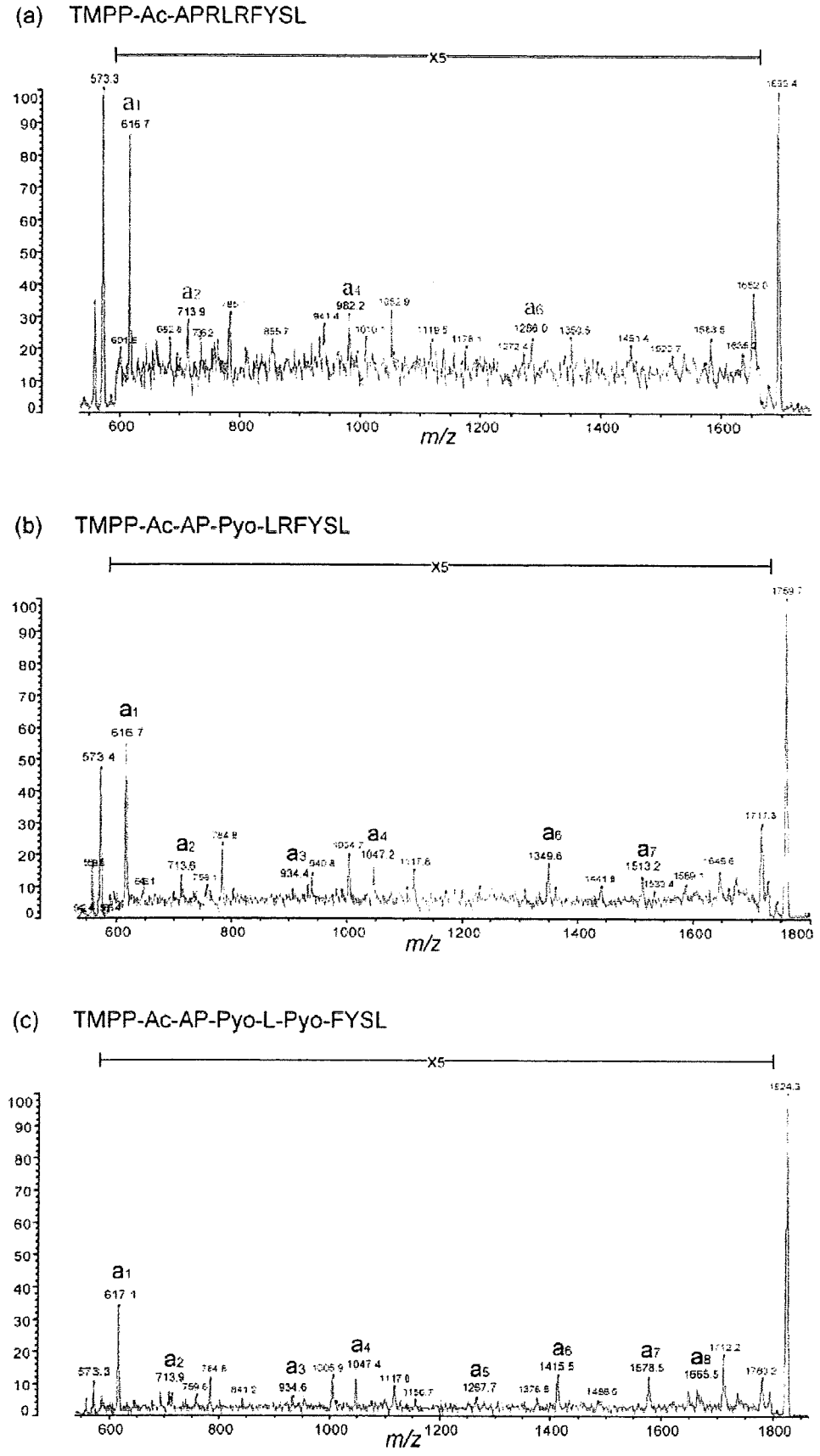


**Figure 6.** Enhancing effect of arginine derivatization using a peptide incorporating two arginine residues (TMPP-Ac- APRLRFYSL): (a) TMPP-Ac peptide with two free arginines; (b) TMPP-Ac peptide with one of the two arginines derivatized with acetylacetone; and (c) TMPP-Ac peptide with both arginines derivatized with acetylacetone.





DOI: 10.1002/rcm

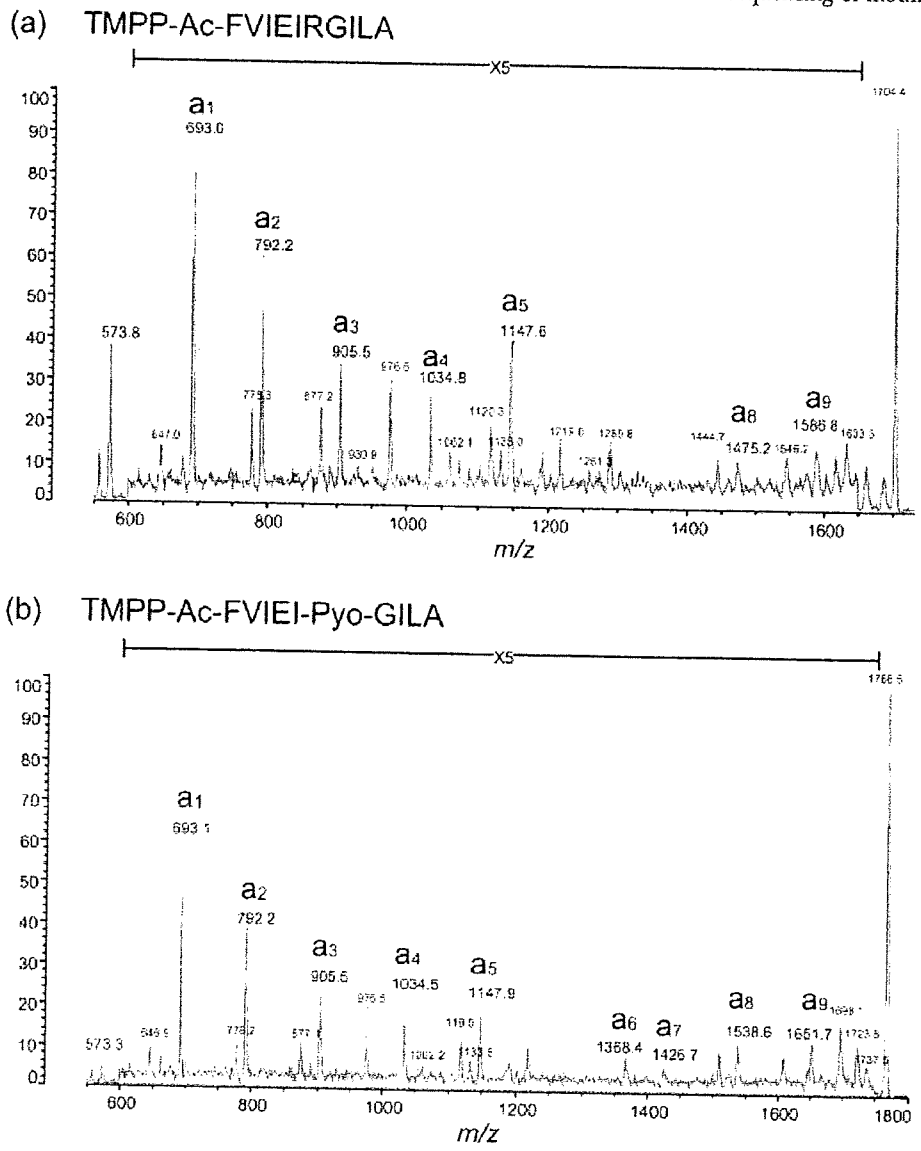


**Figure 7.** Enhancing effect of arginine derivatization using a C-terminal peptide (TMPP-Ac-FVIEIRGILA) isolated from pig albumin. This peptide incorporates an arginine residue in an internal position. (a) Non-derivatized peptide and (b) TMPP-Ac peptide after derivatization of Arg with acetylacetone.

starting peptide (TMPP-Ac-APRLRFYSL) almost disappeared, and the signal of the doubly derivatized peptide slightly exceeded that of the singly derivatized peptide in intensity (Fig. 5(c)). The CID spectrum clearly reveals all a-type ions, the intensities of which are enough for sequencing (Fig. 6(c)).

Selective generation of a-type ions under MS/MS analysis of TMPP-Ac-modified peptides was discussed and the mechanism was illustrated as 'amide H shifts',[7] by which low intensity of ions containing a proline residue at their C-termini (Fig. 2(c)) can be explained.

The MS/MS analyses of the four types of arginine-containing peptides studied indicate that arginine-incorporated fragment ions are reduced in their intensity or not detected. This might be because arginine-containing ions are unstable and prone to undergo secondary reactions during the fragmentation process.[15,22] The detailed investigation of the mechanistic pathway to, or the rationalization of, this phenomenon is outside the scope of this paper and not addressed here.

The studied peptides categorized as having four arginine distributions in their sequences exhibited significant enhancement of structurally informative fragmentation after derivatization of arginine residue(s), except for the peptides having C-terminal arginine residues. Without derivatization, a couple of peptides produced no informative fragmentation (Table 1). It should be noted that all derivatized peptides fragmented well for sequencing (Table 1).




DOI: 10.1002/rcm

## CONCLUSIONS

We have applied the arginine derivatization method using acetylacetone to TMPP-Ac-modified peptides to improve fragmentation. Four types of arginine-containing peptides were derivatized and subjected to MALDI-CID analyses. Although the derivatization did not contribute much to the enhancement of fragmentation of peptides with C-terminal arginine residues, all the derivatized peptides studied produced a complete series of a-type ions as revealed by MALDI-CID analysis.

## REFERENCES

1. Aebersold R, Mann M. *Nature* 2003; **422**: 198.
2. Loboda AV, Krutchinsky AN, Bromirski M, Ens W, Standing KG. *Rapid Commun. Mass Spectrom.* 2000; **14**: 1047.
3. Keough T, Youngquist RS, Lacey MP. *Proc. Natl. Acad. Sci. USA* 1999; **96**: 7131.
4. Keough T, Youngquist RS, Lacey MP. *Anal. Chem.* 2003; **75**: 156A.
5. Marekov LN, Steinert PM. *J. Mass Spectrom.* 2003; **38**: 373.
6. Huang ZH, Wu J, Roth KDW, Yang Y, Gage DA, Watson JT. *Anal. Chem.* 1997; **69**: 137.
7. Liao PC, Huang ZH, Allison J. *J. Am. Soc. Mass Spectrom.* 1997; **8**: 501.
8. Huang ZH, Shen T, Wu J, Gage DA, Watson JT. *Anal. Biochem.* 1999; **268**: 305.
9. Shen TL, Allison J. *J. Am. Soc. Mass Spectrom.* 2000; **11**: 145.
10. Pashkova A, Chen HS, Rejtar T, Zang X, Giese R, Andreev V, Moskovets E, Karger BL. *Anal. Chem.* 2005; **77**: 2085.
11. Peters EC, Horn DM, Tully DC, Brock A. *Rapid Commun. Mass Spectrom.* 2001; **15**: 2387.
12. Chen W, Lee PJ, Shion H, Ellor N, Gebler JC. *Anal. Chem.* 2007; **79**: 1583.
13. Dongre AR, Jones JL, Somogyi A, Wysocki VH. *J. Am. Chem. Soc.* 1996; **118**: 8365.
14. Dikler S, Kelly JW, Russell DH. *J. Mass Spectrom.* 1997; **32**: 1337.
15. Spengler B, Luetzenkirchen F, Metzger S, Chaurand P, Kaufmann R, Jeffery W, Bartlet-Jones M, Pappin DJC. *Int. J. Mass Spectrom. Ion Processes* 1997; **169/170**: 127.
16. Leitner A, Lindner W. *Anal. Chim. Acta* 2005; **528**: 165.
17. Foettinger A, Leitner A, Lindner W. *J. Mass Spectrom.* 2006; **41**: 623.
18. Leitner A, Amon S, Rizzi A, Lindner W. *Rapid Commun. Mass Spectrom.* 2007; **21**: 1321.
19. Kuyama H, Shima K, Sonomura K, Yamaguchi M, Ando E, Nishimura O, Tsunasawa S. *Proteomics* 2008; **8**: 1539.
20. Kuyama H, Sonomura K, Nishimura O. *Rapid Commun. Mass Spectrom.* 2008; **22**: 1109.
21. Zhang J, Zhang L, Zhou Y, Guo YL. *J. Mass Spectrom.* 2007; **42**: 1514.
22. Thorne GC, Ballard KD, Gaskell SJ. *J. Am. Soc. Mass Spectrom.* 1990; **1**: 249.




DOI: 10.1002/rcm

DOI 10.1002/pmic.200700619

*Proteomics* 2008, *8*, 3194–3203

RESEARCH ARTICLE

# Quantitative proteomic analysis to discover potential diagnostic markers and therapeutic targets in human renal cell carcinoma

*Noboru Okamura[1], Taro Masuda[2], Akinobu Gotoh[3], Toshiro Shirakawa[4], Shuji Terao[3], Naoki Kaneko[2, 5], Kazuki Suganuma[2], Makoto Watanabe[2, 5], Toshiya Matsubara[2, 5], Ryota Seto[6], Jun Matsumoto[6], Megumi Kawakami[6], Motohiro Yamamori[6], Tsutomu Nakamura[1, 6], Tatsurou Yagami[7], Toshiyuki Sakaeda[6, 8], Masato Fujisawa[4], Osamu Nishimura[2, 5] and Katsuhiko Okumura[1, 6, 7]*

[1] Department of Clinical Evaluation of Pharmacotherapy, Kobe University Graduate School of Medicine, Kobe, Japan
[2] Life Science Laboratory, Shimadzu Corporation, Kyoto, Japan
[3] Institute for Advanced Medical Sciences, Hyogo College of Medicine, Nishinomiya, Japan
[4] Division of Urology, Department of Organ Therapeutics, Kobe University Graduate School of Medicine, Kobe, Japan
[5] Division of Disease Proteomics, Institute for Protein Research, Osaka University, Osaka, Japan
[6] Department of Hospital Pharmacy, School of Medicine, Kobe University, Kobe, Japan
[7] Faculty of Pharmaceutical Sciences, Himeji Dokkyo University, Himeji, Japan
[8] Frontier Education Center, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto, Japan

Renal cell carcinoma (RCC) is relatively resistant to chemotherapy and radiotherapy. Recent advances in drug development are providing novel agents for the treatment of RCC, but the effects are still minimal. In addition, there is an urgent need to identify diagnostic markers for RCC. In this report, to discover potential diagnostic markers and therapeutic targets, we subjected RCC samples to a quantitative proteomic analysis utilizing 2-nitrobenzenesulfenyl (NBS) reagent. Proteins were extracted from RCC and adjacent normal tissue, obtained surgically from patients, and labeled with NBS reagent containing six $^{12}C$ or $^{13}C$. This was followed by trypsin digestion and the enrichment of labeled peptides. Samples were then subjected to analysis by MALDI-TOF MS. NBS-labeled peptides with a 6 Da difference were identified by MS/MS. Thirty-four proteins were upregulated in more than 60% of the patients of which some were previously known, and some were novel. The identity of a few proteins was confirmed by Western blotting and quantitative real time RT-PCR. The results suggest that NBS-based quantitative proteomic analysis is useful for discovering diagnostic markers and therapeutic targets for RCC.

Received: July 2, 2007
Revised: April 12, 2008
Accepted: April 14, 2008

**Keywords:**
Biomarker / NBS method / Renal cell carcinoma

**Correspondence:** Professor Katsuhiko Okumura, Department of Clinical Evaluation of Pharmacotherapy, Kobe University Graduate School of Medicine, 1-5-6 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan
**E-mail:** okumurak@med.kobe-u.ac.jp
**Fax:** +81-78-304-6047

**Abbreviations: CNDP2,** cytosolic nonspecific dipeptidase 2; **NBS,** 2-nitrobenzenesulfenyl; **RCC,** renal cell carcinoma

## 1 Introduction

Renal cell carcinoma (RCC) can be resistant to chemotherapy and radiotherapy, making nephrectomy the most effective treatment option [1]. Although nephrectomy can be expected to cure patients without metastatic involvement, the five-year survival rate for those with metastasis is less than 10%. Thus, there is an urgent need to discover diagnostic markers for the early detection of RCC. In

addition, recent advances in drug development have led to the approval of novel agents for RCC targeting cancer-specific pathways such as bevacizumab, sorafenib, and sunitinib. Although these drugs prolonged overall survival in patients with advanced RCC [2], the effect is not sufficient. Thus, it is therefore necessary to discover novel targets for the treatment of RCC.

Recent advances in proteomics have provided opportunities to identify new diagnostic or predictive biomarkers. In addition, proteomics is expected to be a powerful tool for finding novel therapeutic targets. Until now, two major proteomic technologies have been used in the clinical setting; 2-DE combined with MS and SELDI-TOF MS. Quantitative analysis of proteins has been achieved using a combination of protein separation with high-resolution 2-DE and MS. SELDI-TOF MS is a newly developed technique for finding biomarkers and has high-throughput ability. However, there are several technical limitations to these proteomic approaches. In 2-DE, it is difficult to detect proteins that are small (<10 kDa), large (>150 kDa), very basic (or acidic), and hydrophobic [3–5]. In SELDI-TOF MS, it is difficult to detect large proteins and identify them directly. Thus, it is clear that other proteomic tools are needed to improve proteome mining. For the relative quantification of protein abundance, the isotope-coded affinity tag (ICAT) method [6] allows for highly accurate quantitative protein profiling.

Recently, we have developed and improved a quantitative proteomic analysis using 2-nitrobenezenefulfenyl (NBS) reagent, labeled with $^{12}C_6$ (light) or $^{13}C_6$ (heavy) [7–9]. Selective introduction of the NBS moiety into tryptophan residues can be achieved and a 6 Da mass difference is generated. This method enables the relative quantification and identification of proteins in the same experiment. Its most prominent advantage is the reduction in the number of peptides to be measured by selecting NBS-labeled tryptophan-containing peptides from enzymatic digests. This is advantageous since tryptophan residues are the least abundant amino acid in proteins, yet they occur in a large proportion of proteins [10]. Thus, this method is expected to sensitively detect differences in protein expression with only small samples by increasing the dynamic range of detection. The good reproducibility and reliability of the NBS method have been shown previously [7, 8, 11, 12]. Thus, the method has been demonstrated as promising for nongel-based quantitative profiling. Here, we subjected RCC samples to this NBS-based proteomics analysis to find potential biomarkers and therapeutic targets.

## 2 Materials and methods

### 2.1 Clinical samples

Samples of RCC and adjacent normal tissue were obtained surgically from 29 patients treated in Kobe University Hospital. Written informed consent was provided by all patients prior to enrollment. The patients were aged 29–81 (60 ± 12, mean ± SD) years and there were 18 males and 11 females. All samples were subjected to mRNA expression analysis and samples from 14 patients (seven males and seven females) were subjected proteomic analysis (Table 1). After resection, samples were separated into RCC and adjacent normal tissue, then frozen immediately and stored at −80°C. All RCC samples were clear cell type. The clinical characteristics of the subjects are listed in Table 1. The protocol was approved by the Institutional Review Board, Kobe University Hospital (Kobe, Japan).

**Table 1.** Profile of patients used for the analysis

| Patient ID | Gender | Age | Histological Subtype | pT class | Proteome Analysis | mRNA Expression |
|---|---|---|---|---|---|---|
| 3 | Male | 54 | Clear cell | pT3a | Yes | Yes |
| 9 | Male | 65 | Clear cell | pT3b | No | Yes |
| 10 | Female | 70 | Clear cell | pT3b | Yes | Yes |
| 11 | Male | 65 | Clear cell | pT3b | No | Yes |
| 12 | Male | 72 | Clear cell | pT1a | No | Yes |
| 13 | Female | 81 | Clear cell | pT3a | Yes | Yes |
| 14 | Female | 69 | Clear cell | pT1a | No | Yes |
| 15 | Female | 52 | Clear cell | pT3b | No | Yes |
| 16 | Male | 77 | Clear cell | pT3a | No | Yes |
| 18 | Male | 75 | Clear cell | pT1a | No | Yes |
| 19 | Male | 54 | Clear cell | pT1a | No | Yes |
| 20 | Male | 54 | Clear cell | pT3b | No | Yes |
| 21 | Female | 71 | Clear cell | pT1a | No | Yes |
| 23 | Female | 51 | Clear cell | pT3b | No | Yes |
| 25 | Male | 29 | Clear cell | pT1a | No | Yes |
| 27 | Male | 64 | Clear cell | pT1b | No | Yes |
| 28 | Female | 49 | Clear cell | pT3a | Yes | Yes |
| 29 | Male | 51 | Clear cell | pT1a | No | Yes |
| 31 | Male | 72 | Clear cell | pT3b | Yes | Yes |
| 33 | Male | 65 | Clear cell | pT1a | Yes | Yes |
| 35 | Male | 67 | Clear cell | pT1b | Yes | Yes |
| 36 | Male | 63 | Clear cell | pT3a | Yes | Yes |
| 37 | Male | 47 | Clear cell | pT1a | Yes | Yes |
| 38 | Male | 58 | Clear cell | pT3b | No | Yes |
| 39 | Female | 33 | Clear cell | pT3b | Yes | Yes |
| 41 | Female | 50 | Clear cell | pT1b | Yes | Yes |
| 42 | Female | 58 | Clear cell | pT1b | Yes | Yes |
| 43 | Male | 66 | Clear cell | pT1a | Yes | Yes |
| 44 | Female | 50 | Clear cell | pT3a | Yes | Yes |

### 2.2 Materials

$^{13}$CNBS stable isotope labeling kit-N was obtained from Shimadzu Corporation (Kyoto, Japan). The protease inhibitor cocktail and 2-D Cleanup Kit were obtained from Sigma–Aldrich (St. Louis, MO) and BioRad (Hercules, CA), respectively. Urea, thiourea, and CHAPS were purchased from GE Healthcare (Buckinghamshire, England). Other chemicals were obtained from Wako (Osaka, Japan).

### 2.3 Preparation of samples for MALDI-TOF MS

Approximately 20 mg amounts of RCC or adjacent normal tissue was homogenized using a mixer mill (MM300, Retch, Haan, Germany) in buffer A (50 mM Tris/HCl, 100 mM NaCl, 100 mM PMSF, and 1% protease inhibitor cocktail, adjusted to pH 8.8). The sample was then centrifuged at 20 000 × *g* for 15 min and the supernatant subjected to labeling with NBS reagents. After two washes with buffer A, the pellet was dissolved with buffer B (6 M urea, 2 M thiourea, 3% CHAPS, 1% Triton X-100, and 1% protease inhibitor cocktail). The supernatant and the pellets were cleaned up using the 2-D Cleanup Kit (BioRad) and the resultant pellets were dissolved with 6 M guanidine hydrochloride containing 5 mM EDTA. Protein concentrations were determined with the BCA Protein Assay Kit (Pierce, Lockford, IL).

NBS reagent labeling was performed according to the manufacturer's protocol. For 14 patient samples, proteins derived from RCC and normal tissue were labeled with the heavy and light forms, respectively, for 1 h at room temperature under darkness. The samples were then combined, reduced, alkylated, and digested with trypsin. NBS-labeled peptides were enriched from the tryptic digests and were fractionated using Phenyl Sepharose. The labeled peptides were eluted in a stepwise manner, with 1 mL of 0.1% TFA containing 10, 15, 20, 25, 30, 35, 40, and 50% ACN solutions, sequentially. The 10 to 15% eluted fractions were pooled and dried in a vacuum concentrator, and then were suspended in 20 μL of 0.1% TFA. The 20 and 25%, 30 and 35%, and 40 and 50% eluted fractions were treated in the same manner. These four pooled fractions were subjected to RP-LC (LC-10ADvp μHPLC System, Shimadzu), as described previously [11]. Eluates were automatically deposited onto MALDI target plates by the LC spotting system (AccuSpot, Shimadzu).

### 2.4 MALDI-TOF MS

Fractionated samples were automatically analyzed by MALDI-TOF MS (AXIMA-CFR Plus; Shimadzu/Kratos, Manchester, UK). Relative quantification of each NBS-tagged peptide pair was performed using TWIP Version 1.0 (Dynacom, Kobe, Japan) as previously described [11, 12]. In this study, peptide pair ratios >2.0-fold, or <0.5-fold, were set as threshold values for upregulation or downregulation, respectively [12]. Two hundred and twenty five paired peaks with common regulation in at least 60% of patients were selected and subjected to MS/MS analysis (AXIMA-QIT-TOF; Shimadzu/Kratos). Peak lists were created from the generated MS/MS spectra by MASCOT Distiller software Version 1.1.2 (Matrix Science, Boston, MA) and analyzed using MASCOT software Version 2.0 (Matrix Science). The MASCOT search parameters were following: trypsin digestion allowing up to one missed cleavage, fixed modifications of 12 CNBS (or 13 CNBS) and carbamidomethyl (C), variable modifications of oxidation (M), peptide tolerance of 0.3 Da, and MS/MS tolerance of 0.5 Da. Search results having *p*-values <0.05 were judged as positive identifications as described previously [12].

### 2.5 Western blot analysis

The obtained proteomics data were confirmed by Western blotting. Approximately 20 mg of RCC or adjacent normal tissue was homogenized as described earlier in the text. Then, each sample was centrifuged at 20 000 × *g* for 15 min and a volume of supernatant containing 20 μg of total protein was diluted two-fold in loading sample buffer (0.125 M Tris-HCl (pH 6.8), 10% 2-mercaptoethanol, 4% SDS, 10% sucrose, and 0.004% Bromophenol blue), incubated at 95°C for 5 min, and loaded onto the gel. Proteins were separated on 12.5% SDS-polyacrylamide gels and then transferred to PVDF membrane (GE Healthcare). Prestained SDS-PAGE standards were used to estimate molecular weights (BioRad). The utilized primary antibodies were rabbit anti-human galectin-1 (Abcam, Cambridge, UK) and cytosolic nonspecific dipeptidase 2 (CNDP2) polyclonal antibody. Anti-human CNDP2 antiserum was raised against the synthetic peptide CKFAELRSPNEFKV, residues 364–376 of the CNDP2 dipeptidase sequence registered in GenBank (accession no. AAH03176). Each PVDF membrane was incubated with diluted galectin-1 and CNDP2 antibody solution for 2 h at room temperature. The probing of blots with these antibodies was carried out according to the manufacturer's protocol (Vectastain ABC kit; Vector Laboratories, Burlingame, CA). Immunostain HRP-1000 (Konica, Tokyo, Japan) was used for visualizing the signal. In addition, used membranes were incubated at room temperature for 1 h with mouse anti-human β-actin polyclonal antibody (Abcam). After that, the membranes were incubated at room temperature for 1 h with HRP-conjugated sheep anti-mouse IgG antibody (GE Healthcare). Protein was then visualized by use of ECL Plus detection reagents (GE Healthcare). The antibodies dilutions were: anti-galectin-1 (1:200); anti-CNDP2 (1:500); anti-β-actin (1:10 000); and anti-mouse IgG (1:5000).

### 2.6 Quantitative real time RT-PCR

The level of mRNA for galectin-1 was measured by the quantitative real time RT-PCR method as described previously [13, 14]. Forward primers, reverse primers and TaqMan probes are listed in Table 2. Concentrations of mRNA are expressed relative to the level of β-actin. Even though Ferugson *et al.* [15] showed the higher expression of β-actin in RCC, it was marginal. Statistical significance of differences between mean values was calculated using the paired *t*-test and a $p<0.05$ (two-tailed) was considered to be significant.

**Table 2.** Forward and reverse primers, and TaqMan probes used for quantitative real time RT-PCR

| Target gene | Accession no | Primers | Probes |
|---|---|---|---|
| Galectin-1 | NM_002305 | Forward AGGCTGTCTTTCCCTTCCA<br>Reverse AGGTTGGCCTGGTCGAA | CCTGGAAGTGTTGCAGAGGTGTGC |
| β-Actin | NM_001101 | Forward TCGTCATACTCCTGCTTGCTGAT<br>Reverse GGCACCCAGCACAATGAAG | AGTACTTGCGCTCAGGAGGAGCAATGATC |

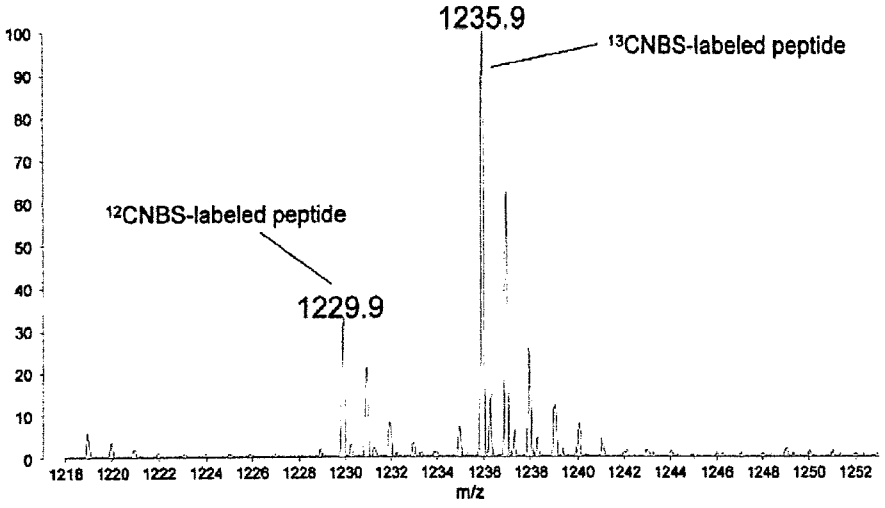


**Figure 1.** Typical MS spectrum from a proteomic analysis of protein extracts from RCC and adjacent normal tissue. Proteins from RCC and normal tissue were labeled with heavy ($^{13}$C) and light ($^{12}$C) NBS, respectively.

## 3 Results

### 3.1 Identification of proteins

To identify the potential diagnostic markers and therapeutic targets, we subjected RCC samples to an NBS-based quantitative proteomic analysis. Approximately 6000–7000 peptide pairs were detected in each sample. The distribution of peptide pair ratios in each sample is shown in Supporting Information. More than half of the detected pair peaks were within the <2-fold difference ratio range (>0.5, <2.0) between RCC and adjacent normal renal tissues. A total of 225 peptides exhibited at least a two-fold difference in over 60% of the patients; these peptides were subjected to MS/MS analysis to identify proteins (Table 3). As a representative example of proteins upregulated in RCC, Fig. 1 shows a typical MS spectrum of an NBS-tagged peptide pair derived from galectin-1 in RCC and adjacent normal renal tissues. The opposite labeling (light for RCC and heavy for normal tissue) had no effect on the spectra (data not shown).

### 3.2 Western blot analysis

Quantitative proteomic analysis revealed the significant upregulation of galectin-1 and CNDP2 in RCC (Table 3). To confirm these results, we further examined the difference in the levels of these protein expressions between RCC and adjacent normal renal tissues by Western blotting. The expression levels of galectin-1 and CNDP2 proteins were confirmed to be higher in RCC (Fig. 2).

### 3.3 Quantitative real time RT-PCR analysis

The mRNA expression of galectin-1 was measured by quantitative real time RT-PCR method. A significantly higher level of galectin-1 was observed in RCC compared with normal renal tissue (Fig. 3).

## 4 Discussion

Surgical resection is still a common treatment for RCC which is known to be resistant to chemotherapy and radiotherapy. In its early stages, RCC has a good prognosis, but in the advanced stage with distant metastasis, the five-year survival rate for RCC is less than 10% [1]. Thus, there is an urgent need to discover potential diagnostic markers for RCC. Recent advances in the molecular study of cancerous pathways have led to the launch of novel drugs for the treatment of RCC, including anti-VEGF humanized antibody, bevacizumab and multikinase inhibitors, sorafenib, sunitinib, which have been approved by the FDA for the

 

Table 3. Differentially expressed proteins in RCC patients

| Protein name | Swiss-Prot accession no. | Peptide sequence identified | MASCOT Score | Expected score (*p* value) | Average T/N ratio[a] | Average $\log_2$ (T/N ratio) ± SD | Frequency (%)[c] | Unwin *et al.* (2003) [16][d] | Perroud *et al.* (2006) [18][d] |
|---|---|---|---|---|---|---|---|---|---|
| **Upregulation more than two-fold** | | | | | | | | | |
| 6-Phosphofructokinase type C | Q01813 | EWSGLLEELAR | 35 | 0.034 | 4.9 | 2.3 ± 0.96 | 8/9 (89%) | | |
| 6-Phosphofructokinase, liver type | P17858 | GQVQEVGWHDVAGWLGR | 77 | 3.40E-06 | 6.9 | 2.79 ± 1.1 | 7/8 (88%) | | |
| α-Crystallin B chain | P02511 | APSWFDTGLSEMR | 28 | 0.012 | 2.9 | 1.6 ± 1.1 | 12/13(92%) | | |
| α-Enolase | P06733 | LAQANGWGVMVSHR | 60 | 0.00019 | 2.3 | 1.2 ± 2.3 | 10/13 (77%) | Up | |
| Annexin A4 | P09525 | KWGTDEVK | 25 | 0.014 | 8.5 | 2.2 ± 2.2 | 11/13 (85%) | Up | |
| Apolipoprotein A-I precursor | P02647 | LLDNWDSVTSTFSK | 63 | 0.00011 | 2.9 | 1.5 ± 0.79 | 8/8 (100%) | | |
| Calnexin precursor | P27824 | IVDDWANDGWGLK | 38 | 0.036 | 2.0 | 0.98 ± 1.8 | 8/11 (73%) | | |
| CNDP 2 | Q96KP4 | WVAIQSVSAWPEKR | 27 | 0.011 | 3.3 | 1.7 ± 1.4 | 9/9 (100%) | | |
| Ezrin | P15311 | IQVWHAEHR | 26 | 0.016 | 1.0 | 0.06 ± 2.6 | 8/12 (67%) | | |
| Fibrinogen γ chain precursor | P02679 | NWIQYK | 23 | 0.018 | 3.1 | 1.6 ± 1.5 | 8/9 (89%) | | |
| | | VELEDWNGR | 29 | 0.0058 | 1.8 | 3.4 ± 1.2 | 8/8 (100%) | | |
| Fibronectin precursor | P02751 | FLATTPNSLLVSWQPPR | 57 | 0.00025 | 4.8 | 2.3 ± 1.0 | 9/11 (82%) | | |
| Four and a half LIM domains protein 1 | Q13642 | FWHDTCFR | 34 | 0.0018 | 2.6 | 1.4 ± 0.71 | 12/14 (86%) | | |
| Fructose-bisphosphate aldolase A | P04075 | CPLLKPWALTFSYGR | 50 | 0.00098 | 2.9 | 1.6 ± 1.5 | 11/13 (85%) | Up | |
| Galectin-1 | P09382 | DGGAWGTEQR | 53 | 0.0015 | 2.4 | 1.3 ± 0.93 | 11/14 (79%) | | |
| Galectin-3 | P17931 | LDNNWGR | 24 | 0.014 | 4.1 | 1.4 ± 1.5 | 7/10 (70%) | | |
| Glucosidase two subunit β precursor | P14314 | YEQGTGCWQGPNR | 49 | 0.0022 | 1.3 | 0.42 ± 1.3 | 10/14 (71%) | | |
| Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) | P04406 | LISWYDNEFGYSNR | 43 | 0.002 | 2.1 | 1.1 ± 1.6 | 10/14 (71) | Up | Up |
| Heat shock protein β-1 (HspB1) (heat shock 27 kDa protein) | P04792 | GPSWDPFR | 19 | 0.012 | 2.1 | 1.1 ± 2.2 | 9/12 (75%) | | Up |
| Heterogeneous nuclear ribonucleoproteins A2/B1 | P22626 | NYYEQWGK | 27 | 0.012 | 2.3 | 1.2 ± 1.4 | 9/14 (64%) | | |
| HLA class I histocompatibility antigen, A-25 α chain precursor | P18462 | WASVVVPSGQEQR | 39 | 0.00068 | 2.4 | 1.4 ± 3.6 | 10/10 (100%) | | |
| HLA class I histocompatibility antigen, B-7 α chain precursor | P01889 | WAAVVVPSGEEQR | 36 | 0.035 | 4.5 | 1.9 ± 0.83 | 10/10 (100%) | | |
| Lamin-A/C | P02545 | AQNTWGCGNSLR | 28 | 0.0088 | 7.4 | 1.6 ± 2.4 | 8/10 (80%) | | |
| L-lactate dehydrogenase A chain | P00338 | SADTLWGIQK | 24 | 0.025 | 7.8 | 3.0 ± 1.0 | 9/9 (100%) | Up | |
| Phosphoglycerate mutase 1 | P18669 | HGESAWNLENR | 46 | 0.0059 | 3.0 | 1.6 ± 1.1 | 10/12 (83%) | Up | Up |
| | | ALPFWNEEIVPQIK | 27 | 0.0094 | 3.3 | 1.7 ± 0.91 | 9/11 (82%) | | |
| Plectin-1 | Q15149 | ESADPLGAWLQDAR | 20 | 0.035 | 2.3 | 1.2 ± 0.64 | 10/13 (77%) | | |
| PRA1 family protein 3 | O75915 | AWDDFFPGSDR | 36 | 0.0012 | 2.1 | 1.1 ± 0.99 | 9/10 (90%) | | |
| Prostaglandin E synthase 3 | Q15185 | KGESGQSWPR | 42 | 0.014 | 4.0 | 1.6 ± 1.0 | 10/10 (100%) | | |
| Pyruvate kinase isozymes M1/M2 | P14618 | DPVQEAWAEDVDLR | 72 | 1.30E-05 | 4.7 | 2.2 ± 2.0 | 10/13 (77%) | Up | Up |
| | | CDENILWLDYK | 43 | 0.014 | [b]NC | [b]NC | 9/11 (82%) | | |
| Reticulocalbin-1 precursor | Q15293 | HWILPQDYDHAQAEAR | 56 | 0.00015 | 2.8 | 1.5 ± 1.6 | 9/11 (82%) | | |
| Retinal dehydrogenase 1 | P00352 | QAFQIGSPWR | 27 | 0.014 | 1.5 | 0.59 ± 2.2 | 10/13 (77%) | | |
| Triosephosphate isomerase (TIM) | P60174 | DCGATWVVLGHSER | 56 | 0.0004 | 1.8 | 0.87 ± 1.7 | 9/14 (64%) | | |
| | | VVLAYEPVWAIGTGK | 40 | 0.022 | 2.2 | 1.1 ± 1.7 | 9/12 (75%) | | |
| Tripeptidyl-peptidase 1 precursor | O14773 | AYPDVAALSDGYWVVSNR | 43 | 0.025 | 1.7 | 0.78 ± 0.90 | 6/9 (67%) | | |
| Vesicular integral-membrane protein VIP36 precursor | Q12907 | NLHGDGIALWYTR | 22 | 0.015 | 1.6 | 0.63 ± 1.2 | 9/14 (64%) | | |
| Vimentin | P08670 | NLQEAEEWYK | 50 | 0.00026 | 5.6 | 2.5 ± 1.6 | 12/14 (86%) | Up | |
| **Downregulation more than two-fold** | | | | | | | | | |
| 4-Aminobutyrate aminotransferase, mitochondrial precursor | P80404 | IDIPSFDWPIAPFPR | 94 | 6.90E-08 | [b]NC | [b]NC | 9/9 (100%) | Down | |
| Abhydrolase domain-containing protein 14B | Q96IU4 | FSSETWQNLGTLHR | 38 | 0.035 | 0.31 | −1.7 ± 1.3 | 7 /10 (70%) | | |
| Aconitate hydratase, mitochondrial precursor | Q99798 | WVVIGDENYGEGSSR | 53 | 0.001 | 0.14 | −2.8 ± 1.0 | 9/9 (100%) | Up | |

**Table 3.** Continued

| Protein name | Swiss-Prot accession no. | Peptide sequence identified | MASCOT Score | Expected score (*p* value) | Average T/N ratio[a] | Average $\log_2$ (T/N ratio) ± SD | Frequency (%)[c] | Unwin *et al.* (2003) [16][d] | Perroud *et al.* (2006) [18][d] |
|---|---|---|---|---|---|---|---|---|---|
| Acyl-coenzyme A thioesterase 2, mitochondrial precursor | P49753 | ASLHALVGSPIIWGGEPR | 53 | 0.00095 | 0.20 | −2.3 ± 0.50 | 11/11 (100%) | | |
| Adenylate kinase isoenzyme 4, mitochondrial | P27144 | GQHWLLDGFPR | 44 | 0.0092 | 0.36 | −1.5 ± 1.0 | 11/12 (92%) | | |
| ADP/ATP translocase 2 | P05141 | EQGVLSFWR | 62 | 0.00014 | 0.21 | −2.2 ± 1.2 | 14/14 (100%) | | |
| ADP/ATP translocase 3 | P12236 | EQGVLSFWR | 62 | 0.00014 | 0.21 | −2.2 ± 1.2 | 14/14 (100%) | | |
| Aldehyde dehydrogenase, mitochondrial precursor | P05091 | AAFQLGSPWR | 32 | 0.0033 | 0.23 | −2.1 ± 1.0 | 12/12 (100%) | Down | |
| Aldose 1-epimerase | Q96C23 | FPPVLLRPGEEYDHTTWFK | 23 | 0.03 | 0.21 | −2.3 ± 1.6 | 9/10 (90%) | | |
| α-Actinin-4 | O43707 | LASDLLEWIR | 51 | 0.0036 | 0.24 | 2.1 ± 0.88 | 9/9 (100%) | | |
| Aminoacylase-1 | Q03154 | SPWWVR | 23 | 0.025 | 0.041 | −4.6 ± 1.1 | 10/10 (100%) | Down | Down |
| Aminopeptidase N | P15144 | ELWILNR | 20 | 0.016 | 0.20 | −2.3 ± 1.3 | 13/13 (100%) | | |
| Apolipoprotein A-I precursor | P02647 | LLDNWDSVTSTFSK | 58 | 0.00032 | 0.25 | −2.0 ± 1.8 | 8/9 (89%) | | |
| | | WQEEMELYR | 42 | 0.012 | 0.40 | −1.3 ± 1.4 | 9/12 (75%) | | |
| Aquaporin-1 | P29972 | VWTSGQVEEYDLDAD-DINSR | 37 | 0.038 | 0.14 | −2.8 ± 0.77 | 10/10 (100%) | | |
| ATP synthase D chain, mitochondrial | O75947 | LAALPENPPAIDWAYYK | 56 | 0.0002 | 0.29 | 1.8 ± 1.4 | 12/13 (92%) | | Down |
| Basement membrane-specific heparan sulfate proteoglycan core protein precursor | P98160 | AAGVPSATITWR | 37 | 0.023 | 0.23 | −2.1 ± 2.0 | 11/12 (92%) | | |
| | | CQVSGSPPHYFYWSR | 47 | 0.0023 | 0.22 | −2.2 ± 1.8 | 9/9 (100%) | | |
| | | LLSGPYFWSLPSR | 42 | 0.014 | 0.12 | 3.0 ± 1.5 | 13/13 (100%) | | |
| β-Enolase | P13929 | AAVPSGASTGIYEALELR | 83 | 9.00E-07 | [b]NC | [b]NC | 11/11 (100%) | | |
| Betaine–homocysteine S-methyltransferase 1 | Q93088 | AGPWTPEAAVEHPEAVR | 71 | 1.40E-05 | 0.16 | −2.6 ± 1.0 | 12/12 (100%) | Down | Down |
| Carbonyl reductase [NADPH] 1 | P16152 | ILLNACCPGWVR | 21 | 0.053 | 0.21 | −2.2 ± 1.0 | 10/10 (100%) | | |
| Choline dehydrogenase | Q8NE62 | YNWCYHTEVQR | 34 | 0.002 | 0.30 | −1.8 ± 0.59 | 8/8 (100%) | | |
| Collagen α-1(XIV) chain precursor | Q05707 | VSEEWYNR | 18 | 0.024 | 0.18 | −2.5 ± 1.3 | 9/9 (100%) | | |
| Collagen α-1(XVIII) chain precursor | P39060 | SVWHGSDPNGR | 31 | 0.0035 | 0.14 | −2.8 ± 0.90 | 13/13 (100%) | | |
| | | LTESYCETWR | 33 | 0.0027 | 0.16 | −2.6 ± 1.2 | 14/14 (100%) | | |
| Complement component C9 precursor | P02748 | RPWNVASLIYETK | 26 | 0.036 | [b]NC | [b]NC | 10/10 (100%) | | |
| Creatine kinase B-type | P12277 | TFLVWVNEEDHLR | 27 | 0.011 | [b]NC | [b]NC | 8/9 (89%) | Down | |
| Cytochrome *c* oxidase subunit VIb isoform 1 (COX VIb-1) | P14854 | GGDISVCEWYQR | 52 | 0.0014 | 0.15 | −2.7 ± 1.3 | 9/9 (100%) | | |
| | | NCWQNYLDFHR | 35 | 0.044 | 0.21 | −2.2 ± 0.83 | 12/12 (100%) | | |
| Enoyl-CoA hydratase, mitochondrial precursor | P30084 | HWDHLTQVK | 42 | 0.011 | 0.073 | −3.8 ± 1.5 | 14/14 (100%) | | Down |
| Fibrinogen β chain precursor | P02675 | EDGGGWWYNR | 22 | 0.006 | 0.26 | −1.9 ± 1.8 | 11/12 (92%) | Up | |
| Fructose-bisphosphate aldolase B (liver-type aldolase) | P05062 | ALQASALAAWGGK | 47 | 0.0047 | [b]NC | [b]NC | 12/12 (100%) | | Down |
| γ-Glutamyltranspeptidase 1 precursor | P19440 | TAGGWAAASDSR | 39 | 0.022 | [b]NC | [b]NC | 13/13 (100%) | | |
| Glycine amidinotransferase, mitochondrial precursor | P50440 | YWPFYQK | 21 | 0.029 | 0.049 | −4.3 ± 1.4 | 10/10 (100%) | | Down |
| | | RPDPIDWSLK | 21 | 0.035 | 0.15 | −2.7 ± 1.8 | 13/13 (100%) | | |
| | | NANSLGGGFHCWTCDVR | 32 | 0.0039 | NC | NC | 10/10 (100%) | | |
| Glycine *N*-acyltransferase | Q6IB77 | FWHFGGNER | 38 | 0.00071 | 0.31 | 1.7 ± 0.45 | 12/12 (100%) | | |
| Isocitrate dehydrogenase [NADP] cytoplasmic | O75874 | LVSGWVKPIIIGR | 43 | 0.0028 | 0.21 | 2.3 ± 1.3 | 10/10 (100%) | | |
| Isocitrate dehydrogenase [NADP], mitochondrial precursor | P48735 | GRPTSTNPIASIFAWTR | 26 | 0.015 | [b]NC | [b]NC | 10/10 (100%) | | |
| L-lactate dehydrogenase B chain | P07195 | SADTLWDIQK | 35 | 0.0018 | 0.16 | −2.6 ± 0.69 | 9/9 (100%) | Down | |
| Methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial precursor | Q02252 | AFPAWADTSVLSR | 48 | 0.0018 | [b]NC | [b]NC | 14/14 (100%) | | |
| | | SDKWIDIHNPATNEVIGR | 109 | 1.10E-09 | [b]NC | [b]NC | 11/11 (100%) | | |
| Moesin | P26038 | IQVWHEEHR | 23 | 0.015 | 0.32 | −1.7 ± 1.3 | 12/13 (92%) | | |
| NADH dehydrogenase [ubiquinone] 1 α subcomplex subunit 4 | O00483 | LALFNPDVCWDR | 34 | 0.027 | 0.074 | −3.8 ± 0.99 | 14/14 (100%) | | |

Table 3. Continued

| Protein name | Swiss-Prot accession no. | Peptide sequence identified | MASCOT Score | Expected score (*p* value) | Average T/N ratio[a] | Average $\log_2$ (T/N ratio) ±SD | Frequency (%)[c] | Unwin *et al.* (2003) [16][d] | Perroud *et al.* (2006) [18][d] |
|---|---|---|---|---|---|---|---|---|---|
| NADH dehydrogenase [ubiquinone] 1 β subcomplex subunit 8, mitochondrial precursor | O95169 | DPWYSWDQPGLR | 32 | 0.0043 | [b]NC | [b]NC | 12/12 (100%) | | |
| NADH dehydrogenase [ubiquinone] 1 β subcomplex subunit 9 | Q9Y6M9 | HLESWCVQR | 22 | 0.033 | 0.25 | −2.0 ± 1.2 | 10/10 (100%) | | |
| NADH dehydrogenase [ubiquinone] iron–sulfur protein 3, mitochondrial precursor | O75489 | FDLNSPWEAFPVYR | 46 | 0.0009 | [b]NC | [b]NC | 10/10 (100%) | | |
| | | KFDLNSPWEAFPVYR | 83 | 2.60E-06 | [b]NC | [b]NC | 11/11 (100%) | | |
| NADH–ubiquinone oxidoreductase 75 kDa subunit, mitochondrial precursor | P28331 | GLLTYTSWEDALSR | 24 | 0.021 | [b]NC | [b]NC | 9/9 (100%) | Down | |
| Nidogen-1 precursor | P14543 | NIFWTDSNLQR | 29 | 0.0084 | 0.42 | −1.3 ± 1.1 | 8/9 (89%) | | |
| Phosphatidylethanolamine-binding protein 1 | P30086 | NRPTSISWDGLDSGK | 39 | 0.027 | 0.42 | −1.2 ± 1.1 | 12/12 (100%) | | |
| | | APVAGTCYQAEWDDYVPK | 47 | 0.0025 | 0.17 | −2.5 ± 1.2 | 10/10 (100%) | | |
| | | YVWLVYEQDRPLK | 45 | 0.0057 | 0.18 | −2.4 ± 1.2 | 10/11 (100%) | | |
| Phosphoenolpyruvate carboxykinase [GTP], mitochondrial precursor | Q16822 | DEAGHFLWPGFGENAR | 48 | 0.0025 | 0.26 | −1.9 ± 1.6 | 10/10 (100%) | Down | Down |
| | | GQLGNWMSPADFQR | 70 | 1.80E-05 | [b]NC | [b]NC | 11/11 (100%) | | |
| Phosphoenolpyruvate carboxykinase, cytosolic | P35558 | IFHVNWFR | 33 | 0.096 | [b]NC | [b]NC | 13/13 (100%) | Down | |
| Retinol binding protein | P02753 | GNDDHWIVDTDYDTYAV QYSCR | 86 | 7.30E-08 | [b]NC | [b]NC | 12/12 (100%) | | |
| Sodium/potassium-transporting ATPase subunit α-1 precursor | P05023 | TSATWLALSR | 21 | 0.0088 | 0.12 | −3.1 ± 1.3 | 14/14 (100%) | | |
| Sorting and assembly machinery component 50 homolog | Q9Y512 | VTGQFPWSSLR | 26 | 0.017 | [b]NC | [b]NC | 10/10 (100%) | | |
| Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial precursor | P31040 | KPFEEHWR | 41 | 0.019 | 0.30 | −1.7 ± 0.97 | 12/13 (92%) | | |
| Sulfite oxidase, mitochondrial precursor | P51687 | VSVQPEESYSHWQR | 40 | 0.012 | [b]NC | [b]NC | 11/11 (100%) | | |
| Thiosulfate sulfurtransferase | Q16762 | VLDASWYQPGTR | 33 | 0.0029 | 0.12 | 3.0 ± 1.6 | 9/9 (100%) | | |
| Trifunctional enzyme subunit β, mitochondrial precursor | P55084 | FNNWGGSLSLGHPF GATGCR | 54 | 0.00051 | 0.21 | −2.3 ± 1.7 | 8/8 (100%) | | |
| Tubulointerstitial nephritis antigen-like precursor | Q9GZM7 | ITGWGEETLPDGR | 36 | 0.041 | 0.20 | −2.3 ± 1.3 | 13/13 (100%) | | |
| Ubiquinol–cytochrome *c* reductase complex 14 kDa protein | P14927 | KWYYNAAGFNK | 55 | 0.0015 | [b]NC | [b]NC | 10/10 (100%) | | |
| Ubiquinol–cytochrome *c* reductase complex core protein 1, mitochondrial precursor | P31930 | RIPLAEWESR | 39 | 0.016 | [b]NC | [b]NC | 11/11 (100%) | Down | |
| Uromodulin precursor | P07911 | DWVSVVTPAR | 52 | 0.00092 | [b]NC | [b]NC | 11/11 (100%) | | |
| Vitronectin, precursor | P04004 | DVWGIEGPIDAAFTR | 41 | 0.014 | 0.22 | −2.2 ± 1.5 | 10/11 (100%) (91%) | | |
| Voltage-dependent anion-selective channel protein 1 | P21796 | YRWTEYGLTFTEK | 37 | 0.04 | 0.27 | −1.9 ± 0.89 | 7/8 (88%) | | |

a) An average (T/N ratio) is calculated by exponential transformation of an average $\log_2$ (T/N ratio).
b) NC; not calculated because a protein is exclusively detected in RCC (or normal) tissues.
c) Frequency; the number of sample in which a protein is altered (>2-fold) in RCC tissue.
d) Up- or downregulated proteins identified in previous reports (Unwin *et al.* [16] and Perroud *et al.* [18]).

treatment of advanced RCC [2]. Although these drugs have extended the survival rate, they are not sufficient for the treatment.

With the completion of the human genome project and great advances in MS technology, proteomics has been attracting attention as a way to discover potential biomarkers

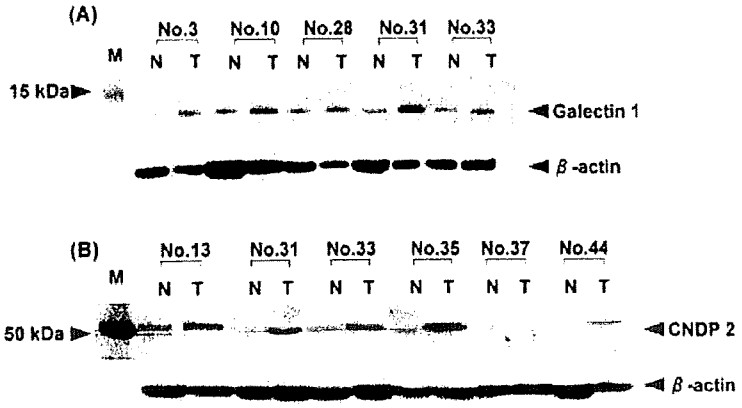


Figure 2. Western blot analysis for galectin-1 (A), CNDP2 (B), and β-actin. M, molecular marker; N, normal tissues; T, cancerous tissues.

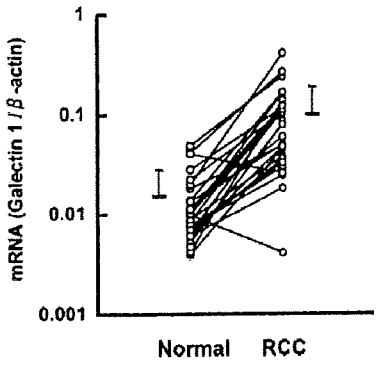


Figure 3. Levels of galectin-1 mRNA in RCC and adjacent normal tissue determined by quantitative real time RT-PCR. Each point from one patient and the bar represents the mean ± SD.

and therapeutic targets. To date, several proteomics methods including 2-DE and SELDI-TOF MS have been applied to clinical samples including cancerous tissues and serum. However, their applications have thus far been limited. In seeking to increase the utility of the proteomics approach, we have developed the NBS method for global quantitative proteome analysis. Here, we used the method to discover potential biomarkers and therapeutic targets for RCC. In the present study, proteins whose expression in tumor and normal tissues were significant different were selected and a final set of 92 proteins (34 upregulated and 58 downregulated proteins) was identified. Of these, 24 proteins have been described previously [16–18] and almost all of the presently obtained results were consistent with the earlier observations. For instance, our finding of higher levels of proteins involved in glycolysis (such as glyceraldehyde-3-phosphate dehydrogenase and fructose-bisphosphate aldolase A) was consistent with earlier reports [16–21].

However, the results for two proteins (aconitate hydratase mitochondrial precursor and fibrinogen β chain precursor) were inconsistent with earlier reports (Table 3). These discrepancies were probably due to differences in the methods used and specimens in each experiment. More interestingly, 68 novel RCC-associated proteins were identified by the NBS-based proteomics analyses.

In this study, we focused on upregulated proteins in cancerous tissues since these proteins are thought to be potential diagnostic biomarkers. Among the upregulated proteins, we selected galectin-1 and CNDP2 and validated their proteomics data since they have not been studied in relation to RCC and have distinct features. Both of the selected proteins showed high expression levels in RCC tissues and occurred with high frequency in RCC patients; galectin-1 in 11 of 14 patients (79%), CNDP2 in all 9 patients (100%).

Galectin-1 is a member of the carbohydrate binding protein family that shows affinity for β-galactosides. It has been thought to have roles in interaction with p27, Ras, Raf, and PI-3K [22–26]. Galectin-1 is extracellularly expressed in many tissues in both normal and pathological conditions, and its expression has been shown to be higher in colorectal, pancreatic, and nonsmall cell lung cancer [12, 27–30]. However, there have been no reports demonstrating an association between galectin-1 and RCC. In this study, we noted high expression of galectin-1 in RCC tissues by proteomics and Western blot analyses.

Furthermore, RT-PCR analyses demonstrated that galectin-1 mRNA was also highly expressed in RCC tissues. Genes that are transcriptionaly upregulated in relation to cancer progression may be targets for gene therapy such as RNA interference. Thus, at present, this molecule offers the potential of being applicable for not only a diagnostic biomarker but also as a target for gene therapy.

CNDP2 is a carnosinase belonging to a member of metallopeptidase M20 family. Carnosinase exists in two isoforms. CNDP2 is expressed in the liver, kidney, and spleen tissue, and CNDP1 is found in serum (plasma), as well as in the brain and spinal fluid [31]. CNDP1 gene is associated with nephropathy [32] and its downregulation is suggested to be associated with cell growth and metastasis of hepatocellular carcinoma [33]. However, the function of CNDP2 remains to be elucidated and there has been no evidence

indicating its association with any types of cancer. In this study, we demonstrated high expression of CNDP2 in RCC tissues and suggested its association with RCC for the first time although the number of tested samples was small. In the future, we will further investigate its expression in normal and cancerous tissues in larger sample sets to confirm the association with RCC by other methods such as immunohistochemistry. In addition, to examine the possibility for a diagnostic marker, we will examine whether this molecule is secreted into sera and detected at significantly higher levels in patients in comparison to healthy volunteers.

In conclusion, the NBS-based proteomics represents a potentially useful tool for mining novel candidates as biomarkers and therapeutic targets for disease. Although more studies are needed, the proteins identified here could be biomarkers and/or therapeutic targets for RCC.

*The authors have declared no conflict of interest.*

## 5 References


[1] Amato, R. J., Renal cell carcinoma: Review of novel single-agent therapeutics and combination regimens. *Ann. Oncol.* 2005, *16*, 7–15.

[2] Garcia, J. A., Rini, B. I., Recent progress in the management of advanced renal cell carcinoma. *CA Cancer J. Clin.* 2007, *57*, 112–125.

[3] Rabilloud, T., Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. *Proteomics* 2002, *2*, 3–10.

[4] Kyte, J., Doolittle, R. F., A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 1982, *157*, 105–132.

[5] Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., Aebersold, R., Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. USA* 2000, *97*, 9390–9395.

[6] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. *et al.*, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 1999, *17*, 994–999.

[7] Kuyama, H., Watanabe, M., Toda, C., Ando, E. *et al.*, An approach to quantitative proteome analysis by labeling tryptophan residues. *Rapid Commun. Mass Spectrom.* 2003, *17*, 1642–1650.

[8] Matsuo, E., Toda, C., Watanabe, M., Iida, T. *et al.*, Improved 2-nitrobenzenesulfenyl method: Optimization of the protocol and improved enrichment for labeled peptides. *Rapid Commun. Mass Spectrom.* 2006, *20*, 31–38.

[9] Matsuo, E., Toda, C., Watanabe, M., Ojima, N. *et al.*, Selective detection of 2-nitrobenzenesulfenyl-labeled peptides by matrix-assisted laser desorption/ionization-time of flight mass spectrometry using a novel matrix. *Proteomics* 2006, *6*, 2042–2049.

[10] Cagney, G., Amiri, S., Premawaradena, T., Lindo, M. *et al.*, In silico proteome analysis to facilitate proteomics experiments using mass spectrometry. *Proteome Sci.* 2003, *13*, 5.

[11] Iida, T., Kuyama, H., Watanabe, M., Toda, C. *et al.*, Rapid and efficient MALDI-TOF MS peak detection of 2-nitrobenzenesulfenyl labeled peptides using the combination of HPLC and an automatic spotting apparatus. *J. Biomol. Technol.* 2006, *17*, 333–341.

[12] Watanabe M., Takemasa I., Kawaguchi N., Miyake M. *et al.*, An application of the 2-nitrobenzenesulfenyl (NBS) method to proteomic profiling of human colorectal carcinoma: A novel approach for biomarker discovery. *Proteomics Clin. Appl.* 2008, *2*, 925–935.

[13] Nakamura, T., Sakaeda, T., Horinouchi, M., Tamura, T. *et al.*, Effect of the mutation (C3435T) at exon 26 of the MDR1 gene on expression level of MDR1 messenger ribonucleic acid in duodenal enterocytes of healthy Japanese subjects. *Clin. Pharmacol. Ther.* 2002, *71*, 297–303.

[14] Kawakami, M., Nakamura, T., Okamura, N., Komoto, C. *et al.*, Knock-down of sorcin induces up-regulation of MDR1 in HeLa cells. *Biol. Pharm. Bull.* 2007, *30*, 1065–1073.

[15] Ferguson, R. E., Carroll, H. P., Harris, A., Maher, E. R. *et al.*, Housekeeping proteins: A preliminary study illustrating some limitations as useful references in protein expression studies. *Proteomics* 2005, *5*, 566–571.

[16] Unwin, R. D., Craven, R. A., Harnden, P., Hanrahan, S. *et al.*, Proteomic changes in renal cancer and co-ordinate demonstration of both the glycolytic and mitochondrial aspects of the Warburg effect. *Proteomics* 2003, *3*, 1620–1632.

[17] Takashi, M., Haimoto, H., Tanaka, J., Murase, T. *et al.*, Evaluation of gamma-enolase as a tumor marker for renal cell carcinoma. *J. Urol.* 1989, *141*, 830–834.

[18] Perroud, B., Lee, J., Valkova, N., Dhirapong, A. *et al.*, Pathway analysis of kidney cancer using proteomics and metabolic profiling. *Mol. Cancer* 2006, *5*, 64.

[19] Seliger, B., Lichtenfels, R., Atkins, D., Bukur, J. *et al.*, Identification of fatty acid binding proteins as markers associated with the initiation and/or progression of renal cell carcinoma. *Proteomics* 2005, *5*, 2631–2640.

[20] Hwa, J. S., Kim, H. J., Goo, B. M., Park, H. J. *et al.*, The expression of ketohexokinase is diminished in human clear cell type of renal cell carcinoma. *Proteomics* 2006, *6*, 1077–1084.

[21] Seliger, B., Fedorushchenko, A., Brenner, W., Ackermann, A. *et al.*, Ubiquitin COOH-terminal hydrolase 1: A biomarker of renal cell carcinoma associated with enhanced tumor cell proliferation and migration. *Clin. Cancer Res.* 2007, *13*, 27–37.

[22] Ingrassia, L., Camby, I., Lefranc, F., Mathieu, V. *et al.*, Anti-galectin compounds as potential anti-cancer drugs. *Curr. Med. Chem.* 2006, *13*, 3513–3527.

[23] Camby, I., Le Mercier, M., Lefranc, F., Kiss, R., Galectin-1: A small protein with major functions. *Glycobiology* 2006, *16*, 137R–157R.

[24] Fischer, C., Sanchez-Ruderisch, H., Welzel, M., Wiedenmann, B. *et al.*, Galectin-1 interacts with the α5β1 fibronectin receptor to restrict carcinoma cell growth via induction of p21 and p27. *J. Biol. Chem.* 2005, *280*, 37266–37277.

[25] Andre, S., Kaltner, H., Lensch, M., Russwurm, R. *et al.*, Determination of structural and functional overlap/divergence of five proto-type galectins by analysis of the growth-regulatory interaction with ganglioside GM1 in silico and in vitro on human neuroblastoma cells. *Int. J. Cancer* 2005, *114*, 46–57.

[26] Elad-Sfadia, G., Haklai, R., Ballan, E., Gabius, H. J., Kloog, Y., Galectin-1 augments Ras activation and diverts Ras signals to Raf-1 at the expense of phosphoinositide 3-kinase. *J. Biol. Chem.* 2002, *277*, 37169–37175.

[27] Grutzmann, R., Pilarsky, C., Ammerpohl, O., Luttges, J. *et al.*, Gene expression profiling of microdissected pancreatic ductal carcinomas using high-density DNA microarrays. *Neoplasia* 2004, *6*, 611–622.

[28] Shen, J., Person, M. D., Zhu, J., Abbruzzese, J. L., Li, D., Protein expression profiles in pancreatic adenocarcinoma compared with normal pancreatic tissue and tissue affected by pancreatitis as detected by two-dimensional gel electrophoresis and mass spectrometry. *Cancer Res.* 2004, *64*, 9018–9026.

[29] Gabius, H. J., Andre, S., Gunsenhauser, I., Kaltner, H. *et al.*, Association of galectin-1- but not galectin-3 dependent parameters with proliferation activity in human neuroblastomas and small cell lung carcinomas. *Anticancer Res.* 2002, *22*, 405–410.

[30] Szoke, T., Kayser, K., Baumhakel, J. D., Trojan, I. *et al.*, Prognostic significance of endogenous adhesion/growth-regulatory lectins in lung cancer. *Oncology* 2005, *69*, 167–174.

[31] Teufel, M., Saudek, V., Ledig, J.-P., Bernhardt, A. *et al.*, Sequence identification and characterization of human carnosinase and a closely related nonspecific dipeptidase. *J. Biol. Chem.* 2003, *278*, 6521–6531.

[32] Bart, J., Daniela, H., Paul, B., Verena, P. *et al.*, Carnosine as a protective factor in diabetic nephropathy association with a leucine repeat of the carcinase gene CNDP1. *Diabetes* 2005, *54*, 2320–2327.

[33] Pingping, Z., David, W. C., Yi, Y. Z., Jin, J. L. *et al.*, Identification of carboxypeptidase of glutamate like-B as a candidate suppressor in cell growth and metastasis in human hepatocellular carcinoma. *Clin. Cancer Res.* 2006, *12*, 6617–6625.

 

RAPID COMMUNICATIONS IN MASS SPECTROMETRY
*Rapid Commun. Mass Spectrom.* 2008; **22**: 3313–3319
Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/rcm.3735

RCM

# Selective isolation of N-terminal peptides from proteins and their *de novo* sequencing by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry without regard to unblocking or blocking of N-terminal amino acids

**Minoru Yamaguchi[1]*, Daisuke Nakayama[1], Keisuke Shima[1], Hiroki Kuyama[2], Eiji Ando[1], Taka-aki Okamura[3], Norikazu Ueyama[3], Takashi Nakazawa[4], Shigemi Norioka[5], Osamu Nishimura[1,2] and Susumu Tsunasawa[1,2]**

[1]Life Science Business Department, Shimadzu Corporation, Kyoto 604-8511, Japan
[2]Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan
[3]Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan
[4]Department of Chemistry, Nara Women's University, Nara 630-8506, Japan
[5]Graduate School of Frontier Biosciences, Osaka University, Suita, Osaka 565-0871, Japan

Received 28 April 2008; Revised 7 July 2008; Accepted 24 August 2008

We have developed a new method to determine the N-terminal amino acid sequences of proteins, regardless of whether their N-termini are modified. This method consists of the following five steps: (1) reduction, S-alkylation and guanidination for targeted proteins; (2) coupling of sulfo-NHS-SS-biotin to $N^{\alpha}$-amino groups of proteins; (3) digestion of the modified proteins by an appropriate protease followed by oxidation with performic acid; (4) specific isolation of N-terminal peptides from digests using DITC resins; (5) *de novo* sequence analysis of the N-terminal peptides by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using the CAF (chemically assisted fragmentation) method or tandem mass spectrometric (MS/MS) analysis according to unblocked or blocked peptides, respectively. By employing DITC resins instead of avidin resins used in our previous method (Yamaguchi *et al.*, Rapid Commun. Mass Spectrom. 2007; 21: 3329), it has been possible to isolate selectively N-terminal peptides from proteins regardless of modification of N-terminal amino acids. Here we propose a universal method for N-terminal sequence analysis of proteins. Copyright © 2008 John Wiley & Sons, Ltd.

Amino acid sequences of proteins contain information regarding the coding genes as well as processing of DNA, RNA or proteins, which is reflected in differences from the amino acid sequences that would be predicted from the corresponding genome by direct application of the genetic code. Coding genes are often edited through RNA splicing, shuffling and other processing mechanisms, resulting in different translation products.[1] In addition, the sequences of many proteins reflect polymorphisms in local regions of the genome. Proteins may also be processed by proteases or post-translationally modified during conversion into active mature forms. The number of proteins in a cell thus exceeds by far the number of genes. More than 80% of proteins are post-translationally processed and/or modified at their N-temini.[2] However, the N- and C-terminal part of each protein is usually processed conservatively, so analysis of the termini of proteins is very important for identification of target proteins and for understanding their roles in cells under various conditions. Several approaches to N-terminal amino acid sequencing of proteins utilizing mass spectrometry (MS) have been reported.[3–10] These technological advances have enabled high-sensitivity and high-throughput analysis, which are essential aspects of the rapidly growing proteomics fields. Of those, McDonald *et al.* suggested a possibility to discriminate N-terminally acetylated peptides among N-terminally free ones by employing stable isotope labeled acetic anhydride in their method,[4] but no experimental results were presented. Here, we have established an effective method for N-terminal sequencing of proteins without regard for blocking or unblocking of their N-terminal amino acids, by isolating N-terminal peptides using DITC resins instead of the standard avidin-biotin technique, and by keeping the superiority of the CAF (chemically assisted fragmentation) method for *de novo* amino acid sequencing.[8] A general scheme for selective isolation of N-terminal peptides from proteins separated by one/two dimensional sodium dodecyl sulfate/polyacrylamide gel electrophoresis (1D/2D SDS-PAGE) is shown in Fig. 1. In this

*Correspondence to: M. Yamaguchi, Life Science Business Department, Shimadzu Corporation, Kyoto 604-8511, Japan.
E-mail: yamagu@shimadzu.co.jp

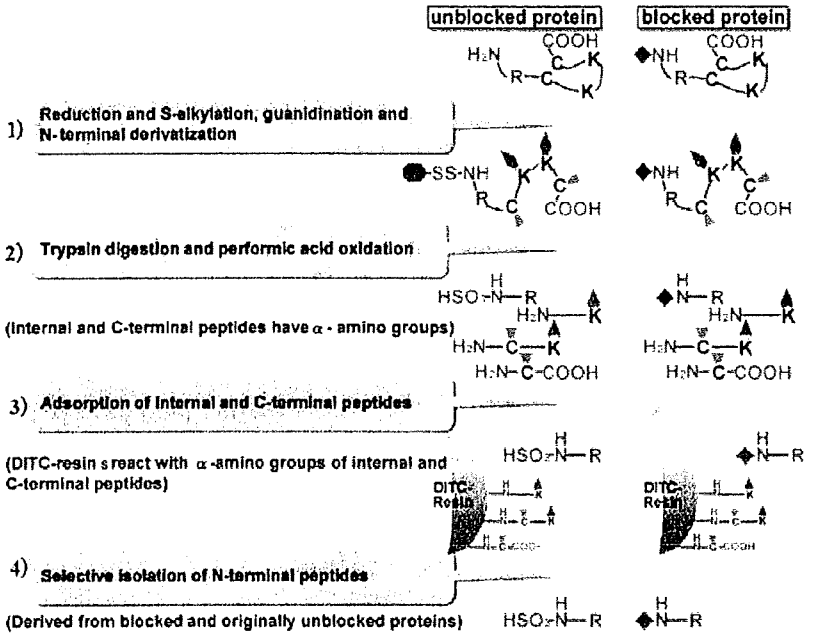


**Figure 1.** Protocol for selective isolation of N-terminal peptides from proteins. The method consists of the following four steps: (1) reduction and S-alkylation, guanidination and coupling with sulfo-NHS-SS-biotin to $N^{\alpha}$-amino groups of targeted proteins; (2) digestion of the modified proteins by trypsin followed by oxidation of N-terminal S–S bonds by performic acid oxidation (for unblocked proteins); (3) adsorption of internal and C-terminal peptides with DITC resins; and (4) selective isolation of N-terminal peptides derived from both blocked and unblocked proteins.

method the N-terminus of an S-reduced and alkylated protein is followed by guanidination of the ε-amino groups of lysines. The protein is then reacted with sulfo-NHS-SS-biotin, followed by tryptic digestion and performic acid oxidation, which introduces a sulfonic acid group only onto the N-terminal peptide. DITC resins are then added and the N-terminal peptides are selectively isolated from the unabsorbed fraction, regardless of whether or not the N-termini of the proteins are blocked. Amino acid sequencing of the peptides is then performed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using the CAF method or tandem mass spectrometric (MS/MS) analysis[11] according to unblocked or blocked states, respectively. This is the first paper to describe a method for isolating selectively N-terminal peptides regardless of their modification and for their *de novo* sequencing by MALDI-TOF MS.

## EXPERIMENTAL

### Chemicals and reagents

Cytochrome *c* (bovine and *Saccharomyces cerevisiae*), phosphorylase B (bovine), lysozyme (chicken egg white), α-cyano-4-hydroxycinnamic acid (CHCA) and a ProteoPrep™ sample extraction kit were purchased from Sigma (St. Louis, MO, USA). A lysate of human epidermoid carcinoma cells (A-431 + EGF (IP) cell lysate) was obtained from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA, USA). Pharmalyte and Immobiline dry strip gels were from GE Healthcare, UK Ltd. *O*-Methylisourea hemisulfate was from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). Sequencing-grade modified trypsin was from Promega (Madison, WI, USA). Sulfo-NHS-SS-biotin was obtained from Pierce Co. (Rockford, IL, USA). DITC resins were used as filler particles for CTFF-1 columns from Shimadzu Corp. (Kyoto, Japan), and ZipTip® μC18 and Microcon® YM3 filter devices were purchased from Millipore Corp. (Bedford, MA, USA). Tris(2-carboxyethyl)-phosphine hydrochloride (TCEP) was from Fluka (Buchs SG, Switzerland). All other chemicals were of analytical reagent grade and were used without further purification.

### Selective isolation of N-terminal peptides

Proteins were subjected to a series of chemical reactions, including S-alkylation, guanidination and N-terminal derivatization. These reactions were usually performed on proteins that had been separated using polyacrylamide gels, as follows: Protein samples separated by 1D or 2D SDS-PAGE were stained with CBB and the corresponding protein spots were excised from the gels. The gel pieces were placed in microtubes and washed successively with 100 μL of 100 mM ammonium bicarbonate solution and acetonitrile and then dried in a SpeedVac evaporator. The dried gel pieces were incubated in 100 μL of 10 mM TCEP/55 mM iodoacetamide/100 mM ammonium bicarbonate solution for 60 min at 56°C in the dark. After centrifugation the reagent solution was removed. A solution consisting of 5 mg *o*-methylisourea hemisulfate in 30 μL of 7N $NH_4OH$ was added and the tubes were incubated at 65°C for 30 min. The




DOI: 10.1002/rcm

gel slices were then washed twice with 350 μL of 0.2 M phosphate buffer (pH 7.2), followed by two successive washes using 0.2 M phosphate buffer (pH 7.2) and acetonitrile, then dried in a SpeedVac evaporator. Next, 30 μL of 0.2 M phosphate buffer (pH 7.6) containing 0.2 mg of sulfo-NHS-SS-biotin were added and the slices were incubated at 37°C for 1 h. The gel pieces were washed with 100 mM ammonium bicarbonate solution followed by three washes in acetonitrile and then dried in a SpeedVac evaporator. After 2 μL of 50 ng/μL trypsin in 25 mM ammonium bicarbonate solution had been added, followed by 15 μL of 25 mM ammonium bicarbonate solution, each slice was incubated at 37°C for 4 h. The resulting tryptic peptides were oxidized using 50 μL of freshly prepared performic acid (950 μL of 99% formic acid and 50 μL of hydrogen peroxide solution was mixed and left to stand at room temperature for 2 h) on ice for 30 min, and the solution was then dried in a SpeedVac evaporator. The dry oxidized tryptic peptides were dissolved in 5 μL of 100 mM sodium hydrogen carbonate, and the solution was applied onto DITC resins, which were then washed with 100 μL of 100 mM sodium hydrogen carbonate. The supernatant was removed from the resins, and they were left to stand at 60°C for 2 h. Then 100 μL of a mixture of 0.1% aqueous trifluoroacetic acid (TFA)/2-propanol/acetonitrile (62:40:18, v/v/v) were added to the resins, and the supernatants were carefully recovered. Next, 50 μL of the above solution were added to the resins. All of the supernatants were then mixed and dried in a Speedvac evaporator.

### 2D PAGE of proteins contained in a crude lysate of human epidermoid carcinoma cells

To remove the detergent contained in a crude lysate of human epidermoid carcinoma cells (IP cell lysates), the solution (500 μL: corresponding to 1.2 mg protein) was poured into centrifugal filter devices, Microcon® YM3, and they were centrifuged at 14000 *g* at 20°C for 30 min. Then, 150 μL of 5% aqueous acetic acid were added to the Microcon® YM3 device and it was centrifuged at 14000 *g* at 20°C for 30min. The lysate was further washed three times with 250 μL of the diluted reagent 3 (in the ProteoPrep™ sample extraction kit) at 1/100 concentration, followed by five washes with 250 μL of 5% aqueous acetic acid. Proteins were dissolved in the diluted reagent 3, and the solution was lyophilized and dissolved again in 200 μL reagent 3 according to the protocol for the ProteoPrep™ sample extraction kit. After 2 μL of Pharmalyte and 2.5 μL bromophenol blue had been added to the solution, it was centrifuged at 15 000 *g* for 20 min at 20°C and the supernatant was rehydrated on an Immobiline dry strip gel with a pH gradient from pH 5.0–8.0, and isoelectric focusing was performed. Proteins thus separated by 2D-PAGE were stained by Coomassie Brilliant Blue (CBB) and subjected to N-terminal analyses after destaining overnight in 25% isopropanol/10% acetic acid.

### Mass spectrometry

Samples were dissolved in 0.1% TFA, desalted with a ZipTip® μC18 and eluted with 2.5 μL 50% acetonitrile/0.05% TFA solution before analysis by MALDI MS. CHCA saturated in 50% acetonitrile/0.05% TFA solution was used as an appropriate matrix to perform MALDI-MS/MS analysis. A portion (0.5 μL) of each sample solution was mixed with 0.5 μL of the matrix solution on the MALDI target and analyzed after drying. The *m/z* values of the spectra were externally calibrated using a bradykinin fragment (*m/z* 757.40) and an ACTH (*m/z* 2465.20) fragment. Mass spectra were measured on an AXIMA-TOF$^2$ mass spectrometer (Shimadzu/Kratos, Manchester, UK). A nitrogen laser (337 nm) was used to irradiate the sample to ionize the fragments. The acceleration voltage of the instrument was set to 20 kV using a gridless-type electrode. The reflectron mode was set to detect positive ions. The MS/MS spectrum was obtained by focusing all of the product ions into a single spectrum by curved field reflectron.[12] Selected product ions generated from angiotensin II were used for external MS/MS calibration.

### Database searches

The data from MS/MS analyses were searched using the program Mascot to search Homo sapiens sequences in the MSDB database. In all cases, guanidination of Lys, oxidation ($O_2$) of methonine, oxidation ($O_2$) of tryptophan and carbamidomethylation and oxidation ($O_2$) of cysteine residues were used as fixed modifications. In the MS/MS analysis, Acetyl (N-term), Formyl (N-term), and Pyro-glu (N-term) were considered as possible modifications of the peptides. Tolerances for peptide mass and MS/MS values were set to 0.5 and 0.5 Da, respectively.

## RESULTS AND DISCUSSION

### Isolation of N-terminal peptides from both N-blocked and unblocked proteins

In previous works,[7,8] we reported a method for selective isolation of N-terminal peptides from N-terminally unblocked proteins followed by their *de novo* sequencing, by effectively using avidin resins and MALDI-TOF MS. In this report we have established a new method for N-terminal sequencing of proteins, regardless of whether or not the N-termini are blocked, using DITC resins instead of avidin resins. Chemical modifications, including S-carboxyamidomethylation, guanidination and N-terminal derivatization by sulfo-NHS-SS-biotin, were carried out for proteins immobilized in gels. Using our previously reported methods,[8] proteins were submitted to tryptic digestion, and all of the N-terminal peptides in the resulting tryptic digests, including those arising from originally blocked proteins, were selectively isolated using DITC resins. The effectiveness of these procedures was confirmed by analyzing two N-terminally blocked (phosphorylase B and bovine cytochrome *c*) and two unblocked (hen egg-white lysozyme and cytochrome *c* from *S. cerevisiae*) proteins as models. Figures 2(a) and (b) show MALDI-TOF MS spectra of tryptic digests of phosphorylase B and of its peptides isolated by DITC resins, respectively. The peak corresponding to the N-terminal peptide of phosphorylase B was the primary final product detected, as shown in Fig. 2(b). Furthermore, MS/MS analysis of the peak in Fig. 2(b) demonstrated that the amino acid sequence of the peptide was Ac-SRPLSDQEKR,




DOI: 10.1002/rcm

RCM

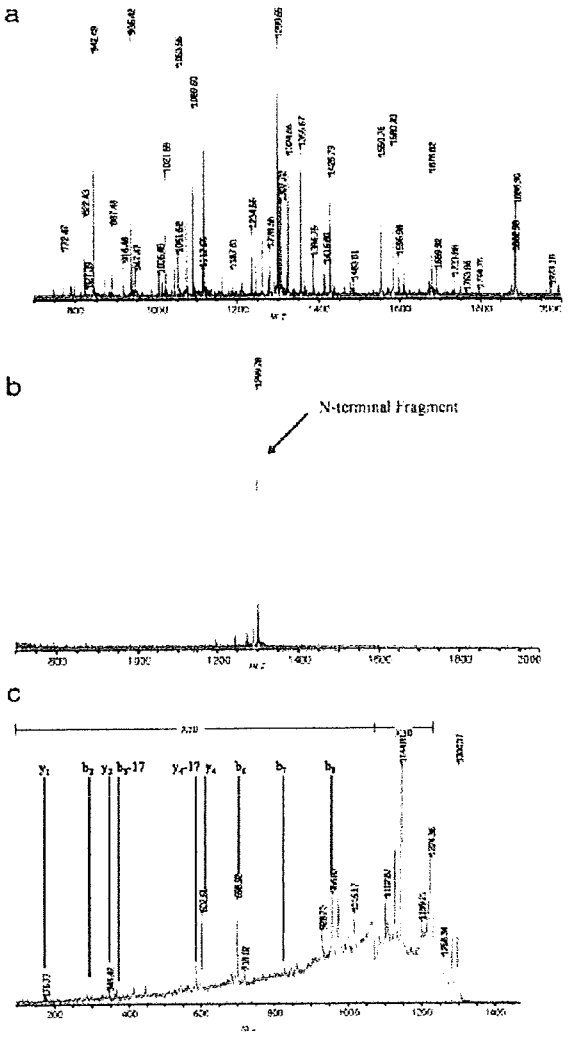


**Figure 2.** MALDI-TOF MS spectra obtained at each step during N-terminal sequencing of phosphorylase B (3 pmol of the protein was loaded on SDS-PAGE gel). (a) MALDI-TOF MS spectrum of trypsin digests of phosphorylase B. (b) MALDI-TOF MS spectrum of the N-terminal peptide specifically isolated from the trypsin digests by DITC resins. (c) MALDI-MS/MS spectrum of the N-terminal peptide.

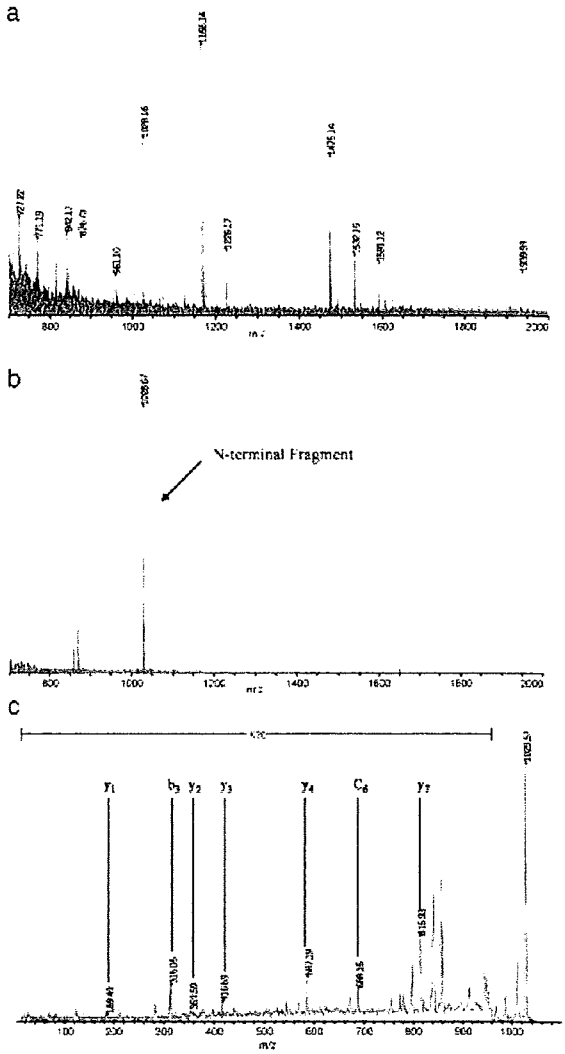


**Figure 3.** MALDI-TOF MS spectra at each step during N-terminal sequencing of bovine cytochrome *c* (CTC-B: 3 pmol): (a) MALDI-TOF MS spectrum of trypsin digests of CTC-B. (b) MALDI-TOF MS spectrum of the N-terminal peptide specifically isolated from trypsin digests by DITC resins. (c) MALDI-MS/MS spectrum of the N-terminal peptide.

which corresponds to the N-terminal peptide of phosphorylase B. Figure 3 shows MALDI-TOF MS spectra obtained at each step when bovine cytochrome *c* was processed using the procedures described above. The N-terminal peptide could be detected in tryptic digests (Fig. 3(a)), and it was enriched by DITC resins (Fig. 3(b)). As shown in Fig. 3(c), the amino acid sequence of the peptide observed in Fig. 3(b) was confirmed as Ac-GDVEKGKK, corresponding to the N-terminus of bovine cytochrome *c*. These results demonstrate that N-terminal fragments can be effectively isolated from N-blocked proteins, and sequence analyses of these peptides can be performed at the pmol level, by using the method described here. We also confirmed the effectiveness of this method for analysis of unblocked proteins, as shown in Fig. 4, which illustrates results obtained when lysozyme and cytochrome *c* from *S. cerevisiae* were used as model proteins. The main peaks in Figs. 4(c) and (d) were analyzed by MALDI-MS/MS, and analyses of their sequences were easily carried out by locating the y-series ions, which were predominant due to the CAF effect.

## N-Terminal analysis of proteins from a crude lysate of human epidermoid carcinoma cells

To demonstrate that this method is practically applicable, we attempted to isolate the N-terminal peptides and perform the *de novo* sequencing for several proteins contained in crude lysates of human epidermoid carcinoma cells. After proteins in the lysate had been separated by 2D SDS-PAGE in the range of pH 5–8, 34 protein spots were randomly excised from the gel stained with CBB (Fig. 5) and subjected to N-terminal sequence analyses by the method described above.

The N-terminal sequences for proteins in 14 of these spots were confirmed by MALDI-MS/MS analysis (Table 1). The





DOI: 10.1002/rcm

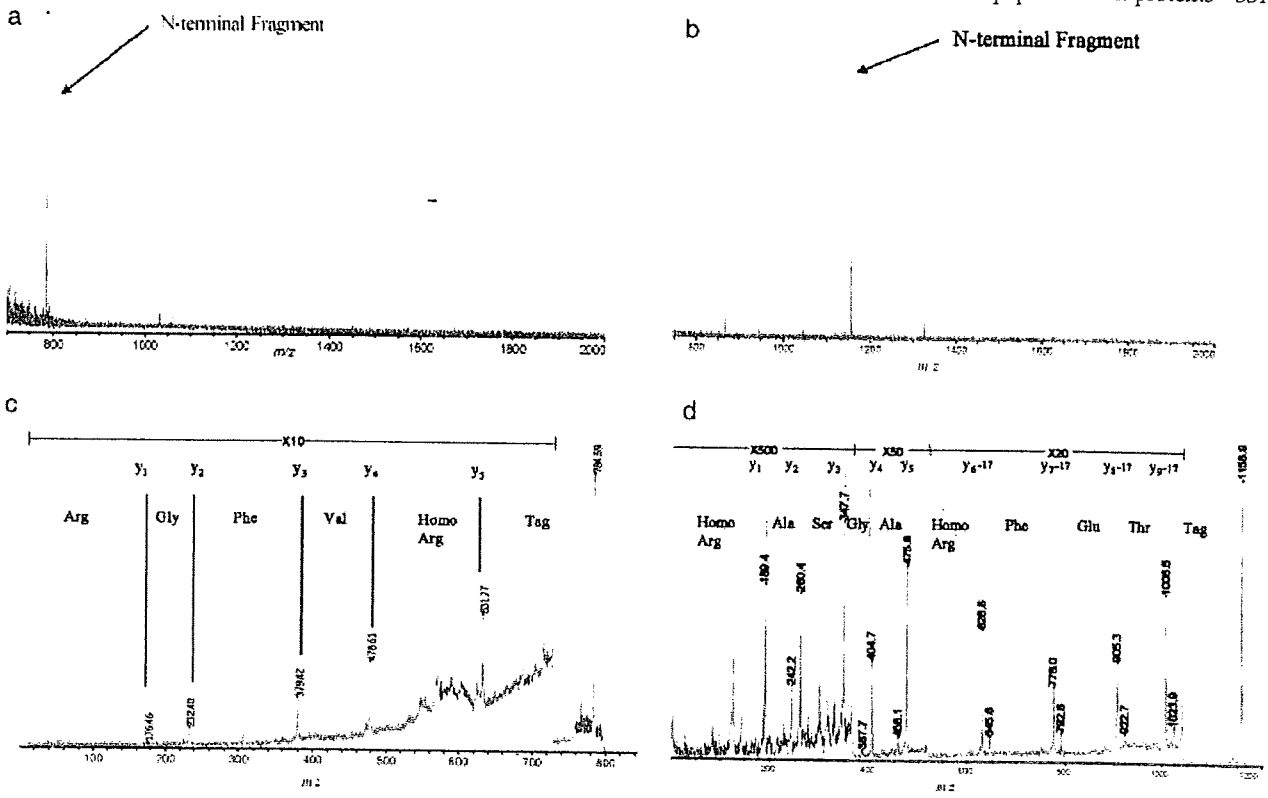


**Figure 4.** MALDI-TOF MS spectra of the N-terminal peptides selectively isolated from lysozyme (3 pmol) (a) and cytochrome *c* from *S. cerevisiae* (CTC-C: 3 pmol) (b). MALDI-MS/MS spectra of the N-terminal peptide from lysozyme (c) and CTC-C (d). Homo Arg denotes homoarginine.

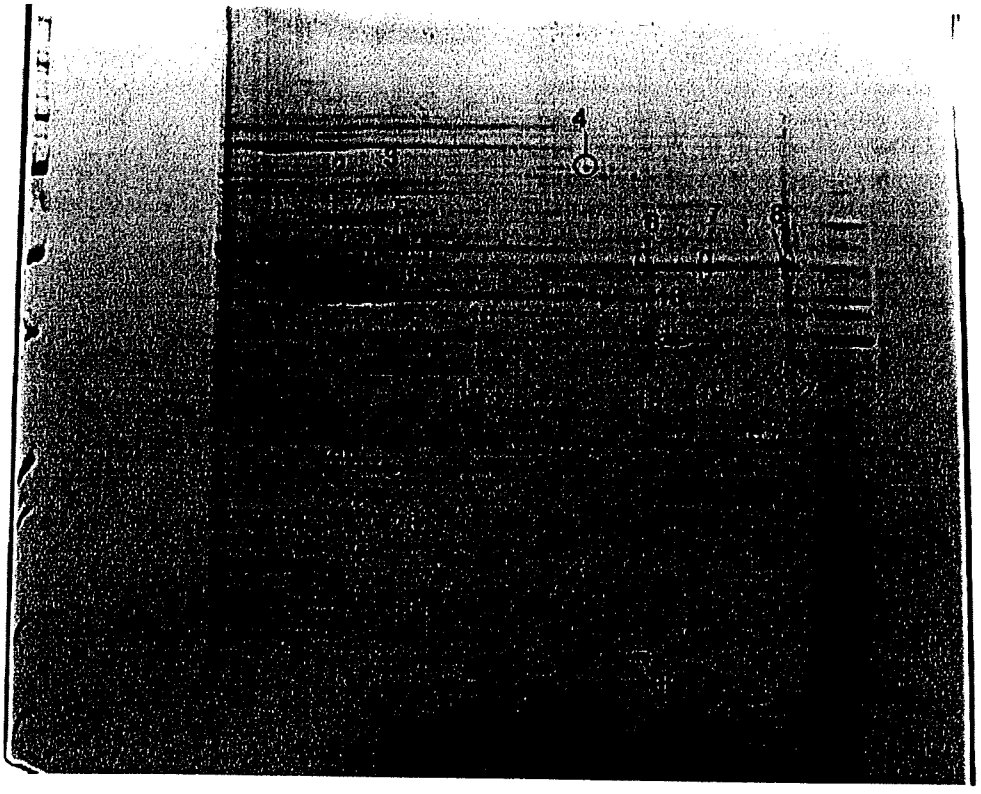

**Figure 5.** 2D SDS-PAGE profile of the human epidermoid carcinoma cell lysate. Spots indicated by numbers were excised and subjected to the N-terminal sequence analyses.



*Rapid Commun. Mass Spectrom.* 2008; 22: 3313–3319
DOI: 10.1002/rcm

**Table 1.** N-Terminal sequences of proteins isolated by 2D SDS-PAGE of the human epidermoid carcinoma cell lysate

| | | N-terminal sequence analysis | | |
|---|---|---|---|---|
| Spot No. | N-terminal peptide *m/z* | *de novo* | MS/MS ion search | Protein |
| 1 | 1041.6 | VJFGADAR | | Chaperonin (heat shock protein) |
| 2 | 1041.4 | VJFGADAR | | Chaperonin (heat shock protein) |
| | 1397.5 | AJDVJFGAD | | Chaperonin (heat shock protein) |
| 3 | 1041.6 | VJFGAD | | Chaperonin (heat shock protein) |
| 6 | 1021.9 | | Acetyl-SILKIHAR | phosphopyruvate hydratase (enolase) |
| 9 | 1069.7 | | Acetyl-ILASQQPDR | KIAA1840 |
| 13 | 1040.6 | | Acetyl-AMVSEFLK | Annexin A1 |
| 14 | 954.6 | | Acetyl-SGFSTEER | Pyrophosphatase 1 |
| 17 | 925.7 | | Acetyl-MDLLFGR | Chromatin-modifying protein 2a |
| 20 | 1130.8 | | Acetyl-AAYKLVLIR | Phosphoglycerate mutase |
| 28 | 1043.5 | | Acetyl-ATVQQLEGR | Fatty acid-binding protein |
| 30 | 1243.8 | | Acetyl-ADEELEALRR | Programmed cell death protein 5 |
| 31 | 1578 | | Acetyl-SQAEFEKAAEEVR | Acyl-CoA-binding protein |
| 32 | 1302.9 | | Acetyl-AkISSPTETDR | calgizzarin |
| 34 | 795.5 | mQiFVJ | | Ubiqutin |

The N-terminal sequences confirmed by *de novo* sequence analysis or MS/MS ion search are listed. m, i and J correspond to Met or Tyr, Leu or Ile, and homoarginine, respectively.

apparent success rate for determining these sequences may be somewhat low. As one possible explanation, we would like to mention the incomplete amino acid sequence coverage of available databases. Post-translational processing, including protein splicing and alternative translation, is still not sufficiently accounted for at the present time, although these defects could be compensated for in the near future by carefully analyzing many proteins that are expressed in cells. For the N-terminal analyses carried out in this study, we also would like to emphasize the fact that 10 peptides of the 15 N-terminal identified peptides were acetylated at their N-termini. The relative success rates in identifying N-acetylated proteins relative to unblocked ones in the present study are consistent with estimated proportions of N-acetylated proteins in a typical eukaryotic cell.[2] Figure 6 shows MALDI-TOF MS and MS/MS spectra of a protein from spot No. 31 as an example. This protein was eventually identified as Acyl-CoA-binding protein after its N-terminal sequence was determined to be Ac-SQAEFEKAAEEVR. In rare cases where the N-terminal residue is Met or Tyr, including in the N-terminal peptides isolated, their sequence analyses by this method might yield ambiguous results. Sequencing of a protein corresponding to spot No. 34 provides an example of this. The main peak (*m/z* 975.5) detected in the MALDI-TOF MS spectrum (Fig. 7) was further analyzed by MALDI-MS/MS. The product ion (*m/z* 839.9) lacked an N-terminal derivative (*m/z* 136.8) and y-series ions were thus predominantly observed from the precursor ion at *m/z* 975.5, so the sequence analysis could be performed *de novo*. Although the sequence thus identified ((M/Y)Q(I/L)FVJ, J: homoarginine) was found to be in agreement with the N-terminal sequence of ubiquitin (as reported in protein databases), we obtained no direct evidence that Met preceded the Tyr residue because the masses of methionine sulfone (which would be produced during the procedures) and Tyr are too close to distinguish between these residues.

As previously described,[7,8] in cases where the N-terminal peptides produced by trypsinization of proteins are too long

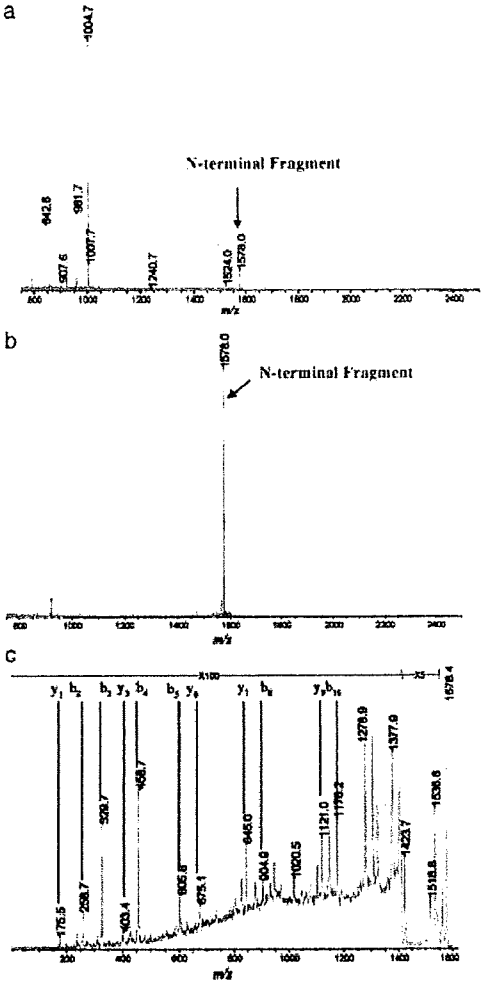


**Figure 6.** MALDI spectra obtained at each step during N-terminal sequencing of spot No. 31 after 2D SDS-PAGE of the human epidermoid carcinoma cell lysate. (a) MALDI-TOF MS spectrum after trypsin digestion. (b) MALDI-TOF MS spectrum of the N-terminal peptide specifically isolated from trypsin digests. (c) MALDI-MS/MS spectrum of the N-terminal peptide.




DOI: 10.1002/rcm

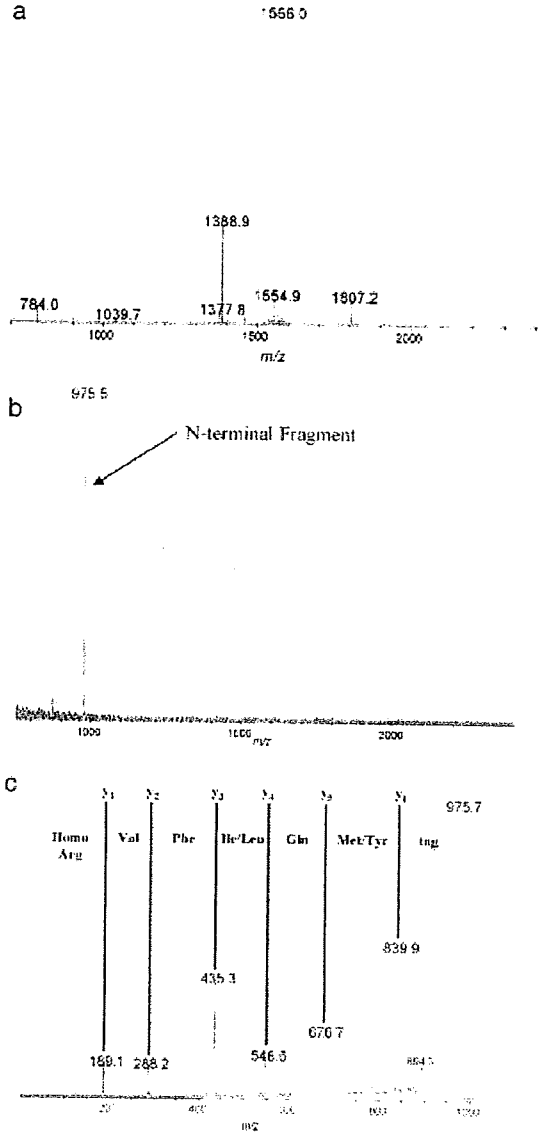


**Figure 7.** MALDI spectra at each step during N-terminal sequencing of spot No. 34 after 2D SDS-PAGE of the human epidermoid carcinoma cell lysate. (a) MALDI-TOF MS spectrum after trypsin digestion. (b) MALDI-TOF MS spectrum of the N-terminal peptides specifically isolated from the trypsin digests. (c) MALDI-MS/MS spectrum of the N-terminal peptide.

or too short, N-terminal analysis by MALDI-TOF MS and MS/MS may be carried out by using proteases with different specificities, such as Glu-C and Asp-N. The sequences of proteins in most publicly available databases correspond to the sequences of precursor proteins, as translated directly from cDNAs or from sequences in genome databases. They therefore do not reflect post-translational modifications such as protein processing and removal of initiator methionine residues. To increase the number of hits in determining the sequences of N-terminal peptides, it would be necessary to update databases, e.g. by including sequentially N-terminally trimmed Arg-C peptides.[3] Alternatively, search engines could consequentially search the possible modifications of N-terminal sequences deduced from precursor (DNA) sequences. Future methods to perform sequence analyses of N-terminal peptides will include improved identification of modified N-terminal residues. The method described in this report could be combined with more massive protein databases and more reliable search engines, providing powerful tools for improved analyses of protein sequences.

## CONCLUSIONS

We have developed a new method using MALDI-TOF MS to determine N-terminal amino acid sequences of proteins, regardless of whether their N-termini are modified. The principle of this method is based on previously published reports,[7,8] but the method developed here represents a significant improvement over earlier sequencing methods to enable the selective isolation of unblocking or blocking of N-terminal peptides for their *de novo* sequencing by MALDI-TOF MS. It represents a significant advance towards a universal method for N-terminal sequence analysis of proteins, regardless of whether or not their N-termini are blocked. This method is highly applicable to high-throughput N-terminal analysis of proteins, because multiple proteins can be analyzed in parallel. In this report we applied the method to determine the N-terminal amino acid sequences of 34 proteins that were chosen at random after 2D SDS-PAGE of a human epidermoid carcinoma cell lysate. Ten N-terminally acetylated and five unblocked proteins were successfully identified. We expect that this method will contribute to the improvement of databases used to determine N-terminal sequences of proteins, including N-blocked ones, which have proven difficult to determine using standard methods.[13]

## REFERENCES

1. Jiang H, Wang Y, Kanost M. *Biol. Chem.* 1994; **269**: 55.
2. Kokame K, Fukuda Y, Yoshizawa T, Takao T, Shimonishi Y. *Nature* 1992; **359**: 749.
3. Gevaert K, Goethals M, Martens L, Damme JV, Sates A, Thomas GR, Vandekerckhole J. *Nat. Biotechnol.* 2003; **21**: 566.
4. McDonald L, Robertson DHL, Hurst JH, Beynon RJ. *Nat. Method.* 2005; **2**: 955.
5. Akiyama TH, Sasagawa T, Suzuki M, Titani K. *Anal. Biochem.* 1994; **222**: 210.
6. Mikami T, Takao T. *Anal. Chem.* 2007; **79**: 7910.
7. Yamaguchi M, Nakazawa T, Kuyama H, Obama T, Ando E, Okamura T, Ueyama N, Norioka S. *Anal. Chem.* 2005; **77**: 645.
8. Yamaguchi M, Obama T, Kuyama H, Nakayama D, Ando E, Okamura T, Ueyama N, Nakazawa T, Norioka S, Nishimura O, Tsunasawa S. *Rapid Commun. Mass Spectrom.* 2007; **21**: 3329.
9. Kuhn K, Thompson A, Prinz T, Müller J, Baumann C, Schmidt G, Neumann T, Hamon C. *J. Proteomics Res.* 2003; **2**: 598.
10. Zhou C, Zhang Y, Oin P, Liu X, Zhao L, Yang S, Cai Y, Qian X. *Rapid Commun. Mass Spectrom.* 2000; **14**: 2348.
11. Keough T, Youngquist RS, Lacey MP. *Proc. Natl. Acad Sci.* 1999; **96**: 7131.
12. Cornish TJ, Citter RJ. *Rapid Commun. Mass Spectrom.* 1994; **8**: 781.
13. Edman P. *Arch. Biochem.* 1949 (July); **22**: 475.




DOI: 10.1002/rcm