CANCER GENOMICS & PROTEOMICS 5: 105-116 (2008)

High-accuracy Prediction of Carcinogenicity by Global Quantitative Analysis of Post-translational Modifications in a 28-Day *In Vivo* Rat Study

HIDENORI YAMANAKA[1], YOSHIKUNI YAKABE[1], KOICHI SAITO[2], MASARU SEKIJIMA[3], HISAYA SATO[4] and TOMOYUKI SHIRAI[5]

[1]*Chemicals Assessment Center, Chemicals Evaluation and Research Institute, Saitama;*
[2]*Environmental Health Science Laboratory, Sumitomo Chemical Co. Ltd, Osaka;*
[3]*Kashima Laboratory, Mitsubishi Chemical Safety Institute Ltd., Ibaraki;*
[4]*Tokyo University of Agriculture and Technology, Tokyo;*
[5]*Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan*

Abstract. *A global quantitative analysis of post-translational modifications (PTMs) of distinct proteins was executed at the proteomic level using two-dimensional fluorescence differential gel electrophoresis. We evaluated the effects of 66 chemical compounds, including 15 genotoxic carcinogens, 28 non-genotoxic carcinogens, and 23 non-carcinogens, in the male F344 rat liver in a 28-day repeated dose study. In the master gel of rat liver protein, we identified 728 spots by hybrid quadrupole time-of-flight mass spectrometry. They collapsed into 356 distinct proteins. Of these, 126 were represented by two or more spots in the 2-D gel. We calculated the logarithmic ratio of volume changes of all 1028 combinations generated from 126 proteins and investigated the relevance to carcinogenicity. This quantitative proteomic study revealed the existence of several PTMs characteristic of carcinogens that may play an important role in early stage of carcinogenicity. Prediction of carcinogenicity from PTM data gave a higher concordance (92.4%) than prediction from protein expression data (74.2%). This novel approach holds great promise as a way of revealing the roles of charge modifications and molecular weight variations of proteins in biological processes.*

Although numerous biological activities are modulated by modifications and variations in proteins, such as post-translational modifications (PTMs) and alternative splicing of mRNA, these cannot be detected by observing the protein or the changes in mRNA expression. Quantification of these modifications or variations in proteins at the proteomic level is a difficult challenge that is currently being addressed by many researchers in the field of proteomics (1-5).

The main cause of charge modifications in proteins is PTMs such as glycosylation and phosphorylation. The molecular weight variations detected by two-dimensional electrophoresis (2-DE) originate mainly from splicing variants, truncation, and some proteolytic products. It is for these reasons that many spots may be detected for the same gene product in 2-DE gels (6, 7).

Two-dimensional fluorescence differential gel electrophoresis (2D-DIGE) is based on a high-resolution separation technique that is able to exploit the charge modifications and molecular weight variations of proteins (8-19). Recent advances in the mass spectrometric analysis of gel-eluted intact proteins have made it possible to characterize the protein variants and PTMs detected by 2-DE (20, 21). However, the exhaustive characterization of these variations and modifications remains a challenge. We focused on the quantitative analysis of these variants and modifications at the proteomic level.

We surmised that it may be possible to use the log ratio of volume changes between the spots generated from the same protein as a quantitative indicator of PTMs. (*Hereinafter, for simplicity PTM is used to describe both molecular weight variations and PTMs. PTM data indicates log ratios of volume changes between the spots generated from the same protein.*)

In this study, we calculated the PTM data for all 1028 combinations generated from 126 distinct proteins to evaluate the effects of 66 chemical compounds, including genotoxic

Abbreviations: PTMs, post-translational modifications; Q-TOF, quadrupole time-of-flight; DIA, difference in-gel analysis; BVA, biological variation analysis.

Correspondence to: Hidenori Yamanaka, 1600, Shimo-Takano, Sugito-machi, Kitakatsushika-gun, Saitama 345-0043, Japan. Fax, +81 480 37 2861, e-mail: yamanaka-hidenori@ceri.jp

Key Words: 2D-DIGE, post-translational modifications, quantitative analysis, rat liver proteins.

Table I. *List of tested chemical compounds and doses*.

Compounds	CAS No.	Carcinogenicity*	Ames test*	Dose (mg/kg/day)
Clofibrate	637-07-0	+	–	250
Di(2-ethylhexyl)phthalate	117-81-7	+	–	300
Carbon tetrachloride	56-23-5	+	–	50
2,4-Diaminotoluene	95-80-7	+	+	10
Quinoline	91-22-5	+	+	25
Phenobarbital	50-06-6	+	–	100
Diethylnitrosamine (DEN)	55-18-5	+	+	20
2-Nitropropane	79-46-9	+	–	40
N-Nitrosomorpholine	59-89-2	+	+	10
Aldrin	309-00-2	+	–	0.3
Di(2-ethylhexyl)adipate (DEHA)	103-23-1	+	–	1000
Ethinylestradiol	57-63-6	+	–	0.5
Hexachlorobenzene (HCB)	118-74-1	+	–	5
α-Hexachlorocyclohexane (HCH)	319-84-6	+	–	20
Trichloroethylene	79-01-6	+	–	700
Butylated hydroxyanisole (BHA)	25013-16-5	+	–	750
Safrole	94-59-7	+	–	300
1,4-Dichlorobenzene (DCB)	106-46-7	+	–	300
1,4-Dioxane	123-91-1	+	–	1000
Furan	110-00-9	+	–	10
Methyl carbamate	598-55-0	+	–	500
Thioacetamide	62-55-5	+	–	20
N-Nitrosodimethylamine (DMN)	62-75-9	+	+	0.2
MeIQx	77500-04-0	+	+	20
PhIP	105650-23-5	+	+	5
Benz[a]anthracene (BA)	56-55-3	+	+	50
7,12-Dimethylbenz[a]anthracene	57-97-6	+	+	1
3-Methylcholanthrene (MC)	56-49-5	+	+	2
4-Nitroquinoline-1-oxide	56-57-5	+	+	2
N-Ethyl-*N*-nitrosourea (ENU)	759-73-9	+	+	3
Trichloroacetic acid	76-03-9	+	–	300
Urethane	51-79-6	+	–	80
Pentachloroethane	76-01-7	+	–	200
Chloroform	67-66-3	+	–	90
Benzo[a]pyrene (BP)	50-32-8	+	+	15
MNNG	70-25-7	+	+	0.5
Tetrachloroethylene	127-18-4	+	–	100
Acetamide	60-35-5	+	–	180
Diethylstilbestrol (DES)	56-53-1	+	–	10
Phenytoin (5,5-Diphenylhydantoin)	57-41-0	+	–	160
dl-Ethionine	67-21-0	+	–	30
4-Dimethylaminoazobenzene (DAB)	60-11-7	+	+	50
Chlorendic acid	115-28-6	+	–	100
2,6-Diaminotoluene	823-40-5	–	+	10
8-Hydroxyquinoline	148-24-3	–	+	25
d-Mannitol	69-65-8	–	–	40
l-Ascorbic acid	50-81-7	–	–	150
2-Chloroetahnol	107-07-3	–	+	1000
2-Chloromethylpyridine HCl	6959-47-3	–	+	250
dl-Menthol	89-78-1	–	–	80
4-Nitro-*o*-phenylenediamine	99-56-9	–	+	500
Benzoin	119-53-9	–	–	1000
Iodoform	75-47-8	–	+	200
Lithocholic acid	434-13-9	–	–	1000
2-Chloro-*p*-phenylenediamine SO_4	61702-44-1	–	+	100
p-Phenylenediamine 2HCl	624-18-0	–	+	60
2,5-Toluenediamine SO_4	6369-59-1	–	+	50
Aspirin	50-78-2	–	–	27

Table I. *continued*

Table I. *continued*

Compounds	CAS No.	Carcinogenicity*	Ames test*	Dose (mg/kg/day)
4-(Chloroacetyl)acetanilide	140-49-8	–	+	250
Phthalamide	88-96-0	–	–	1000
Caprolactam	105-60-2	–	–	375
1-Chloro-2-propanol (technical)	127-00-4	–	+	100
3-Chloro-p-toluidine	95-74-9	–	–	300
Glutaraldehyde	111-30-8	–	+	50
4-Nitroanthranilic acid	619-17-0	–	+	1000
1-Nitronaphthalene	86-57-7	–	+	100

*Carcinogenicity and Ames test; "+" indicates positive, "–" indicates negative.

and non-genotoxic carcinogens, and non-carcinogens, in the male F344 rat liver in a 28-day repeated dose study (22, 23). Statistical analysis was performed to assess whether a relationship exists between carcinogenicity and PTM data. Statistically significant modifications of the distinct proteins that were found were used to predict carcinogenicity.

We have previously investigated and reported the effects of chemical compounds on protein expression in the rat liver after 28 daily doses. Using support vector machines (SVMs), we were able to predict carcinogenicity from these quantitative proteomic data (24). The results of carcinogenicity prediction using this novel PTM data analysis method were compared with the results of conventional analysis of protein expression data.

Materials and Methods

Materials. PBS, DTT, SDS, TEMED, glycine, glycerol (87% w/w), urea, Tris–HCl buffer, acrylamide, IPG strip (pH 3-10 linear), Pharmalytes (pH 3-10), Drystrip cover fluid, bromophenol blue, agarose, Cy2, Cy3, and Cy5 were purchased from Amersham Biosciences (Little Chalfont, Bucks, UK), CHAPS was from Dojin Chemical Japan (Osaka, Japan), and ammonium bicarbonate, formic acid, methanol, acetonitrile, n-butanol, acetic acid, and thiourea were obtained from Wako Pure Chemicals Japan (Osaka, Japan). Iodoacetamide was from Sigma (St. Louis, MO, USA). Protease inhibitor cocktail (Pefabloc SC PLUS; Pefabloc SC and Pefabloc SC protector) was from Roche (Mannheim, Germany). Sequence grade trypsin was obtained from Promega UK (Southampton, UK). Deionized water was prepared with the Milli-Q system (Millipore, Bedford, MA, USA).

Animals. The animals used were described in our previous papers (22-24). A brief summary of the procedure is given here. Five-week-old male Fischer 344 (F-344) rats were obtained from Charles River Laboratories Japan, Inc. (Atsugi, Japan). Two groups of five rats each were administered compounds dissolved in vehicle (water or corn oil) or vehicle only as a control by oral gavage once a day for 28 days. All rats were killed by CO_2 overdose (80% CO_2 and 20% O_2, 10 L/min) 24 h after the last dose. The guidelines of the Japanese Association for Laboratory Animal Science were adhered to in all experiments. Liver samples were collected from the left lateral lobe and kept frozen at –80˚C until analysis.

The methods used for liver homogenization, protein labeling, gel electrophoresis, imaging, image analysis, in-gel digestion, peptide extraction, and mass spectrometric analysis are described in our previous paper (24). A brief summary of these procedures is given in sections 2.3-2.8.

Chemical compounds tested and carcinogenicity databases. The chemical compounds we tested and the dose of each compound are shown in Table I. The doses of carcinogenic compounds used were those known to cause cancer in lifetime carcinogenicity tests. The doses of non-carcinogens were based on the maximum tolerated dose (MTD). We use the term genotoxic compound to refer to compounds giving a positive result in the Ames test.

Liver homogenization and protein labeling. The samples were homogenized in ten volumes of the lysis buffer CHAPS (4% w/v), 2 M thiourea, 8 M urea, 10 mM Tris–HCl, pH 8.8). Centrifugation was then performed at 14,000 rpm for 20 min at 10˚C. Protein concentration was determined using the Bradford method. Lysates were labeled with NHS ester derivatives of the cyanine dyes Cy2, Cy3, and Cy5 (GE Healthcare Bio-Sciences) following the manufacturer's protocol.

Gel electrophoresis and imaging. Immobilized pH gradient (IPG) strips pH 3-10L, 24 cm (GE Healthcare Bio-Sciences, Little Chalfont, UK) were rehydrated and mixed samples were applied by cup loading. Isoelectric focusing was performed using a Multiphor II (GE Healthcare Bio-Sciences) for a total of 54 kVh at 20˚C (25, 26). Equilibrated IPG strips were transferred onto 24 cm × 20 cm, 12% uniform polyacrylamide gels. Gels were run in Ettan DALT twelve (GE Healthcare Bio-Sciences) with 2 W per gel at 20˚C. 2-D gels were scanned directly between glass plates using a 2920 2D-Master Imager (GE Healthcare Bio-Sciences). The images generated were exported as TIF files for further protein profile analysis using DeCyder™ software (GE Healthcare Bio-Sciences).

Image analysis. The differential in-gel analysis (DIA) software of DeCyder™ was used to merge the Cy2, Cy3, and Cy5 images for each gel and to detect spot boundaries. Features resulting from non-protein sources (*e.g.*, dust particles, streaks) and faint spots (*e.g.*, spot area <300, spot volume <10,000) were filtered out. The biological variation analysis (BVA) software of DeCyder™ was then used to match all pairwise image comparisons from DIA for a comparative cross-gel statistical analysis. Comparison of normalized

Cy3 and Cy5 spot volumes with the corresponding Cy2 standard spot volumes within each gel gave a standardized abundance. This value was compared across all gels for each matched spot, and statistical analysis (ANOVA) was performed using triplicate values from each experimental condition.

In-gel digestion and peptide extraction. Gel electrophoresis for mass spectrometric (MS) analysis was performed using 600 μg of pooled liver lysate with the same procedures described in the above section. After gel electrophoresis, the gel was stained with Sypro®Ruby. The gel for MS analysis was matched to the master gel for expression analysis by BVA software. Spots of interest were excised from 2-D gels using an automated spot picker (GE Healthcare Bio-Sciences) following the manufacturer's instructions. The recovered gel pieces were washed with aqueous 50 mM ammonium bicarbonate and acetonitrile, and then incubated with 12.5 ng/μL trypsin (Promega UK) at 30˚C for 15 h. The peptides generated were eluted with 50 mM ammonium bicarbonate followed by 10% v/v formic acid and acetonitrile. The combined fractions were dried and dissolved in 0.1% v/v formic acid.

Mass spectrometric analysis. Mass spectrometric analysis was carried out using LC-MS/MS (27, 28). HPLC (CapLC, Waters, Milford, MA, USA) was performed with a column of internal diameter 75 μm and length 150 mm (L-column micro; CERI, Tokyo, Japan). The eluted peptide was analyzed by Q-TOFmicro MS (Micromass, Manchester, UK). The database search was performed with MASCOT Deamon (Matrix Science, London, UK) (29-31). The pkl files generated were submitted to SWISS-PROT (release 47.4) and NCBInr (14-Jul-2005). Search parameters were as follows: fixed modifications, carbamidomethyl; variable modifications, oxidation (M); missed cleavages, up to 1, monoisotopic, peptide tolerance 1.0 Da, MS/MS tolerance 0.5 Da. The ion score cutoff was set to 20. The automatically identified proteins were checked individually to remove redundancy.

Statistic analysis of protein variants and modifications. Two groups of three rats (a vehicle control group and a group dosed with the test chemical) per compound were used for the protein expression experiments. Comparison of the normalized Cy3-labeled vehicle control sample and the Cy5-labeled sample from the dosed group with the corresponding Cy2-labeled pool sample within each gel gave a standardized abundance. The average ratio of (dosed group sample)/(vehicle control group sample) was calculated from this standardized abundance using the BVA software. The average ratio given by the BVA software was converted to a log ratio. Then the PTM data (differences for all 1028 combinations between the log ratio of the spots *i.e.*, the log ratio of volume changes generated from 126 proteins) were calculated for each of the 64 chemical compounds. The equation used to calculate PTM data was as follows: PTM data=$\log_2$(spot volume$A1_H$)/(spot volume$A1_C$) − $\log_2$ (spot volume$A2_H$)/(spot volume$A2_C$)=$\log_2$[(spot volume$A1_H$)/(spot volume$A1_C$)]/[(spot volume$A1_H$)/(spot volume$A1_C$)]. Spot A1 and spot A2 originated from the same protein. $A1_C$ is the vehicle control sample spot volume for A1, A1H is the sample volume of the chemical-dosed group, A2C is the vehicle control sample spot volume of A2, and A2H is the sample volume of the chemical-dosed group.

Statistical significance was analyzed by Welch's *t*-test. Welch's *t* value was calculated with the following equation:

$$\text{Welch } t = \frac{\text{Sample Mean}_1 - \text{Sample Mean}_2}{\sqrt{\frac{\text{Variance}_1}{\text{Sample Size}_1} + \frac{\text{Variance}_2}{\text{Sample Size}_2}}}$$

Carcinogenicity prediction. We predicted carcinogenicity from the PTM data as follows: (1) Welch's *t* value between carcinogen and non-carcinogen was calculated and the PTM data were ordered according to Welch's *t* value. (2) All PTM data were transformed to an integer score based on the cutoff value X, $-X < \text{data} < X$ to 0, $X \leq \text{data}$ to 1, and $\text{data} \leq -X$ to −1. (3) Scores for all carcinogens and non-carcinogens were summed. If the total score for non-carcinogens was larger than that for carcinogens, all scores were multiplied by −1. (4) The arbitrary number of score for each compound was summed together and calculated. If the summation of the score was above 0, the compound was considered carcinogenic. If the summation of the score was less than 0, it was considered non-carcinogenic.

Performance in the prediction of carcinogenicity was assessed by concordance (total correct predictions/total number of predictions), sensitivity (percentage of correct predictions of carcinogens), and specificity (percentage of correct predictions of non-carcinogens).

Results and Discussion

Exhaustive identification of multiple spots generated from the same gene products in 2-DE gel. To confirm which spots originate from the same gene, we have developed an in-house 2-D database of rat liver proteins by LC-MS/MS analysis using nanoESI on a hybrid Q-TOF mass spectrometer. As described in the Materials and methods section, we used a relatively small amount of liver lysate sample (600 μg of pooled liver lysate) to match precisely the Sypro®Ruby-stained gel for MS analysis and the Cy-labeled master gel of rat liver (Figure 1). Approximately 1000 spots were successfully matched with the master gel. These matched spots were picked from the master gel of rat liver (pI range 3-10) and subjected to LC-MS/MS analysis. A database search resulted in the identification of 728 spots. In all identified proteins, mitochondrial stress-70 protein precursor (Swiss-Prot No. 48721) gave the highest MOWSE score: 1125, from 21 matched peptides. The average score of all identified proteins was 135.6 and the average number of matched peptides was 3.0.

After removing redundancy, 356 different gene products were found in the master gel. One hundred and twenty-six distinct proteins had two or more spots. Fifty-two spots were detected 2 times, 65 spots 3-9 times, and 8 abundant proteins 10 or more times (Figure 2). Protein isoforms originate from many sources, such as PTMs, alternative splicing of mRNA, a truncated/extended reading frame, truncation of proteins, single-point mutations, and proteolytic cleavages. Of these modifications, alternative splicing of mRNA, truncated/ extended reading frame, truncation of protein and proteolytic cleavage cause relatively large variations in molecular weight. In our 2-D

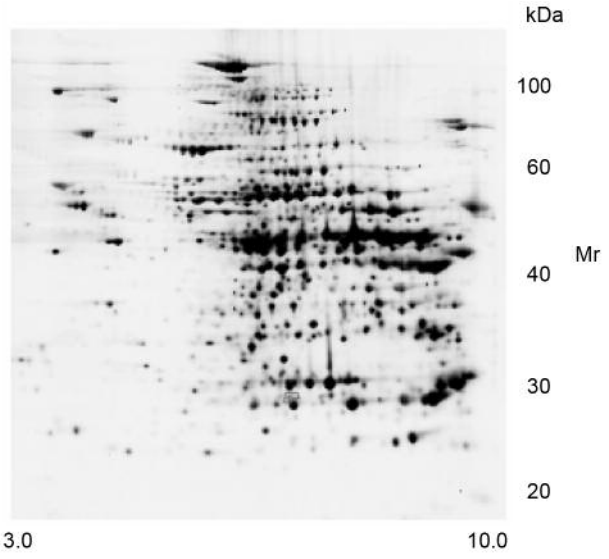

Figure 1. *pI 3-10 master gel (24 cm x 20 cm, 12% T, 7.5% C polyacrylamide gel) image of the rat liver protein.*

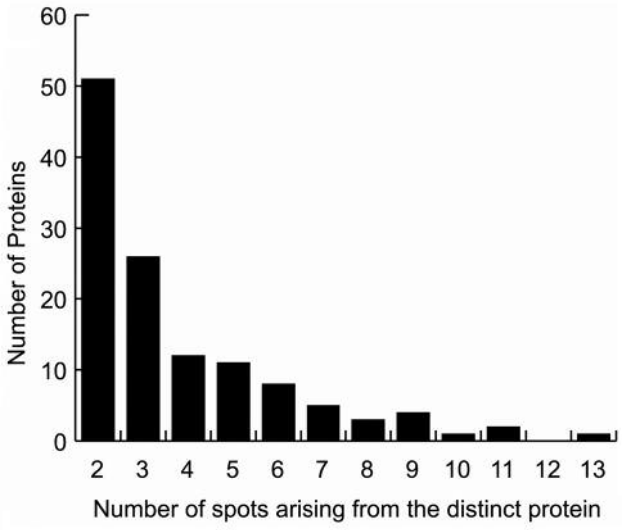

Figure 2. *Distribution of the numbers of proteins that were identified in multiple spots.*

database, 7 proteins showed molecular weight variations. One hundred nineteen proteins found as multiple spots on the 2-D gel showed variation of net charge that originated mainly from PTMs.

Model analysis for quantification of protein variations and modifications. Our aim is to quantify changes in protein expression and PTMs of distinct proteins separately to improve our ability to distinguish carcinogenic from non-carcinogenic compounds. The spot volume changes evaluated by 2D-DIGE are assumed to be the sum of changes in protein expression and PTM changes. We surmised that the contribution of PTMs to changes in abundance could be calculated as the log ratio of volume changes (the differences in logarithmic volume changes) between the different spots generated from the same proteins. (Hereinafter, the term PTM data is used to describe these differences in volume changes between the spots generated from the same proteins.) The ideal model to illustrate this idea is shown in Figure 3A and B. In Figure 3A, spot A1 and spot A2, which originated from the same protein, increased three times as much as the control. In this case, the log2 ratio of the fold change relative to the control in the 2D-DIGE data was the same (1.58) for spots A1 and A2. The difference in volume change between the spots was thus zero. If specific modification occurred after chemical treatment, we were able to detect this modification as the imbalance in the fold change between the spots (Figure 3B). In other words, PTM data indicate the change in modified proteins as a proportion of the total expressed proteins.

The effect of repeated dosing for 28 days on protein expression and PTMs. The average ratio given by BVA software was converted to a log ratio. Then the differences for all 1028 combinations generated from 126 proteins were calculated for data on each of 82 chemical compounds. Figure 4 shows a comparison of the distributions of protein expression data and PTM data (the differences for all 1028 combinations generated from 126 proteins) for *N*-nitrosomorpholine, chloroform, and 1-nitronaphthalene as examples of a genotoxic carcinogen, a non-genotoxic carcinogen, and a non-carcinogen, respectively. Figure 5 shows a histogram for all 66 chemical compounds. If the dose of a chemical compound has only a small effect on the PTM change, the histogram of the PTM data would be narrower than that of the protein data. No significant difference was observed between the distributions of protein expression data and PTM data. This supports the idea that the PTM data are worth analyzing statistically to investigate the relevance to carcinogenicity.

PTMs characteristic of carcinogens in a 28-day repeated dose study. Statistical analysis was carried out to determine whether a particular difference between carcinogens and non-carcinogens in protein variants and modifications of the distinct proteins. Table II shows the 10 proteins ranking highest for Welch's *t* value calculated from the PTM data (when proteins appear more than once in the ranking, only the top-ranked protein is indicated in Tables II-V). Although significant differences were not observed between PTM data and protein expression data in the distribution of the abundance ratio data themselves, 8 out of 10 top-ranked proteins were different from

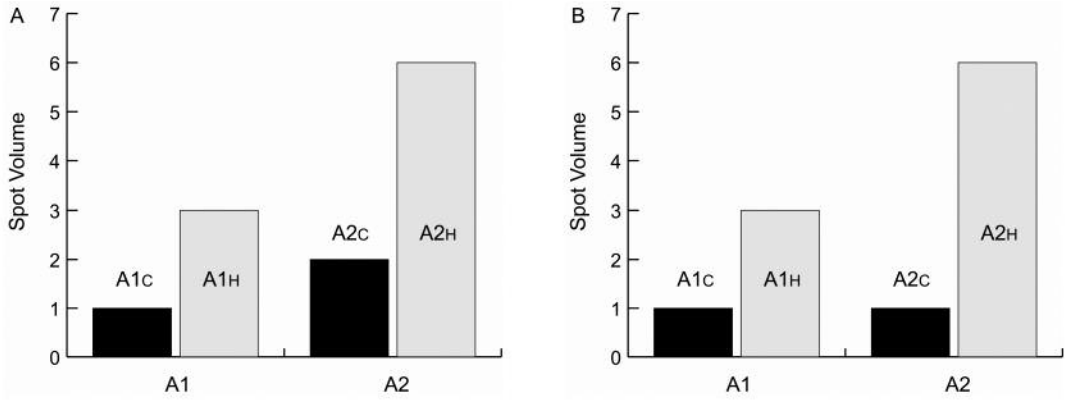

Figure 3. *Schematic graph of the ideal model for the PTM changes in distinct proteins. Spots A1 and A2 originated from the same protein. A1 is sample spot volume for the vehicle control for A1, and $A1_H$ is sample volume for the chemical-dosed group. $A2_C$ is vehicle control sample spot volume of A2 and $A2_H$ is sample volume for the chemical-dosed group. A is an example of no change observed in PTMs, and B is an example of significant change occurring in protein PTMs.*

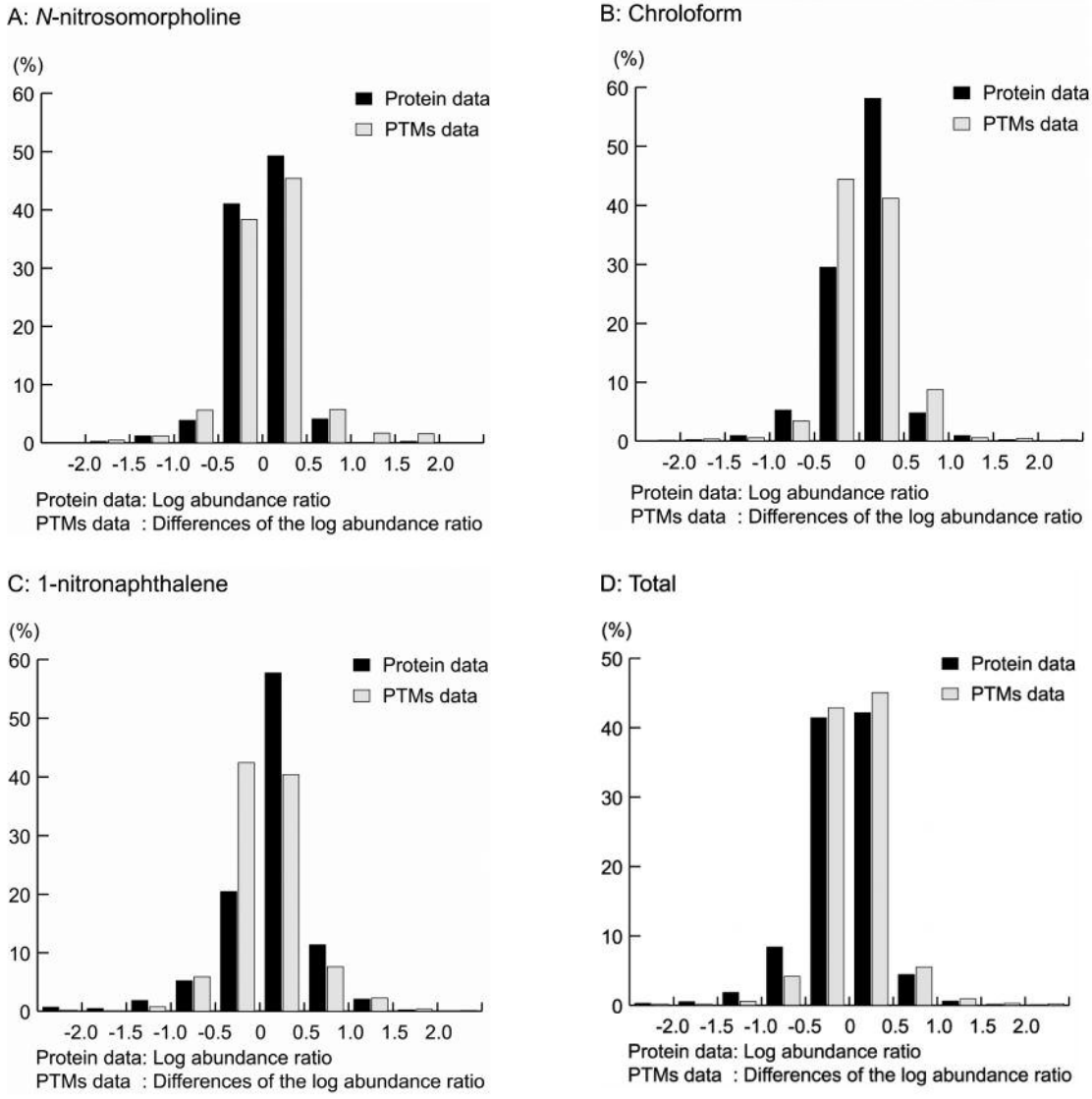

Figure 4. *Comparison of the distribution of log abundance ratio of protein expression data and differences in log abundance ratio of PTM data.*

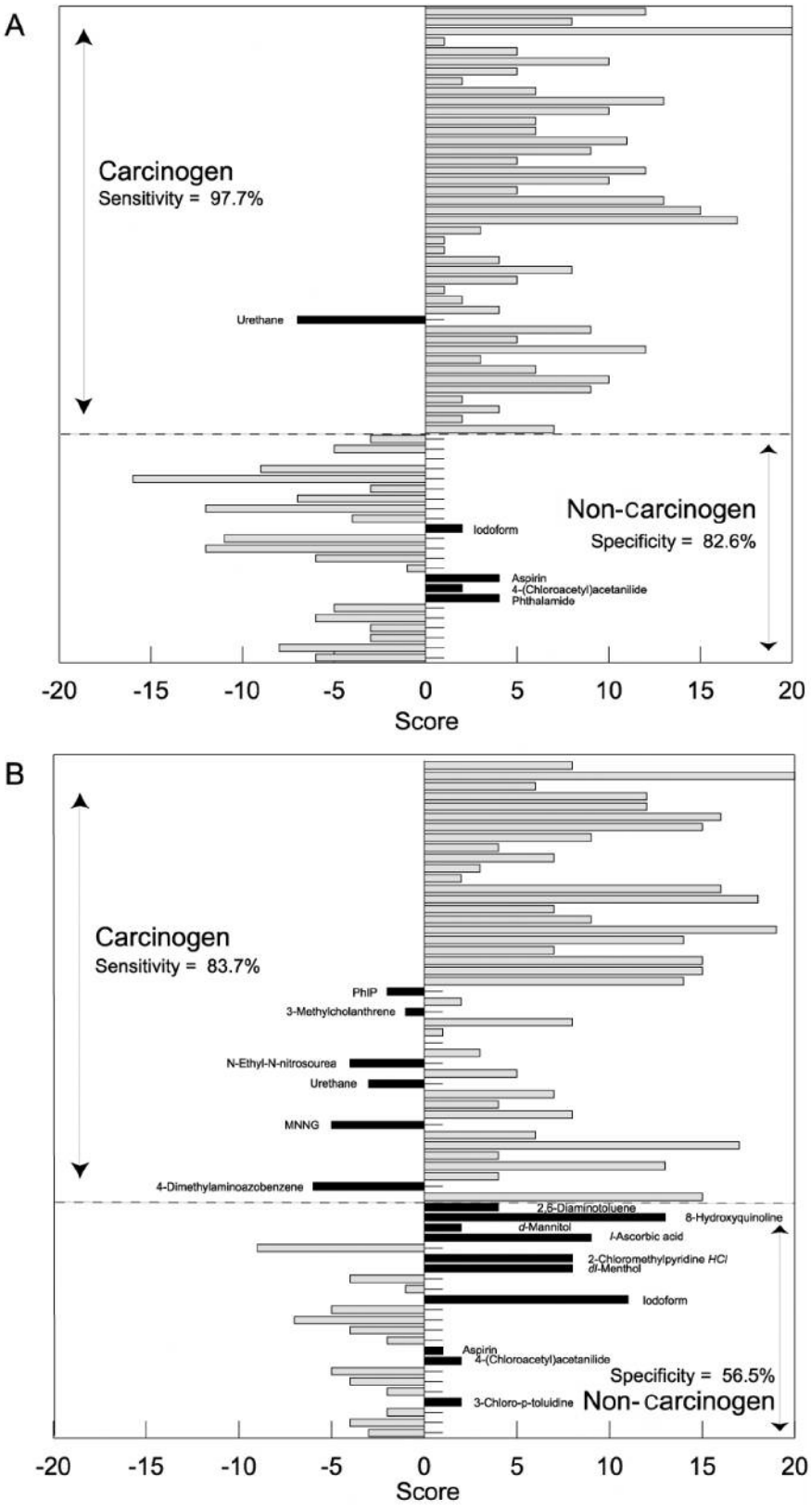

Figure 5. *Results of carcinogenicity prediction using the score method. (A) Carcinogenicity prediction using PTM data. (B) Carcinogenicity prediction using protein data. Black bar indicates the misclassified compound; its name is shown.*

Table II. *List of top ranked 10 proteins in Welch's t-value between carcinogens and non-carcinogens calculated from PTMs data.*

Master Spot No.	Protein name	SWISS-PROT or NCBInr No.	Welch's *t*-value
[1118]-[1135]	Glutamate dehydrogenase, mitochondrial [Precursor]	P10860	3.770
[1504]-[1539]	Fructose-bisphosphate aldolase B	P38918	3.551
[1085]-[1176]	ATP synthase beta subunit	P00884	3.386
[913]-[916]	*Choline dehydrogenase precursor	P10719	3.221
[1096]-[1102]	*ATP synthase alpha chain, mitochondrial precursor	gi/34419913	3.160
[946]-[983]	60 kDa heat shock protein, mitochondrial [Precursor]	P15999	3.107
[1594]-[1602]	Aflatoxin B1 aldehyde reductase member 1	P19227	3.063
[733]-[736]	Heat shock cognate 71 kDa protein	P08109	2.933
[2177]-[2228]	Glutathione S-transferase Mu 2	P08010	2.929
[864]-[872]	Contrapsin-like protease inhibitor 1 [Precursor]	P05544	2.701

*Indicates the protein which also ranked in top 10 in Table III.

Table III. *List of top ranked 10 proteins in Welch's t-value between carcinogens and non-carcinogens calculated from protein expression data.*

Master Spot No.	Protein name	SWISS-PROT or NCBInr No.	Welch's *t*-value
1543	Sorbitol dehydrogenase (L-iditol 2-dehydrogenase)	P27867	3.688
2169	Marapsin [*Rattus norvegicus*]	gi/33438175	3.592
2113	Similar to RIKEN cDNA 0610009116 [*Rattus norvegicus*]	gi/27731305	3.547
1096	ATP synthase alpha chain, mitochondrial precursor	P15999	3.463
2143	Chain D, 2-Enoyl-Coa Hydratase	P14604	3.322
1563	Similar to RIKEN cDNA 2410174K12 [*Rattus norvegicus*]	gi/34874534	3.227
1577	Arginase 1 (Liver-type arginase)	P07824	2.973
991	Nucleobindin 1 [*Rattus norvegicus*]	gi/16758210	2.966
939	Choline dehydrogenase [*Rattus rattus*]	gi/1154950	2.881
302	Pyruvate carboxylase, mitochondrial precursor	P52873	2.852

Table IV. *List of top ranked 10 proteins in Welch's t-value between genotoxic carcinogen and genotoxic non-carcinogen calculated from PTMs data.*

Master Spot No.	Protein name	SWISS-PROT or NCBInr No.	Welch's *t*-value
[784]-[826]	Acyl-CoA dehydrogenase, very-long-chain specific	P45953	3.309
[1550]-[1572]	Fructose-bisphosphate aldolase B	P00884	3.153
[1068]-[1313]	Keratin, type II cytoskeletal 8	Q10758	3.109
[1633]-[1644]	3-oxo-5-beta-steroid 4-dehydrogenase	P31210	2.932
[2394]-[2406]	Peroxiredoxin 1	Q63716	2.712
[1594]-[1602]	Aflatoxin B1 aldehyde reductase member 1	P38918	2.709
[1653]-[1663]	Glyceraldehyde-3-phosphate dehydrogenase	P04797	2.563
[1034]-[1035]	Dihydropyrimidinase	Q63150	2.492
[1361]-[1391]	Betaine–homocysteine S-methyltransferase	O09171	2.472
[1125]-[1129]	Aldehyde dehydrogenase, mitochondrial [Precursor]	P11884	2.417

the proteins calculated from protein expression data (Table III). This result suggests the existence of carcinogen-specific variants and PTMs. Choline dehydrogenase showed a significant change both in protein expression and in PTM data. Choline dehydrogenase has catalytic activity; that is, betaine biosynthesis from choline. Betaine, a metabolite of choline, is an important source of labile methyl groups. Interestingly, choline-deficient rats accumulate fat within the liver and they

Table V. *List of top ranked 10 proteins in Welch's t-value between non-genotoxic carcinogen and non-genotoxic non-carcinogen calculated from PTMs data.*

Master Spot No.	Protein name	SWISS-PROT or NCBInr No.	Welch's *t*-value
[913]-[916]	Choline dehydrogenase precursor	gi/34419913	3.668
[2177]-[2226]	Glutathione S-transferase Mu 2	P08010	3.553
[1504]-[1570]	Fructose-bisphosphate aldolase B	P00884	3.488
[1364]-[1373]	Similar to Succinyl-CoA ligase [GDP-forming] beta chain	gi/34857707	3.397
[1118]-[1135]	Glutamate dehydrogenase, mitochondrial [Precursor]	P10860	3.279
[946]-[938]	60 kDa heat shock protein, mitochondrial [Precursor]	P19227	3.150
[1129]-[1146]	Aldehyde dehydrogenase, mitochondrial [Precursor]	P11884	3.012
[812]-[830]	Transketolase	P50137	2.974
[1096]-[1102]	ATP synthase alpha chain, mitochondrial precursor	P15999	2.888
[1752]-[1785]	Similar to RIKEN cDNA 0610010D20	gi/34865395	2.742

develop hepatocarcinomas in the absence of any known carcinogen (32-34). Hypomethylation of DNA, observed during choline deficiency, is thought to be responsible for carcinogenesis (35).

From another point of view, we can ask whether the comparative statistics reveal a particular difference between a genotoxic carcinogen and a genotoxic noncarcinogen or between a nongenotoxic carcinogen and a nongenotoxic noncarcinogen in. In the analysis of nongenotoxic compounds, choline dehydrogenase showed the most significant difference (Table V). Recently, cancer-associated cleavage of cytokeratin 8/18 heterotypic complexes was reported in human adenocarcinomas (36). In our analysis of genotoxic carcinogens, molecular weight variants of keratin showed a statistically significant difference between genotoxic carcinogens and noncarcinogens (Table IV).

Further investigation should be done to improve our understanding of the biological meaning of these findings. In addition, further characterization of several PTM proteins has been undertaken

Carcinogen prediction from PTM data in a 28-day repeated dose study. Carcinogenicity was predicted by the score method using PTM and protein data. After reordering the PTM data using Welch's *t* value, we evaluated the selected subsets based on the ability to accurately assign the compound as a carcinogen or noncarcinogen. By varying the number of ranked PTM data, we evaluated the PTM data subsets suitable for predicting carcinogenicity. The top 30 PTM data subset gave the highest concordance. Concordance of carcinogenicity prediction by the top 30 PTM data was 92.4% , sensitivity was 97.7% , and specificity was 82.6% (Figure 5A). The PTM data produced a much higher score than the top 30 proteins data. Figure 5B shows that concordance for the protein data was 74.2% , sensitivity was 83.7% , and specificity was 56.5% . In Figure 5, the mislabeled compounds are indicated as black bars. The specificity of PTM data was better than that of protein data. The only false negative was for urethane in the PTM data.

Conclusion

This study demonstrates a novel approach to the quantification and analysis of protein variations and PTMs at the proteomic level. We applied this approach to evaluate the effects of 66 chemical compounds in the male F344 rat liver in a 28-day repeated dose study. We found significant differences in PTMs in several proteins between carcinogens and noncarcinogens by statistical analysis. Furthermore, prediction of carcinogenicity using the PTM data was better than prediction using protein data. This suggests that changes resulting from PTMs are more relevant in the early stages of the development of cancer. Systematic quantitative analysis of PTMs is a promising way of acquiring new insights into the dynamic control of protein activities by PTM, and reveals their roles in biological processes.

Acknowledgements

We thank Kayo Sumida (Sumitomo Chemical Co. Ltd., Osaka, Japan), Kouji Nakayama (Mitsubishi Chemical Safety Institute Ltd., Ibaraki, Japan), Fumiyo Saito, Hiroshi Matsumoto, Masahiro Takeyoshi, Yasushi Minobe, Masanori Otsuka, and Mineo Takatsuki (Chemicals Evaluation and Research Institute, Saitama, Japan) for discussion of the manuscript. We thank Yukiko Kawano (Mitsubishi Chemical Safety Institute Ltd., Ibaraki, Japan), Mitsue Yamanaka, and Satsuki Hoshuyama, (Chemicals Evaluation and Research Institute, Saitama, Japan) for the preparation of rat liver samples and for BVA matching. This work was performed as basic research in a project entitled "Development of a High Precision Simplified Toxicity (Hazard) Assessment System," supported by the New Energy and Industrial Technology Development Organization (NEDO).

References

1 Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M and Jensen ON: Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Mol Cell Proteomics *4*: 310-327, 2005.

2 Cantin GT and Yates JRIII: Strategies for shotgun identification of posttranslational modification by mass spectrometry. Journal of Chromatography A *1053*: 7-14, 2004.

3 Jensen ON: Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. Current Opinion in Chemical Biology *8*: 33-41, 2004.

4 Wells L, Vosseller K, Cole RN, Cronshaw JM, Matunis MJ and Hart GW: Mapping sites of o-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. Mol Cell Proteomics *1*: 791-804, 2002.

5 Mann M, Ong S, Grønborg M, Steen H, Jensen ON and Pandey A: Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends in Biotechnology *20*: 261-268, 2002.

6 Fountoulakis M, Berndt P, Boelsterli UA, Crameri F, Winter M, Albertini S and Suter L: Two-dimensional database of mouse liver proteins: Changes in hepatic protein levels following treatment with acetaminophen or its nontoxic regioisomer 3-acetamidophenol. Electrophoresis *21*: 2148-2161, 2000.

7 Foutoularkis M, Juranville JF, Berndt P, Langen H and Suter L: Two-dimensional database of mouse liver proteins. An update. Electrophoresis *22*: 1747-1763, 2001.

8 Unlu M, Morgan ME and Minden JS: Difference gel electrophoresis. A single gel method for detecting changes in protein extracts. Electrophoresis *18*: 2071-2077, 1997.

9 Tonge R, Shaw J, Middleton B, Rowlinson R, Rayner S, Young J, Pognan F, Hawkins E, Currie I and Davison M: Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology Proteomics *1*: 377-396, 2001.

10 Karp NA, Kreil DP and Lilley KS: Determining a significant change in protein expression with DeCyder during a pair-wise comparison using two-dimensional difference gel electrophoresis. Proteomics *4*: 1421-321, 2004.

11 Weeks ME, Sinclair J, Jacob RJ, Saxton MJ, Kirby S, Jones J, Waterfield MD, Cramer R and Timms JF: Stress-induced changes in the Schizosaccharomyces pombe proteome using two-dimensional difference gel electrophoresis, mass spectrometry and a novel integrated robotics platform. Proteomics *5*: 1669-1685, 2005.

12 Sharma K, Lee S, Han S, Lee S, Francos B, McCue P, Wassell R, Shaw MA and RamachandraRao SP: Two-dimensional fluorescence difference gel electrophoresis analysis of the urine proteome in human diabetic nephropathy. Proteomics *5*: 2648-2655, 2005.

13 Liang CR, Leow CK, Neo JC, Tan GS, Lo SL, Lim JW, Seow TK, Lai PB and Chung MC: Proteome analysis of human hepatocellular carcinoma tissues by two-dimensional difference gel electrophoresis and mass spectrometry. Proteomics *5*: 2258-2271, 2005.

14 Ueda K, Kosako H, Fukui Y and Hattori S: Proteomic identification of Bcl2-associated athanogene 2 as a novel MAP kinase-activated protein kinase 2 substrate. J Biol Chem *279*: 41815-41821, 2004.

15 Bredemeyer AJ, Lewis RM, Malone JP, Davis AE, Gross J, Townsend RR and Ley TJ: A proteomic approach for the discovery of protease substrates. Proc Natl Acad Sci USA *101*: 11785-11790, 2004.

16 Seike M, Kondo T, Mori Y, Gemma A, Kudoh S, Sakamoto M, Yamada T and Hirohashi S: Proteomic Analysis of Intestinal Epithelial Cells Expressing Stabilized β-Catenin. Cancer Res *63*: 4641-4647, 2003.

17 Klenø TG, Leonardsen LR, Kjeldal HØ, Laursen SM, Jensen ON and Baunsgaard D: Mechanisms of hydrazine toxicity in rat liver investigated by proteomics and multivariate data analysis. Proteomics *4*: 868-880, 2004.

18 Gharbi S, Gaffney P, Yang A, Zvelebil MJ, Cramer R, Waterfield MD and Timms JF: Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. Mol. Cell. Proteomics *1*: 91-98, 2002.

19 O'Farrell PH: High resolution two-dimensional electrophoresis of proteins. J Biol Chem *250*: 4007-4021, 1975.

20 Claverol S, Schiltz OB, Gairin JE and Monsarrat B: Characterization of protein variants and post-translational modifications: ESI-MSn analyses of intact proteins eluted from polyacrylamide gels. Mol Cell Proteomics *2*: 483-493, 2003.

21 Schlags W, Lachmann B, Walther M, Kratzel M and Noe CR: Two-dimensional electrophoresis of recombinant human erythropoietin: A future method for the European Pharmacopoeia? Proteomics *2*: 679-682, 2002.

22 Nakayama K, Kawano Y, Kawakami Y, Moriwaki N, Sekijima M, Otsuka M, Yakabe Y, Miyaura H, Saito K, Sumida K and Shirai T: Differences in gene expression profiles in the liver between carcinogenic and non-carcinogenic isomers of compounds given to rats in a 28-day repeat-dose toxicity study. Toxicol Appl Pharmacol *217*: 299-307, 2006.

23 Sumida K, Saito K, Oeda K, Otsuka M, Tsujimura K, Miyaura H, Sekijima M, Nakayama K, Kawano Y, Kawakami Y, Asamoto M and Shirai T: Optimization of an animal test protocol for toxicogenomics studies (I); Requirement study of a protocol. J Toxicol Sci *32*: 19-32, 2007.

24 Yamanaka H, Yakabe Y, Saito K, Sekijima M and Shirai T: Quantitative proteomic analysis of rat liver for carcinogenicity prediction in a 28-day repeated dose study. Proteomics *7*: 781-795, 2007.

25 Sanchez JC, Rouge V, Pisteur M, Ravier F, Tonella L, Moosmayer M, Wilkins MR and Hochstrasser DF: Improved and simplified in-gel sample application using reswelling of dry immobilized pH gradients. Electrophoresis *18*: 324-327, 1997.

26 Rabilloud T, Adessi C, Giraudel A and Lunardi J: Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis *18*: 307-316, 1997.

27 Mann M: A shortcut to interesting human genes: peptide sequence tags, expressed-sequence tags and computers. Trends Biochem Sci *21*: 494-495, 1996.

28 Yates JRIII, Eng JK and McCormack AL: Mining genomes: Correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. Anal Chem *67*: 3202-3210, 1995.

29 Perkins DN, Pappin DJ, Creasy DM and Cottrell JS: Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*: 3551-3567, 1999.

30 Mann M and Wilm M: Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal Chem *66*: 4390-4399, 1994.
31 Eng JK, McCormack AL and Yates JRIII: An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. J Am Soc Mass Spectrom *5*: 976-989, 1994.
32 da Costa KA, Cochary EF, Blusztajn JK, Germer SC and Zeisel SH: Accumulation of 1,2-sn-diradylglycerol with increased membrane-associated protein kinase C may be the mechanism for spontaneous hepatocarcinogenesis in choline-deficient rats. J Biol Chem *268*: 2100-2105, 1993.
33 Chandar N and Lombardi B: Liver cell proliferation and incidence of hepatocellular carcinomas in rats fed consecutively a choline-devoid and a choline-supplemented diet. Carcinogenesis *9*: 259-263, 1988.
34 Lombardi B, Chandar N and Locker J: Nutritional model of carcinogenesis: rats fed a choline devoid diet. J Dig Dis Sci *36*: 979-984, 1991.
35 Dizik M, Christman JK and Wainfan E: Alterations in expression and methylation of specific gene in livers of rats fed a cancer promoting methyl-deficient diet. Carcinogenesis *12*: 1307-1312, 1991.
36 Ditzel HJ, Strik MC, Larsen MK, Willis AC, Waseem A, Kejling K and Jensenius JC: Cancer-associated cleavage of cytokeratin 8/18 heterotypic complexes exposes a neoepitope in human adenocarcinomas. J Biol Chem *277*: 21712-21722, 2002.

Received February 4, 2008
Accepted February 25, 2008