tube were purchased from GL Science Inc. (Tokyo, Japan).

Columns

All the columns were packed in our laboratory. An ion exchange column (2 mm ID x 35 mmL) was packed with Bioassist Q (TOSOH, Tokyo, Japan). A reversed phase column (320 μm ID x 100 mm L) was packed with C 18 bonded silica gel (Mightysil C 18, 3 μm, Kanto Chemicals, Tokyo, Japan). A trap column (1 mm ID x 5 mm L, LC Assist, Tokyo, Japan) was packed with C 18 (Mightysil C 18, 15 μm, Kanto Chemicals).

Installation

The system was composed of six parts, *i.e.* two independent HPLC modules, a trap column assembly, an HPLC controller, a mass spectrometer and a PC−based data analysis system as shown in Figure 1. The ion−exchange chromatograph (LC−1) consisted of two pumps (LC−10 Dvp, Shimadzu, Kyoto, Japan) a dynamic mixer (HG 1580−32, JASCO, Tokyo, Japan) equipped with 80 μL mixing chamber and a sample injector (type 7125, Rheodyne, CA, USA) with a 300 μL loop. The reversed phase chromatograph (LC −2) consisted of two pumps (LC−9 AD, Shimadzu), the same dynamic mixer as used in LC−1 with a 200 μL mixing chamber, and a splitter consisting of a stainless steel tee connector (ZT 1 C, Valco) and a 200−mm long fused silica capillary (30 μmID).

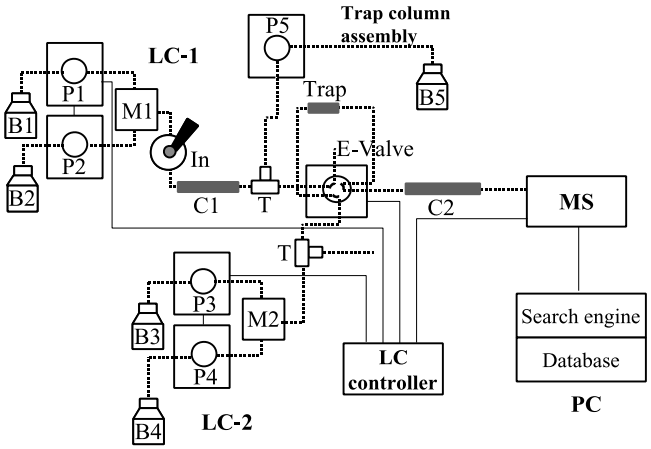

Figure 1. Schematic diagram of the 2 DLC−MS/MS system. The assemblies are explained in the text. B 1−5, eluents 1−5; C 1, ion−exchange column; C 2, reversed phase column; D, drain; E−Valve, electric valve; In, injector; M, dynamic mixer; P 1−5, pumps 1−5; Trap, trap column; T, T connector.

The trap column assembly consisted of a six−way valve (C 2−0006, Valco), a T connector (ZT 1 C, Valco), a pump (LC−10 Advp, Shimadzu) and the trap column. The two HPLC modules were connected via the six−way valve of the trap column assembly with fused silica capillary (50 μm ID) tubing. The HPLC controller (SCL−10 Avp, Shimadzu) synchronized the operation of the HPLC modules and the trap column assembly. The trap column assembly served to remove salts from peptide eluted from the first ion−exchange column C 1. The trap column assembly plays an important role to obtain reproducible analytical results because the salts used in the ion−exchange chromatograph (LC−2) prevent efficient ionization of samples in the MS analysis and often cause malfunction of the system.

A Hybrid type mass spectrometer (Q−TOF 2, Micromass, Manchester, UK) was used for the system as a detector. The mass spectrometer was equipped with an ESI−source composed of a stainless steel capillary tube (130 μm OD, 50 μm ID), a 3−way connector (Valco) and a stainless steel sheath for N_2 gas. The data analysis system consisted of MS/MS data converter software (MassLynx, Micromass), amino acid sequence database and MS/MS ion search software (Mascot, Matrix Science Ltd., London, UK) run on a PC with dual processors (PentiumⅢ, 833 MHz). Performance of the 2 DLC−MS/MS−system components was described previously in ref 23. In brief, columns C 1 and C 2 are equilibrated with B 1 (0.025 M Tris−HCl buffer, pH 8.0) and B 3 (0.2% formic acid in water), respectively, at a flow−rate of 100 μL/min (C 1) or 5 μL/min (C 2) for 20 min by the method program on the HPLC controller. A sample mixture (50−200 μL) is applied to column C 1 through a sample injector, and is eluted with B 1 for 5 min at a flow−rate of 100 μL/min. The eluent is mixed on−line with solvent B 5 (1% formic acid) continuously supplied with P 5. The mixture flows into a reversed phase trap column, where the salts are removed from peptides by continuous washing with solvent B 5. The washing process continues for 2 min to completely remove the salts while the eluent from C 1 is stopped. After the desalination, the six−way valve is switched the line to connect the trap column and C 2. The salt−free samples adsorbed onto the trap column are introduced into the second column (C 2) and are separated by using a 70−min linear gradient elution from B 3 to B 4 (5 to 60% of CH 3 CN in 0.2% formic acid) at a flow−rate of 5 μL/min. Column C 2 is equilibrated again with B 3 for 15 min after the completion of the linear gradient elution. By this step, the whole cycle of the 2 DLC finishes. Then, LC−1 for the first dimension of the chromatography starts again to elute peptides stepwise from the ion−exchange column C 1 by introducing a portion of buffer B 2 (0.4 M NaCl in 0.025 M Tris−HCl buffer, pH 7.5) into B 1. After applying the eluent to C 2, the second dimension of the chromatography is repeated as described above. These separation cycles were repeated 10 times, changing the mixture ratio of B 1 and B 2. Eluent B 2 percentages in steps 1−10 are 0, 5, 10, 15, 20, 25, 30, 40, 50, and 100%. The separated peptides are directly sprayed into the hybrid mass spectrometer. Electrospray ionization is carried out at a voltage of 2.7 kV and tandem mass spectra are automatically acquired in data−dependent mode during the entire run. Under this condition, the total time required for a 2 DLC−MS/MS analysis is 16 hours.

Database search and annotation.

Data generated by mass spectrometry are generally converted to text files listing mass values and intensities of fragment ions and analyzed by the search engine (Mascot) for peptide assignment and protein identification on database (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/Escherichia_coli_K 12/U 00096.faa and ftp://ftp.ncbi.nlm.nih.gov/blast/db/nr.Z, NCBI). The parameters for the database search and protein annotation were as previously described [24].

Preparation of peptide from protein mixture.

Protein mixture (200−500 μg protein) were dissolved in 200 μL of 0.5 M Tris−HCl buffer (pH 8.5) containing 7 M guanidium hydrochloride and 10 mM EDTA, then the solution was bubbled with N_2 gas for 10 min and 100 μg of dithiothreitol (3 mM) was added to the solution. After 2 hours, 250 μg of iodoacetamide (6.5 mM) was added and left the solution in the dark for 2 hours at room temperature. The solution was dialyzed against 10 mM ammonium bicarbonate buffer (pH 8.0), digested overnight at 37℃ with TPCK−treated trypsin (Cooper Biomedicals, Malvern, PA) or Lys−C (Wako Pure Chemicals), and stored frozen at −80℃ until use.

The digests were acidified to pH 2 by addition of an aliquot of 6 M−HCl and, removed precipitates by centrifugation. The supernatant was neutralized with aqueous ammonia to pH 8, the peptide mixture was diluted with an equal volume of water, and applied to the 2 DLC−MS/MS system.

SDS−PAGE and quantitation of separated protein.

SDS−PAGE followed by Coomassie Brilliant Blue staining and quantitation of protein bands were performed as previously described [25].

3. Results and discussion

Separation of tryptic digest from standard protein: coverage of protein digests, dynamic range and sensitivity

As for the evaluation of the system, the author prepared the tryptic digest of human serum albumin, which contains 585 amino acids and has the potential to yield 71 tryptic peptides and free

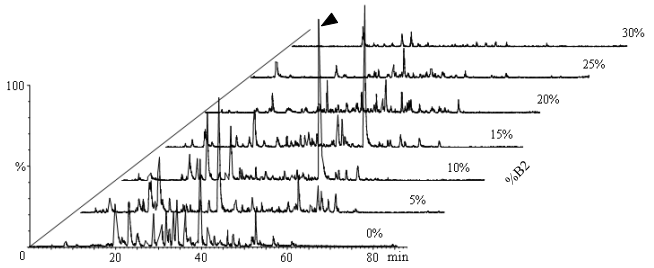

Figure 2. A three−dimensional visualization of 2 DLC profile of the tryptic digest of human serum albumin.
Each horizontal profile represents the single ion chromatogram of the second dimension of chromatography (reversed phase). The percentage of B 2 in B 1 eluent is shown at the right−hand side of each profile.

amino acids. The peptides were introduced into the system and automatically analyzed by data dependent MS/MS mode during the entire run as shown in Figure 2. Using 40 pmol of the tryptic digest, 48 peptides were identified except for the small peptides (molecular mass <700) and free lysine, and the peptides covered 83.9 % of the entire amino acid sequence of the protein (See Supplementary Table 1, http://www.sci.metro−u.ac.jp/proteomicslab/). The small peptides and free amino acids unidentified by the system may not have been retained on the trap column during the first dimension of chromatography (ion−exchange) due to the low hydrophobicity [26]. In addition to this wide coverage, the detection limit of this system was estimated to be 46 fmol at a signal−to−noise ratio (S/N) of 2, based on a fragment with m/z 682.4^{3+} (indicated by a closed arrowhead in Figure 2) detected using a single−ion monitor. These results demonstrate the usefulness of the 2 DLC−MS/MS system for highly sensitive identification of proteins and for analysis of post−translational modifications in proteins available in limited amounts.

In addition to albumin, haptoglobin and transferrin were reproducibly identified from the same mixture by the system (Table 1). These impurities were obviously detected by western blot analysis (data not shown) but barely by SDS−PAGE following Coomassie blue staining as shown in Figure 3 because of low concentration. Optical density of the bands showed that this reagent contained 0.37% of transferrin (130 fmol in 40 pmol albumin) and unquantifiable level of haptoglobin. At this level, transferrin was present at 300 times less than the molar amount of serum albumin indicating that 2 DLC−MS/MS system had broad dynamic range for analyzing protein mixture. This broad dynamic range should be achieved owing to high−resolution separation by 2 DLC and sensitive detection by mass spectrometer, and was also maintained in the analysis of the crude protein mixture such as *E. coli* proteins in late logarithmic phase described below and previously [24].

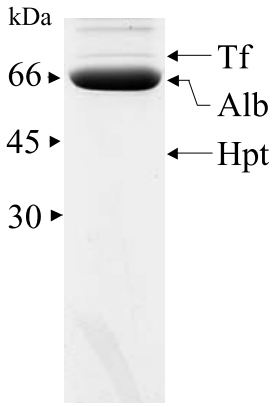

Figure 3. Analysis of the impurities in the albumin reagent.
Two μg of human serum albumin reagent was analyzed in a 17 % of polyacrylamide gel containing 0.1% SDS. Molecular mass markers are indicated on the left−hand side. The arrows indicate the protein bands of transferrin (Tf), serum albumin (Alb) and haptoglobin (Hpt).

Reproducibility and efficiency of the system

A well−characterized mixture of proteins is commonly used to

Table 1. The peptide identified from impurity of the human serum albumin reagent.

Protein	Sequence	Charge	m/z observed	Delta m/z (ppm)	Score[a]
haptoglobin	SCAVAEYGVYVK	2+	673.2	−17.9	17.8
	SPVGVQPILNEHTFCAGMSK	3+	724.6	−30.3	9.9
	VTSIQDWVQK	2+	602.3	−3.0	18.3
	YVMLPVADQDQCIR	2+	854.3	−6.4	27.4
transferrin	EFQLFSSPHGK	2+	638.8	1.9	3.2
	FDEFFSEGCAPGSK	2+	789.3	15.0	19.9
	MYLGYEYVTAIR	2+	739.8	29.6	29.6
	SAGWNIPIGLLYCDLPEPR	2+	1086.0	21.4	10.5
	SASDLTWDNLK	2+	625.3	46.6	30.3
	SVIPSDGPSVACVK	2+	708.3	−22.4	8.1

a. Score indicates individual score minus threshold score ($p<0.05$). Each score was from result of Mascot search.

Table 2. Proteins identified from the standard mixture.
The retention time of the representative and peptide number of peptide identified from each protein were compared among four independent analytical runs under same condition.

Protein [a]	Mr (kDa)	pI	Representative peptide	Retention time (min)[b]		Number of peptide identified [b]			
				exp.1	exp.2	exp.1	exp.2	exp.3	exp.4
MYOH	205	5.6	SELQAALEEAEASLEHEEGK	1136.0	1135.8	10	17	5	3
PHS	97	6.6	VFADYEEYVK	517.6	517.7	4	4	1	2
K 6 PH	84	8.5	GQIEEAGWSYVGGWTGQGGSK	158.9	160.1	2	2	Not identified	1
ALB	66	5.8	TCVADESHAGCEK	372.9	372.1	10	16	13	11
DHE	55	8.4	SEAAADREDDPNFFK	660.7	660.3	5	7	1	4
OVAL	45	5.2	LTEWTSSNVMEERK	150.6	150.0	2	3	2	2
G 3 P	36	8.5	VISNASCTTNCLAPLAK	45.7	44.8	4	4	2	3
CAH	29	7.9	EPISVSSQQMLK	11.6	11.5	5	8	2	4
TRY	24	8.2	DSCQGDSGGPVVCSGK	157.0	156.3	8	7	2	5
ITRA	20	5.0	LVFCPQQAEDNK	157.8	158.0	4	4	4	4
LCA	14	4.9	FLDDDLTDDIMCVK	878.5	879.7	5	5	2	5
IBPS	6.5	9.2	SAEDCMRTCGGA	168.6	168.0	2	2	2	3
BGAL	116	5.3	−−−	−−−	−−−	Not identified	Not identified	2	Not identified

a. Abbreviations were described in Materials & Methods.
b. Lys−C peptide was analyzed in exp. 1 and 2, and tryptic peptide was analyzed in exp. 3 and 4.

verify the consistency of results obtained in a biochemical experiment. Such a mixture should have proteins of widely different molecular weights and isoelectric points. SDS−PAGE protein marker is a readily available protein mixture meeting these criteria. The marker that was constituted of the same amount of proteins listed in Table 2 with wide range of molecular weight (6.5−205 kDa) and isoelectric point (4.9 to 9.2) was used. Five μg of Lys−C and tryptic peptide from the marker were analyzed, respectively, by 2 DLC −MS/MS system two times each. In these analyses, the author selected the most significant peptides identified from each protein and compared the retention time on 2 DLC. All the peptides listed in Table 2 were eluted from 2 DLC within 0.9% variation of the retention times, indicating significant reproducibility of the LC section of the system. In 4 runs, the number of peptide identified from each protein was almost identical, and almost all the proteins were reproducibly identified. These results indicate that the system is capable of reproducibly identifying proteins regardless of pI and molecular weight.

Even though the system offers the high reproducibility and reliable identification of proteins as discussed in the previous section paragraph, beta−galactosidase was only occasionally identified. However, this occasional identification of the protein is not false but is positive. The reliability of the identification was verified by the detailed analysis of MS/MS spectra. This occasional identification may be caused by the protein nature which has the potential to produce only a few peptides covered by the mass range of the system, shows low expression level, and/or by the characteristics of the automatic MS/MS acquisition system which sometimes irregularly pick up peptides [24]. Therefore, the frequency of occasional identifications increases as the complexity of the protein sample becomes higher. For instance, when we analyzed more complex mixtures, such as *E. coli* [24] and embryonic stem cell proteome [27] the occasional identification increased to about 20%. It should be noted that the occasional identifications of such complex proteins appeared to be reliable. We repeatedly analyzed the same peptide prepared under the same conditions to collect occasionally identified proteins and extended the coverage of the proteome [24,27−30].

To examine the efficiency of identification of the 2 DLC−MS/MS system, tryptic peptide of *E.coli* was introduced into the system. Figure 4 shows the relationships between the amounts of peptides and percentages of identified peptides and proteins. By increasing the loading amount of crude peptides of *E.coli*, the percentages of identified peptides and proteins were increased and reached a plateau at around 200 μg. This proves that our standard 2

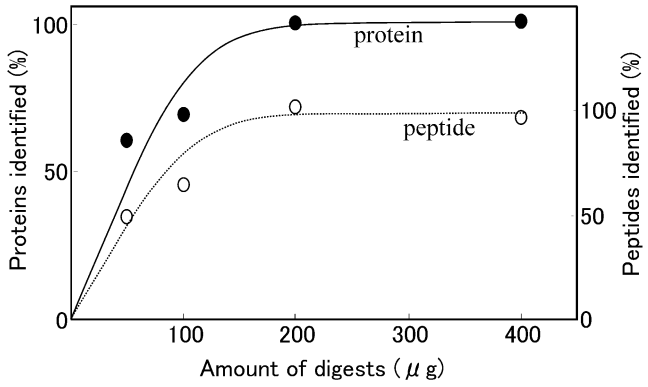

Figure 4. Relationships between the amounts of peptides and percentages of identified peptides and proteins. The data was from 2 DLC−MS/MS by applying crude tryptic digest of *E.coli.* Open and closed circles indicate relative number of identified peptides and proteins, respectively.

DLC–MS/MS system equipped with a 2.0 mm I.D. anion–exchange column and a 0.3 mm I.D. reversed–phase column showed maximal performance even with the loading of complex peptides as low as 200 μg.

Analysis of peptides from crude protein mixture

Now, a typical analysis of the late logarithmic phase of *E.coli* proteome will be presented. Figure 5 illustrates a part of the second dimension of MS and MS/MS chromatograms, which were obtained from the eluent of the first dimension of chromatography at 0% B 2 eluent. In this step, about 2,500 MS/MS spectra were generated and the database search was performed for the spectra through the Mascot algorithm. This step was repeated 10 times with increasing the concentration of sodium chlorides in the eluent of the first dimension of chromatography. Finally, 16,640 MS/MS spectra were obtained which could be assigned to 4,106 peptides leading to the identification of 1,019 proteins, which is about one–fourth of the number of open reading frames predicted from genomic data [31]. Post–translational modifications including 5 N–terminal acetylations and 36 deletions of initial methionine were also detected. A complete list of the proteins and peptides identified by this run was shown in Supplemental Table 2 (http://www.sci.metro–u.ac.jp/proteomicslab/). In this analysis, the portions of cellular proteins such as those with extremes in pI (115 proteins with pIs >9 or <4) and Mr (76 proteins with Mr >100 kDa or <10 kDa), integral membrane proteins (112 proteins) and rarely translated proteins (133 proteins with codon adaptation index below 0.3) were identified. In this work, the number of proteins identified was about 10 times more than that by the 2 D–PAGE based analysis previously reported by Tonella *et.al.* [32].

Using knowledge database, the proteins identified in this work were categorized by the major functional groups (See Figure 6). The proportion between each category was almost the same as that of the genome wide prediction, indicating that this system gave access to *E.coli* proteome in a largely unbiased manner. It should also be noted that this system identified not only abundant cellular proteins such as ribosomal proteins and the elongation factors Tu and G that exist at 10^4 to 10^5 copies per cell [33] but also very minor

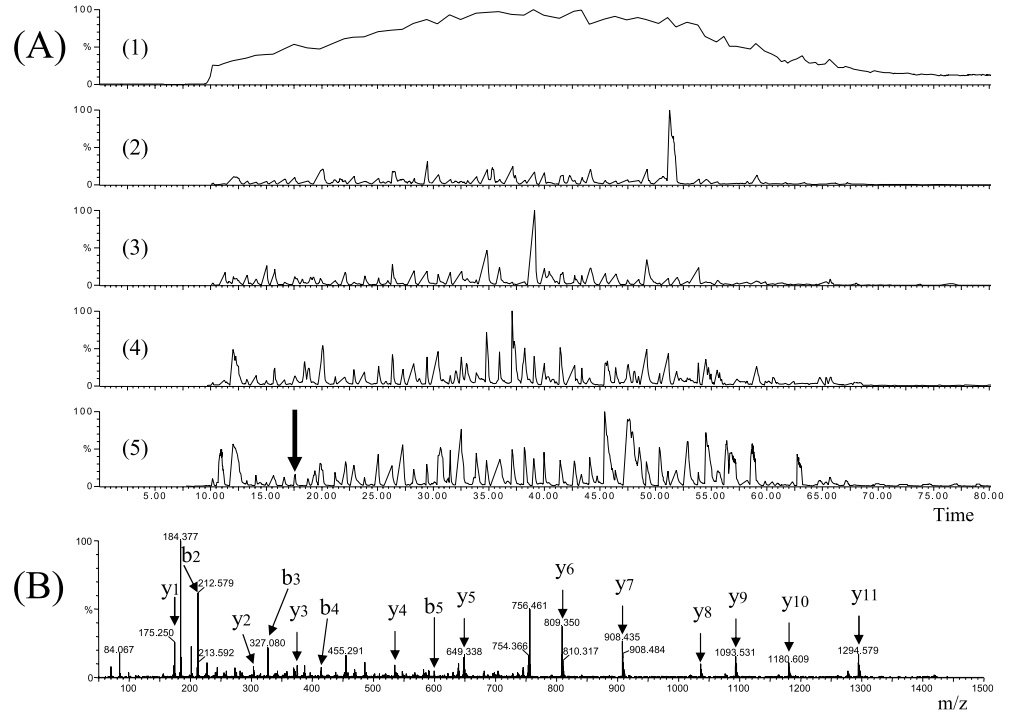

Figure 5. 2 DLC–MS/MS analysis of the tryptic digest of E.coli lysate.
(A 1) A part of typical total ion chromatogram generated in MS mode. The chromatogram shows the analysis of the eluent from the first dimension of chromatography at 0% B 2 eluent.
(A 2–5) Parts of the typical total ion chromatograms generated in MS/MS mode. The four most intense ions in each MS scan were automatically selected for collision–induced fragmentation and analyzed in MS/MS mode. The arrow at 17.6 min correlates with the peptide that was selected as an example of peptide sequence shown in (B).
(B) Corresponding MS/MS spectrum obtained from peak eluting at 17.6 min. A doubly charged ion of mass–to–charge ratio (m/z) of 753.89, corresponding to a peptide of 1505.78, was selected for fragmentation by collision–induced dissociation. The fragment ions that originate either from the N terminus (b type ions) or the C terminus (y type ions) correspond to an amino acid sequence that was identified using the database search program. In this instance, the MS/MS spectrum matched the peptide sequence VINSGQVCNCAER from aldehyde dehydrogenase residues 278–290 (accession number 1787684).

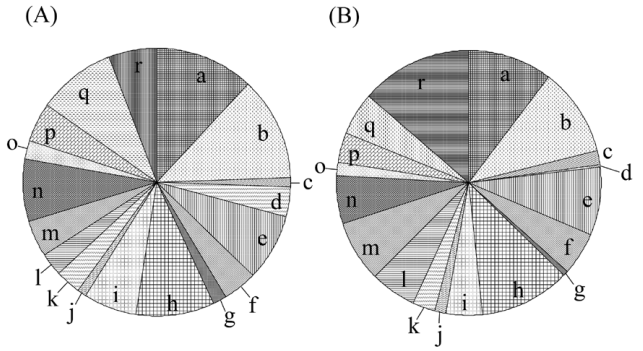

Figure 6. Comparison between functional categories of (A) proteins identified by the 2 DLC−MS/MS analysis and (B) proteins predicted from genome sequence of E.coli.
The proteins identified in this study and predicted from *E. coli* genome database were classified into 18 functional criteria [35]. Protein classified into "Hypothetical, unclassified, unknown" was not included in the graphs. Abbreviations: a, Amino acid transport and metabolism; b, Carbohydrate transport and metabolism; c, Cell cycle control, mitosis and meiosis; d, Cell motility; e, Cell wall/membrane biogenesis; f, Coenzyme transport and metabolism; g, Defense mechanism; h, Energy production and conversion; i, Inorganic ion transport and metabolism; j, Intracellular trafficking and secretion; k, Lipid transport and metabolism; l, Nucleotide transport and metabolism; m, Post−translational modification, protein turnover and chaperones; n, Replication, recombination and repair; o, Secondary metabolites biosynthesis, transport and catabolism; p, Signal transduction mechanisms; q, Transcription; r, Translation.

protein components such as subunits of DNA polymeraseIII (dnaE, dnaX, dnaN, and dnaB), which were present at no more than 100 copies per cell [34]. Thus, the analysis appears to have covered *E. coli* proteins whose cellular abundance over 100 copies per cell in a single run. These results indicate that the 2 DLC−MS/MS technique is a very powerful tool for large−scale and comprehensive protein analysis of crude biological sources.

4. Conclusions

The author presented here a new analytical platform for proteomics, 2 DLC−MS/MS system, which provided a fully integrated approach to non−gel based protein identification. Because of the small amount of sample required, this technique is useful for identification of proteins in various tissues and cells without tedious sample preparation and complicated manipulation. For instance, *E. coli* cells [24], embryonic stem cells [27], whole body of *C. elegans* [30], post synaptic density [28], endoplasmic reticulum of liver, lipid raft, etc. were analyzed and ~2000 proteins were identified from 200−1000 μg of their tryptic peptides. In addition to the identification, information about approximately 1,000 proteins in a single analysis could be given by introducing the strategy using a stable isotope tag into the 2 DLC−MS/MS technology [36]. Therefore, the method presented here is useful for large−scale surveys of differential protein expressions of cells and is also suitable for the medical and pharmaceutical applications such as diagnosis of human diseases and drug discovery and safety assessment.

Acknowledgments

I thank Drs. Toshiaki Isobe, Yoshio Yamauchi, Muneo Saito, Hiroyuki Kaji and Takashi Shinkawa (Tokyo Metropolitan University) for helpful discussions and technical assistance. This work was supported in part by Grants for Integrated Proteomics Project, Pioneer Research on Genome the Frontier from MEXT, Japan.

Refferences

[1] Tyers, M.; Mann, M. *Nature* **2003**, 422, 193−197.

[2] Herbert, B. R.; Harry, J. L.; Packer, N. H.; Gooley, A. A.; Pedersen, S. K.; Williams, K. L. *Trends Biotechnol* **2001**, 19, S 3−9.

[3] Pandey, A.; Mann, M. *Nature* **2000**, 405, 837−846.

[4] Taoka, M.; Wakamiya, A.; Nakayama, H.; Isobe, T. *Electrophoresis* **2000**, 21, 1872−1879.

[5] Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. *Nature* **1996**, 379 , 466−469.

[6] Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. *Proc Natl Acad Sci U S A* **2000**, 97, 9390−9395.

[7] McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R., 3 rd *Anal Chem* **1997**, 69, 767−776.

[8] Ducret, A.; Van Oostveen, I.; Eng, J. K.; Yates, J. R., 3 rd; Aebersold, R. *Protein Sci* **1998**, 7, 706−719.

[9] Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, 20, 3551−3567.

[10] Takahashi, N.; Ishioka, N.; Takahashi, Y.; Putnam, F. W. J *Chromatogr* **1985**, 326, 407−418.

[11] Takahashi, N.; Isobe, T.; Putnam, F. W. In *HPLC of Proteins, Peptides and Polynucleotides* ; Hearn, M. T. W., Ed.; VCH Publishers: NY, 1991; pp 307−330.

[12] Isobe, T.; Uchida, K.; Taoka, M.; Shinkai, F.; Manabe, T.; Okuyama, T. *J Chromatogr* **1991**, 588, 115−123.

[13] Takahashi, N.; Ortel, T. L.; Putnam, F. W. *Proc Natl Acad Sci U S A* **1984**, 81, 390−394.

[14] Takahashi, N.; Takahashi, Y.; Isobe, T.; Putnam, F. W.; Fujita, M.; Satoh, C.; Neel, J. V. *Proc Natl Acad Sci U S A* **1987**, 84, 8001−8005.

[15] Takahashi, N.; Takahashi, Y.; Blumberg, B. S.; Putnam, F. W. *Proc Natl Acad Sci U S A* **1987**, 84, 4413−4417.

[16] Isobe, T.; Takahashi, N.; Putnam, F. W. In *HPLC of Peptides and Proteins; Separation*; Hodges, R. S. et. al., Ed.;

CRC Press: NY, 1991; pp 835−845.

[17] Taoka, M.; Yamakuni, T.; Song, S. Y.; Yamakawa, Y.; Seta, K.; Okuyama, T.; Isobe, T. *Eur J Biochem* **1992**, 207, 615−620.

[18] Taoka, M.; Isobe, T.; Okuyama, T.; Watanabe, M.; Kondo, H.; Yamakawa, Y.; Ozawa, F.; Hishinuma, F.; Kubota, M.; Minegishi, A.; et al. *J Biol Chem* **1994**, 269, 9946−9951.

[19] McDonald, W. H.; Yates, J. R., 3 rd *Curr Opin Mol Ther* **2003**, 5, 302−309.

[20] Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3 rd *Nat Biotechnol* **1999**, 17, 676−682.

[21] Washburn, M. P.; Wolters, D.; Yates, J. R., 3 rd *Nat Biotechnol* **2001**, 19, 242−247.

[22] Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3 rd *Anal Chem* **2001**, 73, 5683−5690.

[23] Isobe, T.; Yamauchi, Y.; Taoka, M.; Takahashi, N. In *Proteins and Proteomics* ; Simpson, R. J., Ed.; Cold Spring Harbor Press: NY, 2003; pp 869−876.

[24] Taoka, M.; Yamauchi, Y.; Shinkawa, T.; Kaji, H.; Motohashi, W.; Nakayama, H.; Takahashi, N.; Isobe, T. *Mol Cell Proteomics* **2004**, 3 , 780−787.

[25] Taoka, M.; Ichimura, T.; Wakamiya−Tsuruta, A.; Kubota, Y.; Araki, T.; Obinata, T.; Isobe, T. *J Biol Chem* **2003**, 278, 5864−5870.

[26] Matsuoka, K.; Taoka, M.; Okuyama, T.; Isobe, T.; Kato, Y. *J Chromatogr* **1990**, 515, 313−320.

[27] Nagano, K.; Taoka, M.; Yamauchi, Y.; Itagaki, C.; Shinkawa, T.; Nunomura, K.; Okamura, N.; Takahashi, N.; Izumi, T.; Isobe, T. *Proteomics* **2005**, 5 , 1346−1361.

[28] Yoshimura, Y.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Donai, H.; Takahashi, N.; Isobe, T.; Yamauchi, T. *J Neurochem* **2004**, 88, 759−768.

[29] Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T. *Nat Biotechnol* **2003**, 21, 667−672.

[30] Mawuenyega, K. G.; Kaji, H.; Yamuchi, Y.; Shinkawa, T.; Saito, H.; Taoka, M.; Takahashi, N.; Isobe, T. *J Proteome Res* **2003**, 2, 23−35.

[31] Blattner, F. R.; Plunkett, G., 3 rd; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado−Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y. *Science* **1997**, 277, 1453−1474.

[32] Tonella, L.; Hoogland, C.; Binz, P. A.; Appel, R. D.; Hochstrasser, D. F.; Sanchez, J. C. *Proteomics* **2001**, 1, 409−423.

[33] Bremer, H.; Dennis, P. P. In *Escherichia coli and Salmonella, Cellular and Molecular Biology, 2 nd edition.*; Neidhardt, F., Curtiss, I. R., Ingraham, J., Lin, E., Low, J. K. B., Magasanik, B., Reznikoff, W., Riley, M., Schaechter, M., Umbarger, H., Eds.; American Society for Microbiology: Washington, D.C., 1996; Vol. 2; pp 1533−1569.

[34] Meeser, W.; Weigel, C. In *Escherichia coli and Salmonella, Cellular and Molecular Biology, 2 nd edition.*; Neidhardt, F., Curtiss, I. R., Ingraham, J., Lin, E., Low, J. K. B., Magasanik, B., Reznikoff, W., Riley, M., Schaechter, M., Umbarger, H., Eds.; American Society for Microbiology: Washington, D.C., 1996; Vol. 2; pp 1579−1601.

[35] Tatusov, R. L.; Natale, D. A.; Garkavtsev, I. V.; Tatusova, T. A.; Shankavaram, U. T.; Rao, B. S.; Kiryutin, B.; Galperin, M. Y.; Fedorova, N. D.; Koonin, E. V. *Nucleic Acids Res* **2001**, 29, 22−28.

[36] Yamauchi, Y. *et. al.* Unpublished results