A New Strategy for Biomarker Discovery

Based on the NBS Method

Makoto Watanabe
2008

Acknowledgements

The study in this thesis has been carried out under the direction of Professor Saburo Aimoto and Professor Osamu Nishimura of Institute for Protein Research, Osaka University. The author wishes to express my gratitude to their cordial guidance, invaluable discussion, and encouragements.

The author would like to thank Professor Morito Monden and Assistant Professor Ichiro Takemasa of the Department of Surgery and Clinical Oncology, Graduate School of Medicine, Osaka University for giving chances to work on proteomic analyses of the colon clinical samples. The author is grateful to Associate Professor Hiroki Kuyama, Guest Professor Susumu Tsunasawa, and Professor Emeritus Katsuya Nagai of Institute for Protein Research, Osaka University for excellent technical advice and useful discussion.

Furthermore, the author greatly thanks Dr. Ei-ichi Matsuo, Dr. Koichi Tanaka, and Dr. Susumu Iwasa of Shimadzu Corporation for helpful discussion.

Makoto Watanabe

August, 2008

Table of contents

Abbreviations

ACTH	adrenocorticotropic hormone
AHCY	S-adenosylhomocysteine hydrolase
3-CHCA	α-cyano-3-hydroxycinnamic acid
CID	collision-induced dissociation
CRC	colorectal carcinoma
DHB	2, 5-dihydroxybenzoic acid
ESI	electrospray ionization
HPLC	high pressure liquid chromatography
3H4NBA	3-hydroxy-4-nitrobenzoic acid
IHC	immunohistochemistry
LGALS1	galectin1
MALDI-TOF	matrix assisted laser desorption/ ionization time-of-flight
MS	mass spectrometry
NBS	2-nitrobenzenesulfenyl
QIT	quadrupole ion trap
RAN	ras related nuclear protein
RCN1	reticulocalbin1
2D-PAGE	two-dimensional polyacrylamide gel electrophoresis
VIM	vimentin
ZYX	zyxin

General Introduction

The human genome sequence has been completely decoded and since then the genomes of many other species have been sequenced. The focus of life science studies has accordingly been shifted to the investigation of structure and function of proteins. Given that the genome is the “blueprint of life”, proteins can be considered the “actual performers of life”. Therefore, elucidating the structures and function of proteins has become particularly important in life science research. The word “protein” is derived from the Greek word “*Proteios (protos)*” which means “champion “ or “of first rank”. This is very suggestive. As the relationship between “genome” and “gene”, the total body of protein expressed in an organism is called its “proteome”, and the research of its field is referred to as “proteomics”. The word “proteome” was coined by Mark Wilkins (1) and it has since been widely used.

Although genomic information is indispensable in studying life science, this information alone is not sufficient to successfully drive research in this field. The author believes that the breakthrough comes from studying proteins because in reality, proteins are the actual “players” in life. Information from proteome, which shows dynamism of proteins in biological systems, is becoming more important and indispensable for pushing research in life sciences ahead. To effectively drive life science research forward, a collaborative approach from genomic and proteomics studies is warranted.

Mass spectrometry has occupied the center of proteome research as the analysis technique of choice. The author recognizes that it was an epoch-making event in protein research when the Nobel Prize was awarded to Koichi Tanaka and J.B. Fenn for their development of soft desorption ionization methods for mass spectrometric analysis of biological macromolecules.(2) This opened the door to the rapid progress of research in the life science, particularly with regards to proteomics.

However, analysis of the proteome is difficult compared to that of the genome, owing to the labile nature of proteins, as well as the larger variety and wider dynamic range of their abundance in a cell or tissue system. Hence, proteomic analysis requires further technical advances.

With this in mind, the author devoted himself to develop novel method for proteomic differential display and to construct the “NBS Biomarker Discovery System”, in which the NBS (2-nitrobenzenesulfenyl) method is placed at the center of the system. This method enables differential protein expression analysis between two states, such as disease and normal states. It is also an important tool for the discovery of proteins associated with disease. The author applied this new system to biomarker discovery using tissues of colorectal carcinoma (CRC) for biomarker discovery and obtained successful results.

In the present thesis, the author describes the development of the NBS method, the construction of new proteome analysis system, and its application for clinical samples.

References

(1) Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I. *Electrophoresis **16***, *1090-1094* (1995)

(2) Tanaka K., Waki H., Ido Y., Akita S., Yoshida Y., Yoshida T. *Rapid Commun. Mass Spectrom. **2**, 151-153* (1988)

Chapter I: An approach to quantitative proteome analysis by labeling tryptophan residues

I-1 Introduction

Quantitative analysis of proteins has been achieved by a combination of protein separation with high-resolution 2D-PAGE (1) and tandem mass spectrometry based sequence identification. From the viewpoint of the efficacy of 2D-PAGE, however, it is labor intensive and difficult to automate. Furthermore, certain classes of proteins are excluded or under-represented in 2D-PAGE. With the development of modern high-performance mass spectrometry using ESI and MALDI, together with separation instrumentation, higher sensitivity and speed are achieved in the identification of proteins after separation with gel-based techniques. However, in the 2D-PAGE/mass spectrometry procedure, quantitative analysis of proteins is still hampered by the inherent flaws of gel-based separation (2). Today, the quantification of relative protein abundance in two proteome samples is an active area in proteomics. A growing number of reports have emerged in this field. The methods involving metabolic labeling of proteins with an isotopically defined growth medium (3), and that using the isotope-coded affinity tag (ICAT) (4), are typical examples. Also, a method of this kind designed specifically for phosphoproteins has been reported (5). The ICAT method differs from the metabolic labeling method in that a stable isotope is incorporated within the labeling agent for the cysteine residues. Very recently, a new method was reported to be capable of high accuracy and precision for protein abundance quantification (6, 7), which has promise as a general approach for quantitative profiling of isotopically labeled protein mixtures. A major difference between our approach and that based on the ICAT strategy is that the former focuses on the tryptophan residue, which is one of the least abundant and most hydrophobic amino acid residues but which plays an important role in biological systems. Our method employs one of the sulfenyl halide reagents, NBSCl (2-nitrobenzenesulfenyl chloride) (8, 9). This type of reagent

has been found to be mild as well as selective for chromophoric modification of tryptophan and cysteine residues in proteins and peptides (10-12). Of the two residues, labeling on tryptophan prevails over cysteine in acidic media (9). On the tryptophan residue, the NBS moiety is introduced at the 2-position of the indole nucleus (**Fig.I-1**). Another difference between the present approach and the ICAT method lies in the use of an avidin affinity column to isolate the resulting cysteine-labeled peptides with the ICAT reagent. In the case of the present method, the tryptophan residue is one of the most hydrophobic residues among those of the 20 amino acids (13). Therefore, tryptic fragments containing a tryptophan residue are hydrophobic and are known to have a relatively stronger affinity to Sephadex media;(13) further, fragments containing the NBS-labeled tryptophan are rendered even more hydrophobic and have still more tendency to be retained on the media, suggesting that isolation is possible with a Sephadex (e.g.LH-20) (13, 14) or reversed-phase column (14). In the former case, the labeled peptides are isolated more efficiently because the Sephadex medium has an affinitive function as well as a sizing function (14). The NBS method described here makes use of the Sephadex (LH-20) column for the separation of tryptophan-containing peptides after labeling. This approach, using the chemically modified and concentrated tryptophan-containing peptides for LC/MS/MS analysis, has generated considerable interest in the field of proteomics. This strategy using the NBS mass tag should be applicable to a wide range of proteomic samples because the presence of the NBS mass tag should not interfere with efficiency in CID in MS/MS experiments, or with the identification of the modified tryptophan-containing peptides using automated database search algorithms. Because the target residue for labeling in the NBS approach is tryptophan, which is less abundant than the cysteine residue, our method generates less complex mass spectra than the ICAT method, thus facilitating analysis. In this chapter, the author describes a method for quantitative

proteome analysis using a type of sulfenyl halide agent, NBSCl.

I-2 Materials and Methods

I-2-1 Chemicals and reagents

Rabbit glyceraldehyde-3-phosphate dehydrogenase, rabbit phosphorylase b, chicken ovalbumin, bovine α-lactalbumin, chicken egg-white lysozyme, iodoacetamide, and a-cyano-4-hydroxycinnamic acid (CHCA) were purchased from Sigma (St. Louis, MO, USA). Human ACTH (1-24) and rat galanin were from the Peptide Institute Inc. (Osaka, Japan). Presenilin-1 (331-349)-Cys (human, mouse) was obtained from BACHEM AG (Bubendorf, Switzerland). Two types of rat serum (GK/Crj and Crj: Wistar, from 5-week-old males) were purchased from Charles River Japan, Inc. (Yokohama, Japan). The sulfenyl halide agent, 2-nitrobenzenesulfenyl chloride (NBSCl-$^{12}C_6$: light form), was purchased from Tokyo Kasei Kogyo Co., Ltd. (Tokyo, Japan). The method of synthesis of the heavy form (NBSCl-$^{13}C_6$) will be published elsewhere. Tris (2-carboxyethyl) phosphine (TCEP) hydrochloride was obtained from Pierce (Rockford, IL, USA). Sephadex LH-20 was purchased fromAmersham Biosciences Corp. (Piscataway, NJ, USA). Acetonitrile, acetic acid, formic acid, and trifluoroacetic acid (TFA) were purchased from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). Hydrochloric acid, sodium dodecyl sulfate (SDS), tris(hydroxymethyl) aminomethane (Tris) and urea were purchased from Nacalai Tesque Inc. (Kyoto, Japan). Sequencing-grade modified trypsin was obtained from Promega (Madison, WI, USA). Water used in all experiments was purified using a MilliQ water purification system (Millipore Corporation, Bedford, MA, USA).

I-2-2 MALDI-TOFMS analysis

An AXIMA-CFR (Shimadzu/Kratos, Manchester, UK) instrument was used to obtain all of the MALDI-TOF mass spectra. The operating conditions were as follows: nitrogen laser (337 nm); reflectron mode; detection of positive ions. The instrument

incorporates a microchannel plate type ion detector. The acceleration potential was set to 20 kV using a gridless-type electrode. CHCA was used as matrix, and a saturated solution was prepared using 50% aqueous acetonitrile containing 0.05% TFA. A portion (2 μL) of each sample solution was mixed with 2 μL of the matrix solution, and an aliquot of 0.5 μL was immediately deposited on the target and dried completely.

I-2-3 ESI-MS analysis

An LCQ Deca XP ion trap mass spectrometer (ThermoFinnigan, San Jose, CA, USA) was used together with a Shimadzu LC-10ADvp solvent delivery system (Shimadzu Corporation, Kyoto, Japan). Proteolytic peptides were separated on a 200 mm×50mm fused-silica capillary column (Magic C18; Michrom BioResources, Inc. Auburn, CA, USA) using an HPLC procedure employing a 60 min linear gradient starting with 5% of mobile phase A and ending with 80% of A (A: 5% aqueous acetonitrile containing 0.1% formic acid; B: 95%aqueous acetonitrile containing 0.1% formic acid) at a flow rate of 5 mL/min. The peptides were dissolved in 1% aqueous formic acid to achieve a concentration of ca. 50 mg/mL. A 1 μL peptide mixture was pressure-loaded onto the column. Eluting peptides were analyzed using μLC/MS and μLC/MS/MS techniques. The intensities of signals from eluting peptide pairs were measured with the scanning mass spectrometer. To determine the amino acid sequence, the mass spectrometer was operated in a data-dependent MS/MS mode (16-18), in which a full-scan mass spectrum was followed by an MS/MS spectrum. The precursor ions were dynamically selected for CID based on their intensity in the MS scan. Peptides were identified by searching the MS/MS spectra against the appropriate databases using the computer program Sequest (19, 20). Search parameters included a static modification of +152.98Da (2-nitrobenzenesulfenyl moiety) on tryptophan and +57.02Da (amidomethyl moiety from iodoacetamide) on cysteine. The rat protein database from NCBI was

used in the search process.

I-2-4 HPLC separation

Retention time shifts of derivatized test peptides relative to non-derivatized peptides were measured using an HPLC system (LC-10ADvp solvent delivery system, Shimadzu Corporation) equipped with a 4.5_250mm C18 column (Shimpack VP-ODS; Shimadzu Corporation). A 50 min binary gradient with 0–100% of solvent B (100% acetonitrile) was used. Solvent A was 0.1M formic acid/triethylamine buffer (pH 4.5); a flow rate of 1 mL/min was used.

I-2-5 General procedure

This method used two mixtures containing sample proteins at different concentrations derived from two different cell states. Each mixture (protein content 100 mg each) was separately denatured using 15 μL of 0.1% SDS containing 5 mM EDTA solution (boiled for 5 min and left to cool down), then modified with a 20-fold molar excess of 2-nitrobenzenesulfenyl chloride (the heavy reagent for one and the light reagent for the other) in 35 mL of acetic acid (left for 4 h in the dark at room temperature). The two resulting labeled protein mixtures were combined, and excess reagent was removed (Sephadex LH-20, eluted with 30% aqueous acetonitrile), and then dried with a vacuum concentrator. The resulting residue was dissolved in 50 μL of 50 mM Tris buffer (pH 8.5) containing 0.1% SDS. To this solution were added 5 μL of 40mM TCEP (left for 30 min at 37℃), followed by 1 μL of 500 mM iodoacetamide (left for 45 min at ambient temperature in the dark). Trypsin (dissolved in 450 μL of 50 mM Tris-HCl buffer, pH 7.8, containing 5 mM $CaCl_2$) was added to the resulting mixture to give an enzyme-to-substrate ratio of 1:10 (w/w). The targeted tagged (tryptophan-containing) fragments in the tryptic peptide mixture were isolated using Sephadex LH-20, and

eluted with 30% aqueous acetonitrile. The modified peptides thus obtained were analyzed with MALDI-TOFMS or ESI-MS.

I-3 Results and Discussion

I-3-1 Principles of this approach

The NBS reagent exhibits selectivity for the indole ring of tryptophan, and can incorporate a six-fold stable-isotopelabeled benzene nucleus (shown along with its general reaction scheme in **Fig. I-1**). This method consists of the following sequential steps (**Fig. I-2**): (1) The tryptophan residues in a denatured protein sample representing one cell state are modified with the isotopically light form of the NBS reagent. Protein samples from two conditions or states are labeled separately with the heavy and the light NBS reagent, combined, digested, concentrated, and then analyzed by mass spectrometry. The equivalent residues in a sample representing a second cell state are modified with the isotopically heavy reagent. (2) The two samples are combined and passed through a Sephadex LH-20 column to remove excess reagent and related small molecules. (3) Disulfide bonds in the sample proteins are reduced, and the generated sulfhydryl groups are alkylated. (4) The sample is then enzymatically cleaved to produce peptide fragments. (5)The labeled (tryptophan containing) fragments are isolated using Sephadex LH-20 column chromatography. (6) In the last step, the isolated peptides are analyzed using MALDI-TOFMS or LC/ESI-MS. The innovation in this approach is based on the selective reactivity of 2-nitrobenzenesulfenyl chloride (8, 9) towards the indole ring of tryptophan. Such reactivity is well known for some types of arylsulfenyl halides, including 4-nitrobenzenesulfenyl chloride (10), 2,4-dinitrobenzenesulfenyl chloride (11), 2-nitro-4-carboxybenzenesulfenyl chloride (12), and 2-nitrobenzenesulfenyl chloride (NBSCl). Among these reagents, NBSCl has been frequently used and thoroughly investigated with respect to its near absence of reactivity with amino acid residues other than tryptophan and cysteine. Of these two residues, the reaction is somewhat selective (but not specific) for the tryptophan residue over the cysteine sulfhydryl group.

Therefore, the cysteine residue is labeled as well with the reagent, in which the sulfenyl moiety attaches to the residue by means of a disulfide bond.

However, if protein samples are processed as described above, any disulfide bonds that may have formed will be cleaved in the reduction step. To confirm this hypothesis the author used a peptide, presenilin-1 (331-349)-Cys (human, mouse), which contains both tryptophan and cysteine residues (**Fig. I-3**). This peptide (**Fig. I-3 (a)**, m/z 2252.3) was modified with the sulfenyl chloride following the above procedure. The molecular mass (**Fig. I-3(b)**, m/z 2558.5) was increased by 306.2 Da, which indicates double labeling (i.e. on both residues, 153.1 Da for each 2-nitrobenzenesulfenyl moiety). The doubly labeled peptide was then reduced and alkylated (iodoacetamide). The molecular mass (**Fig. I-3(c)**, m/z 2462.4) of the resultant peptide indicates that one NBS moiety was replaced by the amidomethyl moiety (57.0 Da). In the reduction stage using TCEP, the disulfide bond is cleaved but the sulfide bond remains intact. Thus, the author determined that the final peptide was derivatized with a NBS group on the tryptophan residue and an amidomethyl group on the cysteine residue. From this result, the author reached the conclusion that the specific modification on the tryptophan residue was successfully achieved using the sequential procedure described above.

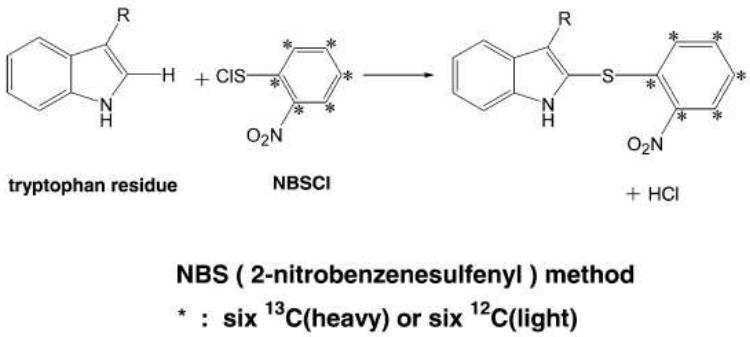


Figure I-1. Structure and reaction of the NBS reagent. The reagent exists in two forms, heavy (contains six ^{13}C) and light (contains no ^{13}C).

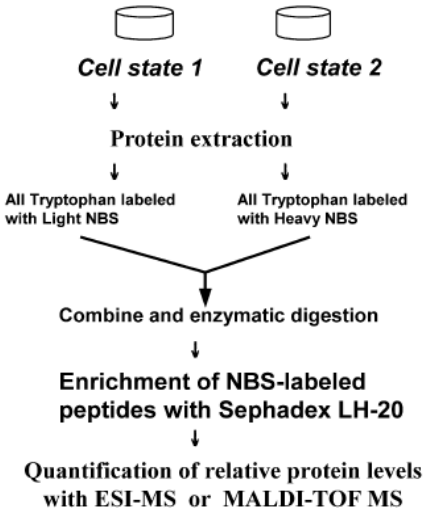


Figure I-2.
NBS strategy for quantifying differential protein expression.
Protein samples from two conditions or states are labeled separately with the heavy and the light NBS reagent, combined, digested, concentrated, and then analyzed by mass spectrometry.

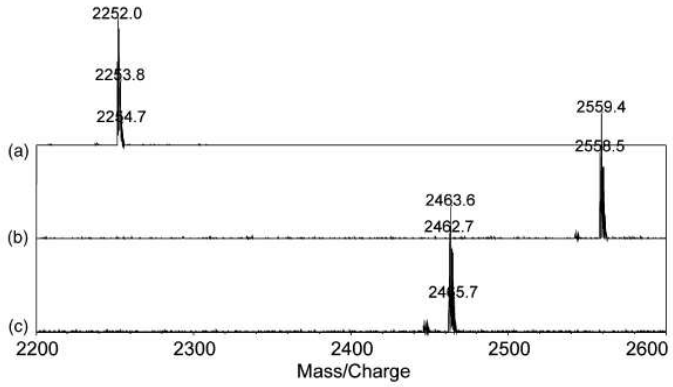


Figure I-3. Specific introduction of the NBS moiety onto tryptophan residues.
(a) Presenilin 1 (331–349)-Cys as the starting peptide; (b) a doubly derivatized starting peptide after treatment with NBSCl (light form); (c) the starting peptide with an NBS moiety on the tryptophan residue and an amidomethyl moiety on the cysteine residue after derivatization, reduction and alkylation.

I-3-2 Application of this method to standard mixtures

First, the author used human ACTH (1–24) and rat galanin as model tryptophan-containing peptides, and assumed two different cell states: state A consisting of ACTH/galanin.1:1 (molar ratio), and state B consisting of ACTH/galanin.1:2 (molar ratio). The molar ratios for each of the two peptides

individually, between the samples, are 1:1 (ACTH) and 1:2 (galanin). The peptide mixture representing state A was dissolved in 70% acetic acid. A 20-fold molar excess of 2-nitrobenzenesulfenyl chloride (light reagent) was added to the solution, and the mixture was stirred for 4 h in the dark at room temperature. The mixture representing state B was labeled with the heavy reagent using the same procedure. The two postlabeling solutions were combined and purified with a Sephadex LH-20 column or a reversed-phase HPLC system. The MALDI-TOF mass spectrum (**Fig. I-4**) shows that the observed ratios of the two components reflect the expected ratios well. The changes in their respective retention times were 16.57–34.78 min (ACTH-labeled ACTH) and 21.78–35.27 min (galanin-labeled galanin). The next application of this approach was performed with two mixtures containing the same four proteins at different concentrations (mg/15 mL): rabbit glyceraldehyde-3-phosphate dehydrogenase (12.5, 25); rabbit phosphorylase b (12.5, 25); chicken ovalbumin (12.5, 25); and bovine alactalbumin (12.5, 25). These two samples were processed sequentially according to the general procedure described above. The isolated labeled peptides were quantified by their peak intensities in the MALDI-TOF mass spectrum, and their sequences were assigned by comparison with the theoretically digested fragments of the four proteins using trypsin. The MALDI-TOF mass spectrum is shown in **Fig. I-5** and the assigned peptide fragments are summarized in **Table I-1**. A plurality of labeled peptides was encountered for each protein. The mean difference between the measured and predicted abundance ratios for light-labeled and heavy-labeled peptides (**Table I-1**) ranged between 2 and 12%, demonstrating the accuracy of quantification. If some of the proteins are to be non-expressed in one or other of the control and perturbed samples (e.g. the ratios of the proteins are 1:0 or 0:1 in the healthy cell and afflicted cell, respectively), the entire procedure should be performed separately for the two samples and analyzed using mass spectrometry.

The peptides of interest can then be discerned more clearly in the measured spectrum.

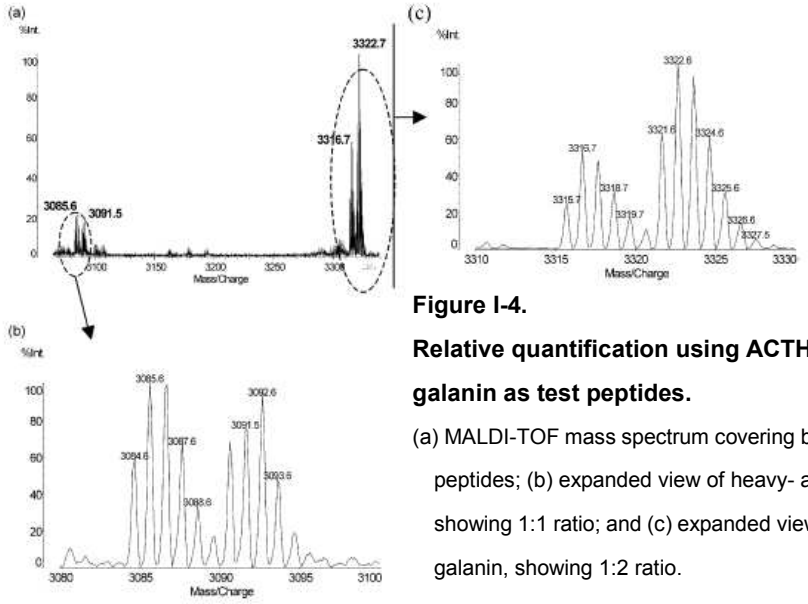


Figure I-4.

Relative quantification using ACTH and galanin as test peptides.

(a) MALDI-TOF mass spectrum covering both heavy- and light-labeled peptides; (b) expanded view of heavy- and light-labeled ACTH, showing 1:1 ratio; and (c) expanded view of heavy- and light-labeled galanin, showing 1:2 ratio.

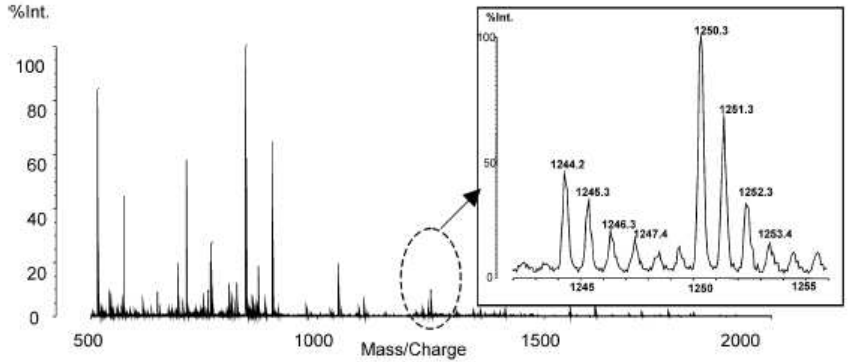


Figure I-5. Relative quantification of a tryptic peptide in a four-protein mixture.

The MALDI-TOF mass spectrum of tryptic peptides of the four-protein mixture is shown; the inset shows an expanded view of a pair of labeled peptides (from rabbit phosphorylase b, EWTRMVIR).

Gene name[a]	Peptide sequence identified[b]	Observed ratio ($^{13}C_0/^{13}C_6$)	Expected ratio	Mean ± SD	% Error
LCA_BOVIN	IWCK	0.48	0.5	0.44 ± 0.03	12
	LDQWLCEK	0.43			
	VGINYWLAHK	0.40			
OVAL_CHICK	GLWEK	0.45	0.5	0.54 ± 0.09	8
	GLWEKAFK	0.63			
G3P_RABIT	LWR	0.48	0.5	0.49 ± 0.01	2
	LISWYDNEFGYSNR	0.50			
PHS2_RABIT	WIR	0.48	0.5	0.45 ± 0.03	10
	EWTR	0.46			
	EWTRMVIR	0.42			

[a]Gene names are according to Swiss-Prot nomenclature.[21]
[b]Each tryptophan residue in the peptides shown is labeled with a NBS moiety.

Table I-1. Sequence identification and quantification of components of a four-protein mixture

I-3-3 Spike test

The spike test was performed to demonstrate that our method is applicable to biologically derived samples. Two spiked samples were prepared as follows: state A contained 300 ng (21 pmol) of egg-white lysozyme in 5 mL of rat serum (Wistar), and state B contained 600 ng (42 pmol) of egg-white lysozyme in 5 mL of the same serum. The serum (5 μL) was assumed to contain 0.28 mg of total protein, of which the albumin content was 0.14 mg (2.1 nmol). These two test samples were processed using the general procedure. In the MALDI-TOF mass spectrum, two signals, separated by the expected 6Da (m/z 1198.74, 1204.72) and derived from the added lysozyme, were clearly observed, and the intensity ratio (100:48) was in good agreement with the theoretical value (100:50; **Fig. I-6**).

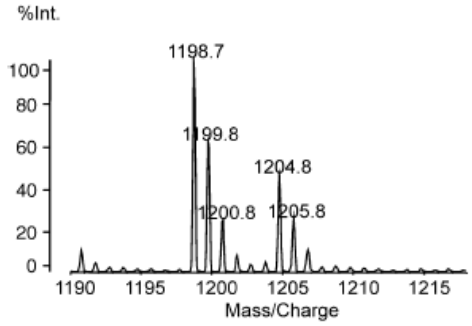


Figure I-6. MALDI-TOF mass spectrum of spiked sample. The spectrum shown is that of a derivatized synthetic tryptic peptide (GTDVQAWIR, 117-125 from egg-white lysozyme) which was added to rat serum (Wistar) as a spike.

I-3-4 Efficiency of Sephadex LH-20 isolation

The increment in the hydrophobicity of a tryptophan-containing fragment after modification using NBS reagent enabled us to employ Sephadex LH-20 for the isolation of labeled fragments. Two samples, each of 5 μL of Wistar rat serum, were used and assumed to represent two biological states. The two were processed using the general procedure described above. Half of the resultant peptide mixture was applied to the Sephadex LH-20 column (bed volume 2 mL). The eluent was 30% aqueous acetonitrile and the fractionation volume was 130 μL. The elution profile is partly shown in **Figs. I-7(b)–(e)**, with the paired peaks clearly discernable with a 6 Da

difference in **Figs. I-7(c) and (d)**; the target tryptophan-labeled peptides eluted between fraction numbers (19-26).

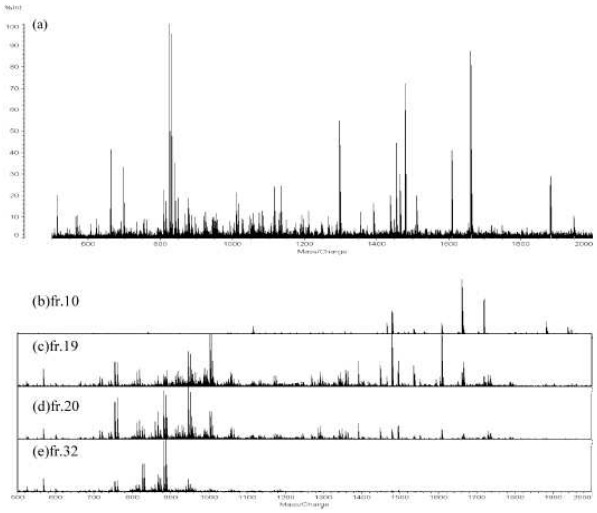


Figure I-7. Elution profile of derivatized Wistar rat serum using LH-20.

(a) MALDI-TOF mass spectrum of the tryptic peptide mixture before fractionation and concentration with LH-20.

(b)–(e) MALDI-TOF mass spectra for some LH-20 fractions. Pair peaks (6 Da separation) are clearly seen for fractions 19 and 20.

I-3-5 Co-elution of heavy- and light-labeled peptides in reversed-phase HPLC

The differential elution of the light- and heavy-isotope labeled peptides could cause significant difficulty in identifying or quantifying the peptide pairs. For this reason, the author investigated the elution profile of NBS-labeled (light and heavy) peptide fragments. The peptide used was a tryptic fragment of egg-white lysozyme, GTDVQAWIR (117–125), which was synthesized by the Fmoc method using a PSSM-8 peptide synthesizer (Shimadzu Corporation). The peptide was then modified to give the same amount of light- and heavy-labeled peptides. These two samples were mixed and passed through the Sephadex LH-20 column. The resultant peptide mixture was analyzed using reversed-phase HPLC as described in “Section I-2-3”. The heavy- and light-labeled peptides co-eluted (**Fig. I-8**), and sequencing using the MS/MS technique was successfully achieved (**Fig. I-9**).

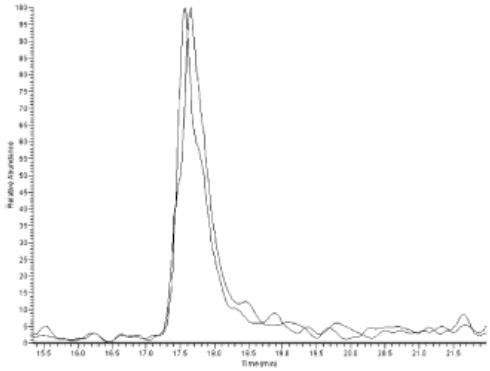


Figure I-8. Co-elution LC/MS profile of a derivatized tryptic peptide (GTDVQAWIR, 117-125 from egg-white lysozyme).
Heavy- and light-labeled peptides practically co-elute from a Magic C18 (Michrom BioResources, Inc.) fused-silica capillary column.

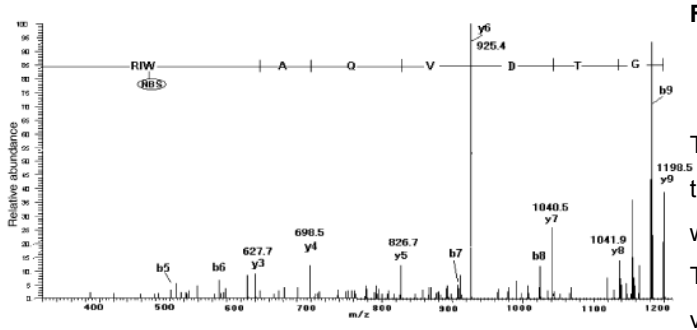


Figure I-9. Sequence identification of a tryptic peptide (GTDVQAWIR, 117-125 of egg-white lysozyme).
The example shown is the identification of the amino acid sequence for the peptide for which the abundance was measured in Fig.I-6. The y-ions are shown together with their m/z values; b-ions are also annotated.

I-3-6 Application of the method in quantitative analysis of protein expression under different conditions

The NBS method was applied to study the differences in protein expression in rat serum (Rattus norvegicus). The author compared protein expression in the serum of a control rat (Crj: Wistar) and a hyperglycemic rat (GK/Crj). The blood glucose level of the hyperglycemic rat was 372 mg/dL. A 2-μL sample (total protein content 100 μg) of each serum was treated using the procedure described in “Section I-2-5” (light-labeling for the serum from Wistar, heavy-labeling for GK). The isolated labeled peptides were quantified and sequenced by μLC/MS and μLC/MS/MS experiments using an ESI iontrap mass spectrometer. Three peptides were sequenced as follows: (1) AWAVAR (m/z 826.5, 832.5); (2) WKIRK (m/z 883.5, 889.5); and (3) NTAAWAK (m/z 914.5, 920.4). The observed intensity ratios (Wistar/GK) of the three peptides, calculated from the integration of the corresponding extracted mass chromatogram peaks, were all

close to 1. The identified proteins were albumin, purinergic receptor P2X, and transferrin, respectively. The ESI mass spectrum (range m/z 650–1050) obtained for the LC peak for the two labeled forms of the peptide AWAVAR (m/z 826.5 and 832.5 for peptides in which the tryptophan residue is respectively light- and heavy-NBS-labeled), is shown in **Fig. I-10(a)**. Reconstructed mass chromatograms for the peptide pair (m/z 826.5 and 832.5) are shown in **Fig. I-10(b)**, where the peaks appear serrated because the mass spectrometer switched between the MS and the MS/MS modes once per second to generate the CID mass spectrum of a selected peptide ion for the identification. This methodology has been developed to avoid SDS-PAGE separation because of its inherent drawbacks, as stated above. However, in treating complex protein mixtures (e.g. rat serum), pre-separation seems to be desirable. In this regard, the author is now investigating the effectiveness of pre-separation procedures (e.g. SDS-PAGE, etc.) in order to obtain better results.

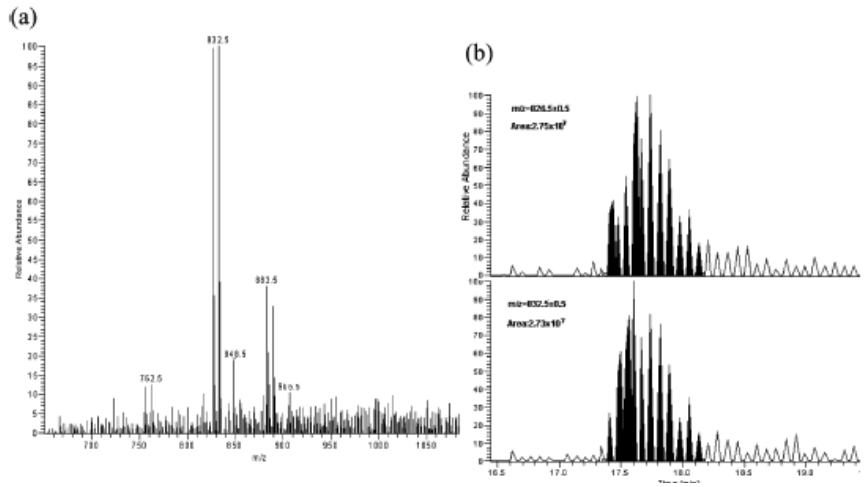


Figure I-10. Quantitative analysis of a protein from a rat serum sample.

(a) Full-scan (m/z500–1000) mass spectrum obtained at retention time 17.40 min in the mLC/MS/MS mixture analysis. Three different peptide doublets (6 Da separations), eluting from the column, are shown. (b) Expanded view of extracted mass chromatograms, showing the relative abundance of the peptide pair (m/z 826.5, 832.5). The ratio of the calculated areas (0.99:1) was used to determine the relative peptide quantities.

In conclusion, the author demonstrated the feasibility of using the NBS reagent for quantitative analysis of protein mixtures and identification of the protein components. The potential of this method lies in the following aspects: (1) the ability to specifically introduce the NBS moiety onto the tryptophan residue; (2) the efficient and simplified isolation of NBS-labeled peptides owing to their highly hydrophobic nature; (3) the co-elution of the isotopically light- and heavy-labeledpeptides; (4) the accurate quantification of a complex protein mixture using the stable isotope dilution approach; (5) successful sequence identification owing to the compatibility of the NBS label with this technique; and (6) the reduction of data complexity because of the relative scarcity of tryptophan residues in proteins. Several methods have been published using isotope dilution for protein quantification. Each method has its strengths and limitations. Thus, the ^{15}N-labeling method is only applicable to cells grown in culture. The ICAT method may have difficulties if applied to other separation techniques because it was developed specifically for use with affinity purification; also the labeling on cysteine residues implies that it is not applicable to cysteine-free proteins. The present NBS method has its own strengths and limitations. Tryptophan-free proteins cannot be the target of this method. The scarcity of tryptophan residues in proteins might in turn be the cause of poor results, in particular with proteins of low abundance. To overcome these limitations of each methodology, and to obtain more reliable outcomes, one method should be employed in combination with another method, or with conventional separation techniques such as SDS-PAGE. The present method presents a universal approach for the analysis of biological systems and pathways in cells, tissues, and organisms.

In this chapter, the author developed a novel quantitative proteome technique (the NBS method) and demonstrated that it is useful for quantitative proteome analysis. However, the procedure had to be optimized for the improvement in the efficiency of the

analysis. Then, in Chapter II, the author optimized the protocol of the method and improved the efficiency of enrichment for NBS-labeled peptides.

References

1. O'Farrell PH. J. Biol. Chem. **250**: 4007-4021 (1975)

2. Gygi SP, Rist B, Aebersold R. Curr. Opin. Biotechnol. **11**: 396-401 (2000)

3. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. Proc. Natl. Acad. Sci. USA. **96**: 6591-6596 (1999)

4. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. NatureBiotech **17**: 994-999 (1999)

5. Weckwerth W, Willmitzer L, Fiehn O. Rapid Commun. Mass Spectrom. **14**: 1677-1681 (2000)

6. Vogt JA, Schroer K, Hoelzer K, Hunzinger C, Klemm M, *et al.*, Rapid Commun. Mass Spectrom. **17**: 1273 (2000)

7. Cahill MA, Wozny W, Schwall G, Schroer K, Hoelzer K, *et. al.*, Rapid Commun. Mass Spectrom. **17**: 1283 (2000)

8. Scoffone E, Fontana A, Rocchi R. Biochem. Biophys. Res. Commun. **25**: 170- (1996).

9. Scoffone E, Fontana A, Rocchi R. Biochemistry **7**: 971- 1968

10. Zincke Th, Leinhardt S. Ann. Chem. **400**: 2 (1913)

11. Kharash N, Gleason GI, Buess CM. J. Am. Chem. Soc. **72**: 1796- (1950)

12. Havlik AJ, Kharash N. J. Am. Chem. Soc. **77**: 1150 (1955)

13. Tanaka S, Mohri N, Kihara M, Ohno M. J. Biochem. (Tokyo) **97**: 1377 (1985)

14. Sasagawa T, Titani K, Walsh KA. Anal. Biochem. **134**: 224 (1983)

15. Waki M, Izumiya N. Bull. Chem. Soc. Jpn. **41**: 1909 (1968)

16. Link AJ, Hays LG, Carmack EB, Yates JR. Proc. Natl. Acad. Sci. USA **93**: 14440 (1996)

17. Gygi SP, Rochon Y, Franza BR, Aebersold R. Mol. Cell. Biol. **19**: 1720 (1999)

18. Gygi SP, Han DKM, Gingras AC, Sonenberg N, Aebersold R. Electrophoresis **20**: 310-319 (1999)

19. Eng J, McCormack AL, Yates JR. J. Am. Soc. Mass Spectrom. **5**: 976 (1994).

20. Mann M, Wilm M. Anal. Chem. **66**: 4390 (1994).

21. Swiss-Prot. Available: www.expasy.ch.

Chapter II: Improved the NBS method: optimization of the protocol and improved enrichment for labeled peptides

II-1 Introduction

Both the feasibility of the NBS reagent for quantitative analysis and the potential of this method were shown in Chapter I ; however, the procedure has since been optimized. If the protocol is performed exactly as described in Chapter I, a considerable amount of the sample may be lost during the procedure. Furthermore, peaks that were increased by 57 m/z units from their original values were observed in the mass spectra, and several unlabeled peptides were eluted along with the labeled peptides. The author therefore revised and improved each reaction throughout the protocol. Because the use of detergents is often poorly compatible with trypsin digestion and/or mass spectrometric measurement,(1) the concentration of sodium dodecyl sulfate (SDS) was kept low (0.1%) in the original protocol (Chapter I). However, it is obvious that this concentration of SDS was too weak to denature all of the proteins in the sample, and it is also difficult to keep denatured proteins soluble. Thus, limitations of the original denaturing conditions resulted in some sample loss. Instead, the author used 8 M urea and/or 6 M guanidine hydrochloride (GdnHCl), both of which are strong denaturants widely used for various protein studies that can be easily removed by desalting or by reversed-phase liquid chromatography (RP-LC). This concentration range is strong enough to denature and solubilize most of the proteins, and is also high enough to keep those denatured proteins from aggregation. In addition, even a high concentration (4 M) of urea is compatible with trypsin digestion. (1) In the first half of this study, the sample loss and other defects seen in the original protocol were examined. In the original NBS method, targeted peptides having NBS-labeled tryptophan residue(s) were enriched by an LH-20 column. This procedure took advantage of the relatively stronger affinity of tryptophan-containing peptides for Sephadex media (LH-20), as well as additional changes in hydrophobicity and chromatographic mobility induced by the

reaction with NBSCl. (2) However, several non-labeled peptides were also co-eluted with the labeled peptides. Here, the author utilized a phenyl resin, which is generally used for hydrophobic interaction chromatography and RP-LC, in an attempt to enrich for the NBS-labeled peptides. Because the NBS-tryptophan moiety is not only hydrophobic but is also aromatic, π-electron interactions between phenyl groups of the column and the NBS-tryptophan moiety of the labeled peptides are likely to be more specific, and hence the use of a phenyl column should be advantageous for the enrichment step. In fact, the importance of π-π interactions in the special selectivity of a phenyl column for an aromatic solute, and in addition an increase in the selectivity achieved by nitro substitutions in aromatic rings, have been reported. (3) In the latter half of this study, the efficiency of enrichment for NBS-labeled peptides using a phenyl column was examined.

II-2 Materials and Methods

II-2-1 Chemicals and reagents

Chemicals and reagents used in this study were mostly the same as described in Chapter I. HiTrap phenyl FF (high sub), HiTrap phenyl HP and phenylsepharose HP were purchased from Amersham Biosciences (Uppsala, Sweden) and were used as a 'phenyl column'. Urea and guanidine hydrochloride (GdnHCl) were purchased from Bio-Rad (Hercules, CA, USA) and Sigma, respectively. The protease inhibitor cocktail and phosphatase inhibitor cocktails I and II were purchased from Sigma. Sample preparation '4 protein mix' and 'liver extract' samples Mixtures of proteins containing α-lactalbumin (bovine), glyceraldehyde-3-phosphate dehydrogenase (rabbit), lysozyme (chicken) and ovalbumin (chicken) (25 μg each, purchased from Sigma) were prepared and used as '4 protein mix' samples. Mouse (C57BL/6JJcl, 10 weeks, male) liver tissues purchased from CLEA Japan (Tokyo, Japan) were crushed to powder by a mortar and pestle, with the samples kept frozen by liquid nitrogen. A portion of the crushed powder was suspended in lysis buffer (50 mM TrisHCl (pH 8.0), 50 mM NaCl, 1 mM EDTA containing the protease inhibitor cocktail (1/200 v/v) and phosphatase inhibitor cocktail I and II (1/100 v/v each)) and homogenized using a sample grinding kit (Amersham Bioscience) on ice. The lysate was separated by centrifugation at 20,000 $\times$g for 45 min to yield a supernatant that was used as the 'liver extract' sample. Protein concentration was determined by a BCA kit (Pierce, Rockford, IL, USA) using bovine serum albumin as a standard protein.

II-2-2 For the "sample recovery assay"

Pairs of '4 protein mix' or 'liver extract' samples were suspended in 0.1% SDS solution containing 5 mM EDTA followed by boiling for 3 min, or suspended in 8 M urea solution containing 5 mM EDTA or 6 M GdnHCl solution containing 5 mM EDTA followed by vortex mixing at room temperature. Pairs of denatured proteins were labeled with

either NBSCl (heavy) or NBSCl (light) reagent, followed by mixing of heavy- and light-labeled samples and desalting with an LH-20 column. A portion of each sample was separated by sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS-PAGE).

II-2-3 For the "tryptic digestion optimization assay"

NBS-labeled and desalted samples prepared as in 'Sample recovery assay' were lyophilized and resuspended in 50 μL of 8M urea or 6M GdnHCl, each containing 50mM Tris-HCl (pH8.8) (the same denaturant as used in 'sample recovery assay' was used). After being reduced and alkylated, each sample was adjusted to a final concentration of 0.6, 1.2, or 2.0 M GdnHCl or 0.8, 1.6, or 4.0 M urea, and tryptic digestion (trypsin/substrate.1:20 w/w) was carried out in a 500 μL volume at 37℃ for 16 h. An equivalent sample containing 40 μg of each digest was separated by SDS-PAGE.

II-2-4 For the "+57 Da peak assay"

A pair of '4 protein mix' samples was labeled with NBS reagent and desalted according to the improved protocol using 8 M urea as denaturant. These samples were reduced and alkylated followed by tryptic digestion according to either the original protocol (in a 0.01% SDS solution) or the improved protocol (in an 8 M urea solution). Acetonitrile was added to a final concentration of 25% in the original protocol to improve the efficiency of digestion. Both tryptic digests were fractionated using a phenyl column according to the improved protocol.

II-2-5 For the "enrichment column evaluation assay"

A tryptic digest of the NBS-labeled '4 protein mix' samples was prepared according to the improved protocol using 8 M urea (in both the first denaturation and the second

denaturation steps). This digest, containing both NBS-labeled and non-labeled peptides, was fractionated using either a Sephadex LH-20 column (2 mL) or a phenyl sepharose HP column (1 mL) to enrich for the labeled peptides. For the LH-20 column, the digest was loaded onto a column pre-equilibrated with 30% aqueous acetonitrile and was eluted by continuous addition of 500 mL of 30% aqueous acetonitrile (Fr.1–Fr.10). For the phenyl column, the digest was loaded onto a column pre-equilibrated with 0.1% trifluoroacetic acid (TFA) and washed with 1 mL of 0.1% TFA three times (W1–W3). The labeled peptides were then eluted as 1mL fractions using a step-wise gradient of increasing acetonitrile (10, 15, 20, 25, 30, 35, 40%; EL1–EL7) containing 0.1% TFA. As eluted fractions from LH-20 contain both peptides and salts, desalting using a Zip Tip μ-C18 was performed, and 0.5 μL of each desalted sample was subjected to analysis by MALDI-TOF MS. In the case of the phenyl column, however, only wash fractions (W1–W3) contain salts, and these samples were desalted before MALDI-TOFMS analysis, whereas 0.5 μL of each eluted fraction (EL1–EL7) was directly analyzed by MALDI-TOF MS.

II-2-6 For the 'overall comparison assay'

A pair of 'liver extract' samples was labeled with NBS reagent, desalted, reduced, alkylated, digested by trypsin, and fractionated on an enrichment column according to the original NBS protocol[2] and the improved NBS protocol described here. Each fraction was concentrated ten-fold and 0.5 mL was subjected to MALDI-TOF MS analysis. Desalting using a Zip Tip μ-C18 was performed if needed.

II-2-7 SDS-PAGE

In the 'sample recovery assay', a portion of the desalted sample was lyophilized and

resuspended in SDS sample buffer (1% SDS, 50 mM Tris-HCl (pH 6.8), 50mM NaCl, 5% β-ME). In the 'tryptic digestion optimization assay', reactions were stopped by the addition of a nine times volume of acetone. The precipitates were washed with acetone and then solubilized in SDS sample buffer by boiling for 5 min. These samples were separated by SDS-PAGE using 12.5% acrylamide gels and the discontinuous buffer system of Laemmli to separate the '4 protein mix' samples, and 15% gels to separate the 'liver extract' samples. (4) The Precision Plus Dual Standard (Bio-Rad) was used as a molecular weight marker in electrophoresis. Protein bands were stained with Coomassie Brilliant Blue R-250. (4)

II-2-8 MALDI-TOFMS analysis

All of the mass spectra were obtained using a MALDI-TOF mass spectrometer, AXIMA-CFR plus (Shimadzu, Kyoto, Japan/Kratos, Manchester, UK). Detailed conditions were described in "Chapter I". CHCA and 3H4NBA were dissolved in 0.1% TFA containing 50% acetonitrile at the concentration of 10 mg/mL to make matrix solutions. The CHCA solution was used to analyze the '4 protein mix' samples, whereas a co-matrix containing a CHCA and 3H4NBA solution (1:1 v/v) was used to analyze the 'liver extract' samples.

II-2-9 Improved NBS procedure

Two sample mixtures containing the same set of proteins were prepared. Each mixture (100 mg of proteins) was dried separately and then denatured by dissolving in 25 mL of denaturation solution (6 M GdnHCl or 8 M urea, each containing 5 mM EDTA). After dissolving the dried samples by using a vortex mixer, the NBSCl heavy solution (0.17 mg of heavy reagent ($^{13}C_6$-NBSCl) in 25 mL of acetic acid) was added to one protein solution and the NBSCl light solution containing the light reagent ($^{12}C_6$-NBSCl)

was added to the second protein solution. These protein solutions were then incubated for 1 h at room temperature in the dark. The two labeled protein mixtures were then combined, and excess reagent and denaturant were removed using Sephadex LH-20. The protein mixtures (100 mL) were loaded onto an LH-20 column (500 μL) preequilibrated using 30% aqueous acetonitrile, and the flow-through was discarded. Then an additional 100 μL of 30% aqueous acetonitrile was loaded, and again the flow-through was discarded. The labeled proteins were eluted by adding another 200 mL of 30% aqueous acetonitrile. The eluate (200 μL) was lyophilized and resuspended in 48 mL of resuspension solution (6 M GdnHCl or 8 M urea, containing 50mM Tris-HCl (pH 8.8)). After adding 1 mL of 200 mM tris(2-carboxyethyl) phosphine hydrochloride (TCEP) as a mild reducing reagent, the solution was incubated for 30 min at 37℃, then 1 μL of 500 mM iodoacetamide (IAA) was added and the solution was incubated for 45 min at room temperature. Trypsin was dissolved in 450 mL of 50 mM Tris-HCl (pH7.8) containing 5 mM $CaCl_2$ and added to the reduced and alkylated protein mixture to give an enzyme-to-substrate ratio of 1:20 (w/w), followed by incubation at 37℃ for 4–16 h. NBS-labeled tryptophan-containing peptides in the tryptic digest were isolated using a phenyl column (HiTrap phenyl column or 1mL of phenyl sepharose HP) pre-equilibrated with a 0.1% TFA solution. Next, 5 μL of a 10% TFA solution were added to the sample and it was loaded onto the column, then washed three times with 1mL of 0.1% TFA. Elution was performed with 1 mL of the 0.1% TFA solution containing increasing concentrations of acetonitrile, from 10% to 40% with a 5% interval. Each eluted fraction was analyzed directly by MS.

II-3 Results and discussion

As mentioned above, this improved protocol differs from our original procedure primarily at two points. One is that we changed the denaturant from SDS to urea or GdnHCl,

and the second is that the enrichment is now on a phenyl column rather than an LH-20 column. These changes necessitated concomitant changes of reaction conditions for labeling of protein samples, reduction, alkylation, and tryptic digestion of the labeled proteins, and also for enrichment of the labeled peptides. To determine the best conditions for each reaction and to confirm the improved efficiency, the author now divides the protocol into three steps and examine each step separately. (**Fig. II-1**)

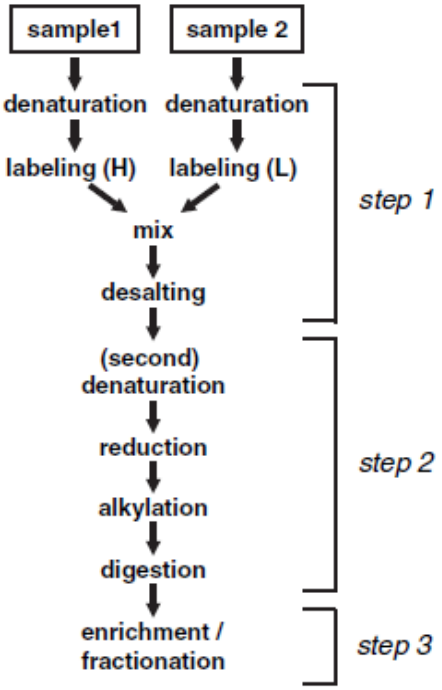


Figure II-1. Flow chart showing the NBS procedure. The procedure was divided into three steps and each step was separately optimized and evaluated for sample recovery (step 1), for efficiency of digestion and existence of the 57 Da modification (step 2), and for efficiency of enrichment for NBS-labeled peptides (step 3).

II-3-1 Improved sample recovery using urea or GdnHCl as denaturant

In step 1, the author examined the procedure from the denaturation of the protein samples to the desalting. (**Fig. II-1**) In the improved method, the reagent for solubilization and denaturation was changed from 0.1% SDS to 8 M urea or 6 M GdnHCl to keep the denatured proteins soluble and also to improve the labeling efficiency. The main purpose of the next desalting step is to remove excess labeling reagent. At the same time, exchange of buffer and/or solvent is also important to control the conditions for the following reactions. If these denaturing solutions are

used as a running solution in the desalting step, too much denaturant will remain in the eluted sample, because the volume at the reduction step is four times less than that of the desalted eluate. Furthermore, if a solution without denaturant is used as a running solution, the denatured proteins will aggregate and be lost. Therefore, a 30% aqueous acetonitrile solution, which can easily be evaporated but can keep denatured proteins soluble, was used as a running solution in the desalting step. As a result, the loss of samples was negligible in this step, regardless of whether the denaturant was urea or GdnHCl, and samples without salt or denaturant were thus obtained. (**Fig. II-2**) The labeling efficiency was almost complete (higher than 95%) and the reaction was almost completed in less than 10 min in the case of the '4 protein mix' (data not shown). It should be noted that one must use a freshly prepared urea solution made of high-grade reagent and never incubate at high temperature, otherwise carbamylation will occur and +43 m/z unit peaks will appear. (5) The use of thiourea should be avoided to solubilize protein samples, because the labeling reaction will be incomplete (probably because the NBS reagents are quenched by thiourea). One advantage of the NBS method is that it reduces the complexity of the analysis, because the number of tryptophan residues in one protein is small, and hence the number of the labeled peptides to be analyzed is also small. On the other hand, this means that accurate quantitation of each paired peak is required. By using this improved protocol, suppression of the sample loss was achieved, as mentioned above, and as a result the degree of quantitation was also improved (The standard deviation of ratios from a data set obtained using the improved protocol decreased to about a half to a third of values obtained using the original protocol; data not shown.).

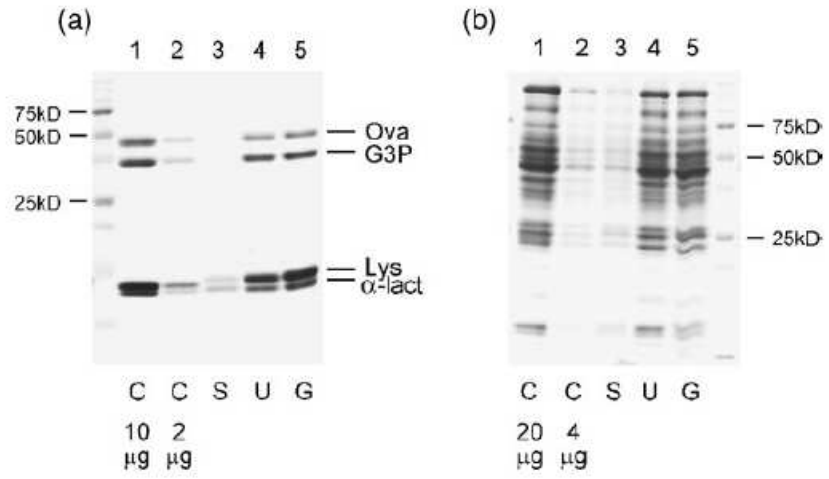


Figure II-2. Recovery of the protein samples after labeling and desalting.

(a) Protein samples were prepared using the '4 protein mix' as described in Sample preparation for 'sample recovery assay'. The denaturants used were 0.1% SDS (S; lane 3), 8M urea (U; lane 4), or 6M guanidine hydrochloride (G; lane 5), respectively. One twentieth of the desalted sample was applied to a gel and subjected to SDS-PAGE analysis (lanes 3–5). Starting samples corresponding to the same amount (10 mg; lane 1) and one-fifth of this amount (2 mg; lane 2) were also applied. It should be noted that migration of each protein was retarded, as the molecular mass increases upon NBS labeling. Abbreviations: Ova, ovalbumin; G3P, glyceraldehyde-3-phosphate dehydrogenase; Lys, lysozyme; a-lact, a-lactalbumin. (b) The same set of experiments was carried out using 'liver extract' as a starting sample. One-tenth of the desalted sample (lanes 3–5) and a corresponding starting sample of 20 mg (lane 1) and 4 mg (lane 2) were applied to a gel and subjected to SDS-PAGE analysis.

II-3-2 Optimization of the conditions for trypsin digestion

In step 2, the author examined the procedure from the second denaturation to tryptic digestion (**Fig. II-1**). The reaction conditions for reduction, alkylation and digestion of the labeled proteins were optimized. First, the desalted and dried sample proteins were re-solubilized using denaturing buffer (8 M urea or 6 M GdnHCl, containing Tris-HCl). In this condition, both the reduction of disulfide bonds by TCEP and the alkylation of free sulfhydryl groups by IAA were confirmed to occur satisfactorily (data not shown). However, under the same condition, proteases are denatured or their activities are affected by such a high concentration of denaturant, resulting in partial cleavages and/or mistakes in the recognition of cleavage sites. Therefore, the author

examined how many times the solution should be diluted to optimize the digestion. As shown in **Fig. II-3**, a final concentration of 0.6 M for GdnHCl and 0.8 to 4.0 Mfor urea seemed to be favorable. From the MS analysis, 0.8 and 1.6 M urea is recommended, because partially cleaved fragments and unidentified peaks were sometimes observed for 0.6 M GdnHCl and 4.0 M urea (data not shown). The author concluded that urea is an appropriate denaturant for tryptic digestion of proteins, and its optimum dilution is 5–10 times.

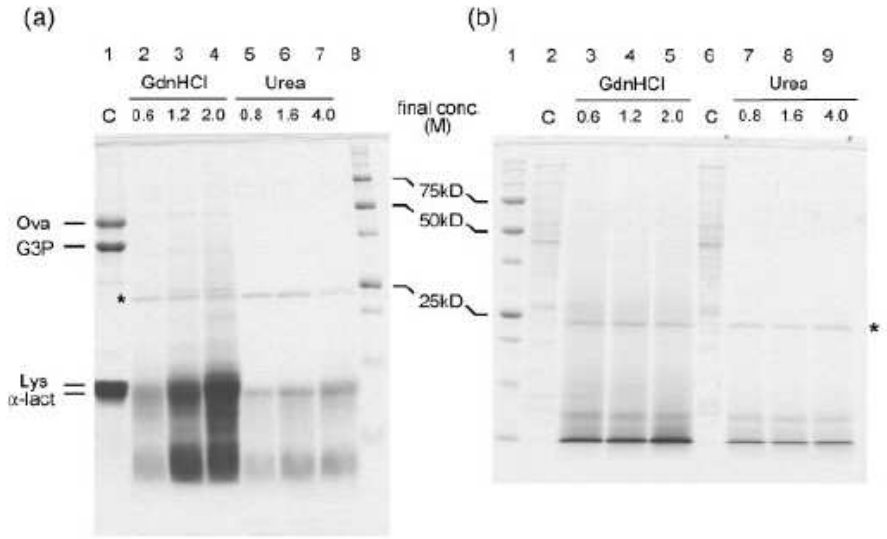


Figure II-3. Determination of optimum conditions for trypsin digestion. (a) Protein samples were prepared using the '4 protein mix' as described in Sample preparation for 'tryptic digestion optimization assay'. An equivalent sample containing 40 mg of each digest under the denaturing conditions shown in the figure was separated by SDS-PAGE (lanes 2–7); 10 mg of the undigested samples were also applied to the gels as a control (C; lane 1). Asterisk indicates the modified trypsin used for digestion. Abbreviations used are the same as in Fig. 2. (b) The same set of experiments was carried out using 'liver extract' as a starting sample.

II-3-3 Decrease in 57 *m/z* larger paired peaks caused by side reactions of the alkylating reagent

In mass spectra obtained from the original NBS protocol, paired peaks larger by 57 *m/z* units than the theoretical paired peaks were observed (**Fig. II-4**, upper lane). In the

original protocol, a step for removal or quenching of excess IAA, a reagent used for the alkylation reaction, was omitted. IAA reacts mainly with free sulfhydryl groups of cysteine, but is also known to react with histidine, lysine, tyrosine, and methionine under some conditions. (6) From this information and our observations, the author assumed that the peaks larger by 57 *m/z* units corresponded to the peptides in which an unexpected alkylation had occurred on a side chain other than the cysteine SH group. The author found that the appearance of these larger peaks could be suppressed by quenching with SH reagents or by removing the alkylation reagent by desalting or acetone precipitation (data not shown).

On the other hand, in mass spectra obtained from the improved NBS protocol, such paired peaks larger by 57 *m/z* units were no longer seen, and additional manipulations were unnecessary (**Fig. II-4**, lower lane). The author speculates that the amino groups of the denaturants urea or GdnHCl acted as buffers for the alkylation reaction, thus suppressing the side reactions of IAA.

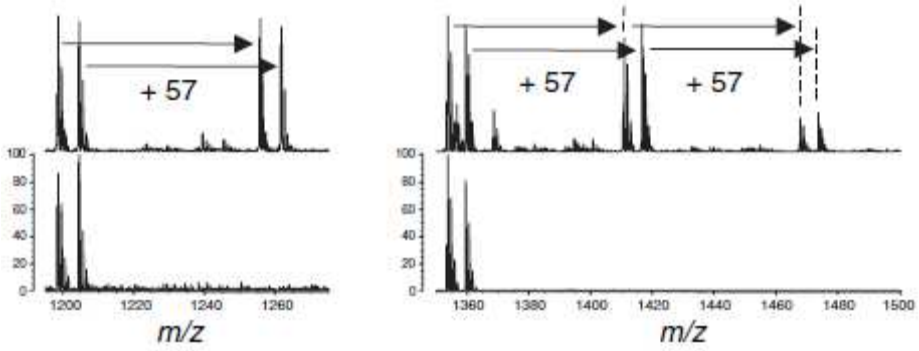


Figure II-4. Elimination of the byproduct peaks corresponding to molecules with mass increased by 57 Da from the original mass. Samples were prepared as described in Sample preparation for 'þ57 Da peak assay' using 0.01% SDS (upper panel) or 8M urea (lower panel), respectively, and were subjected to MALDI-TOFMS analysis. Parts of the mass spectra for the two eluted fractions corresponding to EL3 (left panel) and EL6 (right panel) are shown. Arrows indicate the peaks shift by 57 m/z units (upper panels). Paired peaks observed at m/z 1198.5 and 1204.5 in the left panel correspond to a peptide (GTDVQAWIR) derived from lysozyme, and those observed at m/z 1353.6 and 1359.6 in the right panel correspond to a peptide (VGINYWLAHK) derived from a-lactalbumin.

II-3-4 Improved enrichment for NBS-labeled peptides using a phenyl column

In step 3, the author examined the last part of the procedure, enrichment for labeled peptides (**Fig. II-1**). In the original NBS method, an LH-20 column was used for enrichment and many labeled peptides were eluted in the mid to late fractions (Fr. 4–9 in **Fig. II-5(a)**). However, some non-labeled peptides were co-eluted with the labeled peptides in the mid fractions (Fr.4–Fr.6 in **Fig. II-5(a)**). In addition, recovery of the labeled peptides was not sufficient, so peptides 3 and 9 were only marginally detected in Fr.6 and Fr.5, respectively.

Here, the author searched for other chromatographic media and tried to utilize a phenyl resin to enrich for the labeled peptides. As shown in **Fig. II-5(b)**, the enrichment of labeled peptides was apparently improved. Advantageous points of the phenyl column are listed below. First, contamination with unlabeled peptides was reduced. Most of the unlabeled peptides were eluted in the washing step (W1–W3), and the eluted fractions (EL1–EL7) contained almost exclusively the labeled peptides. This clearly contributed to the discrimination of single peaks as non-paired labeled peptides rather than unlabeled contaminating peptides. Second, recovery of the labeled peptides was increased. Even if only 'Elute' fractions (EL1–EL7) were collected, the recovery rate of the labeled peptides reached 80–90%, as determined by the NBS absorbance at 360 nm (data not shown). Third, as small molecules like salts and buffers were removed in the 'Wash' fractions (W1–W3), the eluted fractions (EL1–EL7) could be directly measured by MS. In the previous LH-20 purification, desalting was needed before measurement because the labeled peptides and salts were co-eluted. Fourth, individual NBS-labeled peptides were separated from each other to some extent. Using a capillary-type phenyl column, both the isolation of labeled peptides from unlabeled peptides and the fractionation between labeled peptides can be achieved at the same time. Fifth, the potential capacity of the phenyl column is expected to be

much higher than that of an affinity column, because a small ligand such as a phenyl group can be attached to the resin at a high concentration. This feature is more significant in the purification of peptides compared to that of proteins, because the mass ratio of the target molecule and tag moiety is relatively low. In fact, a similar level of purification was obtained, with the same efficiency and the same recovery rate, when four times the original amount of sample was loaded (data not shown).

The aromaticity of the NBS-tryptophan moiety is exploited in the phenyl column chromatography. One might think that peptides containing tyrosine and phenylalanine residues would preferentially interact with the phenyl resin and be eluted in the relatively late fractions. However, unlabeled peptides containing tyrosine and/or phenylalanine residues were all found in the 'Wash' fractions, and they were well separated from the labeled peptides (among the 14 peaks encircled by red dashed lines in **Fig. II-5**, five peaks corresponded to peptides containing two of these residues (Tyr and Phe) and six peaks to peptides containing one of them). The author mentions that the author has succeeded in making a monoclonal antibody specific to the NBS moiety (unpublished results). However, an affinity column made using this monoclonal antibody was not superior to the phenyl column in that its recovery rate was not higher. Therefore, use of the phenyl column is more efficient and economical.

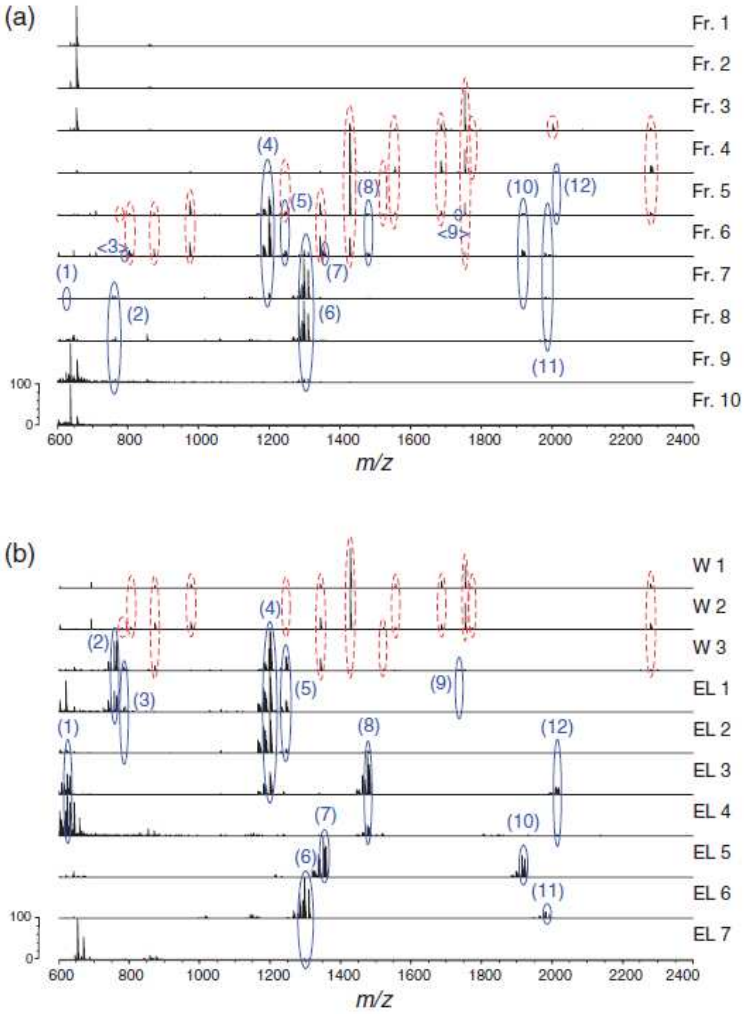


Figure II-5. Enrichment for the NBS-labeled peptides by Sephadex LH-20 versus phenyl sepharose. Samples were prepared as described in Sample preparation for 'enrichment column evaluation assay' using an LH-20 column (a) or a phenyl column (b), respectively. A portion of each eluted fraction (Fr.1–Fr.10) from the LH-20 column and each washed fraction (W1–W3) and eluted fraction (EL1–EL7) from the phenyl column was subjected to MALDI-TOFMS analysis. Paired peaks corresponding to NBS-labeled peptides are encircled by blue lines and single peaks corresponding to non-labeled peptides are encircled by red dashed lines. Numbers in parentheses indicate theoretical 12 NBS labeled peptides observed between m/z 600 and 2400. The monoisotopic masses of the light derivatives, the peptide amino acid sequences, and the proteins from which they originated are listed below. Numbers in angled brackets (<3> and <9> in (a)) indicate weak detection of the peaks. (1) 627.27, LWR (glyceraldehyde-3-phosphate dehydrogenase;G3P); (2) 759.29, EWTR (phosphorylase b); (3) 785.33, GLWEK (ovalbumin); (4) 1198.53, GTDVQAWIR (lysozyme); (5) 1244.51, LDQWLCEK (a-lactalbumin); (6) 1299.38, WWCNDGR (lysozyme); (7) 1353.64, VGINYWLAHK (a-lactalbumin); (8) 1478.62, GYSLGNWVCAAK (lysozyme); (9) 1734.71, LTEWTSSNVMEER (ovalbumin); (10) 1916.79, LISWYDNEFGYSNR (G3P); (11) 1981.78, IVSDGNGMNAWVAWR (lysozyme); (12) 2011.95, ELINSWVESQTNGIIR (ovalbumin).

II-3-5 Increase in the total number of observed paired peaks

Finally, the author compared the protocol described here with the original one in the analysis of mouse liver extract as a model biologically derived sample. **Fig. II-6** shows the mass spectra of three representative fractions from each protocol. It is clear that the total number of paired peaks detected was much larger and that the number of contaminated unlabeled peaks became much smaller when using the improved protocol described here (**Fig. II-6**). In fact, 12 identical paired peaks were observed in total from Fr.1–Fr.10 using the original protocol, whereas 68 identical paired peaks were observed in total from EL1–EL7 using the improved protocol (data not shown). The author observed several unlabeled single peaks that were co-eluted with labeled peptides even in the 'Elution' fractions (for example, see the asterisk in **Fig. II-6**). This is because, in the case of biologically derived samples, both the number of the proteins contained in the samples and the diversity among the peptide sequences was much larger. However, even using an avidin-biotin affinity purification system, it has been reported that some untagged peptides can co-elute in the elution fractions.(7) Regardless of this, it is clear that the efficiency of enrichment is much improved by the use of a phenyl column and that the NBS method is sufficiently sophisticated to be applied on a practical level. By further fractionating each eluate from the phenyl column using HPLC and a C18 capillary column, the author observed more than 1000 paired peaks in total derived from one sample (unpublished results).

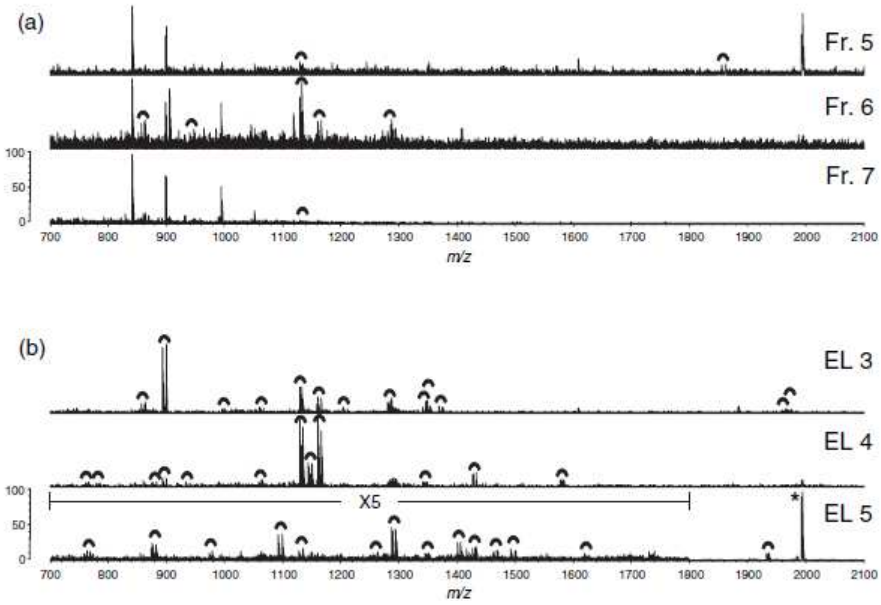


Figure II-6. Overall comparison of the original and improved protocols. Samples were prepared as described in Sample preparation for 'overall comparison assay' according to the original protocol (a) or the improved protocol (b), respectively, and subjected to MALDI-TOFMS analysis. Mass spectra of three fractions of each (Fr.5–Fr.7 in (a) and EL3–EL5 in (b), respectively) are shown. Thick half circles indicate NBS-labeled paired peaks with differences of 6 m/z units. The asterisk indicates a peak corresponding to a peptide with amino acid sequence VTYVDFLAYDILDQYR, derived from the tryptic digest of mouse glutathione S-transferase Mu 1.

In conclusion, The author has analyzed improvements to the NBS protocol and demonstrated the effectiveness of the new protocol. There were two basic improvements to the original procedure: the use of urea and/or guanidine hydrochloride as a denaturant and improved enrichment of labeled peptides by the use of a phenyl column. The improved protocol resulted in reduction of sample loss throughout the procedure (which should result in improvements in the sensitivity), reduction of peaks larger than 57 *m/z* units without any additional sample manipulation, improvement in the quantitative recovery, and reduction of overall working time. The use of a phenyl column resulted in reduction of contamination by unlabeled peptides in the elution fractions, increased recovery of the labeled peptides, elimination of the desalting step by differential elution of salts and labeled peptides, larger capacity compared to an affinity or antibody column, and fractionation of individual labeled peptides to some extent. In addition to the improvements detailed above, the derivatized MALDI peaks that were previously observed at positions lower by 16 or 32 *m/z* units, which seemed to be derived from the NBS moiety, can be suppressed by the use of the new matrix. (Chapter I) The use of this matrix as a co-matrix with 4-CHCA also improved the sensitivity with which NBS labeled peptides were detected. In conclusion, the combination of all of these improvements allowed us to achieve a much more accurate and thorough analysis of samples using the NBS method. However, it was necessary to improve the sensitivity of the NBS method further to analyze biological samples efficiently. Then, in Chapter III, the author describes a novel approach for enhancing the sensitivity and the selectivity of the NBS-labeled peptide detection by using a new matrix (3-hydroxy-4-nitrobenzoic acid) for MALDI-TOF MS analysis.

REFERENCES

1. Coligan JE, Dunn BM, Ploegh HL, Speicher DW, Wingfield PT. Current Protocols in Protein Science. John Wiley: New York.
2. Sasagawa T, Titani K, Walsh KA. Anal. Biochem. **134**: 224- (1983).
3. Marchand DH, Croes K, Dolan JW, Snyder LR, Henry RA, Kallury KMR, Waite S, Carr PW. J. Chromatogr. A; **1062**: 65 (2005)
4. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. Current Protocols in Molecular Biology. John Wiley: New York.
5. Stark GR. J. Biol. Chem. **239**: 1411- (1964)
6. Gurd FRN. Methods Enzymol. **25**: 424- (1972)
7. Hansen KC, Schimitt-Ulms G, Chalkley RJ, Hirsch J, Baldwin MA, Burlingame AL. Mol. Cell. Proteomics **2**: 299- (2003)

Chapter ///: Selective detection of NBS-labeled peptides by MALDI -TOF MS using a novel matrix

III-1 Introduction

In vitro labeling methods, such as the NBS method, are becoming increasingly useful for quantification of proteins expressed in cells. However, two problems still must be overcome. One is that the number of peptides to be analyzed is on the order of 105 or more after tryptic digestion of a cell lysates. (1) The second is that if there is an extreme difference in the amounts of a protein between two samples to be compared, the peaks will appear in mass spectra as a single peak instead of a pair of peaks labeled with either a heavy or light reagent. These false singlet peaks are difficult to distinguish fromunlabeled single peaks without sequencing, because more than 80% of the peaks correspond to unlabeled peptides. (1) Thus, an enrichment system for the labeled peptides is required to reduce the complexity of the analysis and to improve its efficiency. The author presents here a simple method to significantly improve the sensitivity of the NBS method, in which selective detection of labeled peptides is achieved by changing the matrix from CHCA to 3H4NBA, or to a comatrix consisting of 3H4NBA plus CHCA, for MALDI-TOF MS analysis. The predominant peaks detected in samples without purification using these matrices correspond to labeled peptides, which are clearly detected in samples after purification. This method also effectively suppresses peaks corresponding to unlabeled peptides, which were not removed during the enrichment process.

III-2 Materials and methods

III-2-1 NBS-labeling of model peptides

Ten micrograms of ACTH (5–10) peptide (EHFRWG; purchased from BACHEM, Bubendorf, Switzerland) was suspended in 50 mL of 70% acetic acid solution containing 20 equiv. of NBSCl ($^{12}C_6$-NBSCl, light form; purchased from Tokyo Kasei Kogyo, Tokyo, Japan), and incubated at room temperature for 1 h with gentle mixing. After the reaction, the NBS-labeled peptides were desalted by ZipTip μ-C18 (Millipore, Bedford, MA, USA) according to the manufacturer, and were eluted with 2 μL of 0.1% TFA solution containing 50% ACN twice. As a control sample the unlabeled peptide was prepared in the same way without NBSCl. Both peptides were mixed and subjected to MALDI-TOF MS analysis.

III-2-2 NBS-labeling of proteins followed by tryptic digestion

The mixture of proteins containing bovine a-lactalbumin, rabbit glyceraldehyde-3-phosphate dehydrogenase, chicken lysozyme, and chicken ovalbumin: (25 mg each, purchased from SIGMA, Steinheim, Germany) was labeled with NBSCl according to the improved method (17). In brief, each mixture containing 100 mg of proteins was denatured by an 8 M urea solution containing 5 mM EDTA and labeled with either $^{13}C_6$-NBSCl (heavy form) or 12C6-NBSCl (light form) at room temperature for 1 h, respectively. Next, both NBS-labeled (heavy and light) proteins were mixed and desalted. After lyophilization, both samples were dissolved in an 8 M urea solution (50 mM Tris-HCl, pH 8.8, 8 M urea), reduced, and alkylated with 4 mM tris (2-carboxyethyl) phosphine and 10 mM iodoacetamide, respectively, and then submitted to tryptic digestion. The digests were used for MALDI-TOF MS analysis with or without further purification of samples. For MALDI-TOF MS, 2 μg of tryptic digest was desalted with ZipTip μC18, and one-eighth of the eluate was analyzed.

III-2-3 Nitrotyrosine-containing peptides derived from nitrated lysozyme

One milligram of lysozyme (ca. 70 nmol) was reacted with a 20 times molar excess of tetranitromethane (270 μg, 1.4 nmol) for 1 h at 47C [9, 10]. The reaction mixture was passed through a column of Sephadex G-25 (1.0 cm 15 cm) to remove the tetranitroformate that was produced. The nitrated lysozyme was subjected to trypsin digestion and a portion of the digest corresponding to 14 μg was lyophilized and dissolved in 50 μL of a 0.1% TFA solution, desalted by ZipTip μC18, and then subjected to MALDI-TOF MS analysis.

III-2-4 MALDI-TOF MS analyses

Mass spectra were obtained by AXIMA-CFR plus, AXIMACFR (MALDI-TOF mass spectrometers; Shimadzu, Kyoto, Japan/Kratos, Manchester, UK), and AXIMA-quadrupole IT (QIT) (MALDI-QIT TOF mass spectrometer; Shimadzu/ Kratos). These mass spectrometers are equipped with a nitrogen laser emitting at 337 nm, and analyses are performed in the positive ion reflectron mode. External calibration was carried out using both ACTH (1–7) and bradykinin (18–39) (both purchased from SIGMA). On each well of the MALDI target plate, 0.5 mL of sample solution was spotted and 0.5 μL of matrix solution was spotted over the dried sample. Each matrix solution was prepared by dissolving each chemical in 0.1% TFA solution containing 50% ACN at a concentration of 10 mg/mL (2,5-dihydroxybenzoic acid (DHB), a-cyano-3-hydroxycinnamic acid (3-CHCA), CHCA, 3H4NBA, 4-hydroxy-3-nitrobenzoic acid (4H3NBA), 4-nitroaniline (4NA), 2-nitrophenol (2NP), and 2,5-dinitrophenol (25DNP)) or at a saturated concentration (,10 mg/mL: 4-nitrobenzoic acid (4NBA), 2-hydroxy-4-nitrobenzoic acid (2H4NBA), 2,4-dinitroaniline (24DNA), and 2-bromo-4,6-dinitroaniline (2B46DNA)). To prepare a comatrix solution of CHCA and 3H4NBA, equal volumes of each corresponding solution prepared as above were mixed.

All of the m/z values presented in this report are theoretical monoisotopic masses of single positive-charge ions, which correspond to the lowest-mass isotope.

III-2-5 Definition of selectivity index

We defined a “selectivity index” for each matrix as the relative signal intensity ratio of NBS-labeled peptide to nonlabeled peptides, based on the standard signal intensity ratio obtained using CHCA. The following formula is used for calculation of a selectivity index. For example, the selectivity index of 3H4NBA (SH) is determined as SH = Av. (LH/LC)/Av. (NH/NC). Here, each symbol represents the following: SH, the selectivity index of 3H4NBA; Av. (LH/LC), an average of LH/LC; Av. (NH/NC), an average of NH/NC; LH, absolute intensity of each NBS-labeled peptide detected using 3H4NBA; LC, absolute intensity of each NBS-labeled peptide detected using CHCA; NH, absolute intensity of each nonlabeled peptide detected using 3H4NBA; and NC, absolute intensity of each nonlabeled peptide detected by CHCA. The selectivity index of DHB (SD) is also calculated as SD = Av. (LD/LC)/Av. (ND/NC).

III-3 Results and Discussion

III-3-1 Analysis of NBS-labeled peptides using a MALDI-QIT mass spectrometer (AXIMA-QIT) and selected matrices

AXIMA-QIT is a MALDI-TOF mass spectrometer equipped with a QIT for MSn analysis. In this type of mass spectrometer, the molecules that have been ionized by the MALDI method are introduced into a quadrupole trap, and then isolation and collision of precursor ions followed by MS/MS analyses are performed. Therefore, the time required for a generated ion to reach the detector and undergo measurement is two or three orders of magnitude longer than that required by MALDI-TOF MS without IT equipment. Therefore, when a “hot” matrix such as CHCA is used for the analysis, the ionized molecules tend to have excess internal energy that causes them to gradually self-degrade while they are held in the trap. Because of this, it is recommended that DHB should be used as a “cool” matrix instead of CHCA. When NBS-labeled peptides derived from a tryptic digest of model proteins were analyzed by QIT using DHB as a matrix, the mass spectrum differed from that obtained by AXIMA-CFR using CHCA as a matrix. In the spectrum obtained using QIT, pairing peaks with a difference of 6 or 12 m/z units identified the paired peaks representing NBS-labeled peptides, but most of these paired peaks appeared at m/z values that were lower than the expected values by 14, 16, 30, or 32 m/z units, and this difference seemed to be a consequence of fragments of the nitro groups (**Fig. III-1A**). The use of DHB was deemed inappropriate for measuring NBS-labeled peptides for QIT, so the author searched for other cool matrices that could be used with QIT for the detection of NBS-labeled peptides. After evaluating several candidate compounds, the author found that 3-CHCA and 3H4NBA are effective matrices, and they detect NBS-labeled peptides with comparable sensitivity (**Fig. III-1-B and C**).

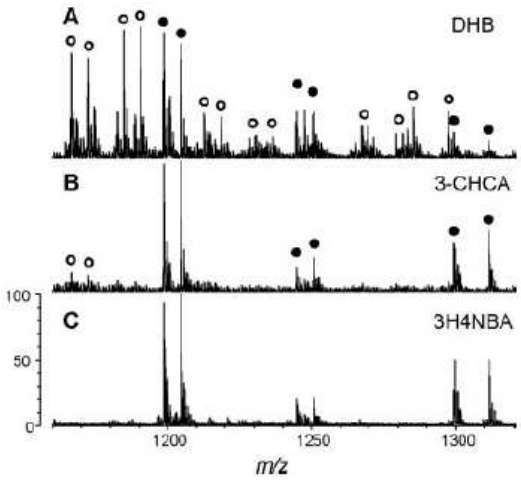


Figure III-1. Different effect of matrices on MALDI-QIT-TOF MS analysis of NBS-labeled peptides. NBS-labeled peptides were enriched from tryptic digests of a mixture of four model proteins (ovalbumin, glyceraldehyde-3-phosphate dehydrogenase, lysozyme, and a-lactalbumin) according to the NBS improved protocol, and a portion of the eluate was subjected to MALDI-QITTOF MS analysis with AXIMA-QIT. Matrices used were DHB (A), 3-CHCA (B), and 3H4NBA (C). Closed circles indicate the peaks corresponding to NBS-labeled peptides detected at the expected positions, and open circles indicate the peaks corresponding to fragments of NBS-labeled peptides.

III-3-2 Specific detection of NBS-labeled peptides using 3H4NBA as a matrix

The author examined the effectiveness of the matrices described above, including CHCA and DHB, in analyzing NBS-labeled peptides by AXIMA-CFR, a MALDI-TOF mass spectrometer without IT equipment. A mixture was prepared containing equivalent amounts of an NBS-labeled ACTH peptide (residues 5–10) and its unlabeled counterpart as test samples. When CHCA was used as a matrix, both labeled and unlabeled peptides were detected with almost the same intensity (**Fig. III-2B**). In the case of DHB, the unlabeled peptide gave the dominant signal, whereas the labeled peptide was hardly detectable, except for small derivatized peaks that were smaller than the expected m/z by 14 or 30 units, as shown in **Fig. III-2A**. This phenomenon was identical to what the author observed with QIT (**Fig. III-1A**). The incompatibility of DHB and NBS-labeled peptides is thought to result from MALDI ionization mechanisms, rather than from the IT technique. Both labeled and unlabeled peaks were detected with similar intensities using the 3-CHCA matrix, as in the case of CHCA matrix (**Fig. III-2C**). However, in the case of 3H4NBA, completely opposite results were obtained in comparison with the DHB matrix results (**Fig. III-2D**). The labeled peptide peak was dominantly detected, but the unlabeled peak and the 14 and/or 30 *m/z* units smaller

peaks were only weakly detected or were barely discernible. A similar tendency was observed for QIT (data not shown), showing that these selective ionizations and detections occur during the MALDI process, but not during ion trapping. On the basis of the above results, the author next analyzed tryptic digests of NBS-labeled proteins using 3H4NBA as a matrix. The sample was a peptide mixture derived from four proteins, as shown in Fig. 1, but the tryptic peptides were not purified to make the NBS-labeled peptides sample. A portion of the sample desalted with ZipTip was analyzed by MALDITOF MS. Both NBS-labeled paired peaks and unlabeled single peaks were detected using the CHCA matrix (**Fig. III-3B**), whereas unlabeled single peaks and labeled paired peaks that were smaller by 14 and/or 30 m/z units were mainly detected using the DHB matrix (**Fig. III-3A**). When the matrix 3H4NBA was used for detection, NBS-labeled paired peaks were observed almost exclusively, while unlabeled single peaks and the smaller labeled paired peaks were hardly observed at all (**Fig. III-3C**). These results suggest that selective detection of NBS-labeled peaks from a mixture of labeled and unlabeled peptides can be achieved by using 3H4NBA as a matrix. The author proposes that such specificity can be quantified by use of a "selectivity index" (details in Section 2). The selectivity index for each matrix was determined using four peaks (one for NBS-labeled and three for nonlabeled peptides) from Fig. 2 and 19 peaks (ten for NBS-labeled and nine for nonlabeled peptides) from Fig. 3: SH = 8.46 (Fig. 2) and 9.28 (Fig. 3), SD =0.043 (Fig. 2) and 0.22 (Fig. 3). That is, the selectivity index of 3H4NBA is nearly 10 whereas that of DHB is approximately 0.1. This indicates that 3H4NBA (or DHB) enables the labeled peptides to be observed at ten times larger (or smaller) intensity than CHCA, assuming that unlabeled peptides are detected at the same intensity by both matrices.

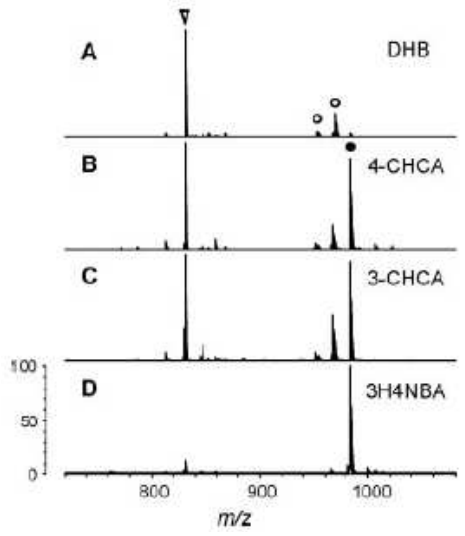


Figure III-2. Specific detection of NBS-labeled peptides by MALDI-MS analysis using the 3H4NBA matrix.

A mixture of 12C6-NBS (light)-labeled and unlabeled ACTH (5–10) peptides was measured by MALDI-TOF MS (AXIMA-CFR) using DHB (A), CHCA (B), 3-CHCA (C), and 3H4NBA (D) as a matrix, respectively. Marks indicated are as follows: closed circle, a peak corresponding to the NBS (light)-labeled ACTH (5–10) peptide (m/z 984.55); open circles, peaks that correspond to 16 or 32 m/z units smaller than NBS-labeled ACTH (5–10); open arrow head, peaks corresponding to the unlabeled ACTH (5–10) peptide (m/z 831.39).

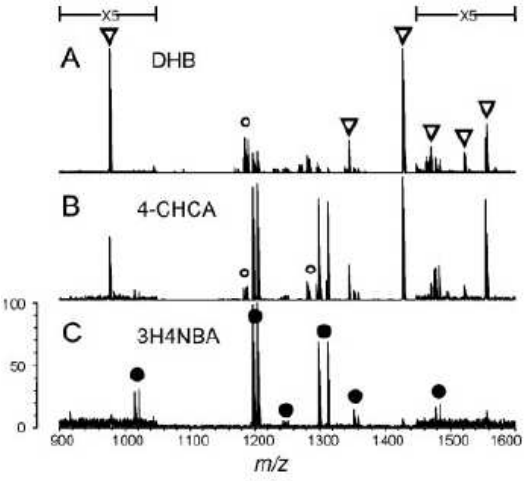


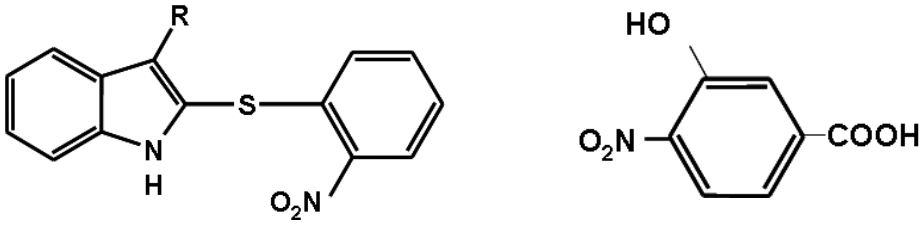
Figure III-3. Specificity of 3H4NBA for selective detection of NBSlabeled paired peaks in a mixed sample.

A mixture of NBS labeled and unlabeled peptides was prepared from the tryptic digest of four model proteins and was subjected to MALDI-TOF MS analysis. Matrices were DHB (A), CHCA (B), and 3H4NBA (C). Mass spectra ranging from m/z 900 to 1050 and from m/z 1450 to 1600 are amplified by five-fold along the y-axis. Closed circles indicate the paired peaks of NBS-labeled peptides having a difference of 6 or 12 m/z units at the expected positions, and open circles indicate the pairs of peaks derivatized from NBS labeled paired peaks. Open arrowheads indicate the unlabeled peptides.

III-3-3 Characteristics of nitrobenzene derivatives as matrices

In the MALDI process, the matrix plays a key role by incorporating and isolating analyte molecules in its bulk crystal; thereafter molecules are released into the gas phase concurrently with matrix desorption upon laser-energy absorption, and finally they are ionized by ion-molecule reactions in the laser plume. Therefore, a good combination of matrix and analyte and/or an appropriate preparation of the sample mixture are key requirements for a successful MALDI-TOF MS analysis (for review [11–13]). The author next searched for the basis of the selectivity of this specific combination of the NBS-labeled derivatives and 3H4NBA, on the basis of their structural aspects and noted that nitrobenzene is the common skeleton for both NBS and 3H4NBA, as shown in **Fig.**

III-4. To address whether this could be a key factor in the selective detection seen in MALDI-TOF MS analysis, the author examined other nitrobenzene derivatives as matrices. The author chose a number of mono- and dinitro aromatic compounds having functional groups, such as carboxyl, hydroxyl, amino, sulfate, or aldehyde or their combinations, that are capable of transferring protons in the MALDI ionization process. Our results indicated that many nitrobenzene derivatives could function as matrices, and they were effective in the selective detection of NBS-labeled peptides in MALDI-TOF MS analysis, although their sensitivities were rather low (**Fig. III-5**). Among the nitro-chemicals tested so far, 3H4NBA is the best matrix for detecting NBS peptides, due to its sensitivity and specificity.



NBS-Trp 3H4NBA

Figure III-4. Structures of the NBS-labeled indole ring of tryptophan and of 3H4NBA.
Moieties corresponding to nitrobenzene are highlighted in bold.

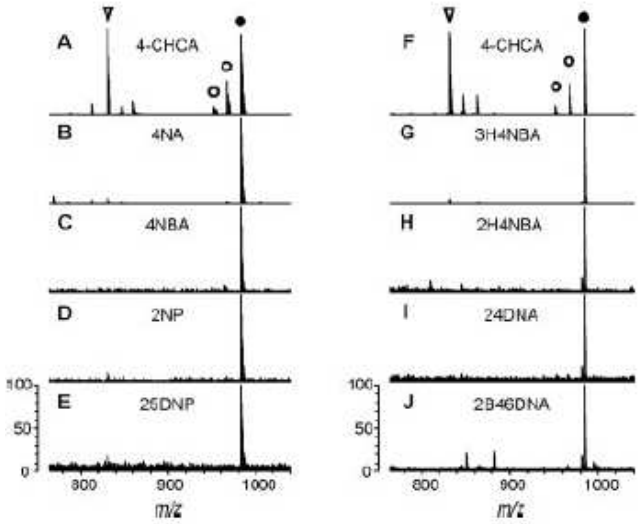


Figure III-5. Specific detection of NBS-labeled peptides using various nitrobenzene derivatives as a matrix.

4.5 mg of ACTH (5–10) peptide was labeled with 12C6-NBS (light) and purified on a C18 RP column (Michrome) using HPLC equipment (Shimadzu). NBS-ACTH (5–10) and unlabeled ACTH (5–10) were dissolved in a 0.1% TFA solution containing 50% ACN, respectively. A mixture of NBS-labeled and unlabeled ACTH (5–10) peptides (A–E, 2.5 pmol each; F–J, 0.25 pmol each) was prepared and subjected to MALDI-TOF MS analysis using CHCA (A, F); 4NA (B); 4NBA (C); 2NP (D), 25DNP (E); 3H4NBA (G); 2H4NBA (H); 24DNA (I); and 2B46DNA (J) as a matrix, respectively. Marks indicated are as follows: closed circle, a peak corresponding to the NBS (light)-labeled ACTH (5–10) peptide (m/z 984.55); open circles, peaks that correspond to 16 or 32 m/z units smaller than NBS-labeled ACTH (5–10); open arrowhead, peaks corresponding to the unlabeled ACTH (5–10) peptide (m/z 831.39).

III-3-4 Specific detection of a nitrotyrosine-containing peptide using the 3H4NBA matrix

The author next examined whether nitrobenzene-containing compounds other than NBS-labeled peptides can also be detected with specific selectivity using the 3H4NBA matrix. The peptides containing a nitrotyrosine (3-nitrotyrosine) residue, which is one of the most common nitrobenzene-containing substances that exists in living organisms, were examined to see whether they were selectively detected by MALDITOF MS analysis using 3H4NBA as a matrix. To create these types of peptides, lysozyme nitrated by tetranitromethane was digested with trypsin and prepared for analysis. In preparing the sample, the nitration reaction was intentionally not allowed to proceed to

completion, so the sample included nitrotyrosine-containing peptides, tyrosine-containing peptides, and other peptides without tyrosine residues. As a result, the mass spectrum obtained using 3H4NBA showed that the peaks corresponding to peptides containing nitrotyrosine residues were stronger than those of non-nitrated peptides, and the selectivity was the same or somewhat higher compared to the case of the NBS-labeled peptides (**Fig. III-6** lower panels). The author also noted that peaks smaller by 16 or 32 *m/z* units than the primary nitrotyrosine-containing peaks, which are apparently caused by loss of oxygen from nitro groups, as in the case of NBS-peptides, were also only weakly observed in the mass spectrum of 3H4NBA. Highly reactive nitrogen species (RNS) such as peroxynitrite react with phenyl groups of tyrosine residues in a protein to generate 3-nitrotyrosines [14]. A tyrosine residue may also be a target for phosphorylation, which plays a critical role in signal transduction, cell death, and various other cellular events. For these reasons, nitrotyrosine residues have received attention not only as targeted biomarkers indicating RNS but also as functional groups correlating with diseases and cancer. The methods described here should prove advantageous in studies of nitrotyrosine or RNS.

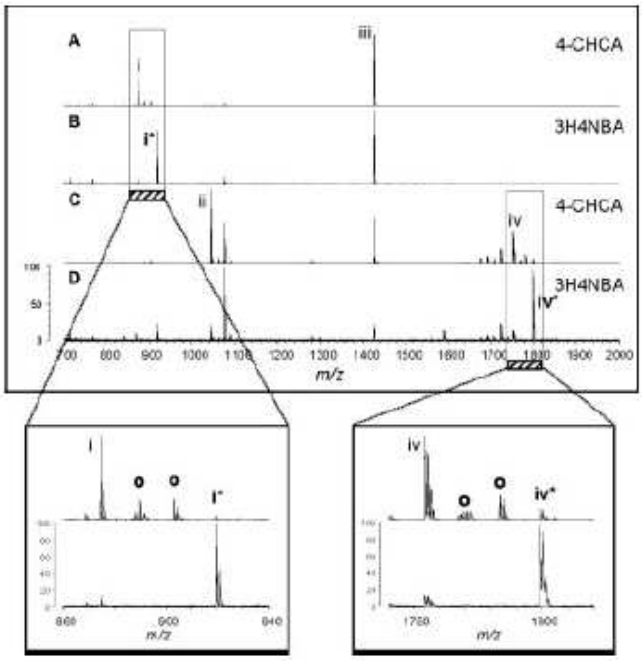


Figure III-6. Selective detection of nitrotyrosine-containing peptides using 3H4NBA as a matrix.
A tryptic digest of nitrated lysozyme was desalted using ZipTip m-C18 and peptides were eluted with 4 mL of a stepwise gradient of increasing ACN (10, 20, 30, 40, and 50%) containing 0.1% TFA. Of these eluates, 10% (A, B) and 30% (C, D) ACN eluates were subjected to MALDI-TOF MS analysis using CHCA (A, C) and 3H4NBA (B, D) as a matrix, respectively. Greek numbers indicate the peaks corresponding to the peptides with sequence and theoretical monoisotopic masses as follows: (i) HGLDNYR, 874.42; (i*) HGLDNY*R (Y* represents a nitrotyrosine residue), 919.41; (ii) GTDVQAWIR, 1045.54; (iii) FESNFNTQATNR, 1428.65; (iv), NTDGSTDYGILQINSR, 1753.84; (iv*) NTDGSTDY*GILQINSR, 1798.83. Asterisks denote the peaks that correspond to nitrotyrosine-containing peptides. Open circles indicate the peaks that correspond to 16 or 32 m/z units smaller than the peaks with an asterisk. Lower two panels contain expanded views of A and B ranging from m/z 850 to 960 (left) and of C and D ranging from m/z 1740 to 1820 (right).

III-3-5 Practical use of 3H4NBA as a matrix

Although the 3H4NBA matrix has advantages in the detection of peptides containing a nitrobenzene ring by MALDI-TOF MS, it also has drawbacks of lower sensitivity and shotto-shot reproducibility, which resulted in quick signal decay. The CHCA matrix has no selectivity for a peptide with a nitrobenzene ring, but it is highly sensitive, it is easy to search for “sweet” spots, and it is stable for signal detection even when multiple shots are taken from one spot. The author next used a comatrix of CHCA and 3H4NBA,

because it is known that some combination of two or more matrices (chemicals) can result in improved performance in MALDI-TOF MS analysis [15]. As shown in **Fig. III-7**, the signal intensity improved in comparison with the CHCA matrix, while the selectivity of 3H4NBA for labeled peptides was roughly retained. Somewhat continuous signal detection from one spot has also been achieved. The comatrix of CHCA and 3H4NBA is especially effective for detecting samples containing a small amount of protein. For example, the author picked up ten NBS-labeled peaks (1198.5, 1204.5, 1244.5, 1250.5, 1299.4, 1311.4, 1353.6, 1359.6, 1478.6, and 1484.6), which were clearly detected in all three MS spectra, and obtained their S/N values. The average S/N value for the comatrix (175.2) is about four times higher than that of CHCA (42.3) and that of 3H4NBA (47.8). This indicates that the sensitivity of detection was much improved by use of the comatrix. The recommended ratio for mixtures is 1:1 v/v, because the author does not want the characteristics of either comatrix component to predominate over the other. The author next examined comatrices of CHCA with various other nitrobenzene compounds. When each compound was used as a single matrix, its sensitivity for detection was very low (**Fig. III-5**), but the comatrices with CHCA resulted in higher sensitivity and selectivity, similar to the results for the comatrix of CHCA and 3H4NBA (data not shown).

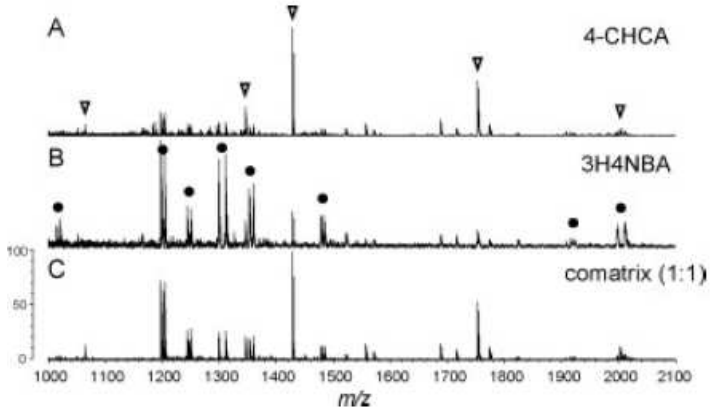


Figure III-7. Improvement of utility using a comatrix of 3H4NBA with CHCA.

A tryptic digest of four-model proteins containing NBS-labeled and unlabeled peptides was prepared and desalted by using ZipTip m-C18. One-thousandth of the ZipTip eluate was subjected to MALDI-TOF MS analysis. Matrices were CHCA (A), 3H4NBA (B), and a comatrix of CHCA, and 3H4NBA (mixed in a 1:1 ratio) (C). Closed circles indicate the paired peaks of NBS labeled peptides having a difference of 6 or 12 m/z units at the expected positions. Open arrowheads indicate the unlabeled peptides.

III-3-6 Discussion

The method developed here of using 3H4NBA and combining it with CHCA to make a comatrix has several advantages. Firstly, the selective detection of NBS-labeled peptides from an unpurified tryptic digest can now be achieved (**Fig. III-3**). Secondly, nonpaired labeled peptides derived from a protein that exists only in either sample can be discriminated from unpurified nonlabeled single peaks (see **Fig. III-8**). Thirdly, when used as a comatrix with CHCA, the sensitivity of peak detection is increased in comparison with the use of 3H4NBA alone, and even with that of CHCA alone (**Fig. III-7**). Fourthly, peaks that are 16 or 32 units smaller than the NBS labeled and nitrobenzene-containing peaks are suppressed, thus simplifying the mass spectra (Figs. 1–3, 5, and 6). Fifthly, it can be determined which peaks correspond to nitrotyrosine-containing peptides, which are found in many proteins (**Fig. III-6 (ii)**). Finally, the combinations of matrices and analytes that confer selectivity of detection for specific substances in MALDI-TOF MS analysis might prove useful in basic studies of

the MALDI process. In MALDI-TOF MS analysis, peptides containing tryptophan residues may be prominently detected by conventional methods using the CHCA matrix (**Fig. III-6**). However, the detection of these peptides was not enhanced by use of the 3H4NBA matrix. In contrast, although the NBS moieties of NBS-labeled cysteine are removed by reduction and replaced by carbamidomethyl groups in the NBS method [4], specially constructed peptides containing NBS-labeled cysteines were detected selectively using the 3H4NBA matrix (data not shown). These results indicate that selectivity of the detection originates not merely in the aromaticity of tryptophan or NBS but also in the nitrobenzene structure of NBS or nitrotyrosine. One interesting observation is that DHB has a “reverse” or “negative” specificity for detection of NBS-labeled peptides. That is, NBS-labeled peptides are detected with lower intensity than unlabeled peptides. This indicates that factors other than the nitro group contribute to this specificity, because both CHCA and DHB lack a nitro group. One possible reason for this specificity is that DHB has only two hydroxyl groups in addition to a benzene ring and a carboxyl group, whereas CHCA has unsaturated bonds in both the cyano group and the olefin moiety. Thus, these unsaturated electrons can make resonance with electrons in benzene ring, and similar interactions could occur as in the case of nitrobenzene. Regarding this speculation, the author found an interesting rule that all compounds having nitro groups, which showed clear selectivity as listed in **Fig. III-5**, have at least one functional group in either the ortho- or para-position against the nitro group. These results suggest that electrons of the substituted groups influence the electron cloud of the nitrobenzene, and these may eventually influence the co-crystallization of matrix and analyte, which must be a key step in the specific selectivity of MALDI-TOF MS analysis. Therefore, it may be interesting to calculate the electron density of each chemical and examine its relevance to the selectivity, as this seems to be significant. In the MALDI-TOF MS analysis of NBS-labeled and

nitrotyrosine-containing peptides, several peaks were observed at positions smaller by 16 or 32 m/z units than the original peaks (in the case of CHCA) and by 14, 16, 28, 30, 32 m/z units (in the case of DHB). However, these peaks were not observed on ESI-MS (LCQ Deca XP plus, Thermoelectron), and therefore the author expects that the fragments were generated in the process of MALDI. Because the nitrobenzene and its substituents absorb the energy of an UV laser commonly used in MALDI apparatus, analytes having nitrobenzene groups are thought to be energized by UV-laser irradiation and degraded. When the nitrobenzene compounds such as 3H4NBA are used as the matrix, the matrix surrounding the analytes may perform a buffering action to prevent analytes against absorption of the laser energy. Regarding the assignment of these fragment peaks, the author speculates that the peak smaller by 16 m/z units arises from production of the relatively stable nitrosyl group (-NO) by deletion of oxygen from the nitro group, but the author cannot determine the molecular species corresponding to the other fragment peaks.

It has been reported that some matrices and/or their combinations are useful for detection of specific substances, as seen in the case of DHB to detect phosphorylated peptides and oligosaccharides. (13) However, the method reported here, using 3H4NBA as a matrix, is significantly more specific in that target substances are easily detected in mixed samples, and are often the predominant peaks. Moreover, even in cases where 3H4NBA or some other nitrobenzene chemicals by themselves are not ideal matrices for a particular application, they can be used as comatrices with CHCA, and this may confer selectivity for specific substances. Therefore, if one chooses a chemical having affinity to both CHCA and the analyte, as in the case of 3H4NBA, it may effectively function as a “selective comatrix” with CHCA. It should be noted that the method described here has some limitations. The detection is selective for labeled over unlabeled compounds by a factor of about 10, but it is not exclusive. Therefore, if

an unlabeled peptide is present in the sample with a ten-fold excess over the labeled peptides, it may be detected as the strongest peak. Furthermore, the selectivity decreases if the laser power is stronger than the optimum level, or if the mass of a labeled analyte to be focused is relatively large. Nevertheless, this method will provide useful tools for analysis of specific compounds and will allow development of new applications in protein analysis.

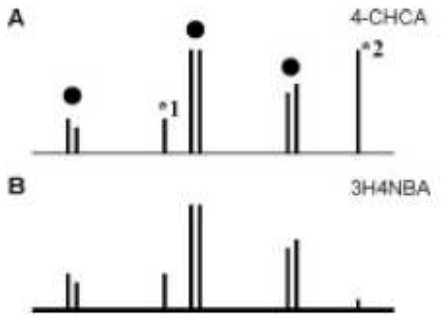


Figure III-8. Schematic mass spectrum showing the discrimination of single peaks between NBS-labeled and unlabeled peptides by using a 3H4NBA matrix.

Closed circles indicate the paired peaks of NBS-labeled peptides. Asterisk 1 represents an NBS-labeled single peak derived from a protein that exists only in either sample, while asterisk 2 represents a contaminating unlabeled single peak.

The improvements shown in this chapter and Chapter II resulted in an expansion of dynamic range of measurement. By combining the NBS method with HPLC, an automatic spotter, MALDI-TOF MS, and an analytical software, the author constructed a system called the “NBS Biomarker Discovery System”. In Chapter IV, the author applied the analysis system to CRC (colorectal carcinoma) clinical samples to the discovery of drug target and/or biomarker candidates for CRC.

References

1. Julka, S., Regnier, F., J. Proteome Res. **3**, 350–363 (2004)

2. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. et al., Nat. Biotechnol. **17**, 994–999 (1999)

3. Hansen, K. C., Schimitt-Ulms, G., Chalkley, R. J., Hirsch, J. et al., Mol. Cell. Proteomics **2**, 299–314 (2003)

4. He, T., Alving, K., Field, B., Norton, J. et al., J. Am. Soc. Mass Spectrum. **15**, 363–373 (2004)

5. Zhang, H., Li, X.-J., Martin, D. B., Aebersold, R., Nat. Biotechnol. **21**, 660–666 (2003)

6. Riordan, J. F., Christen, P., Biochemistry **7**, 1525–1530 (1968)

7. Sokolovsky, M., Riordan, J. F., Vallee, B. L., Biochemistry **5**, 3582–3589 (1966)

8. Dreisewerd, K., Chem. Rev. **103**, 395–425 (2003)

9. Karas, M., Krüger, R., Chem. Rev. **103**, 427–439 (2003)

10. Knochenmuss, R., Zenobi, R., Chem. Rev. **103**, 441–452 (2003).

11. Oldreive, C., Rice-Evans, C., Free Rad. Res. **35**, 215–231 (2001).

12. Laugesen, S., Roepstorff, P., J. Am. Soc. Mass Spectrom. **14**, 992–1002 (2003).

13. Asara, J. M., Allison, J., J. Am. Soc. Mass Spectrom. **10**, 35–44 (1999).

Chapter IV: An application of the NBS method to proteomic profiling of human colorectal carcinoma for biomarker discovery

IV-1 Introduction

Colorectal carcinoma (CRC) is the third most common type of cancer and the second leading cause of cancer death in developed countries. Over the past two decades, the clinical test for CRC has utilized carcinoembryonic antigen (CEA) as a marker protein. However, most positive cases are found in patients with advanced cancers or even metastases. For example, the positive detection ratio of CEA in patients with metastatic cancer generally ranges from 70% to 80%, whereas it decreases in patients with both locally recurrent and early cancers. (1-3) CEA has not been proven effective as a screening marker for early-stage cancers, and its applications have been limited to the detection of advanced cancers. This supports the need to develop novel CRC biomarkers to improve the accuracy of diagnosing CRC. Recently, various approaches involving transcriptome analysis have been extensively applied in an effort to identify novel diagnostic markers for CRC. (4, 5) However, mRNA expression levels do not necessarily correlate with protein expression. Hence, direct analysis of proteins is indispensable for the discovery of novel biomarkers.

Most proteomic approaches involving CRC tissues have been performed by 2D-PAGE in combination with MS, and some successful results have been obtained (6-10). However, it is still difficult to perform comprehensive proteome analysis, as this method has several technical limitations (11). Therefore, the author developed and improved the NBS method for global quantitative proteome analysis as shown in Chapter I, Chapter II and Chapter III. In this method, labeled peptides after enzymatic digestion are subjected to HPLC separation, while intact proteins are analyzed directly using the 2D-PAGE. These two different methodologies can detect different sets of proteins, so the NBS method can complement other methods such as 2D-PAGE.

The primary advantage of this new method is that it reduces the number of peptides, measured by selecting NBS-labeled tryptophan-containing peptides from bulk tryptic digests. This is advantageous because tryptophan residues are the least abundant amino acid in proteins, yet they occur in a large proportion of proteins. (12) Another advantage of this method is the special matrix used for MALDI-TOF MS measurement, which can detect the NBS-labeled peptides and with high sensitivity. (Chapter III) From these reasons, the author believes this method can improve proteome mining by increasing the dynamic range of detection and that it shows promise for quantitative proteome analysis. (13-15) Here, the author applied this NBS method to analyze clinical samples from CRC patients in order to discover novel biomarker candidates.

IV -2 Materials and Methods

IV-2-1 Tissue samples

Twelve primary colorectal cancer specimens and corresponding normal colonic mucosal specimens were obtained from surgical resections from March, 2003 to November, 2004. All patients with tumors were diagnosed at advanced stages, and none of the adenomas were contained in a cancerous component. All normal tissues were histopathologically confirmed as cancer-free. None of the patients were treated with preoperative chemotherapy or radiotherapy. The samples were stored in RNAlater (Qiagen, Valencia, CA) at -20˚C after sampling. This study was approved by the Institutional Review Board of Osaka University and informed consent was obtained from each patient.

IV-2-2 Sample preparation

Frozen tissue samples were homogenized in 500 μL of lysis buffer A (50 mM Tris-HCl at pH 8.0, 100 mM NaCl, 5 mM EDTA, 1 mM PMSF, 1 μg/mL leupeptin, and 5 μg/mL aprotinin) on ice using a Sample Grinding Kit (GE Healthcare, Buckinghamshire, UK). Homogenates were centrifuged at 100,000 × *g* for 60 min and supernatants were obtained as the cytosolic fraction (CF). Pellets were washed twice with lysis buffer A and homogenized in 500 μL of lysis buffer B (2% CHAPS, 9 M urea, 50 mM Tris-HCl at pH 8.0, 100 mM NaCl, 5 mM EDTA, 1 mM PMSF, 1 μg/mL leupeptin, and 5 μg/mL aprotinin); homogenates were centrifuged at 100,000 × *g* for 60 min. Supernatants were obtained as the 2% CHAPS-soluble fraction (CSF). These fractionated samples were precipitated using the 2D-Clean-Up Kit (Bio-Rad, Hercules, CA) and resuspended in 8 M urea and 5 mM EDTA. After centrifugation at 10,000 × *g* for 5 min, supernatants were recovered and subjected to NBS reagent labeling. Protein concentration was determined by BCA Protein Assay (PIERCE, Rockford, US) using BSA as a standard.

IV-2-2 NBS reagent labeling, peptide fractionation and MS measurement

NBS reagent labeling was performed according to the manufacturer's protocol (^{13}CNBS stable isotope labeling kit-N; Shimadzu Biotech, Kyoto, Japan). Normal and tumor tissue samples (100 μg each) were labeled with isotopically light and heavy NBS reagent, respectively. NBS-labeled samples were then mixed, reduced, alkylated and digested by trypsin. NBS-labeled peptides were enriched from tryptic digests and fractionated using Phenyl sepharose, as described previously (Chapter II). The resulting seven fractions were combined into three fractions and subjected to reverse-phase liquid chromatography (LC-10ADvp μHPLC System; Shimadzu), as described previously (13). Eluates were automatically deposited onto MALDI target plates by the LC spotting system (AccuSpot; Shimadzu). These samples were automatically analyzed by MALDI-TOF MS (AXIMA-CFR Plus; Shimadzu / Kratos, Manchester, UK). (13)

IV-2-3 Relative quantification and identification of differentially expressed proteins in CRC

Relative quantification of each NBS-labeled peptide pair was performed using TWIP Version 1.0 (Dynacom, Kobe, Japan), referring to a monoisotopic mass list from MASCOT Distiller Ver. 1.1.2 (Matrix Science). The author previously demonstrated that quantification errors (%) using a model protein mixture were less than 4%. (13) Thus, peptide pair ratios larger than 1.5-fold, or smaller than 0.66, were set as threshold values for significant differences. A threshold value for the occurrence was set to 70% of all the CRC patient samples in which peptide pairs were detected. In this manner, candidate peptides were selected and further subjected to MS/MS analysis (AXIMA-QIT-TOF; Shimadzu / Kratos). (13) Proteins were identified by the MASCOT MS/MS Ion Search algorithm (Version 2.0; Matrix Science) using mass lists generated by MASCOT Distiller. The Mascot search parameters were as follows: trypsin digestion allowing up to two missed cleavages, fixed modifications of 12CNBS (or

13CNBS) and carbamidomethyl (C), variable modifications of oxidation (M), peptide tolerance 0.3 Da and MS/MS tolerance of 0.5 Da. Search results having p-values less than 0.05 were judged as positive identifications.

IV-2-4 Western blot analysis

Total protein extracts (20 μg; CF or CSF) from the tumor and corresponding normal tissue samples of each patient were separated on 10% or 15% SDS-polyacrylamide gels. Proteins were then transferred to a nitrocellulose membrane and prestained SDS-PAGE standards (Bio-Rad) were used to estimate their molecular weights. The following primary antibodies were used: mouse anti-human Zyxin (ZYX), polyclonal (Abnova, Taipei, Taiwan), mouse anti-human RAN, monoclonal (Abcam, Cambridge, UK), mouse anti-human S-adenosylhomocysteine hydrolase (AHCY), polyclonal (Abnova), rabbit anti-human reticulocalbin 1 (RCN1), monoclonal (Abnova), rabbit anti-human galectin1 (LGALS1), polyclonal (Abcam), and rabbit anti-human Vimentin (VIM), polyclonal (Abcam). Nitrocellulose membranes were incubated with diluted antibody solution for 2 h at room temperature. After washing in PBS, the membranes were incubated at room temperature for 1 h with HRP-conjugated sheep anti-mouse IgG antibody (GE Healthcare) for ZYX, RAN and AHCY, or HRP-conjugated donkey anti-rabbit IgG antibody (GE Healthcare) for RCN1, LGALS1 and VIM. Primary antibody dilutions were: anti-ZYX (1:500); anti-RAN (1:1000); anti-AHCY (1:1000); anti-RCN1 (1:1000); anti-LGALS1 (1:1000); and anti-VIM (1: 1000). Secondary antibody dilutions were anti-mouse IgG (1:4000) and anti-rabbit IgG (1:10000). Proteins were then visualized by ECL Plus detection reagents (GE Healthcare), exposed to X-ray film (Kodak, U.S), and the protein band densities were quantified using “CS Analyzer v3.0” software (ATTO, Tokyo). Used membranes were stained with 0.2% CBB R-250 in 40% MeOH, 10% AcOH for 5 minutes and destained with 90% MeOH, 2% AcOH for 15 minutes to confirm equal protein loading and blotting (data not shown).

IV-2-5 Immunohistochemical staining

Ten percent buffered formalin-fixed paraffin-embedded sections were prepared from 10 surgically resected cancers. Tissue specimens from these same cancers were also used for proteomics analyses. The streptavidin-biotin immunoperoxidase complex method was used for immunohistochemical analysis. Briefly, 4-μm slices of tissue section were deparaffinized and incubated with 0.03 mol/L citrate buffer (pH6.0) and heated to 98°C for 40 min for antigen retrieval. Endogenous peroxidase activity was blocked with 0.3% hydrogen peroxide and 0.1% sodium azide in distilled water for 15 min. After three rinses in phosphate-buffered saline pH 7.2 (PBS), 10% bovine serum (Wako, Osaka, Japan) was applied for 10 min to block non-specific reactions. Sections were incubated with the primary antibody for 60 min at room temperature. Primary antibodies for immunohistochemical staining were the same as those used in the WB analyses. After washing in PBS, the sections were treated with biotinylated sheep anti-mouse IgG (Amersham, London, U.K.) for ZYX, RAN and AHCY or biotinylated anti-rabbit IgG (Nichirei, Tokyo, Japan) for RCN1, LGALS1 and VIM for 15 min. After washing in PBS, the sections were reacted with streptavidin-biotin peroxidase complex (Dako, Copenhagen, Denmark) at 1:300 dilution for 15 min. The peroxidase reaction was visualized by incubating the sections with 0.02% 3,3'-diaminobenzidine tetrahydrochloride in 0.05M Tris buffer (pH 7.6) with 0.01% hydrogen peroxide for 3 min. Sections were then counterstained with hematoxylin. Negative control sections were tested using normal mouse or rabbit serum instead of the primary antibody. Tissue sections of normal liver (for AHCY), skin (for VIM), lymph node (for LGALS1) and testis (for RAN, RCN1, and ZYX) were prepared as positive controls according to the manufacturers' recommendations or previous publications. All slides were evaluated by a blinded pathologist. For each immunohistochemical analysis the mean intensity of the tumor cells or stromal cells was evaluated in

comparison with the positive controls as follows: (a) weak, 1+; (b) moderate, 2+; (c) strong, 3+.

IV-3 Results and Discussion

IV-3-1 Proteomic profiling and identification of differentially expressed proteins in CRC tissues

Differential proteome analysis between CRC and normal tissues from each patient was performed using the NBS method (**Fig. IV-1-(A)**). This analysis was performed using CF and CSF samples from each of the 12 patients. After a series of experiments, including NBS labeling, peptide fractionation and MS measurement, 2600– 3000 paired peaks were observed per analysis. In this method, the relative ratio of expression for each protein is calculated from the relative ratio of peak intensity (or area) in each pair-peak (NBS-labeled peptides). Following this relative quantification, pair-peak lists were evaluated (see Materials and Methods) and 320 pairs were judged to have significant differences in protein expression and to occur with significant frequency in patients.

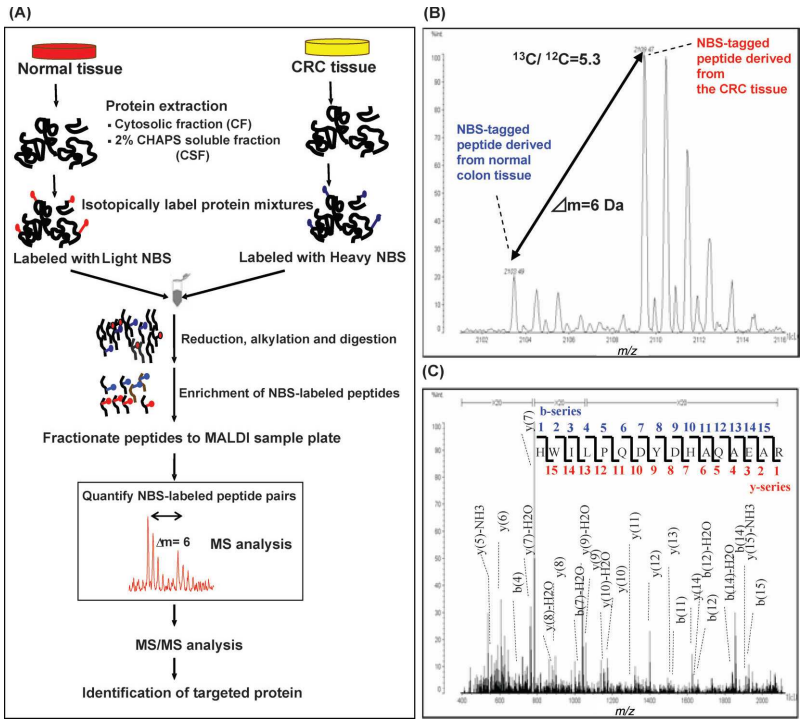


Figure IV-1 Proteomic analysis of CRC tissue samples by the NBS method.

A schematic drawing of NBS analysis (A) and a typical example of MS (B) and MS/MS (C) spectra of an NBS-labeled peptide (HWILPQDYDHAQAEAR) from RCN1 are shown.

After these peaks were subjected to MS/MS analysis, 226 decent MS/MS spectra were obtained, and 156 search results (138 identical peptides) were judged to be positive identifications. This corresponded to 128 proteins, with 71 up-regulated and 57 down-regulated, as listed in **Table IV-1.** Their subcellular localizations are shown in **Fig. IV-2**. The proportion of cytoplasmic proteins identified using CF analysis was nearly half (45.2%), whereas CSF analysis indicated a proportion of less than a quarter (22.2%). In contrast, the proportion of extracellular and plasma membrane proteins identified using CSF analysis (27.0%) was much larger than that identified using CF analysis (16.4%). Most of the proteins were identified from either CF analysis or CSF analysis, and only a few from both (65, 55 and 8 proteins were identified from CF only, CSF only or both fractions, respectively). This means that CF / CSF fractionations were successfully achieved and that these fractionations improved the proteome coverage. NBS labeling followed by MS analyses was carried out twice for all samples to evaluate experimental variations (**Fig. IV-3**). Correlation coefficients calculated for each of the 12 patient samples were all above 0.95, indicating this quantitation method was reliable.

In this study, the author focused on 23 up-regulated proteins whose mRNA expression was also up-regulated, as determined by cDNA microarray analysis (unpublished data). The author selected six of these proteins (ZYX, RAN, RCN1, AHCY, LGALS1, and VIM) for further characterization, using the criteria that they had not been well studied in relation to CRC and that they have distinct features.

Table IV-1 Differentially expressed proteins in CRC tissues

Protein name	[a)]Average T/N ratio	Average Log_2 (T/N ratio) ± SD	Occurrence in patients (%)	[b)]Fraction	Previously reported in CRC
Up-regulated proteins in CRC tissues					
Alpha1 acid glycoprotein	2.5	1.3±1.0	9 /11 (81%)	CF	
Alpha 1 acid glycoprotein type 2	[c)]NC	[c)]NC	10	CF, CSF	(8)
Alpha-tubulin	2.0	1.0±0.54	9 /9 (100%)	CSF	
Beta-tubulin	2.1	1.1±0.36	8 /9 (89%)	CSF	
Apurinic endonuclease	2.0	1.0±0.47	12 /12 (100%)	CF	
Calumenin	2.7	1.4±0.86	10 /10 (100%)	CSF	
Chaperonin1	2.3	1.2 ±0.46	11 /11(100%)	CSF	(6), (9)
Clathrin heavy polypeptide	1.7	0.78 ±0.78	7 /9 (89%)	CSF	
Clathrin light polypeptide A	3.2	1.7 ±1	10/ 11 (91%)	CSF	
Complement factor H	2.1	1.1±0.33	9 /10 (90%)	CF,CSF	
Cysteine rich intestinal protein 1	1.7	0.78±0.51	8 /9 (89%)	CSF	(6), (9)
Cytokeratin 18	2.8	1.5 ±0.99	10 /12 (83%)	CSF	(10)
Enolase 1	2.1	1.1 ±0.92	9 /10 (90%)	CF,CSF	(7)
Ezrin	2.3	1.2±0.6	8 /10 (80%)	CF	
F-box protein 40	2.5	1.3±0.61	11 /12 (91%)	CSF	
Fibrinogen gamma	2.2	1.2 ±0.5	9 /10 (90%)	CF	
Fk506 Binding Protein 1A	2.2	1.2 ±0.5	9 /10 (90%)	CF	
Galectin 1	2.1	1.1±0.39	9 /11(81%)	CSF	
Glutathione peroxidase 1	1.8	0.81±0.37	8 /10 (80%)	CF	
Glycyl tRNA synthetase	[c)]NC	[c)]NC	8	CF	(10)
Glyceraldehyde-3-phosphate dehydrogenase	2.0	1.0±0.82	8 /10 (80%)	CF	
Golgi complex-associated protein 1	2.2	1.1±0.73	8 /9 (89%)	CF	
Heat shock 70kD protein 9B	2.8	1.5±0.85	8 /9 (89%)	CF	
Heat shock protein 27	2.1	1.1±0.34	10 /12 (83%)	CF	
Heparan sulfate proteoglycan 2	2.5	1.3±0.59	8 /9 (89%)	CSF	
Heterogeneous nuclear ribonucleoprotein H2	[c)]NC	[c)]NC	9	CSF	
High density lipoprotein binding protein	2.1	1.1 ±0.25	8 /8 (100%)	CSF	
HLA-C	2.1	1.1±0.55	7 /8 (88%)	CF	
Hypothetical protein FLJ38663	2.2	1.1 ±0.87	7 /9 (78%)	CF	
Inorganic pyrophosphatase	2.5	1.3±0.76	10 /11 (90%)	CF	(6), (8)
Membrane-bound C2 domain-containing protein	2.0	0.98±0.31	8 /9 (89%)	CSF	
Mitogen inducible gene 2 protein	1.9	0.94±0.7	7 /9 (78%)	CF	

6-Phosphogluconolactonase	2.0	0.99±0.78	9 /10 (90%)	CF	
Plastin 2	2.4	1.2 ±0.49	10 /10 (100%)	CF	
Plectin 1	2.0	1.0±0.8	8 /10 (80%)	CSF	
Proteasome subunit p58	2.1	1.1±0.4	8 /8 (100%)	CSF	
Protein tyrosine phosphatase, receptor type c	[c)]NC	[c)]NC	8	CSF	
Protein tyrosine phosphatase, receptor type, α	[c)]NC	[c)]NC	8	CSF	
Pyruvate kinase 3	1.9	0.93±0.4	11 /12 (92%)	CF	(6)
RAB18, member RAS oncogene family	2.7	1.5 ±0.86	10 /12 (83%)	CSF	
RAB22A	1.9	0.94±0.73	7 /10 (70%)	CSF	
RACK1	2.0	1.0±0.32	10 /10 (100 %)	CF	
Radixin	1.8	0.84±0.76	8 /12 (73%)	CF	
RAN, member RAS oncogene family	2.0	0.99±0.67	9 /11 (81%)	CF	
Reticulocalbin 1	3.4	1.8±0.96	9 /10 (90%)	CF	
Rhodanese; thiosulfate sulfurtransferase	1.9	0.95±0.66	7 /10 (70%)	CF	
Ribonuclease RNase A family 3	[c)]NC	[c)]NC	8	CSF	
Ribosomal protein L13	3.4	1.8 ±0.98	10 /10 (100%)	CSF	
Ribosomal protein L27a	2.1	1.0 ±0.71	8 /11 (73%)	CSF	
Ribosomal protein L4	2.0	1.0 ±0.64	9 /11 (82%)	CSF	
Ribosomal protein S18	2.8	1.5 ±0.46	10 /10 (100%)	CSF	
Ribosomal protein S29	2.0	1.0 ±0.47	7 /8 (88%)	CSF	
Ribosome binding protein 1	1.8	0.87±0.34	10 /11(91%)	CF	(7)
S adenosylhomocysteine hydrolase	2.3	1.2 ±0.79	10 /11 (90%)	CF	
S100 calcium binding protein A9	2.2	1.2 ±0.7	9/ 11 (82%)	CF	(9)
Solute carrier family 25, member 5	2.2	1.1±0.7	8 /9 (89%)	CSF	
Solute carrier family 3, member 2	2.0	1.0±0.31	8 /9 (89%)	CSF	
Splicing factor 3B, subunit 3	2.8	1.5±0.99	9 /10 (90%)	CF	(7)
Splicing factor, arginine/serine-rich 3 (SRp20)	2.3	1.2±0.52	7 /8 (88%)	CF	
TLS protein	[c)]NC	[c)]NC	9	CSF	
Transgelin	2.1	1.1 ±0.78	7 /9 (78%)	CF	(8)
Transgelin 2	2.3	1.2 ±0.32	10 /10 (100%)	CF	(6)
Triosephosphate isomerase 1	2.0	0.99±0.92	10 /12 (83%)	CF, CSF	(6), (8)
Tumor rejection antigen 1	[c)]NC	[c)]NC	8	CSF	(6)
Ubiquitin activating enzyme 1	2.1	1.1 ±0.47	9 /10 (90%)	CF	
Ubiquitin isopeptidase T	2.3	1.2±0.6	9 /10 (90%)	CSF	
U5 snRNP-specific protein, 116 kD	2.5	1.3±0.71	9 /11 (82%)	CSF	
Vimentin	2.5	1.3 ±0.39	9 /9 (100%)	CSF	(6)
Vitronectin	2.1	1.1 ±0.8	7 /8 (88%)	CSF	
XTP3 transactivatied protein A	2.4	1.3 ±0.52	9 /10 (90%)	CF	
Zyxin	2.2	1.1 ±0.7	9 /10 (90%)	CF	

Down-regulated proteins in CRC tissues					
Abhydrolase domain containing 14B	0.39	-1.4± 0.55	8 /8 (100%)	CF	
ADP ribosylation factor like 10C	0.27	-1.9±1.1	7 /8 (88%)	CSF	
Aldehyde dehydrogenase 2	0.42	-1.3±0.54	7 /8 (88%)	CF	
Alpha 1 antitrypsin	0.37	-1.5±0.78	8 /11 (73%)	CF,CSF	(8)
Alpha-actinin 1, 4	0.46	-1.1±0.52	7 /8 (88%)	CF	
Annexin A2 isoform 2	0.41	-1.5±0.90	9 /11 (82%)	CSF	(6)
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	0.34	-1.6±1.2	8 /11 (73%)	CSF	
ATP binding cassette transporter subfamily A member 12	0.36	-1.5±0.68	9 /9 (100%)	CSF	
Calnexin	0.23	-2.2±0.82	10 /10 (100%)	CF,CSF	
Calreticulin	0.48	-1.1±0.53	7 /7 (100%)	CSF	(6)
Carbonic Anhydrase Ⅱ	0.1	-3.3±1.1	7 /7 (100%)	CF	(10)
Carbonyl reductase 1	0.54	-0.9±0.24	8 /8 (100%)	CF	
Cathepsin S	0.43	-1.2 ±0.68	10 /10 (100%)	CSF	(7)
Collagen type 12 alpha-1	0.43	-1.2 ±00.63	7 /7 (100%)	CSF	
Collagen type 14 alpha-1	[c)]NC	[c)]NC	7	CSF	
Creatine Kinase-B	0.50	-1.0± 0.33	8 /8 (100%)	CF	(8)
Cysteine rich protein 1	0.33	-1.6±0.47	7 /7 (100%)	CF	
Desmoglein 2	[c)]NC	[c)]NC	8	CSF	
Dynein light chain 1	0.48	-1.1±0.53	10 /10 (100%)	CSF	
Endoplasmic-reticulum-lumenal protein 29	0.42	-1.2±0.46	7 /7 (100%)	CSF	
Endoplasmic-reticulum-lumenal protein 46	0.42	-1.2±0.75	8 /9 (89%)	CF	
Enoyl CoA hydratase 1	0.34	-1.6±0.94	9 /9 (100%)	CF	
Eukaryotic translation elongation factor 2	0.51	-0.98±0.39	8 /8 (100%)	CF	
Eukaryotic translation initiation factor 3 subunit 6	0.33	-1.6±0.65	12 /12 (100%)	CF	
FHL1 (skeltal muscle LIM-protein)	0.34	-1.6±0.84	7 /9 (78%)	CF	(7)
Filamin A	0.4	-1.3±0.99	8 /9 (89%)	CF, CSF	(7)
Filamin B	0.39	-1.4±0.62	11 /11(100%)	CF	
Gelsolin isoform a	0.44	-1.2±0.45	9 /10 (90%)	CF	
Glucosamine-fructose-6-phosphate aminotransferase 1	0.34	-1.5±0.77	7 /7 (100%)	CF	
GTP-binding protein Rab1	0.59	-0.77±0.33	7 /10 (70%)	CF	
Haptoglobin	0.46	-1.1±0.69	8 /11 (73%)	CF	(7)
Heterogeneous nuclear ribonucleoprotein A2 /B1	0.42	-1.3±0.39	7 /7 (100%)	CF	(7)
Hydroxymethylglutaryl-CoA synthase,	0.30	-1.7±0.83	9 /10 (100%)	CF	

mitochondrial					
Isocitrate dehydrogenase 1	0.09	-3.5±1.6	9 /10 (100%)	CF	
JM5 protein	0.55	-0.86±0.32	7 /9 (78%)	CSF	
Major vault protein	0.47	-1.1±0.44	7 /7 (100%)	CSF	
MHC class I antigen	0.53	-0.92±0.50	7 /9 (78%)	CF	
Myeloperoxidase	0.46	-1.1±0.34	7 /7 (100%)	CSF	
Myosin, heavy polypeptide 14	0.38	-1.4±0.74	7 /7 (100%)	CSF	
Myozenin 3	0.47	-1.1±0.57	9 /9 (100%)	CSF	
NADH Ubiquinone oxidoreductase subunit B13	0.38	-1.4±0.71	8 /10 (80%)	CF	(8)
Normal mucosa of esophagus specific 1	0.25	-2.0±1.0	8 /8 (100%)	CSF	
Olfactomedin 4	0.46	-1.1±0.55	7 /7 (100%)	CSF	
Phosphoenolpyruvate calboxykinase 2	0.44	-1.2±0.58	7 /8 (88%)	CF	(10)
Phosphoglycerate mutase 1	0.23	-2.1±1.5	8 /9 (89%)	CF	
Proline arginine-rich end leucine-rich repeat protein	0.34	-1.6±0.84	11 /11 (100%)	CSF	
Protein kinase C and casein kinase substrate in neurons 2	0.49	-1.0±0.48	8 /8 (100%)	CSF	
Pyridoxine 5-prime-phosphate oxidase	0.27	-1.9±1.6	9 /12 (75%)	CF	
Raf kinase inhibitor protein	0.41	-1.3±0.62	10 /12 (83%)	CF	
Ras associated protein Rab5B	0.42	-1.2±0.40	7 /7 (100%)	CF	
Retinoblastoma binding protein 4, 7	0.43	-1.2±0.56	7 /7 (100%)	CSF	
Succinate dehydrogenase complex,subunit A,,flavoprotein	0.43	-1.2±0.56	7 /7 (100%)	CF,CSF	(7), (8)
TNRC15 protein	0.29	-1.8±1.1	8 /9 (89%)	CF	
Transferrin	0.46	-1.1±0.61	9 /9 (100%)	CF	
UDP-glucose dehydrogenase	0.36	-1.5±0.62	7 /7 (100%)	CF	
Valosin containing protein	0.49	-1.0±0.33	10 /11(91%)	CF	
Villin 1	0.13	-2.9±1.9	8 /9 (89%)	CF	(7)

a) An average (T/N ratio) is calculated by exponential transformation of an average Log_2 (T/N ratio)

b) Fraction; CF: cytosolic fraction, CSF: 2 % CHAPS-soluble fraction

c) NC; not calculated. This is because a protein is exclusively detected in CRC (or normal) tissues.

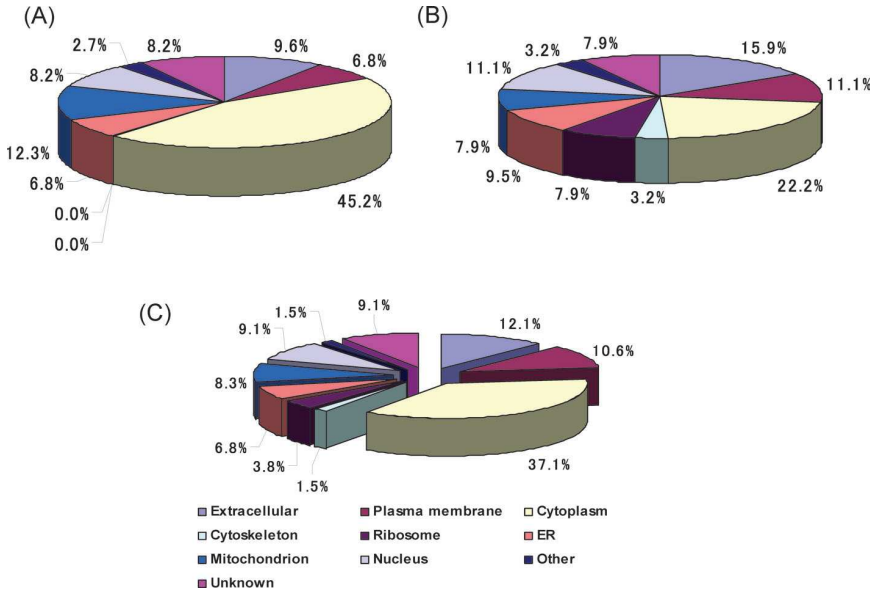


Figure IV-2 Classification of differentially expressed proteins in CRC tissues.

Identified proteins that were differentially expressed in CRC tissues are classified into categories based on their subcellular localization. Proteins identified in the cytosolic fraction (CF) (A), and those identified in the 2% CHAPS-soluble fraction (CSF) (B) and the total of 128 proteins (C) are represented graphically.

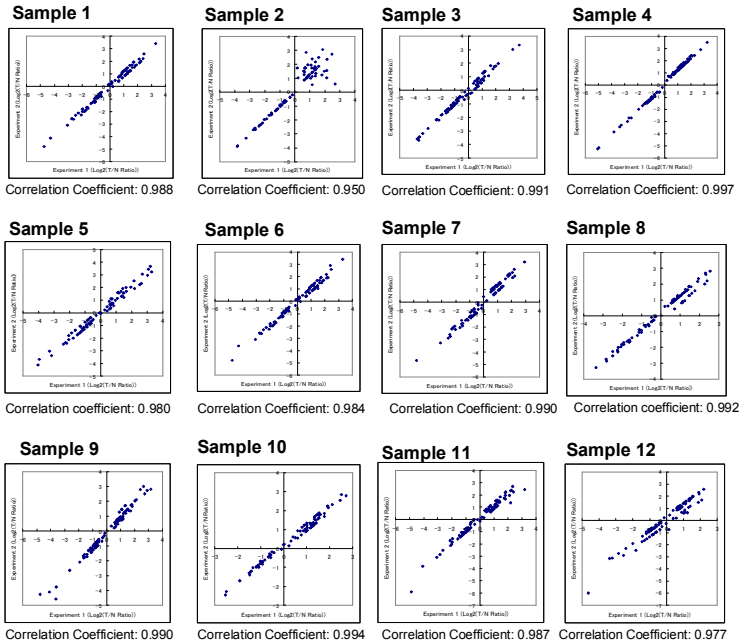


Figure IV-3 Correlation among the two tests by the NBS method.

The x-axis shows the ratio of protein quantity in colorectal carcinoma (T) to normal colon (N) in experiment 1, in Log_2 scale. The y-axis shows the ratio in experiment 2.

IV-3-2 Verification of results obtained from global quantitative proteome analysis by the NBS method

Although the good reproducibility and reliability of the NBS method have been demonstrated previously (Chapter I, Chapter II, 13) and were also demonstrated in the above analyses, the author carried out western blot (WB) analysis to verify our results using an independent method. Using the same fraction (CF or CSF) and the same CRC patient samples used in the NBS analyses, WB analyses were performed for the six selected proteins (**Fig. IV-4**). Generally, high T/N ratios were again observed in WB analyses, though the precise ratio was sometimes unmatched in one-to-one data comparison with the NBS analysis. The inconsistency sometimes observed between NBS and WB data may be due to the different assay methods (see Discussion). Because the high T/N ratios observed in the NBS analyses were confirmed by WB analyses for all six selected proteins, these proteins were further evaluated by immunohistochemical staining analysis.

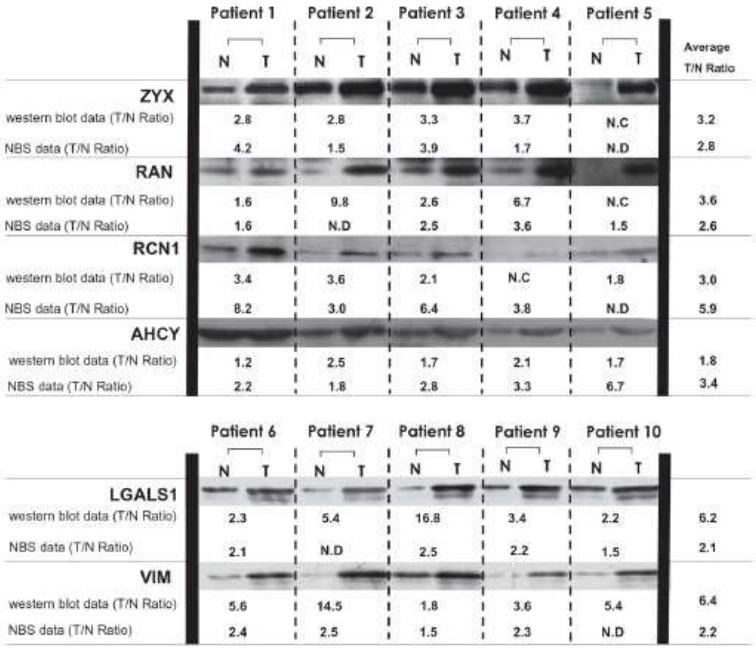


Figure IV-4 Western blot analysis for data verification.

Western blot (WB) analyses were performed using matched pairs of normal (N) and tumor (T) tissues to detect the following proteins: ZYX, RAN, RCN1, AHCY, LGALS1, and VIM. CF samples were used to detect the four proteins in the upper panel and CSF samples were used to detect the two proteins in the lower panel. The T/N ratios obtained from WB and NBS analyses are shown below each blot. The average T/N Ratio for each protein was calculated from the ratio in each sample where both NBS and WB data were obtained. N.C.: not calculated. N.D.: not detected.

IV-3-3 Further validation by immunohistochemical staining

Finally, immunohistochemical staining was performed to investigate the localization of the six selected proteins and to examine their expression patterns in whole CRC tissues (**Fig. IV-4**). All of these proteins, in all samples tested, were immunohistochemically detected more frequently in CRC tissues than in adjacent normal tissues. Cancer cells expressed five proteins (ZYX, RAN, RCN1, AHCY, and LGALS1); RCN1 and AHCY were strongly detected and ZYX, RAN and LGALS1 were found at moderate levels, on average, in all samples tested, although their expression patterns were not homogenous. In particular, RAN, RCN1 and LGALS1 were localized partially in cancer cells. On the other hand, normal colorectal epithelial cells expressed only ZYX, RCN1 and AHCY, and their expression was generally weak. Various stromal cells were positive for five proteins (ZYX, RAN, AHCY, LGALS1, and VIM), including leukocytes, blood vessels, nerves and fibroblasts. VIM expression was strongly and specifically expressed in stromal cells close to the cancer cells. ZYX and LGALS1 were moderately expressed in stromal cells generally, and their expression here was as strong as that of cancer cells. Stromal cells close to the cancer cells expressed these three proteins more intensely than those distant from cancer cells. Five of the six selected proteins (ZYX, RAN, AHCY, LGALS1, and VIM) were expressed not only in cancer cells, but also in the surrounding stromal cells.

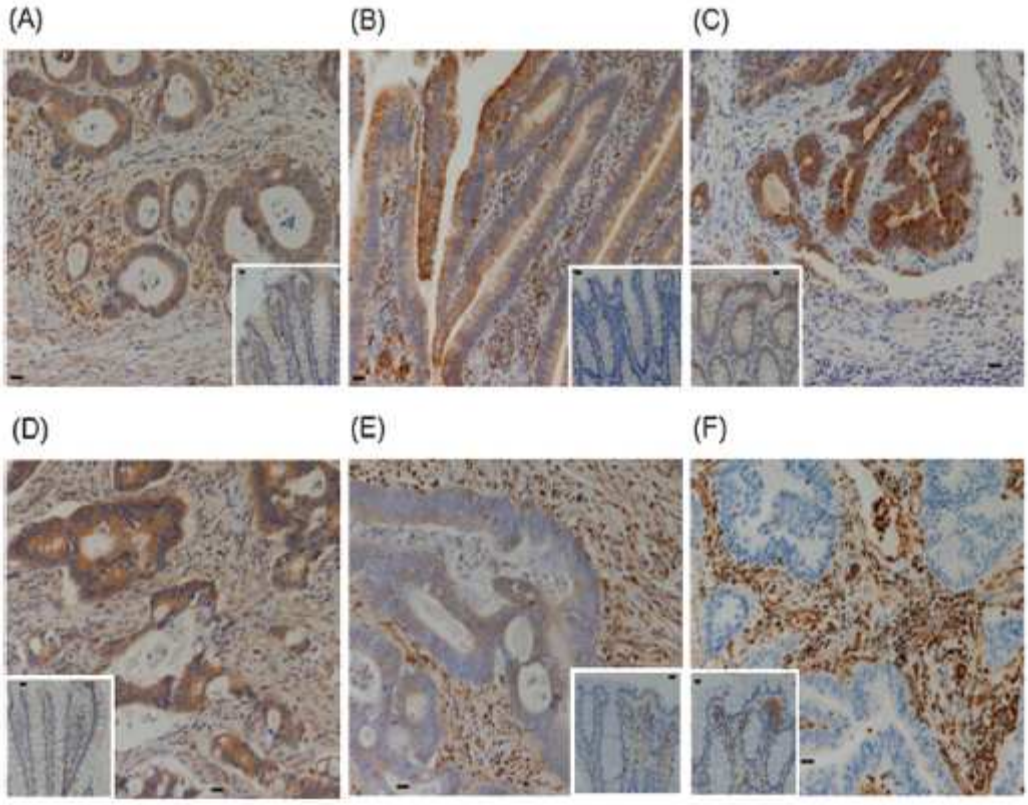


Figure IV-5 Immunohistochemical staining (IHC) for the six selected proteins in CRC tissues.
Insets show IHC staining of normal colonic tissue. Bars indicate 20 μm in all figures, including insets (Original magnification X140, insets X70). Each displayed picture (A -F) is a typical example of ten tested samples.

(A) ZYX was expressed in cancer cells as well as in stromal cells close to cancer cells, including leukocytes and blood vessels. Weak ZYX expression was observed in normal epithelial cells and in stromal cells distant from cancer cells (Inset).

(B) RAN was expressed in cancer cells, while normal epithelial cells were negative. Some leukocytes adjacent to cancer cells also expressed RAN.

(C) RCN1 was expressed in cancer cells and weakly in normal epithelial cells. No RCN1 expression was detected in stromal cells.

(D) AHCY was expressed in cancer cells and weakly in normal epithelial cells. Weak AHCY expression was observed in stromal cells such as leukocytes and blood vessels.

(E) LGALS1 was expressed in stromal cells including leukocytes, nerve cells and fibroblasts. Expression in stromal cells adjacent to cancer cells was much stronger. Some cancer cells expressed LGALS1, while normal epithelial cells were negative.

(F) VIM was strongly expressed in stromal cells including leukocytes, fibroblasts and blood vessels. Expression in stromal cells adjacent to cancer cells was much stronger. Neither normal epithelial cells nor cancer cells expressed VIM at all.

IV-3-4 Discussion

The author performed proteomic profiling of CRC tissue samples to identify novel biomarker candidates by the NBS method. Proteins that were differentially expressed between tumor and normal tissues, with significant differences in expression and affecting most of the patients, were selected and a final set of 128 proteins was identified. Of these, 23% (30 proteins) were reported in earlier proteomic studies on CRC. (6-10) Interestingly, the present study has led to identification of about 100 novel CRC-associated proteins that have not been reported to associate with CRC before. In this study, two fractions (CF and CSF) were prepared in order to increase the range of analyses. Many extracellular and plasma membrane proteins were recovered simultaneously and identified in CSF fractions. CEA, a well-known biomarker used for clinical detection of CRC, is a GPI-anchored membrane glycoprotein. It has been suggested that CEA is cleaved by glycosylphosphatidy -linositol-phospholipase D (GPI-PLD) and then secreted into blood. (16) Thus, plasma membrane proteins have the best chance of being secreted into circulatory systems, along with extracellular proteins. In other words, plasma membrane proteins and extracellular proteins are excellent candidates for biomarker development. Viewed in this context, our report seems to contain many potential CRC biomarker candidates. There are some discrepancies between the NBS and WB results (Fig. IV-3; for example, AHCY for Patient 5, LGALS1 for Patient 8, VIM for Patient 7). This probably was due to differences in the two analytical methods. In WB analysis, whole proteins, including post translational modifications (PTMs) or alternative splicing variants, can be detected by a shift in migration, although subtle differences of molecular weight change caused by a small modification or a point mutation are overlooked. On the other hand, in NBS analysis, peptides are detected with a resolution power of less than one Da, although PTMs, mutations, and splice variants not present in the NBS-labeled peptides will be

overlooked. Thus, the NBS method by itself is not suitable for a comprehensive analysis of PTM moieties of proteins because the coverage of each protein is quite low. This is indeed a drawback and a limitation of the NBS method, which must be compensated by other methods. However, global proteome analyses with respect to PTMs-proteins can be performed by the NBS method if it is combined with prior enrichment of the sample. (15) Six selected proteins showed high expression levels in CRC tissues and occurred with high frequency in CRC patients (ZYX (90% (9/10)), RAN (81% (9/11)), RCN1 (90% (9/10)), AHCY (90% (10/ 11)), and LGALS1 (81% (9/11), respectively). These proteins were also selected because they have not been well studied in relation to CRC, and so they were considered to be novel biomarker candidates. Each of these is discussed in detail below.

ZYX is a zinc-binding phosphoprotein that is a member of the LIM protein family. This protein is widely expressed in human tissues and is most prominent in lung and colon tissues. (17) It has been suggested that ZYX might enhance hepatocellular carcinoma cell migration and intravasation through its action on the actin cytoskeleton, as it promotes cell dissemination as a part of integrin-signaling pathways. (18) However, there have been no reports demonstrating an association of ZYX with CRC. In this study, we have demonstrated high expression of ZYX in CRC tissues for the first time. Moreover, ZYX was highly expressed in both cancer cells and in the surrounding stromal cells. Because this molecule is also strongly expressed in blood vessels in close proximity to cancer cells, it might be secreted into the blood of CRC patients.

RAN is a small GTP-binding protein belonging to the RAS superfamily. This molecule has also proven essential for various mammalian cellular processes, such as nuclear-cytoplasmic transport, cell cycle progression, nuclear organization, nuclear envelope assembly, mitotic control and genomic instability. (19, 20) Because genetic instability is a major factor in carcinogenesis and development of cancer, this protein is

considered to be associated with carcinogenesis. It has been reported that high-level expression of RAN is strongly associated with the prognosis of epithelial ovarian tumors [24]. In this study, the author demonstrated its up-regulation in both cancer cells and in the surrounding stromal cells. Taking into consideration these data and previous reports, it appears that this protein might be a key molecule in CRC carcinogenesis.

RCN1 is a calcium-binding protein expressed in the ER. This protein contains six repeats of a domain containing an EF-hand motif, which is considered to play a role in Ca2+-dependent cell adhesion. (22) This protein regulates cadherin expression in breast carcinoma and colorectal carcinoma cells (SW480) (23, 24), and high expression of this protein in hepatoma cells has been demonstrated. (25) The author noted for the first time high expression levels of this protein in CRC tissue. In addition, the author demonstrated that RCN1 was overexpressed with high frequency only in cancer cells. Our data suggest that RCN1 might be a promising candidate for biomarker development.

AHCY catalyzes hydrolysis of S-adenosylhomocysteine (AdoHcy) to adenosine and homocysteine. This enzyme is crucial in the control of transmethylation reactions, and a deficiency of this molecule induces hypermethioninemia. (26) There is little evidence indicating any association of this molecule with cancer. Another proteomics study demonstrated up-regulation of AHCY in CRC tissue (8); however, this was not characterized further. The present study utilized IHC to show, for the first time, high expression of AHCY in cancer cells and weaker expression in the surrounding stromal cells.

Galectins regulate pleiotropic biological functions involved in cell growth, differentiation, adhesion, RNA processing, apoptosis and malignant transformation. (27) Galectin-1 is found extracellularly in many tissues in both normal and pathological conditions, and several reports have demonstrated an association of galectin-1 with cancer. (27)

Up-regulation of LGALS1 in CRC tumor tissue was found using another proteomic approach (28); however, no follow-up studies have been reported to date. In the present study, this protein was strongly expressed in stromal cells adjacent to cancer cells, in addition to its up-regulation in tumor tissue.

VIM is an intermediate filament that represents a third class of well-characterized cytoskeletal elements, along with the actins and tubulins. This protein is considered to be a key protein in cell physiology, cellular interactions, and organ homeostasis. (29) In previous reports, IHC examination has shown that most cases of primary colonic malignant melanoma, which is a rare tumor, were positive for VIM. (30) In the present study, this molecule was strongly expressed in stromal cells in the vicinity of cancer cells, and was associated with CRC.

Analyses of whole tissues including stroma are important to understand cancer biology and to discover novel biomarker candidates. This is because as much as a half of tumor tissues are composed of stroma, and cancer cells are frequently influenced by this. Also, stroma cells may be influenced by neighboring cancer cells, and changes of the stroma could constitute a “cancer signal”. IHC experiments in this study revealed that two proteins (LGALS1 and VIM) were up-regulated in CRC tissues, but they are located primarily in stroma, not in tumor cells. This observation supports the importance of whole tissue analysis. In addition, it suggests that a proteomic study should be compensated and validated by IHC, because it alone does not reveal detailed information regarding protein localization, which is required for precise functional analyses.

In this study, proteins up-regulated in CRC tissues, whose mRNA expression was also up-regulated, were primarily selected for further studies. It is natural to focus on up-regulated proteins in CRC to identify potential diagnostic biomarkers to achieve early detection of CRC. This set of proteins may include a key protein that is responsible for

carcinogenesis. Considering the above possibility, genes with high mRNA expression in CRC are the best targets for gene therapy, such as RNA interference. This is why the author prioritized investigating these proteins over others in this study. However, up-regulated proteins that showed a negative correlation with mRNA expression require further investigation because they were detected only by proteomic studies.

To identify biomarker candidates for early diagnosis, examinations of specimens from patients diagnosed at an early stage of disease may be ideal. However, advanced tumor specimens were used in this study for the first screening, as previously reported [6-10], partly because it is difficult to obtain enough early stage specimens. Although many proteins identified in this experiment may not be applicable to early diagnosis, several of them are anticipated to be already altered and to be applicable as early stage markers, as proved in earlier studies. (7, 9) This is why proteins identified from advanced stage specimens may still be used as early diagnostic markers. The author is now examining whether selected candidate proteins are detected at significantly higher levels in sera of CRC patients in comparison with healthy volunteers.

The author originally assumed that the term “biomarker” is simply used in clinical diagnostic scenes, but it could be used in clinicopathological ones as well. For example, there are few cases in which the pathological diagnosis for cancer is difficult. However, it is sometimes difficult to specify the localization of the primary tumor when a metastatic focus is first detected in cancer screening. (31) In such a case, the proteins identified in CRC tissues in this study would be very useful to identify the primary site, if they were applicable. The author believes that the present study will contribute to future improvements in diagnostic/prognostic applications, understanding of CRC carcinogenesis, and the discovery of new therapeutic targets and drugs.

References

1. Carpelan-Holmstorom, M. A., Haglund, C. H., Roberts, P. J., *Dis. Colon Rectum.*, **39**, 799-805 (1996)

2. Von Kleist, S., Hesse, Y., Kananeeh, H., *Anticancer Res.* **16**, 2325-2332 (1996).

3. Carpelan-Holmstrom, M. A., Haglund, C., Javinen, H. J., Roberts, P. J. *Anticancer Res*., ***16***, 981-986 (1996)

4. Takemasa, I., Higuchi, H., Yamamoto, H., Sekimoto, M. *et al.*, *Biochem. Biophys. Res. Commun.*, ***285***, 1244-1249 (2001).

5. Komori, T., Takemasa, I., Higuchi, H., Yamasaki, M. *et al.*, *J. Exp. Clin. Cancer Res.* ***23***, 521-527 (2004)

6. Tomonaga, T., Matsushita, K., Yamaguchi, S., Oh-Ishi, M. *et al.*, *Clin. Cancer Res. **10***, 2007-2014 (2004).

7. Roessler, M., Rollinger, W., Palme, S., Hagmann, M. *et al.*, *Clin. Cancer Res.*, ***11***, 6550-6557 (2005)

8. Friedman, D. B., Hill, S., Keller, J. W., Merchant, N. B. *et al.*, *Proteomics*, ***4***, 793-811 (2004).

9. Stulik, J., Hernychova, L., Porkertova, S., Knizek, J. *et al.*, *Electrophoresis*, ***22***, 3019-25 (2001).

10. Bi, X., Lin, Q., Foo, T. W., Joshi, S. *et al.*, *Mol. Cell. Proteomics **5***, 1119-30 (2005).

11. Rabilloud T., *Proteomics **2***, 3-10 (2002).

12. Cagney, G., Amiri, S., Premawaradena, T., Lindo, M. et al., *Proteome Sci. **13***, 5 (2003)

13. Iida, T., Kuyama, H., Watanabe, M., Toda, C. *et al.*, *J. Biomol. Tech.* **17**, 333-341 (2006).

14. Ou, K., Kesuma, D., Ganesan, K., Yu, K. *et al.*, *Proteome Res.**5***, 2194-206 (2006).

15. Ueda, K., Katagiri, T., Shimada, T., Irie, S. *et al.*, J Proteome Res **6** :3475-83 (2007).

16. Naghibalhossaini, F., Ebadi, P., *Cancer lett.* ***234***, 158-67 (2006).

17. Macalma, T., Otte, J., Hensler, M. E., Bockholt, S. M. *et al.*, *J. Biol. Chem.* ***271***, 31470-31478 (1996).

18. Sy, S. M., Lai, P. B., Pang, E., Wong, N. L. *et al.*, *Modern Pathology* ***19*****,** 1108-1116 (2006)

19. Ren, M., Drivas, G., D'Eustachio, P., Rush, MG., *J. Cell Biol*. **120**, 313-323 (1993)

20. Di Fiore, B., Ciciarello, M., Lavia, P., *Cell Cycle* ***3***, 305-313 (2004)

21. Ouellet, V., Guyot, M. C., Le Page, C., Filali-Mouhim, A. *et al.*, *Int. J. Cancer* ***119***, 599-607 (2006)

22. Tachikui, H., Navet, A. F., Ozawa, M., *J. Biochem*. ***121***, 145-149 (1997)

23. Liu, Z., Brattain, M. G., Appert, H., *iochem. Biophys. Res. Commun.* ***231***, 283-289 (1997).

24. Vermeulen, S. J., Bruyneel, E. A., Bracke, M. E., De Bruyne, G. K. *et al.*, *Cancer Res*. ***55***, 4122-4728 (1995)

25. Yu, L.R., Zeng, R., Shao, X.X., Wang, N. *et al.*, *Electrophoresis* ***21***, 3058-3068 (2000).

26. de la Haba, G., Cantoni, G. L., *J. Biol. Chem.* ***234***, 603-608 (1959).

27. Perillo, N. L., Marcus, M. E., Baum, L. G., *J. Mol. Med.* ***76,*** 402-412 (1998)

28. Uhlen, M., Bjorling, E., Agaton, C., Szigyarto, C. A. *et al.*, *Mol. Cell. Proteomics* ***4***, 1920-32 (2005).

29. Ivaska, J., Pallari, HM., Nevo, J., Eriksson, JE., *Exp. Cell Res.* **313**, 2050-62 (2007)

30. Mori, D., Satoh, T., Nakafusa, F., Tanaka, M. *et al.*, *Pathol. Int.* ***56***, 744-748 (2006)

31. N. Pavilidis, *Annals of Oncology*, **14**, 11-18 (2003)

General conclusion

In this thesis, the author developed a novel quantitative proteome technique (NBS method) in order to discover drug target and/or biomarker candidates for diseases such as cancers. The method is based on labeling with the chemical reagent NBSCl in conjunction with tandem mass spectrometry. In this method, selective introduction of the NBS moiety onto tryptophan residues is achieved, and a six Da mass differential is generated using $^{13}C_6$-labeled NBSCl (NBSCl-$^{13}C_6$) and $^{12}C_6$-labeled NBSCl (NBSCl-$^{12}C_6$). The six Da mass differential between the NBS-$^{12}C_6$-labeled and the NBS-$^{13}C_6$-labeled peptides assigns a mass signature to all tryptophan-containing peptides in any pool of proteolytic digests for protein identification through peptide mass mapping.

In Chapter I, using this strategy, the author compared the protein expression in rat sera using a normal (control) rat (Crj:-Wistar) and a hyperglycemic rat (GK/Crj). As a result, the stable isotope dilution techniques used in this method provide highly accurate relative quantification. The approach offers a widely applicable means of analyzing protein mixtures derived from biological samples, and the method described here presents an effective and simplified approach to proteome analysis.

In Chapter II, the author optimized the procedure in the NBS method by utilizing urea or guanidine hydrochloride as a protein denaturant, in conjunction with an improved chromatographic enrichment method for the NBS-labeled peptides using a phenyl resin column. By using this new protocol, both sample loss throughout the protocol and the elution of unwanted unlabeled peptides was minimized, improving the efficiency of the analysis significantly.

In Chapter III, it was demonstrated that 3H4NBA is a more suitable matrix than DHB for detecting NBS-labeled peptides by MALDI-QIT-TOF MS. Furthermore, NBS-labeled peptides were selectively ionized and detected in a mixture of

NBS-labeled and unlabeled peptides. Labeled paired peaks were easily detected without enrichment, and non paired labeled peaks were clearly distinguished from unlabeled contaminating peptides. Furthermore, nitrotyrosine-containing peptides were selectively detected on the 3H4NBA matrix, while by-product-peaks arising from nitrobenzene moieties were suppressed. The use of 3H4NBA as a comatrix with CHCA improved the sensitivity of detection while essentially retaining the selectivity of 3H4NBA. The 3H4NBA matrix offers great advantages in terms of simplicity, sensitivity, and utility when used for the NBS method and for MALDI-TOF MS analysis applied to compounds having a nitrobenzene ring. The improvements described in Chapters II and III resulted in reduced sample loss and increased sensitivity, and as a result, an expansion of dynamic range of measurement. By combining NBS reagents with HPLC, an automatic spotter, MALDI-TOF MS, and analytical software, the author constructed a system called the “NBS Biomarker Discovery System”.

In Chapter VI, the author applied the system to proteomic profiling of CRC to identify novel proteins with altered expression in CRC. Each pair of tumor and normal tissue specimens from 12 CRC patients was analyzed, and approximately 5000 NBS-labeled paired peaks were quantified. Peaks with altered signal intensities (>1.5-fold) and occurring frequently in the samples (>70%) were selected, and 128 proteins were identified by MS/MS analyses as differentially expressed proteins in CRC tissues. Many proteins were newly revealed to be CRC related; 30 were reported in earlier studies of CRC. Six proteins that were up-regulated in CRC (ZYX, RAN, RCN1, AHCY, LGALS1, and VIM) were further characterized and validated by Western blot and immunohistochemistry. All six were found to be CRC-localized, either in cancer cells or in stroma cells near the cancer cells. These results indicate that the proteins identified by the system can be novel cancer biomarkers.

As a result, in this thesis, it was demonstrated that the NBS method provide highly

accurate relative quantification using clinical samples and the analysis system is useful for the novel biomarker discovery.

List of Publications

1. Watanabe M., Takemasa I., Kawaguchi N., Miyake M., Nishimura N., Matsubara T., Matsuo E., Sekimoto M., Nagai K., Matsuura N., Monden M. and Nishimura O.

"An Application of the 2-Nitrobenzenesulfenyl (NBS) Method to Proteomic Profiling of Human Colorectal Carcinoma: A Novel Approach for Biomarker Discovery."

Proteomics-Clinical Applications, **4**: 925-931 (2008)

2. Matsuo E., Toda C., Watanabe M., Iida T., Masuda T., Minohata T., Ando E., Tsunasawa S. and Nishimura O.

"Improved 2-nitrobenzenesulfenyl method: optimization of the protocol and improved enrichment for labeled peptides."

Rapid Commun. Mass Spectrom., **20**: 31–38 (2006)

3. Matsuo E., Toda C., Watanabe M., Ojima N., Izumi S., Tanaka K., Tsunasawa S. and Nishimura O.

"Selective detection of 2-nitrobenzenesulfenyl-labeled peptides by matrix-assisted laser desorption/ ionization-time of flight mass spectrometry using a novel matrix."

Proteomics, **6**:2042-2049 (2006)

4. Kuyama H., Watanabe M., Toda C., Ando E., Tanaka K. and Nishimura O.

"An approach to quantitative proteome analysis by labeling tryptophan residues."

Rapid Commun. Mass Spectrom., **17**: 1642–1650 (2003)

Related Publications

1. Okamura N., Masuda T., Gotoh A., Shirakawa T., Terao T., Kaneko N., Suganuma N., Watanabe M., Matsubara T., Seto R., Matsumoto J., Kawakami M., Yamamori M., Nakamura T., Yagami T., Sakaeda T., Fujisawa T., Nishimura O. and Okumura K. **"**Quantitative proteomic analysis to discover potential diagnostic markers and therapeutic targets in human renal cell carcinoma"

Proteomics, in press, (2008)

2. Iida T., Kuyama H., Watanabe M., Toda C., Matsuo E., Kido A., Ando E., Tsunasawa S. and Nishimura O.

"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 Tech., **17**:333-41 (2006)

3. Kita K., Okumura K., Takao T., Watanabe M., Matsubara T., Nishimura O. and Nagai K. "Evidence for phosphorylation of rat liver glucose-regulated protein 58, GRP58/ERp57/ER-60, induced by fasting and leptin"

FEBS Lett., **580**(1):199-205 (2006)

4. Kuyama H., Toda C., Watanabe M., Tanaka K. and Nishimura O.

"An efficient chemical method for dephosphorylation of phosphopeptides"

Rapid Commun Mass Spectrom., **17**:1493-6 (2003)