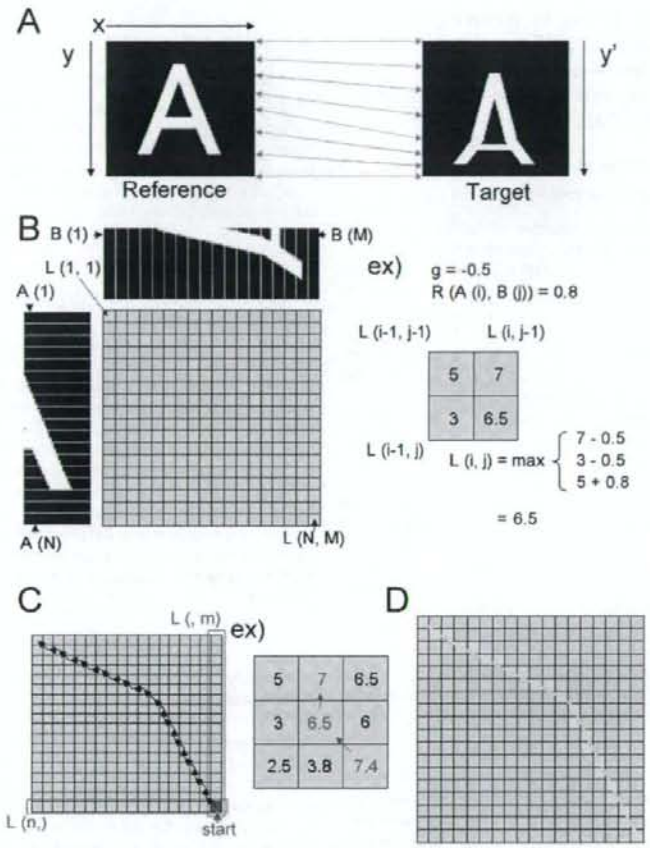


FIG. 1. **Adjustment of LC time jitter with a dynamic algorithm.** *A*, concept underlying compensation of RT changes along the *y axis*. This algorithm finds a function that compensates for time jitter *y'* by calculating $y = f(y')$ from target to reference. *B*, dynamic algorithm. The function is the path giving the optimal correspondence position (*yellow box* in two-dimensional lattice coordinates) in each cycle number (ascending RT order) of *A* and *B*. When the gap penalty is 0.5, the coefficient for the correlation between the mass spectra in the *i* cycle and the *j* cycle is 0.8, and the coordinate values are $L(i-1, j-1) = 5$, $L(i-1, j) = 3$, and $L(i, j-1) = 7$; then $L(i, j)$ becomes 6.5. The coordinate values of *L* in all lattices are calculated in this manner. *C*, optimal correspondence position. The optimal correspondence position starts at the lattice coordinates (k, l) where

$$(k, l) = \arg\max L(i, j) \text{ subject to: } \begin{cases} i = N, j = 1, \ldots, M \\ i = 1, \ldots, N, j = M \end{cases} \quad \text{(Eq. 6)}$$

When (k, l) is defined as V_0, the lattice coordinate is traced back. Thus

$$V_i = \arg\max L(V_{prev}),\ V_{prev} \in \begin{cases} V_{i-1} - (1,1) \\ V_{i-1} - (0,1) \\ V_{i-1} - (1,0) \end{cases} \quad \text{(Eq. 7)}$$

As shown in the example (*ex*), the next box to be selected has the maximum value among the *upper side*, *left side*, and *left upper side* of the former box. *D*, spline interpolation. To give the natural curve for the optimal correspondence coordinates, the coordinates arrangement of *V'*, which gives $V_{i+1} = V_i + (1, 1)$, is extracted (the combination in the *yellow boxes*). The function is the spline interpolation to the arrangement of *V'*.

Peak Matching and Protein Identification

Peaks in different LC-MS runs were matched with a tolerance of ±0.25 *m/z* and ±0.5 min after alignment of the RT. Matching was confirmed by visual inspection of enlarged two-dimensional views.

MS/MS was performed on peaks having an RT and *m/z* within that range in the preparatory LC run for protein identification. Peptide fragmentation data were analyzed with Mascot software (Matrix Sciences, London, UK).

Immunoblot Analysis

Cell lysates were separated by SDS-PAGE and electroblotted onto polyvinylidene difluoride membranes (Millipore). Anti-hemagglutinin monoclonal antibody was purchased from Roche Applied Science, anti-breast cancer-specific gene-1 (BCSG1) goat polyclonal antibody (C-20) was from Santa Cruz Biotechnology (Santa Cruz, CA), anti-cytokeratin 18 (RCK16) and anti-cytokeratin 19 (BA17) monoclonal antibodies were from Chemicon (Temecula, CA), and anti-β-actin monoclonal antibody (AAN02) was from Cytoskeleton (Denver, CO). The membranes were incubated with horseradish peroxidase-conjugated secondary antibodies and visualized with an enhanced chemiluminescence kit (Amersham Biosciences) (25).

RESULTS

Strategy for Quantitative Proteomics Using 2DICAL—To improve the reproducibility of LC without reducing its separation capacity we developed the nanoflow HPLC system utilizing a splitless direct gradient pump. Concentration of protein samples into small volumes and separation by a low flow rate HPLC with small sized reversed-phase columns significantly increase the sensitivity of ESI-MS (6, 26). In our HPLC system the dead volume was minimized, and the flow rate was reduced to the 50–200 nl/min range, but the elution gradient did not fluctuate significantly during runs of over 3 h (data not shown). Multidimensional LC does not always separate peptides in the same manner among experiments. The complexity of protein samples can be sufficiently reduced by such long runs even by one-dimensional LC separation.

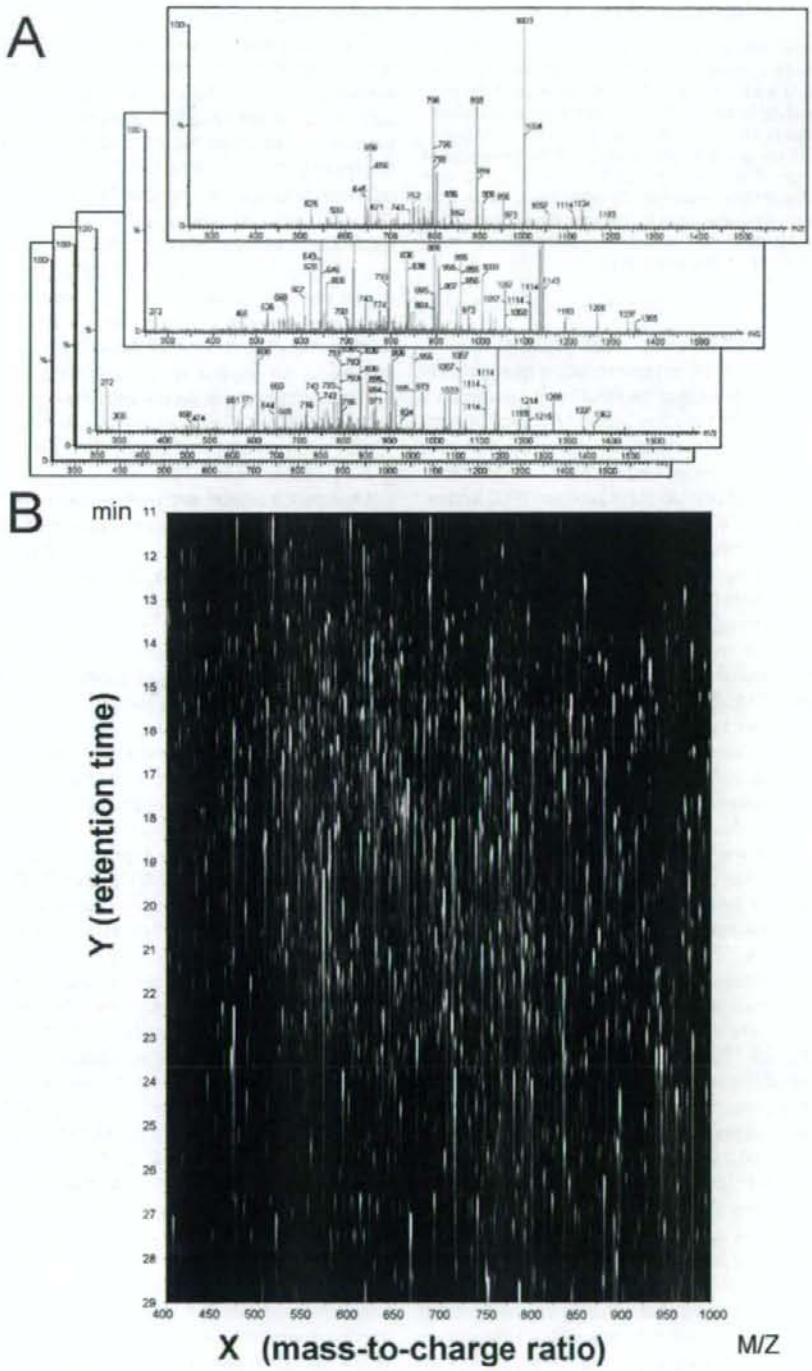
We also eliminated MS/MS and tried to detect all the peptides present in a given sample by high speed survey scanning as frequently as every 1 s. MS/MS scanning is a time-consuming procedure and often overlooks minor protein fragments (10). The automatic selection of precursor ions for MS/MS is not always constant, and such uncontrolled selection is likely to reduce the reproducibility.

Three thousand six hundred mass spectrum data (Fig. 2*A*) were obtained from a 1-h LC-MS run and converted into a single two-dimensional image with *m/z* values along the *x* axis and RT along the *y* axis (Fig. 2*B*). Fig. 2*C* shows a flowchart of the experimental procedures and subsequent data processing (Fig. 2, *D–I*). Signals with the same *m/z* appearing in at least two sequential RTs were grouped and considered as a peak (described under "Experimental Procedures") (Fig. 2, *F* and *G*). After normalizing to the total ion intensity of all of the peaks, the intensity of each peak was converted to a log value and expressed as a digital spot image (Fig. 2, *H* and *I*). We were able to detect 68,243 independent peaks (spots) in a 60-μg lysate of DLD1 Tet-Off ACTN4 cells cultured in the presence of doxycycline (Dox) (19).

Detection of Actinin-4 as a Differentially Expressed Protein in the Entire Proteome of DLD1 Tet-Off ACTN-4 Cells Cultured in the Absence of Dox—We previously established a colorectal cancer cell line that is capable of inducing an actin-binding protein, actinin-4, under the strict control of the tetracycline-regulatory system (designated DLD1 Tet-Off ACTN-4) (19). Upon removal of Dox from the culture medium, DLD1 Tet-Off ACTN-4 cells showed expression of actinin-4 (Fig. 3*A*), extended filopodia, and increased their motility (19). In a model experiment we investigated whether 2DICAL could pinpoint actinin-4 as a differentially expressed protein among the entire protein content of DLD1 Tet-Off ACTN-4 cells. Whole-cell lysates of DLD1 Tet-Off ACTN-4 cultured in the presence and absence of Dox were analyzed by 2DICAL in duplicate. We noted LC time jitter among the four runs (Fig. 3*B*) with a maximum difference in RT of 296 s and an average RT difference among the four runs of 36.0 s (Fig. 3, *B* and *C*). However, it was possible to adjust the LC time jitter with the newly developed software based on the dynamic programming algorithm (Figs. 1 and 3*D*), which had been developed to search sequences for regions of similarity and to align large DNA sequences (18). Because the *m/z* value (*x* axis) of each peak did not fluctuate among runs because of the high mass accuracy and constancy of Q-TOF MS, only adjustment of RT (*y* axis) was necessary.

After the RT adjustment, the average coefficient of variance (CV) of peak intensities between duplicates reached 0.37 (Dox (+)) (Fig. 3*E*) and 0.37 (Dox (−)), and the dynamic range of peak intensity calculated by our system exceeded 10^3 (Fig. 3*E*). There were 106 peaks that were more than 10-fold more highly expressed in DLD1 Tet-Off ACTN4 cells in the absence of Dox and 68 peaks that were more than 10-fold more highly expressed in the presence of Dox (Fig. 3*F*). We further confirmed the differential expression by visual inspection (Fig. 3*F*) and by repeating the same LC-MS experiment several times (data not shown). MS/MS analysis identified 15 (Fig. 3*G*) of the 106 peaks as having been derived from actinin-4. The biological significance of the other proteins whose expression was affected by induction of actinin-4 will be described elsewhere.

Detection of Differentially Expressed Proteins in Poorly Motile Capan-1 and Highly Motile BxPC3 Pancreatic Cancer Cells—Finally we compared the protein expression profiles of two pancreatic cancer cell lines, Capan-1 and BxPC3 (20), to determine whether 2DICAL is applicable to comparisons of proteomes with large differences because most of the protein (Fig. 3) as well as mRNA (data not shown) content of DLD1 Tet-Off ACTN-4 cells was unchanged after removing Dox, and that may have simplified the adjustment of the sample-to-sample time jitter. We noted that a significant number of spots were detected equally in Capan-1 cells and BxPC3 cells (Fig. 4*A*, *yellow spots*), making it possible to perform time adjustment of similar quality to that for DLD1 Tet-Off ACTN4 cells. In addition, a total of 15,407 spots that were differentially expressed between Capan-1 and BxPC3 were detected (9692 spots that were expressed more abundantly in Capan-1 than in BxPC3 and 5715 spots that were expressed more abundantly in BxPC3 than in Capan-1, $p < 0.01$, Student's *t* test, between triplicates) (Fig. 4*B*, *red* and *green spots*) after subtracting spots that were equally expressed by both (Fig. 4*A*, *yellow spots*). A representative differentially expressed pep-



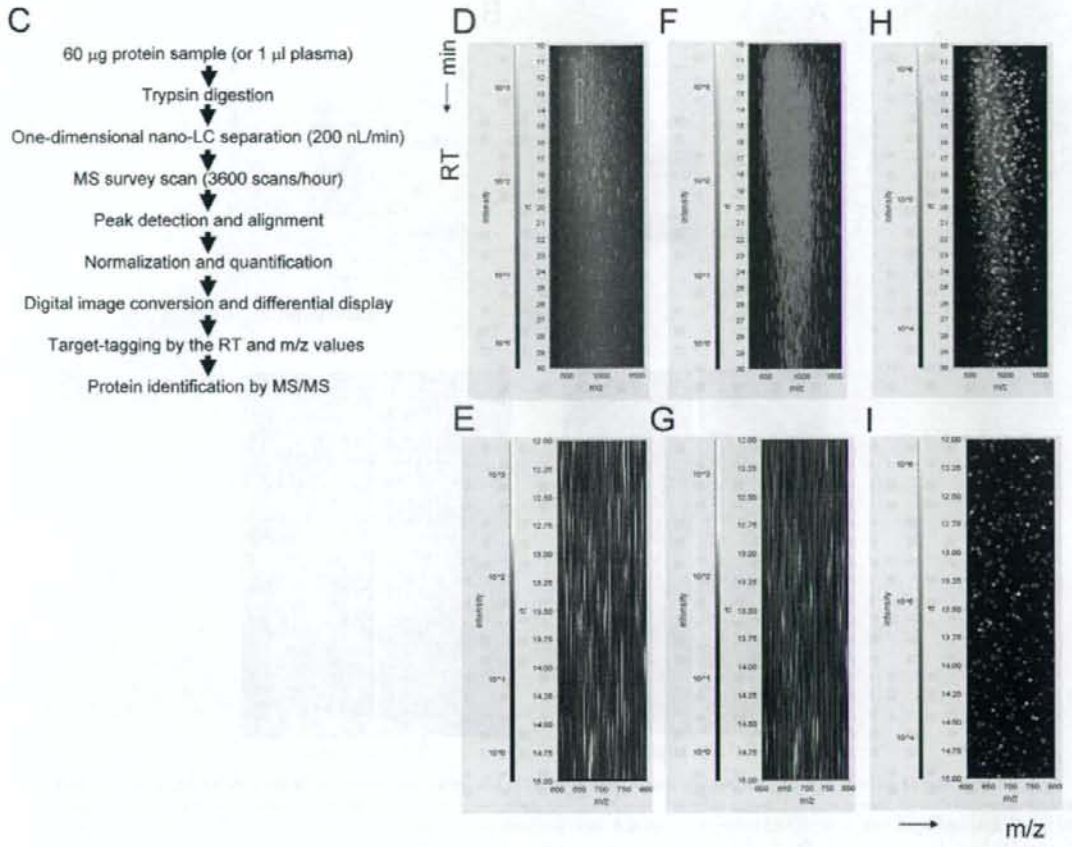


FIG. 2—*continued*

tide is shown in Fig. 4*C* (*red arrow*). Peaks were tagged by their intrinsic RT and *m/z* values, and MS/MS was performed on peaks having the same RT and *m/z* in the preparatory LC run for protein identification (Fig. 2*C*). Protein identification and differential expression were confirmed by immunoblotting with available antibodies. Fig. 4*D* shows representative data: BCSG1 (27), cytokeratin 19, and cytokeratin 18 were expressed more abundantly in the poorly motile Capan-1 cells than in the highly motile pancreatic cancer BxPC3 cells.

DISCUSSION

We reviewed various aspects of LC-MS for application to large scale quantitative proteomics and eliminated factors that reduce reproducibility and/or comprehensiveness, such

FIG. 2. **Strategy for quantitative proteomics using 2DICAL.** *A*, raw LC-MS mass spectra obtained from 1 μl of plasma from a healthy volunteer. A 1-h LC-MS run yielded 3600 mass spectra, which were used for two-dimensional image analysis. *B*, two-dimensional display of a plasma peptide array with the *m/z* values (400–1,000 *m/z*) along the *horizontal* (*x*) *axis* and RT (11–29 min) along the *vertical* (*y*) *axis*. *C*, flowchart for experimental procedures and data processing. *D*, two-dimensional raw image of the proteome of DLD1 Tet-Off ACTN4 cells with *m/z* values (250–1,600 *m/z*) along the *x axis* and RT (10–30 min) along the *y axis*. *E*, enlargement of the *light blue square* area in *D*. *F* and *G*, peak detection. The peaks appearing in *D* and *E* were picked up by using the algorithm described under "Experimental Procedures," and these are shown in the *red squares*. In *G* only peaks having intensity greater than 50,000 are *highlighted*. Peak intensity is the sum of the intensities of grouped signals. *H* and *I*, digital image conversion of the peaks detected in *F* and *G*. The virtual spots are located at *m/z* as monoisotopic molecular weights and at RT as the gravity center of ion intensity. The *brightness* of the spots corresponds to the peak intensity, defined as the integral of ion intensity of grouped signals, as described under "Experimental Procedures." As a result, the spot intensity exceeds 10^3, although the Q-TOF detector is saturated at thousands of counts per second (14, 16). The areas of *F* and *H* corresponding to the area in the *light blue square* in *D* have been enlarged, and these are shown in *G* and *I*, respectively.

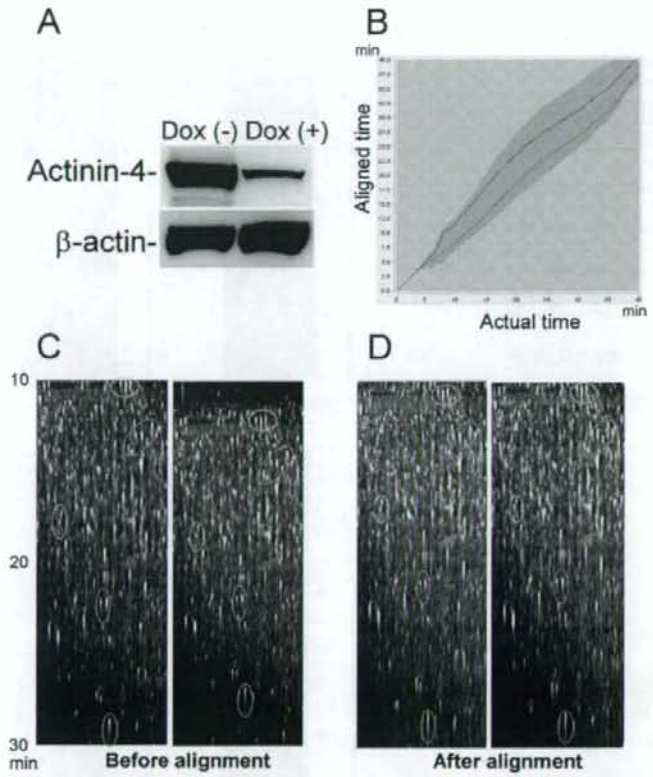


FIG. 3. **Detection of actinin-4 as a differentially expressed protein in the entire proteome of DLD1 Tet-Off ACTN-4 cells cultured in the absence of Dox.** *A*, immunoblot analysis of DLD1 Tet-Off ACTN4 cells cultured in the presence (+) and absence (−) of 0.1 μg/ml Dox for 72 h with anti-hemagglutinin antibody (to detect induced actinin-4) and anti-β-actin antibody (internal control). *B*, RT calibration curves of four 40-min LC runs. Lysates of DLD1 Tet-Off ACTN4 cells cultured in the presence (*blue* and *red lines*) and absence (*green* and *light blue lines*) of Dox were analyzed in duplicate (total, four runs). The *horizontal* (*x*) *axis* represents actual RT, and the *vertical* (*y*) *axis* represents adjusted RT. The average and maximum RT differences from a reference run of DLD1 Tet-Off ACTN4 cells cultured in the presence of Dox (*straight blue line* with a *slope* of 45°) of the three other runs were 36.0 and 296 s, respectively. *C*, 2DICAL images of duplicate runs of DLD1 Tet-Off ACTN4 cultured in the presence of Dox before the RT alignment. Representative peaks are highlighted in *light blue*, *yellow*, *green*, *pink*, and *red circles*. *D*, 2DICAL images of duplicate runs of DLD1 Tet-Off ACTN4 cells cultured in the presence of Dox after the RT alignment. Representative peaks are highlighted in *light blue*, *yellow*, *green*, *pink*, and *red circles*. *E*, reproducibility between duplicate runs of DLD1 Tet-Off ACTN4 cells cultured in the presence of Dox after the RT alignment. The *horizontal* (*x*) *axis* represents the distribution of peak intensities of the first run, and the *vertical* (*y*) *axis* represents that of the second run. The intensity CC between the two runs is 0.97. The average CV of all 68,243 peaks was 0.37. The average CV for DLD1 Tet-On ACTN4 cells cultured in the absence of Dox was 0.37 (data not shown). More than 70% of the duplicate peaks were plotted within a 2-fold difference (*blue lines*), and more than 90% were plotted within a 3-fold difference (*red lines*). We are able to detect 6.3-fold changes with 95% confidence. *F*, a representative peptide differentially expressed in DLD1 Tet-Off ACTN4 cells after removal of Dox and appearing at 717.4 *m/z* and 13.30 min. *G*, 15 peptide fragments appearing in DLD1 Tet-Off ACTN4 cells after removal of Dox and identified as derived from actinin-4 by MS/MS analysis. The *m/z* values, RT, and amino acid sequences of these 15 peptides are shown at the *right*.

as sample labeling, multidimensional LC separation, and MS/MS (Fig. 2*C*). However, the key technology of our platform is the application of the dynamic algorithm developed to align large DNA sequence data to the alignment of large peptide peak data generated by nano-LC-MS. Aebersold and colleagues (10, 28) also claimed that RT alignment is necessary to compare different LC-MS data and reported a new software suite. They first reduced the complexity of serum/plasma samples by extracting *N*-glycoproteins and identified 1000–5000 peaks from one set of LC-MS data. They reported a mean CV of 0.31 over four runs, and ~12% of the peptides were missed by the procedure (10). Wang *et al.* (14) detected ~3400 molecular ions in 25 human serum samples with a median CV of 25.7%. Our alignment method is capable of analyzing a much larger number of peaks with reasonable computing speed (2 h for the comparative analysis of two

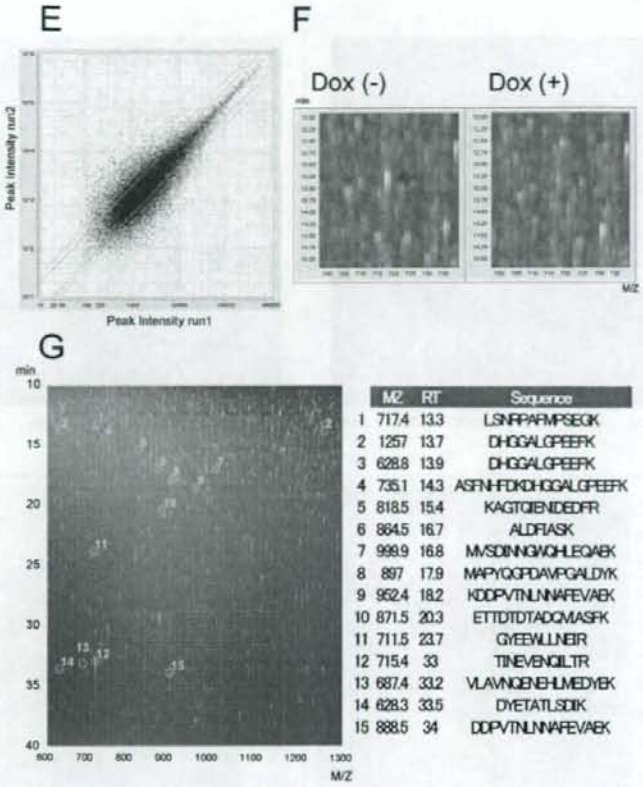


FIG. 3—*continued*

LC-MS experiments). We were able to detect more than 100,000 peaks in unlabeled and unfractionated protein samples and obtained equivalent but slightly higher CV values (0.35–0.39) than in their studies (Fig. 3*E*). We confirmed the efficacy of 2DICAL for quantitative protein analysis using two model experiments: an experiment that compared proteomes with small differences (Fig. 3) and another that compared proteomes with large differences (Fig. 4).

Bogdanov and Smith (29) reported two-dimensional display of capillary LC-FTICR analysis in which it was possible to detect more than 100,000 peaks and 1000 proteins in a single run and yielded a dynamic range of peak intensities of 10^3. We were able to obtain a comparable level of comprehensiveness using an easy-to-use and common MS instrument. Our alignment software seems ideal for comparative analysis of large data sets generated by LC-FTICR-MS. However, reduction of sample complexity in both platforms still seems necessary for detection of low abundance serum or plasma proteins because the concentration ranges of serum/plasma proteins span an estimated >10 orders of magnitude (30). Furthermore decreased sample complexity significantly reduces computing time and improves the accuracy of peak matching.

We consider the development of 2DICAL to still be in the early stage. Its reliability in regard to matching and quantification of low intensity peaks is not expected to be as high as for high intensity peaks. Because mismatching of peaks has a significant adverse effect on quantification, deliberate effort must be made to eliminate mismatching by visual inspection (Figs. 3*F* and 4*C*) and by recalculation. Differential protein expression cannot be identified based on statistical data alone. At a p value of <0.01, one would expect 1% of measured values to appear regulated due to variance in the data alone. Confirmatory reruns of different lots of samples were always necessary before targeted LC-MS/MS. Protein identification and differential expression also need to be confirmed by Western blotting whenever antibodies are available (Fig. 4*D*). We are now accumulating 2DICAL data to construct a two-dimensional map linking the m/z and RT of peptides to their amino acid sequences.

The amount of data obtained in one LC-MS experiment reached 2 gigabytes. The optimal number of samples for comparing seems to be determined by computer capacity.

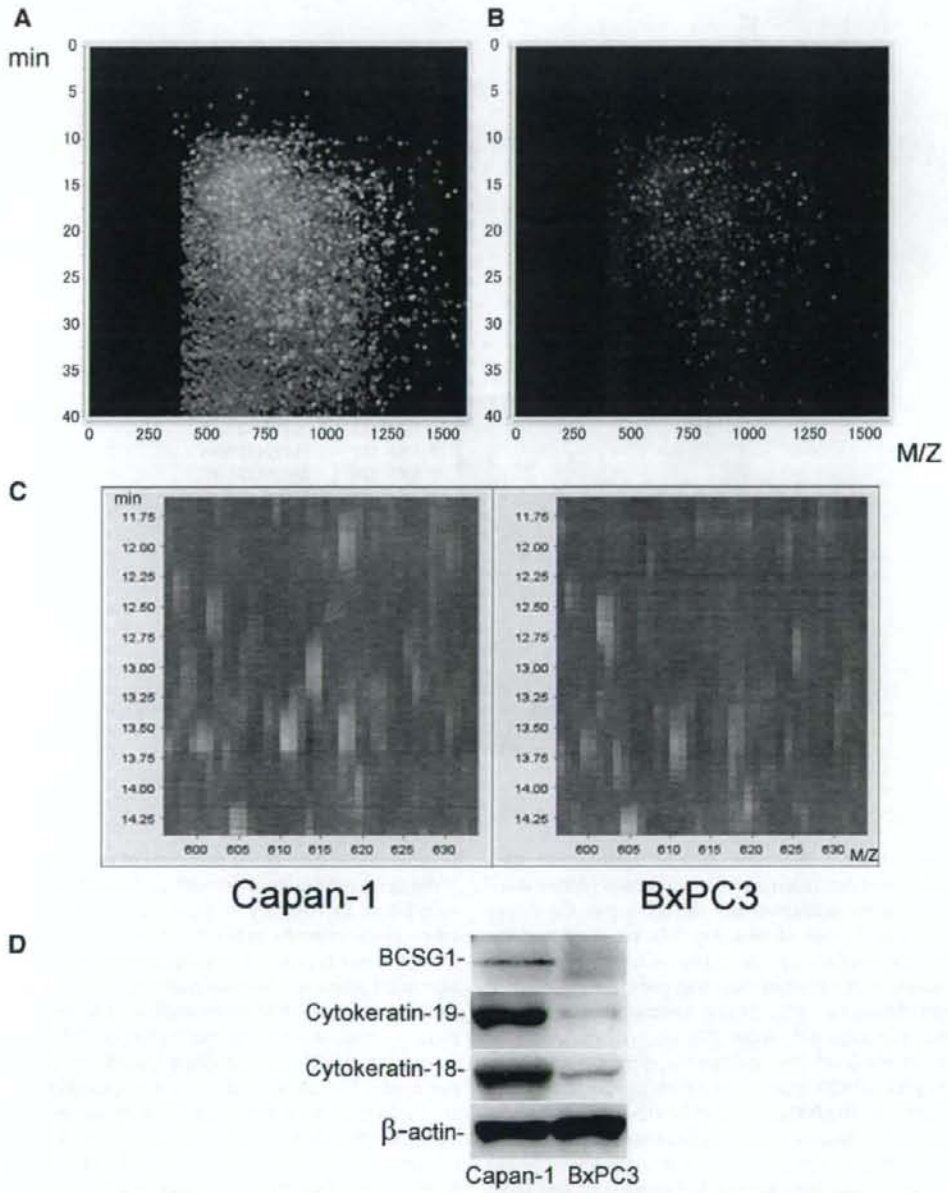


FIG. 4. **Detection of differentially expressed proteins in poorly motile Capan-1 and highly motile BxPC3 pancreatic cells.** *A*, overlap of spots appearing in Capan-1 cells (*red*) and BxPC3 cells (*green*) in the 400–16,000 *m/z* and 0–40-min range. *B*, differentially expressed spots in Capan-1 cells (*red*) and BxPC3 cells (*green*) in the 400–16,000 *m/z* and 0–40-min range. The levels of expression of 15,407 peaks (*spots*) were concluded to differ significantly between the two cell lines ($p < 0.01$, Student's *t* test, between triplicates). *C*, a peptide differentially expressed in Capan-1 cells (*left*) and BxPC3 cells (*right*) at 614.0 *m/z* and 13.0 min. This peptide was subsequently identified as being derived from the product of the BCSG1 by MS/MS. *D*, immunoblot confirmation of the differential expression of BCSG1, cytokeratin 19, and cytokeratin 18 in Capan-1 cells (*left*) and BxPC3 cells (*right*).

We currently use a high performance computer cluster consisting of four processors with a clock speed of 3.6 GHz connected in parallel to overcome this limitation. We conclude that 2DICAL is a simple, high throughput, and highly reproducible platform that will provide a new paradigm of quantitative proteomics. The software will be made available to the scientific community.

* This work was supported by the "Program for Promotion of Fundamental Studies in Health Sciences" conducted by the National Institute of Biomedical Innovation of Japan and the "Third-Term Comprehensive Control Research for Cancer" conducted by the Ministry of Health, Labor and Welfare of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Chemotherapy Division, National Cancer Center Research Inst., Tokyo 104-0045, Japan. Tel.: 81-3-3542-2511; Fax: 81-3-3547-6045; E-mail: tyamada@gan2.res.ncc.go.jp.

REFERENCES

1. Herbert, B. (1999) Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis* **20,** 660–663
2. Oh-Ishi, M., Satoh, M., and Maeda, T. (2000) Preparative two-dimensional gel electrophoresis with agarose gels in the first dimension for high molecular mass proteins. *Electrophoresis* **21,** 1653–1669
3. Swanson, S. K., and Washburn, M. P. (2005) The continuing evolution of shotgun proteomics. *Drug Discov. Today* **10,** 719–725
4. Wolters, D. A., Washburn, M. P., and Yates, J. R., III (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **73,** 5683–5690
5. Li, X. J., Pedrioli, P. G., Eng, J., Martin, D., Yi, E. C., Lee, H., and Aebersold, R. (2004) A tool to visualize and evaluate data obtained by liquid chromatography-electrospray ionization-mass spectrometry. *Anal. Chem.* **76,** 3856–3860
6. Natsume, T., Yamauchi, Y., Nakayama, H., Shinkawa, T., Yanagida, M., Takahashi, N., and Isobe, T. (2002) A direct nanoflow liquid chromatography-tandem mass spectrometry system for interaction proteomics. *Anal. Chem.* **74,** 4725–4733
7. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422,** 198–207
8. Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B. B., Simpson, R. J., Eddes, J. S., Kapp, E. A., Moritz, R. L., Chan, D. W., Rai, A. J., Admon, A., Aebersold, R., Eng, J., Hancock, W. S., Hefta, S. A., Meyer, H., Paik, Y. K., Yoo, J. S., Ping, P., Pounds, J., Adkins, J., Qian, X., Wang, R., Wasinger, V., Wu, C. Y., Zhao, X., Zeng, R., Archakov, A., Tsugita, A., Beer, I., Pandey, A., Pisano, M., Andrews, P., Tammen, H., Speicher, D. W., and Hanash, S. M. (2005) Overview of the HUPO Plasma Proteome Project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **5,** 3226–3245
9. Fujii, K., Nakano, T., Kanazawa, M., Akimoto, S., Hirano, T., Kato, H., and Nishimura, T. (2005) Clinical-scale high-throughput human plasma proteome analysis: lung adenocarcinoma. *Proteomics* **5,** 1150–1159
10. Li, X. J., Yi, E. C., Kemp, C. J., Zhang, H., and Aebersold, R. (2005) A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography-mass spectrometry. *Mol. Cell. Proteomics* **4,** 1328–1340
11. 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. *Nat. Biotechnol.* **17,** 994–999
12. DeSouza, L., Diehl, G., Rodrigues, M. J., Guo, J., Romaschin, A. D., Colgan, T. J., and Siu, K. W. (2005) Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. *J. Proteome Res.* **4,** 377–386
13. 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
14. Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T. A., Hill, L. R., Norton, S., Kumar, P., Anderle, M., and Becker, C. H. (2003) Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* **75,** 4818–4826
15. Radulovic, D., Jelveh, S., Ryu, S., Hamilton, T. G., Foss, E., Mao, Y., and Emili, A. (2004) Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. *Mol. Cell. Proteomics* **3,** 984–997
16. Silva, J. C., Denny, R., Dorschel, C. A., Gorenstein, M., Kass, I. J., Li, G. Z., McKenna, T., Nold, M. J., Richardson, K., Young, P., and Geromanos, S. (2005) Quantitative proteomic analysis by accurate mass retention time pairs. *Anal. Chem.* **77,** 2187–2200
17. Anderle, M., Roy, S., Lin, H., Becker, C., and Joho, K. (2004) Quantifying reproducibility for differential proteomics: noise analysis for protein liquid chromatography-mass spectrometry of human serum. *Bioinformatics* **20,** 3575–3582
18. Weir, B. S. (1988) Statistical analysis of DNA sequences. *J. Natl. Cancer Inst.* **80,** 395–406
19. Honda, K., Yamada, T., Hayashida, Y., Idogawa, M., Sato, S., Hasegawa, F., Ino, Y., Ono, M., and Hirohashi, S. (2005) Actinin-4 increases cell motility and promotes lymph node metastasis of colorectal cancer. *Gastroenterology* **128,** 51–62
20. Shimamura, T., Yasuda, J., Ino, Y., Gotoh, M., Tsuchiya, A., Nakajima, A., Sakamoto, M., Kanai, Y., and Hirohashi, S. (2004) Dysadherin expression facilitates cell motility and metastatic potential of human pancreatic cancer cells. *Cancer Res.* **64,** 6989–6995
21. Honda, K., Hayashida, Y., Umaki, T., Okusaka, T., Kosuge, T., Kikuchi, S., Endo, M., Tsuchida, A., Aoki, T., Itoi, T., Moriyasu, F., Hirohashi, S., and Yamada, T. (2005) Possible detection of pancreatic cancer by plasma protein profiling. *Cancer Res.* **65,** 10613–10622
22. Breen, E. J., Hopwood, F. G., Williams, K. L., and Wilkins, M. R. (2000) Automatic Poisson peak harvesting for high throughput protein identification. *Electrophoresis* **21,** 2243–2251
23. Gras, R., Muller, M., Gasteiger, E., Gay, S., Binz, P. A., Bienvenut, W., Hoogland, C., Sanchez, J. C., Bairoch, A., Hochstrasser, D. F., and Appel, R. D. (1999) Improving protein identification from peptide mass fingerprinting through a parameterized multi-level scoring algorithm and an optimized peak detection. *Electrophoresis* **20,** 3535–3550
24. Giegerich, R. (2000) A systematic approach to dynamic programming in bioinformatics. *Bioinformatics* **16,** 665–677
25. Seike, M., Kondo, T., Mori, Y., Gemma, A., Kudoh, S., Sakamoto, M., Yamada, T., and Hirohashi, S. (2003) Proteomic analysis of intestinal epithelial cells expressing stabilized beta-catenin. *Cancer Res.* **63,** 4641–4647
26. Reinders, J., Lewandrowski, U., Moebius, J., Wagner, Y., and Sickmann, A. (2004) Challenges in mass spectrometry-based proteomics. *Proteomics* **4,** 3686–3703
27. Ji, H., Liu, Y. E., Jia, T., Wang, M., Liu, J., Xiao, G., Joseph, B. K., Rosen, C., and Shi, Y. E. (1997) Identification of a breast cancer-specific gene, BCSG1, by direct differential cDNA sequencing. *Cancer Res.* **57,** 759–764
28. Zhang, H., Yi, E. C., Li, X. J., Mallick, P., Kelly-Spratt, K. S., Masselon, C. D., Camp, D. G., 2nd, Smith, R. D., Kemp, C. J., and Aebersold, R. (2005) High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol. Cell. Proteomics* **4,** 144–155
29. Bogdanov, B., and Smith, R. D. (2005) Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom. Rev.* **24,** 168–200
30. Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1,** 845–867

DOI 10.1002/pmic.200500883 *Proteomics* 2006, *6*, 3938–3948

RESEARCH ARTICLE

Plasma proteomics of lung cancer by a linkage of multi-dimensional liquid chromatography and two-dimensional difference gel electrophoresis

Tetsuya Okano[1, 4], Tadashi Kondo[1], Tatsuhiko Kakisaka[1], Kiyonaga Fujii[1], Masayo Yamada[1], Harubumi Kato[2, 3], Toshihide Nishimura[2], Akihiko Gemma[4], Shoji Kudoh[4] and Setsuo Hirohashi[1]

[1] Proteome Bioinformatics Project, National Cancer Center Research Institute, Tokyo, Japan
[2] Clinical Proteome Center, Tokyo Medical University, Tokyo, Japan
[3] Department of Surgery, Tokyo Medical University, Tokyo, Japan
[4] Fourth Department of Internal Medicine, Nippon Medical School, Tokyo, Japan

Received: December 6, 2005
Revised: February 19, 2006
Accepted: March 12, 2006

To investigate aberrant plasma proteins in lung cancer, we compared the proteomic profiles of serum from five lung cancer patients and from four healthy volunteers. Immuno-affinity chromatography was used to deplete highly abundant plasma proteins, and the resulting plasma samples were separated into eight fractions by anion-exchange chromatography. Quantitative protein profiles of the fractionated samples were generated by two-dimensional difference gel electrophoresis, in which the experimental samples and the internal control samples were labeled with different dyes and co-separated by two-dimensional polyacrylamide gel electrophoresis. This approach succeeded in resolving 3890 protein spots. For 364 of the protein spots, the expression level in lung cancer was more than twofold different from that in the healthy volunteers. These differences were statistically significant (Student's *t*-test, *p*-value less than 0.05). Mass spectrometric protein identification revealed that the 364 protein spots corresponded to 58 gene products, including the classical plasma proteins and the tissue-leakage proteins catalase, clusterin, ficolin, gelsolin, lumican, tetranectin, triosephosphate isomerase and vitronectin. The combination of multi-dimensional liquid chromatography and two-dimensional difference gel electrophoresis provides a valuable tool for serum proteomics in lung cancer.

Keywords:
Lung cancer / Multi-dimensional liquid chromatography / Two-dimensional difference gel electrophoresis

1 Introduction

Lung cancer is a leading cause of cancer death in Japan, claiming 55 000 lives annually, and is a major health problem in many countries. The prognosis of patients with lung cancer is generally poor, with an overall 5-year survival rate for patients receiving treatment of only 14%. In contrast, overall 5-year survival for patients diagnosed with stage I adenocarcinoma approaches 63% [1]. As non-small-cell lung cancer accounts for almost 80% of lung cancers, of which 40% are adenocarcinoma, a substantial number of patients with lung cancer have the potential to be cured successfully by early treatment. However, the majority of lung tumors have reached locally advanced stage III (33%) or metastatic stage IV (41%) by the time of diagnosis [2]. Therefore, early diagnosis of lung cancer is necessary to improve patient survival. Plasma is a preferred specimen for the early diagnosis of lung cancer because samples are easily available by non-

Correspondence: Professor Tadashi Kondo, Proteome Bioinformatics Project, National Cancer Center Research Institute, 5–1-1 Tsukiji, Chuo-ku, Tokyo 104–0045, Japan
E-mail: takondo@gan2.res.ncc.go.jp
Fax: +81-3-3547-5298

invasive methods. However, the currently available plasma tumor markers such as CEA, NSE, TPA, chromogranin, CA125, CA19–9, Cyfra 21–1, and ProGRP have limited sensitivity and specificity for early diagnosis [3], and novel plasma markers are required.

Multi-dimensional separation techniques based on the combination of LC and gel electrophoresis have been applied for plasma proteomics [4]. Typically, LC with immuno-affinity, size-exclusion, ion-exchange and RP columns is used to fraction the proteins, which are then subjected to high-resolution 2-D-PAGE for quantitative expression studies. Sample fractionation prior to 2-D-PAGE can separate gene products present in low copy numbers from highly abundant proteins, increasing the number of observable proteins. Indeed, less abundant proteins, including tissue-leakage proteins and regulatory proteins, have been identified in plasma by the combination of LC and 2-D-PAGE [4]. As the results of 2-D-PAGE can be quantified and stored in a database, 2-D-PAGE is a powerful tool for biomarker development. However, the intrinsic limitations of 2-D-PAGE resulting from gel-to-gel variations can hinder accurate comparisons of protein expression levels. To cancel such experimental differences, 2-D-DIGE has been developed [5–7]. In 2-D-DIGE, the experimental samples and an internal control sample are labeled with different fluorescent dyes, mixed together and co-separated in identical gels. The fluorescent dyes are designed so that the electrophoretic migration of proteins labeled with the different dyes is almost identical. Therefore, the intensity of the spots of the experimental sample can be normalized to the intensity of the corresponding spots of the internal control sample in the same gel. In 2-D-DIGE the amounts of proteins are measured as fluorescence signals, thus the dynamic range is wider than with conventional silver staining and spot detection can be achieved by simple laser scanning in a high-throughput manner. 2-D-DIGE has been used to characterize the proteome of plasma from patients with disease [8]. Recently, novel highly sensitive fluorescent dyes, CyDye DIGE Fluor saturation dyes (GE Healthcare Amersham Biosciences, Uppsala, Sweden), referred to here as 'saturation dye', have been developed [9]. The high sensitivity of saturation dye has enabled proteomic studies of primary cultured human hepatocytes [10] and laser-microdissected tumor tissues [11–14], where only limited amounts of proteins were available.

In this study, we used the combination of multi-dimensional chromatography and 2-D-DIGE with saturation dye to conduct a proteomic comparison of serum from patients with lung cancer and from healthy volunteers. The high sensitivity of saturation dye generated 3890 protein spots from only 30 μL of plasma. We identified 364 spots as aberrantly regulated plasma proteins in lung cancer. MS revealed that these spots corresponded to 58 gene products. The identified proteins included previously uncharacterized tissue-leakage proteins in addition to acute phase plasma proteins. These results demonstrate the utility of multi-dimensional LC and 2-D-DIGE for plasma proteomics in cancer.

2 Materials and methods

2.1 Serum samples

Blood samples by venipuncture and informed consent were obtained from five patients with lung cancer and four healthy volunteers at Nippon Medical School. Clinical information for the donors of serum proteins is summarized in Table 1. A 10-mL blood sample was obtained with a VENOJECT II (10 mL; TERUMO, Tokyo, Japan) and allowed to clot for 2 h at 4°C. The clotted material was removed by centrifugation at 3000 rpm for 10 min. The supernatant sera obtained from the blood samples were recovered and stored at −80°C until use.

Table 1. Patients' characteristics

	Histology [29]	Gender	Age (years)	Stage [30]
1.	Adenocarcinoma	Male	48	T2N2M0 IIIA
2.	Adenocarcinoma	Male	62	T4N2M0 IIIB
3.	Squamous cell carcinoma	Female	75	T4N3M1 IV
4.	Squamous cell carcinoma	Male	76	T4N1M1 IV
5.	Small cell carcinoma	Male	40	T2N2M0 IIIA
Four healthy volunteers		4 Males	29–38	

2.2 Immuno-affinity and anion-exchange chromatography

Serum proteins were separated by immuno-affinity and ion-exchange chromatography with the AKTA Explore system (GE Healthcare Amersham Biosciences). The immuno-affinity column (4.6 × 100 mm; Agilent Technologies) contained anti-albumin, anti-transferrin, anti-haptoglobin, anti-alpha-1-anti-trypsin, anti-IgA and anti-IgG resins. A serum sample (30 μL) was diluted with 120 μL of a neutral buffer (buffer A; Agilent Technologies), and filtered with a spin filter (pore size 0.22 μm; Agilent Technologies) by centrifugation at 16 000 × *g* for 1 min prior to use. The filtered sample was applied to the immuno-affinity column at a flow rate of 0.5 mL/min for 10 min. The flow-through fraction was recovered and the proteins retained in the column were eluted with a low-pH urea buffer (buffer B; Agilent Technologies) at a flow rate of 1.0 mL/min for 10 min. The column was recycled for further use by washing with buffer A at a flow rate of 1.0 mL/min for 12 min. The flow-through fractions were concentrated to 500 μL in a Spin Concentrator, 5K MWCO (4 mL capacity, Agilent Technologies).

The concentrated sample was diluted with 4.5 mL of 25 mM Tris-HCl, pH 9.0, and applied to a Resource Q column (1.0 mL resin, 6.4 mm id × 30 mm; GE Healthcare

Amersham Biosciences) at a flow rate of 4.0 mL/min. The separations were performed with a step-wise gradient of NaCl as follows: 0 mM for 12.5 min, 100, 150, 200, 250, 300, 350, and1000 mM for 2.5 min each. All elution buffers contained 25 mM Tris-HCl, pH 9.0. The anion-exchange column was then recycled for further use by thorough washing with 25 mM Tris-HCl, pH 9.0, containing 2 M NaCl. In the interval between the elution steps, the pump system was washed with 10 mL of the next elution buffer. Protein peaks were monitored at 280 nm. The eluted proteins were concentrated with Amicon Ultra PL-10 (a molecular mass cut-off 10 kDa) in an Amicon Ultra-15 filter unit (Amicon Bredford, MA).

2.3 2-D DIGE

The fractionated proteins were precipitated by addition of four volumes of acetone at −20°C for 20 min. After centrifugation at 13 000 rpm for 10 min, the supernatant was discarded and the pellet was air-dried for 10 min. The dried pellet was dissolved in 50 μL of lysis buffer containing 6 M urea, 2 M thiourea, 3% CHAPS, 1% Triton X-100 and 40 mM Tris (pH 8.0). The protein concentration was measured with a Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA). The fluorescence labeling was performed as described previously with some modifications [10]. In brief, samples (50 μg of protein) were reduced by incubation with 2 μM Tris-(2-carboxethyl)phosphine hydrochloride (TCEP; Sigma, St. Louis, MO) at 37°C for 60 min. The protein samples were fluorescence labeled by incubation with 40 nM of saturation Cy3 or Cy5 dye (GE Healthcare Amersham Biosciences) at 37°C for another 30 min. For 2-D-PAGE, the labeling reaction was terminated by addition of an equal volume of the lysis buffer containing 130 mM DTT and 2.0% Pharmalyte (Amersham Biosciences). In the fraction eluted with 1000 mM NaCl, the amount of protein was so small that its concentration could not be measured. Therefore, all the protein in these fractions was labeled and used for the subsequent studies. As the trains of spots in the vertical dimension were not observed on 2-D images, we concluded that the proteins were saturatedly labeled with fluorescent dyes.

The 2-D-PAGE was carried out as described previously in our reports [11]. In brief, the first dimension separation was carried out with Immobiline Drystrips (24 cm, pH 3–10; GE Healthcare Amersham Biosciences). Each strip was rehydrated for 12 h at 30 V with 420 μL of protein sample, and IEF was performed in an IPGphor unit (GE Healthcare Amersham Bioscience). After electrophoresis, the strips were equilibrated in equilibration buffer (6 M urea, 2% SDS, 50 mM Tris HCl, pH 8.8, 30% glycerol w/v) for 20 min. The second dimension separation was performed on homemade 9–15% gradient polyacrylamide gels with an EttanDalt II system (GE Healthcare Amersham Biosciences) at a constant wattage of 17 W at 20°C for 17 h. The image was acquired by scanning the gels with a laser scanner (2D MasterImager; GE Healthcare Amersham Biosciences). Spot detection, quantification and image matching were performed with DeCyder software (GE Healthcare Amersham Biosciences). One portion of the labeled proteins was separated by SDS-PAGE using 10% polyacrylamide gel to monitor the contents of the fractions. Gel electrophoresis was performed in the dark.

2.4 Protein identification by MS

In-gel digestion was performed for protein spots excised by an automated spot collector (SpotPicker; GE Healthcare Amersham Biosciences), according to our previous report [10]. Then, the gel pieces were extensively washed with ammonium bicarbonate. The protein in the dried gel plug was digested overnight at 37°C with sequencing-grade modified trypsin (Promega, Madison, WI). The tryptic digests were recovered by incubation with 50% ACN/0.1% TFA, and the isolated peptides were concentrated under nitrogen gas and subjected to LC-MS/MS. The MS study was carried out as described previously [15]. In brief, the LC-MS/MS system comprised a Paradigm MS4 dual solvent delivery system (Michrom BioSciences, Auburn, CA, USA) for HPLC, an HTS PAL auto sampler with two 10-port injector valves (CTC Analytics, Zwingen, Switzerland), and a Finnigan LTQ linear ITMS (Thermo Electron, San Jose, CA, USA) equipped with NSI sources (AMR, Tokyo, Japan). A database search against Swiss-Prot was performed with MASCOT software. When multiple proteins were identified in a single spot, the proteins with the highest numbers of peptides were considered as those corresponding to the spots. When multiple protein candidates were listed with an equal number of the identified peptides, the proteins with a higher MASCOT score were selected.

2.5 Western blotting

Protein samples were separated by SDS-PAGE and transferred onto NC membranes. Differential expression of plasma proteins was examined by specific antibodies against leucine-rich alpha 2 glycoprotein (x1000 dilution; Abnova, Taiwan), inter-alpha-trypsin inhibitor heavy chain H4 (x1000 dilution; Santa Cruz Biotechnology, Santa Cruz, CA), plasma retinol binding protein (x50 dilution; Biomeda), haptoglobin (x2000 dilution; Sigma Aldrich, St. Louis, MO, USA), complement component C4 (x200 dilution; Antibody Shop, Gentofte, Denmark), complement component C3 (x200 dilution; Antibody Shop), and prothrombin (x100 dilution; BD Transduction laboratories). The anti-prothrombin antibody could recognize both thrombin and prothrombin. A second antibody against goat IgG (Santa Cruz Biotechnology) was used for inter-alpha-trypsin inhibitor heavy chain H4 at a dilution of 1:2000. A second antibody against mouse IgG (GE Healthcare Amersham Biosciences) was used for leucine-rich alpha 2 glycoprotein, plasma retinol-binding protein, haptoglobin, complement component C4, complement component C3 and prothrombin at a dilution of 1:1000. Immune complexes were detected with an enhanced

chemiluminescence system (GE Healthcare Amersham Biosciences) and monitored with an LAS-1000 (Fuji Film, Tokyo, Japan).

3 Results

3.1 Evaluation of the fractionation process

In the first dimension chromatography, we used a recently developed immuno-affinity column to deplete the most abundant plasma proteins, including albumin, immunoglobulin, transferrin, anti-trypsin, and haptoglobin. This immuno-affinity column has been employed for plasma proteomics to enrich the less abundant plasma proteins [4, 16–18]. Whole serum was separated into two fractions: the flow-through fraction containing low abundance serum proteins and the bound fraction containing the abundant proteins listed above (Fig. 1A). Proteins in the flow-through fraction were subjected to anion-exchange chromatography (Fig. 1B). To ensure reproducibility of protein overlap between neighboring fractions, we employed step-wise separation instead of gradient separation. In addition, we washed the column with the elution buffer at the end of each fraction step to reduce the carry-over of elution buffer between the steps. To evaluate the reproducibility of the fractionation process, the 250-mM NaCl fraction of three independently prepared protein samples was separated by RP chromatography (Fig. 1C). The three chromatograms showed that the fractionation process was highly reproducible.

To examine the effects of fractionation, the protein samples were labeled with fluorescent dyes and separated by SDS-PAGE (Fig. 2). The major bands in the unfractionated serum proteins, as indicated by arrows (lane 1), were observed in the bound fraction of the immuno-affinity column (lane 3). Proteins in the flow-through fraction of the immuno-affinity column generated a higher number of bands (lane 2). The other portion of this fraction was further separated by anion-exchange chromatography, labeled with fluorescent dyes and subjected to SDS-PAGE (lanes 4–11). The fractions from anion-exchange chromatography showed distinct protein migration patterns (lanes 4–11) compared with the unfractionated serum (lane 1) and the flow-through fraction of the immuno-affinity column (lane 2). Anion-exchange chromatography combined with immuno-affinity chromatography reduced the complexity and dynamic range of the protein content and separated the less abundant proteins from the neighboring highly abundant proteins, resulting in an increase in the number of observable proteins.

We conducted 2-D-PAGE to separate further the fractionated proteins. The 2-D-PAGE allows detection of post-translational differences, observed as a shift in p*I* that are difficult to recognize by SDS-PAGE. To validate the effects of fractionation on the number of observable proteins, we labeled a whole-serum sample and a fractionated protein

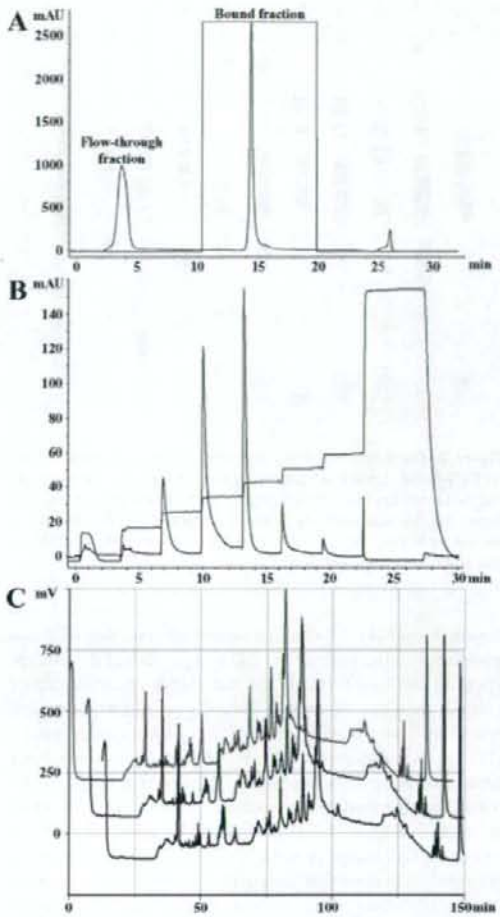


Figure 1. Process of multi-dimensional chromatography separation. (A) The immuno-affinity column separated plasma proteins into two fractions: flow-through and bound fraction. The column was recycled by washing after elution of the bound proteins. (B) The flow-through fraction from the immuno-affinity column was subjected to anion-exchange chromatography and separated into eight fractions by elution with a step-wise gradient of NaCl. (C) Three independent samples of the 250 mM NaCl fraction were prepared and separated by RP chromatography to examine the reproducibility of the immuno-affinity and anion-exchange procedures.

sample with Cy3 and Cy5, respectively, mixed them together and co-separated them by 2-D-PAGE. Figure 3 shows the two-color images of whole-serum (red) and fractionated proteins (green). It is apparent that most of the protein spots that bound to, and were recovered from the immuno-affinity column, matched the spots with high intensity in whole

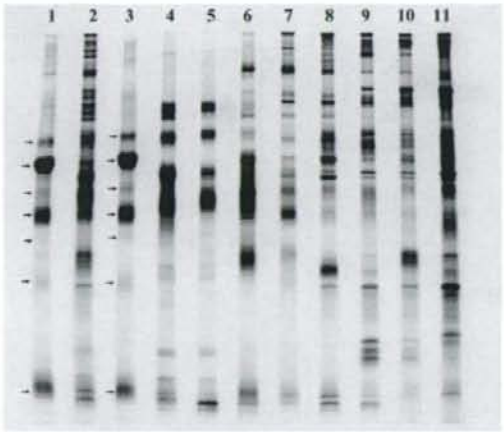


Figure 2. Separation of whole and fractionated plasma samples by SDS-PAGE. Lane 1, whole serum sample; lane 2, flow-through fraction from the immuno-affinity column; lane 3, bound fraction from the immuno-affinity column; lanes 4–11, the fractions eluted by 0, 100, 150, 200, 250, 300, 350 and 1000 mM NaCl from the anion-exchange column.

serum (panel A). In contrast, immuno-depletion of these proteins resulted in a distinct 2-D image; most of the protein spots in the flow-through fraction of the immuno-affinity column were not observed in the 2-D image of whole serum (panel B). These results are consistent with previous reports [17, 19]. We also separated the Cy5-labeled fractions from anion-exchange chromatography by 2-D-PAGE (panels C–J). Visual comparison of the location and intensity of spots through the series of 2-D images can be achieved easily by using the Cy3 image of whole serum as a common internal standard. It is clear that the number of spots was increased by fractionation with anion-exchange chromatography. However, it is unclear how many proteins repeatedly appeared as different spots in the gels.

3.2 Comparison of serum from patients with lung cancer and healthy volunteers

We compared serum proteins in patients with lung cancer and healthy volunteers (Table 1). We prepared a mixture of the paired protein samples to make an internal standard and labeled it with Cy3. The individual samples were labeled with Cy5 and mixed with the Cy3-labeled internal standard sample. The mixture of labeled protein samples was separated by 2-D-PAGE. Figure 4 shows a representative gel image of Cy5-labeled proteins. The number of spots observed is described in the lower right corner of the 2-D image panel. The number of spots in the mixture of unfractionated serum of patients with lung cancer and of healthy volunteers was 124, and fractionation increased the total number of spots to 3766. The arrows indicate the 364 spots whose expression in lung cancer differed significantly by more than twofold from that in the healthy volunteers (Student's *t*-test, *p*-value less than 0.05). MS identified the proteins from the spots. We observed multiple proteins from single spots in several cases. These observations were consistent with the previous report [20]. Tentatively, the proteins with the highest number of peptides were considered to be the proteins corresponding to the spots, although this is not always the case. Based on this criterion, we concluded that MS study of these 364 spots resulted in the identification of 58 distinct gene products. The results of protein identification are summarized in Suppl. Table 1.

We found that certain proteins appeared on the 2-D images as multiple protein spots with aberrantly controlled expression in lung cancer. In addition, different protein spots translated from the same gene showed different regulation, so that isoforms of a particular plasma protein could be up- or down-regulated in lung cancer. Figure 5 summarizes these observations. Among 58 proteins identified, 18 were detected as a single spot. These were alpha-1-acid glycoprotein, apolipoprotein D, apolipoprotein M, carbonic anhydrase II, catalase, ceruloplasmin, complement C5, corticosteroid-binding globulin, Ig alpha-1 chain C region, Ig gamma-1 chain C region, Ig kappa chain V-III region SIE, inter-alpha-trypsin inhibitor heavy chains H1, H2 and H4, kininogen, tetranectin, triosephosphate isomerase, and vitronectin. Spots for 28 proteins showed consistent up- or down-regulation (Fig. 5A) and the spots for 12 proteins showed inconsistent regulation (Fig. 5B). For example, the intensity of all 67 haptoglobulin spots was increased in lung cancer, whereas for the 38 spots of complement component C3 the intensity of 20 spots was up-regulated and that of 18 spots was down-regulated in lung cancer. Thus, isoform-specific regulation may exist for these proteins.

We then used specific antibodies to validate the differential expression of the plasma proteins whose isoforms showed consistent aberrations in lung cancer (Fig. 6), because clinical application of our results would require differential expression to be monitored by methods more convenient than the combination of multi-dimensional LC and 2-D-DIGE. We examined individual samples by SDS-PAGE followed by Western blotting with specific antibodies. The higher intensity of the lower bands of leucine-rich alpha-2-glycoprotein in lung cancer was consistent with the results of multi-dimensional chromatography followed by 2-D-PAGE (Fig. 6A). The intensity of the upper bands of leucine-rich alpha-2-glycoprotein did not differ between the two groups and the spots corresponding to these bands were not recognized as those showing aberrant intensity in lung cancer (Fig. 6A). The higher intensity of haptoglobulin bands was also consistent with the results of 2-D-PAGE (Fig. 6D). However, the intensities of the bands of inter-alpha-trypsin inhibitor heavy chain H4 (Fig. 6B) and plasma retinol binding protein (Fig. 6C) were also higher in lung cancer, and these results were inconsistent with those from multi-

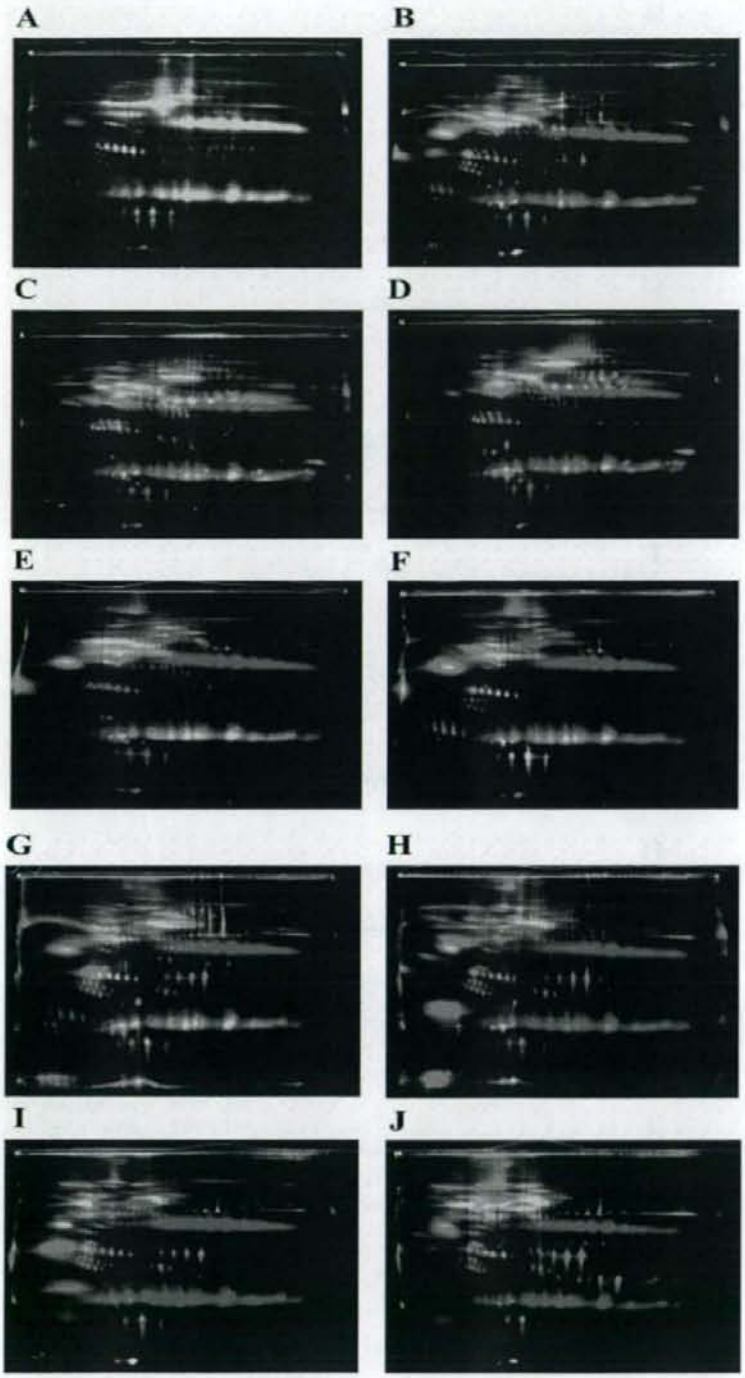


Figure 3. 2-D-PAGE of plasma proteins. Whole serum and fractionated proteins were labeled with Cy3 (red) and Cy5 (green), respectively. The labeled protein samples were mixed together and co-separated by 2-D-PAGE. The fractionated proteins showed distinguishable protein profiles, suggesting the efficient enrichment of low abundance proteins by fractionation. The sources of the protein samples are as follows: (A) bound fraction from the immuno-affinity column; (B) flow-through fraction from the immuno-affinity column; (C–J) 2-D images of the fractions eluted with 0, 100, 150, 200, 250, 300, 350 and 1000 mM NaCl from the anion-exchange column.

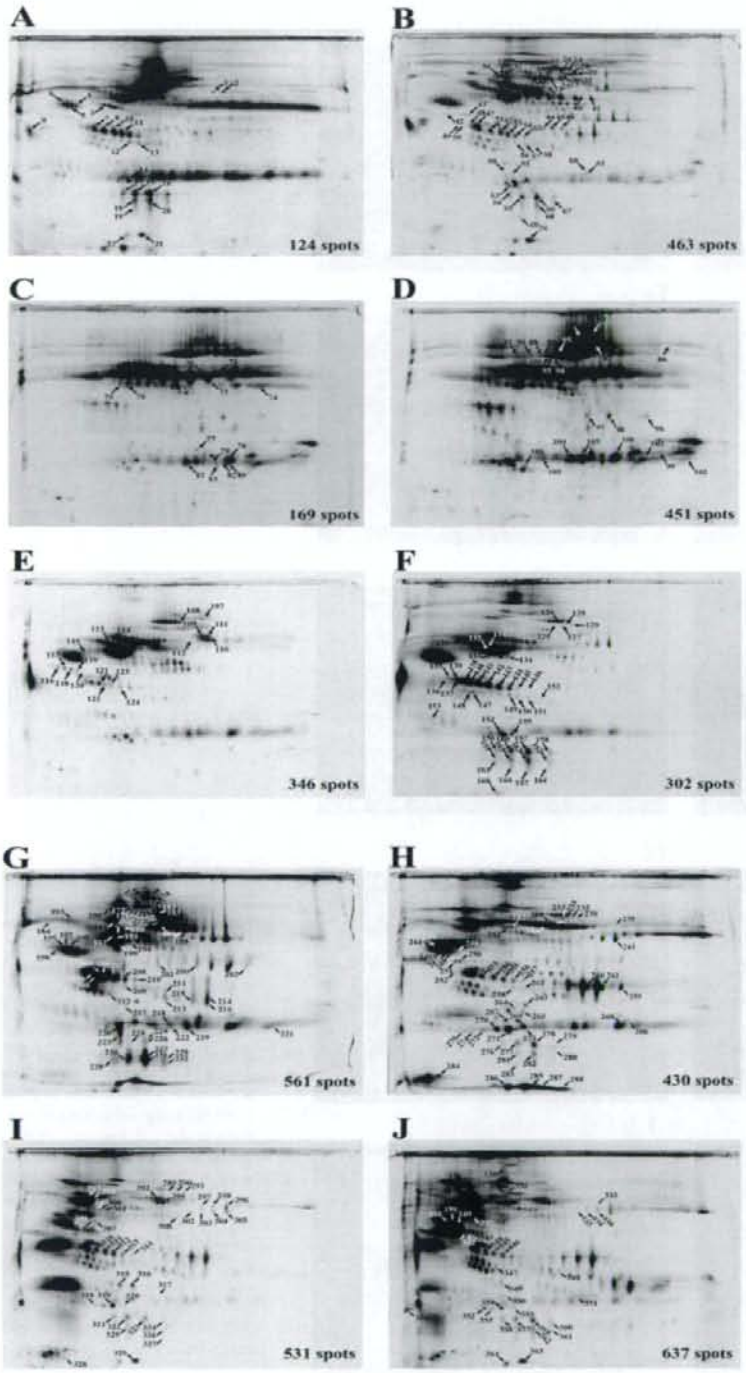


Figure 4. Localization of spots on the 2-D map of Cy5-labeled proteins. The arrows indicate the spots showing aberrant expression in lung cancer. The spot numbers correspond to those in Suppl. Table 1. The sources of the protein samples are as follows: (A) unfractionated serum; (B) flow-through fraction from the immuno-affinity column; (C–J) 2-D images of the fractions eluted with 0, 100, 150, 200, 250, 300, 350 and 1000 mM NaCl from the anion-exchange column.

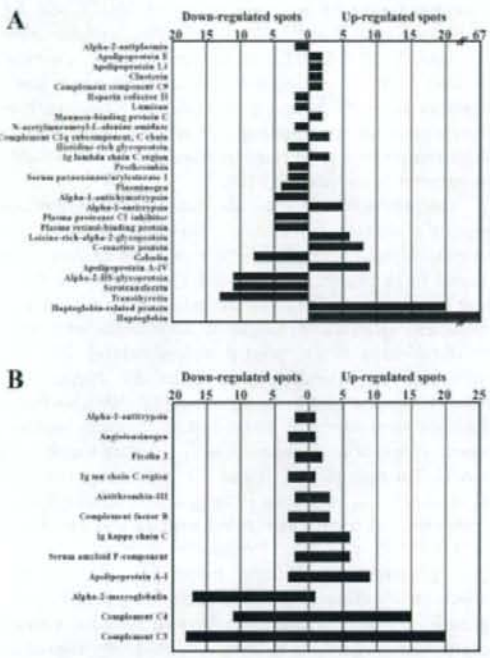


Figure 5. Analysis of spots showing aberrant expression levels in lung cancer. (A) Proteins for which the corresponding spots showed consistent differences in lung cancer; (B) proteins for which the corresponding spots showed inconsistent differences in lung cancer. X-axis representss the number of protein spots identified as those showing a significantly aberrant expression level in lung cancer. The 18 proteins identified as single spots were described in Section 3.

dimensional chromatography followed by 2-D-PAGE. We also studied the expression of complement component C4, whose isoforms showed differential expression. Complement component C4 consists of three subunits with molecular masses of 97, 75 and 33 kDa. The antibody we used visualized the 97-kDa band (Fig. 6E). On 2-D gel images, the expression level of the C4 spots of molecular mass approximately 97 kDa was increased in lung cancer. Western blotting identified increased expression of complement component C3, with a molecular mass of 45 kDa, though the C3 spots with this molecular weight showed both up- and down-regulated intensity in lung cancer (Fig. 6F). These results may indicate that complement component C3 is up-regulated as a whole in lung cancer and that 2-D-DIGE detects minor isoforms that are regulated in a different way from the major (abundant) isoforms. Western blotting also revealed an increased expression of prothrombin, with a molecular mass of 50 kDa, in lung cancer, and this result was inconsistent with the data obtained by 2-D-DIGE (Fig. 6G). Western blotting demonstrated increased expression of pro-

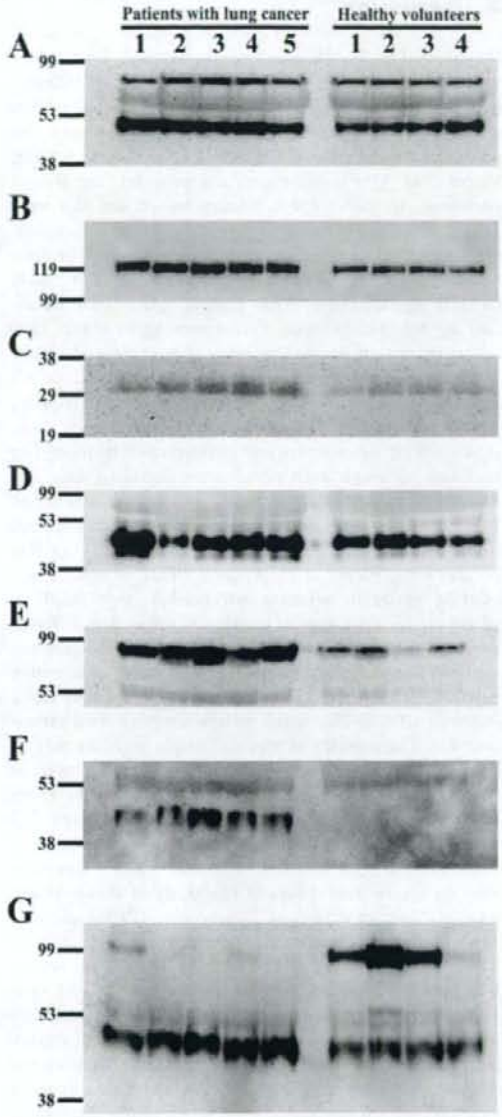


Figure 6. Validation of the differential expression of plasma proteins between individuals. Western blotting using (A) anti-leucine-rich alpha 2 glycoprotein, (B) anti-inter-alpha-trypsin inhibitor heavy chain H4, (C) anti-plasma retinol-binding protein, (D) anti-haptoglobin, (E) anti-complement component C4, (F) anti-complement component C3, and (G) anti-prothrombin antibody.

thrombin with a molecular mass of 99 kDa in three of four healthy volunteers (Fig. 6G), and this increase was not observed by 2-D-DIGE (Fig. 4).

4 Discussion

We applied a highly sensitive fluorescent dye-labeling technique to multi-dimensional chromatography followed by 2-D-PAGE for lung cancer plasma proteomics. In this study, the proteins were separated in three dimensions according to their net charge, p*I* and molecular weight. Pieper *et al.* [4] reported a similar approach to the plasma proteome. In their report, human serum samples were separated into 74 fractions by immuno-affinity, anion-exchange, and size-exclusion chromatography. The fractionated samples were then subjected to 2-D-PAGE, yielding approximately 3700 protein spots. They identified by MS 325 proteins corresponding to about 1800 spots. In our study, we observed 3890 spots and identified 58 proteins corresponding to 364 spots. As we focused on proteins with aberrant expression in lung cancer and did not perform global protein identification, it is difficult to compare the performance of these two methods, although most proteins we identified were also identified in [4]. However, our use of highly sensitive fluorescent dyes has several advantages over the method used in this report, as exemplified by the following four points. First, Pieper *et al.* [4] used 20 mL of serum as a starting material, whereas we needed only 30 μL of plasma to obtain a similar number of spots. The different amount of plasma protein required for the analysis was probably a result of the sensitivity of the spot detection method, as Pieper *et al.* visualized the proteins by staining with CBB G-250, which is less sensitive than saturation dye. The amount of plasma sample available may be limited, and a method requiring less sample may be more suitable for clinical proteomics. Alternatively, our protocol may have the potential to visualize even less abundant plasma proteins by increasing the sample volume. Secondly, gel staining with CBB G-250 is time-consuming (more than 3 days in the study of Pieper *et al.*), whereas spot detection of fluorescence-labeled proteins can be completed within 30–60 min by a laser scanner. This feature is especially advantageous because several 2-D gels are required for each of the fractionated samples. Thirdly, the fluorescent dye-labeling method may allow more quantitative and reproducible protein expression profiling by running an internal control sample and generating multiplex images. Forthly, the linear dynamic range of protein expression level measured as fluorescent intensity in 2-D-DIGE is wider than that in conventional 2-D-PAGE based on a colorimetric method such as CBB G-250.

Wang *et al.* [17] reported the utility of the fluorescent dye-labeling technique in the linkage of IEF, LC and SDS-PAGE for the study of intact plasma proteins. They identified plasma proteins associated with acute graft-versus-host disease. In their protocol, the highly abundant plasma proteins were depleted with an immuno-affinity column. Low abundance proteins were then labeled with fluorescent dyes, mixed, and subjected to IEF and RP chromatography. Finally, the fractionated proteins were separated by SDS-PAGE and detected by laser scanning. To label the proteins, they used minimal dye, which labels lysine residues [9]. Although minimal dye is less sensitive than saturation dye, minimal dye is advantageous when three samples are to be compared because Cy2 is available in addition to Cy3 and Cy5 [18].

The results of this study are consistent with previous reports of aberrant expression of plasma proteins in lung cancer. Alpha-1-acid glycoprotein in plasma has been shown to be highly sensitive and specific for the detection of lung cancer [21], and its levels are correlated with treatment effects and prognosis in patients with non-small-cell lung cancer treated with docetaxel [22]. The utility of retinol-binding protein for the diagnosis of lung cancer has also been described [23]. We also found aberrant expression of tissue-leakage plasma proteins whose expression in tumor cells is correlated with lung cancer. For example, we found that the expression level of clusterin was increased in lung cancer. Clusterin is a component of membrane-coated vesicles and has a protective function in cells. Previous studies have revealed that inhibition of clusterin expression enhances the effects of paclitaxel or gemcitabine in delaying the growth of A549 tumors [24]. We also found decreased expression of gelsolin in lung cancer. Down-regulation of gelsolin in tumor cells was correlated with poor survival of patients with non-small-cell lung cancer [25]. Although the pathophysiological significance of these tissue-leakage proteins is largely obscure, their clinical application as candidate tumor markers should be considered.

In this study, we did not identify proteins showing very low abundance such as growth factors, suggesting that further development of the current protocol is required for a comprehensive understanding of the lung cancer proteome. Block *et al.* [26] reported that they observed low abundance tissue-leakage proteins as spots on 2-D images when they fractionated plasma samples with a lectin-affinity column. Thus, additional variations of multi-dimensional separation including multiple specific immuno-affinity columns will facilitate the study of lower abundance plasma proteins.

The proteins retained in the immuno-affinity column were not limited to the six target plasma proteins (Fig. 2). Albumin is known to interact with a variety of proteins [27], and as the proteins were separated as intact forms in our study, depletion of albumin may cause the removal of unexpected proteins. We also identified alpha-1-anti-trypsin, haptoglobin, and transferrin in the fractionated samples as aberrantly regulated plasma proteins in lung cancer, even though our procedure was designed to deplete these proteins at the initial step of fractionation by use of the immuno-affinity column. Together, these observations suggest that the immuno-depletion procedure requires

further optimization. For example, it may be possible to dissociate protein complexes prior to the immuno-depletion procedure so that albumin-bound proteins are recovered in the flow-through fraction. For better depletion of target proteins, the flow-through fraction could be applied repeatedly to the immuno-affinity column or two immuno-affinity columns could be connected in tandem. These variations are compatible with our current protocol based on 2-D-DIGE. An alternative explanation of the failure of the immuno-affinity procedure to completely deplete the target plasma proteins is that these proteins display aberrant features in lung cancer and are unable to bind to an immuno-affinity column containing antibodies raised against normal plasma proteins. These issues should be considered further when immuno-affinity procedures are applied to disease proteomics.

We found isoform-specific aberrations of plasma protein expression in lung cancer: the protein spots corresponding to 12 gene products were both up- and down-regulated. In contrast, the spots of 28 proteins showed consistent up- or down-regulation. Because the spots of haptoglobin, which showed the largest identified number of spots from a single gene product in our study, were consistently up-regulated in lung cancer, a high number of spots does not always result in inconsistent regulation, suggesting isoform-specific regulation of certain proteins. These isoform-specific aberrations may not be identified by analysis of peptide subsets from complex digests as accomplished using multidimensional protein identification technology, suggesting the importance of analysis of intact plasma proteins. Although isoform-specific alterations could be a potential source of biomarkers, it might be difficult to use such isoforms as biomarkers in a clinical setting if their monitoring requires time-consuming proteomic technology such as the combination of multidimensional chromatography and 2-D-PAGE. More conventional and high-throughput methods will be required for clinical application. Although use of a specific antibody is always an early consideration when developing a high-throughput method, we demonstrated that the results of Western blotting following SDS-PAGE did not always match those from 2-D-PAGE. This discrepancy probably arises from the facts that the antibodies used were not isoform specific and the samples were prepared differently. To solve this problem, a novel technique such as the use of DNA aptamers [28] may be required to generate molecules with high affinity for each protein variant. This issue is a generic problem using 2-D-PAGE, in which various isoforms are visualized, quantified and selected as candidate biomarkers. However, this does not mean that a 2-D-PAGE approach is impractical. We believe that 2-D-PAGE is a powerful tool for providing detailed proteomic information that cannot be obtained from other proteomic methods, and that novel technologies are required to utilize the output of 2-D-PAGE for clinical application.

We are grateful to Dr. Teruhiko Yoshida for his collaboration and to Kano Nishiyama for technical assistance. This work was supported by a grant from the Ministry of Health, Labor and Welfare and by the Program for Promotion of Fundamental Studies at the National Institute of Biomedical Innovation of Japan. Tetsuya Okano and Tatsuhiko Kakisaka are awardees of Research Resident Fellowships from the Foundation for Promotion of Cancer Research (Japan) for the 3rd Term Comprehensive 10-Year Strategy for Cancer Control.

5 References

[1] Hoffman, P. C., Mauer, A. M., Vokes, E. E., *Lancet* 2000, *355*, 479–485.

[2] Fry, W. A., Phillips, J. L., Menck, H. R., *Cancer* 1999, *86*, 1867–1876.

[3] Tarro, G., Perna, A., Esposito, C., *J. Cell Physiol.* 2005, *203*, 1–5.

[4] Pieper, R., Gatlin, C. L., Makusky, A. J., Russo, P. S. *et al.*, *Proteomics* 2003, *3*, 1345–1364.

[5] Tonge, R., Shaw, J., Middleton, B., Rowlinson, R. *et al.*, *Proteomics* 2001, *1*, 377–396.

[6] Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M. J. *et al.*, *Mol. Cell. Proteomics* 2002, *1*, 91–98.

[7] Unlu, M., Morgan, M. E., Minden, J. S., *Electrophoresis* 1997, *18*, 2071–2077.

[8] Chen, J. H., Chang, Y. W., Yao, C. W., Chiueh, T. S. *et al.*, *Proc. Natl. Acad. Sci. USA* 2004, *101*, 17039–17044.

[9] Shaw, J., Rowlinson, R., Nickson, J., Stone, T. *et al.*, *Proteomics* 2003, *3*, 1181–1195.

[10] Fujii, K., Kondo, T., Yokoo, H., Yamada, T. *et al.*, *Proteomics* 2005, *5*, 1411–1422.

[11] Kondo, T., Seike, M., Mori, Y., Fujii, K. *et al.*, *Proteomics* 2003, *3*, 1758–1766.

[12] Sitek, B., Luttges, J., Marcus, K., Kloppel, G. *et al.*, *Proteomics* 2005, *5*, 2665–2679.

[13] Greengauz-Roberts, O., Stoppler, H., Nomura, S., Yamaguchi, H. *et al.*, *Proteomics* 2005, *5*, 1746–1757.

[14] Seike, M., Kondo, T., Fujii, K., Okano, T. *et al.*, *Proteomics* 2005, *5*, 2939–2948.

[15] Fujii, K., Nakano, T., Kanazawa, M., Akimoto, S. *et al.*, *Proteomics* 2005, *5*, 1150–1159.

[16] Pieper, R., Su, Q., Gatlin, C. L., Huang, S. T. *et al.*, *Proteomics* 2003, *3*, 422–432.

[17] Wang, H., Clouthier, S. G., Galchev, V., Misek, D. E. *et al.*, *Mol. Cell. Proteomics* 2005, *4*, 618–625.

[18] Misek, D. E., Kuick, R., Wang, H., Galchev, V. *et al.*, *Proteomics* 2005, *5*, 3343–3352.

[19] Bjorhall, K., Miliotis, T., Davidsson, P., *Proteomics* 2005, *5*, 307–317.

[20] Campostrini, N., Areces, L. B., Rappsilber, J., Pietrogrande, M. C. *et al.*, *Proteomics* 2005, *5*, 2385–2395.

[21] Ganz, P. A., Baras, M., Ma, P. Y., Elashoff, R. M., *Cancer Res.* 1984, *44*, 5415–5421.

[22] Bruno, R., Olivares, R., Berille, J., Chaikin, P. *et al.*, *Clin. Cancer Res.* 2003, *9*, 1077–1082.

[23] Edes, T. E., McDonald, P. S., *Cancer Detect. Prev.* 1991, *15*, 341–344.

[24] July, L. V., Beraldi, E., So, A., Fazli, L. *et al.*, *Mol. Cancer Ther.* 2004, *3*, 223–232.

[25] Yang, J., Tan, D., Asch, H. L., Swede, H. *et al.*, *Lung Cancer* 2004, *46*, 29–42.

[26] Block, T. M., Comunale, M. A., Lowman, M., Steel, L. F. *et al.*, *Proc. Natl. Acad. Sci. USA* 2005, *102*, 779–784.

[27] Mehta, A. I., Ross, S., Lowenthal, M. S., Fusaro, V. *et al.*, *Dis. Markers* 2003, *19*, 1–10.

[28] Fang, X., Sen, A., Vicens, M., Tan, W., *Chembiochem.* 2003, *4*, 829–834.

[29] Travis, W., Colby, T., Corrn, B., Shimosato, Y., Brambilla, E., *Histological Typing of Tumours of Lung and Pleura: World Health Organization International Classification of Tumors*, Springer Verlag, New York 1999.

[30] Sobin L. H., W. C., *UICC: TNM Classification of Malignant Tumours*, John Wiley & Sons, New York 1997, pp. 91–100.

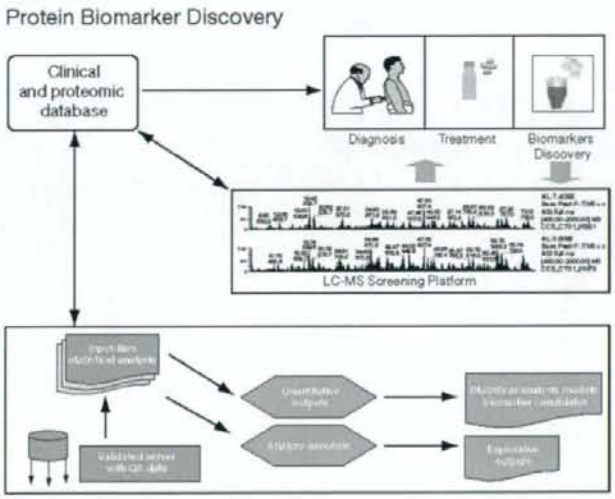
Journal of proteome research

Perspectives

Personalized Medicine and Proteomics: Lessons from Non-Small Cell Lung Cancer

Gyrgy Marko-Varga, Atsushi Ogiwara, Toshihide Nishimura, Takeshi Kawamura, Kiyonaga Fujii, Takao Kawakami, Yutaka Kyono, Hsiao-kun Tu, Hisae Anyoji, Mitsuhiro Kanazawa, Shingo Akimoto, Takashi Hirano, Masahiro Tsuboi, Kazuto Nishio, Shuji Hada, Haiyi Jiang, Masahiro Fukuoka, Kouichiro Nakata, Yutaka Nishiwaki, Hideo Kunito, Ian S. Peers, Chris G. Harbron, Marie C. South, Tim Higenbottam, Fredrik Nyberg, Shoji Kudoh, and Harubumi Kato

J. Proteome Res., **2007**, 6 (8), 2925-2935• DOI: 10.1021/pr070046s • Publication Date (Web): 17 July 2007



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 3 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article