

Figure 5. MALDI-mass spectra of tryptic digests of AP. The acquired profiles were focused on mass ranges from 1970 to 2100 (A), from 2100 to 2300 (B), from 2450 to 2540 (C) and from 3600 to 5000 (D). The purified AP was treated under three following conditions: (a) reduction and S-alkylation, (b) S-alkylation, (c) no treatment. Arrows and numerical values indicate specific peptide peaks and observed masses, respectively.

Table 4. Theoretical and observed mass values for disulfide-linked peptides (AP)

Mass value	Position	MC ^a	Modification(s)	Modified mass value	Observed mass value	Peptide sequence ^b
2167.63 ^{c1}	331–353	0	Disulfide bond: 288–338	4649.06	4649.03	QDHAANPCGQIGI-TYDLDEAVQR
2184.43 ^{c1}	275–294	0				ATYHGNIDKPWTCTPNPQR
1982.93 ^d	169–187	2	Disulfide bond: 170–180	1980.93 ^b	1980.97	KCYGPSAISI-KCPGNALFK

a) MC stands for the number of missed cleavages

b) Disulfide linkages are indicated by lines

c) Average mass values. The mass spectrum of this disulfide-linked peptide was acquired in linear positive ion mode

d) Monoisotopic mass values. The mass spectrum of this disulfide-linked peptide was acquired in reflectron positive ion mode

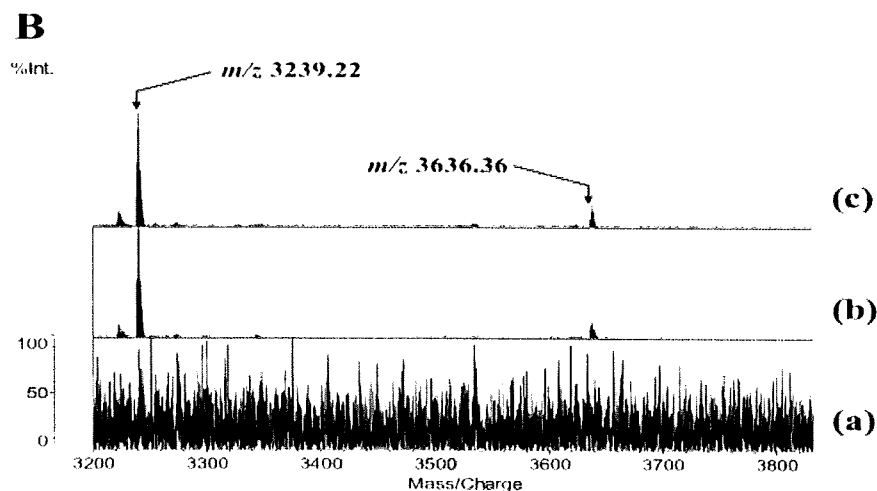
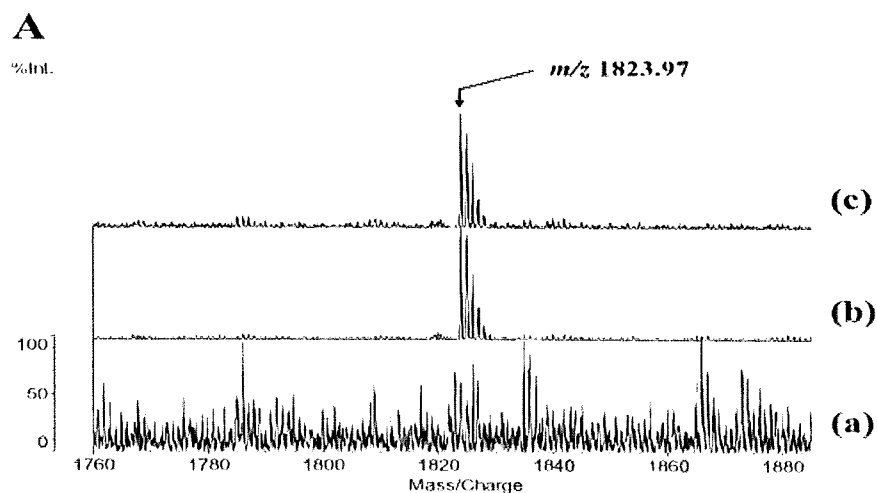


Figure 6. MALDI-mass spectra of tryptic digests of h-LYZ. The acquired profiles were focused on mass ranges from 1760 to 1880 (A) and from 3200 to 3800 (B). The purified h-LYZ was subjected to the three following conditions: (a) reduction and S-alkylation, (b) S-alkylation, (c) no treatment. Arrows and numerical values indicate specific peptide peaks and observed masses, respectively.

Table 5. Theoretical and observed monoisotopic mass values for disulfide-linked peptides (h-LYZ)

Mass value	Position	Modification(s)	Modified mass value	Observed mass value	Peptide sequence ^{a)}
2756.32	71–98	Disulfide bond:	3636.67	3636.36	TPGAVNACHISCSALLQDNIADAVACAK
885.36	64–70	66–82, 78–96			YWCNDGK
2651.17	124–131	Disulfide bond:	3239.46	3239.22	QYVQGCQVGLGSGGGGGGWSHPQELK
591.29	7–11	7–129			CFLAR
1306.67	23–34	Disulfide bond:	1823.89	1823.97	GISLANWMLCAK
520.23	117–120	31–117			CQNR

a) Disulfide linkages are indicated by lines.

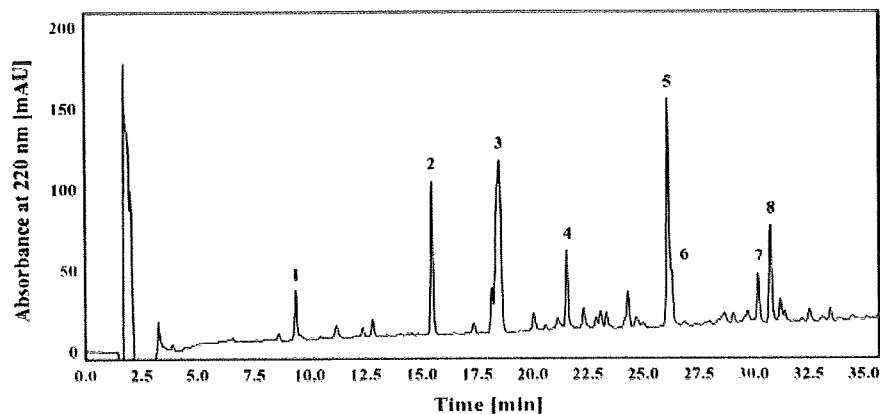


Figure 7. Separation of tryptic digests of h-LYZ synthesized in insect cell-free system using reverse-phase HPLC. The tryptic digests of h-LYZ were separated by reverse-phase HPLC with a 50-min linear gradient increasing from 5 to 50% ACN containing 0.1% TFA. The flow rate was 1.0 mL/min. Numerical values indicate peak IDs. Each peptide in the eight peaks was identified by MALDI-TOF MS, as shown in Table 6.

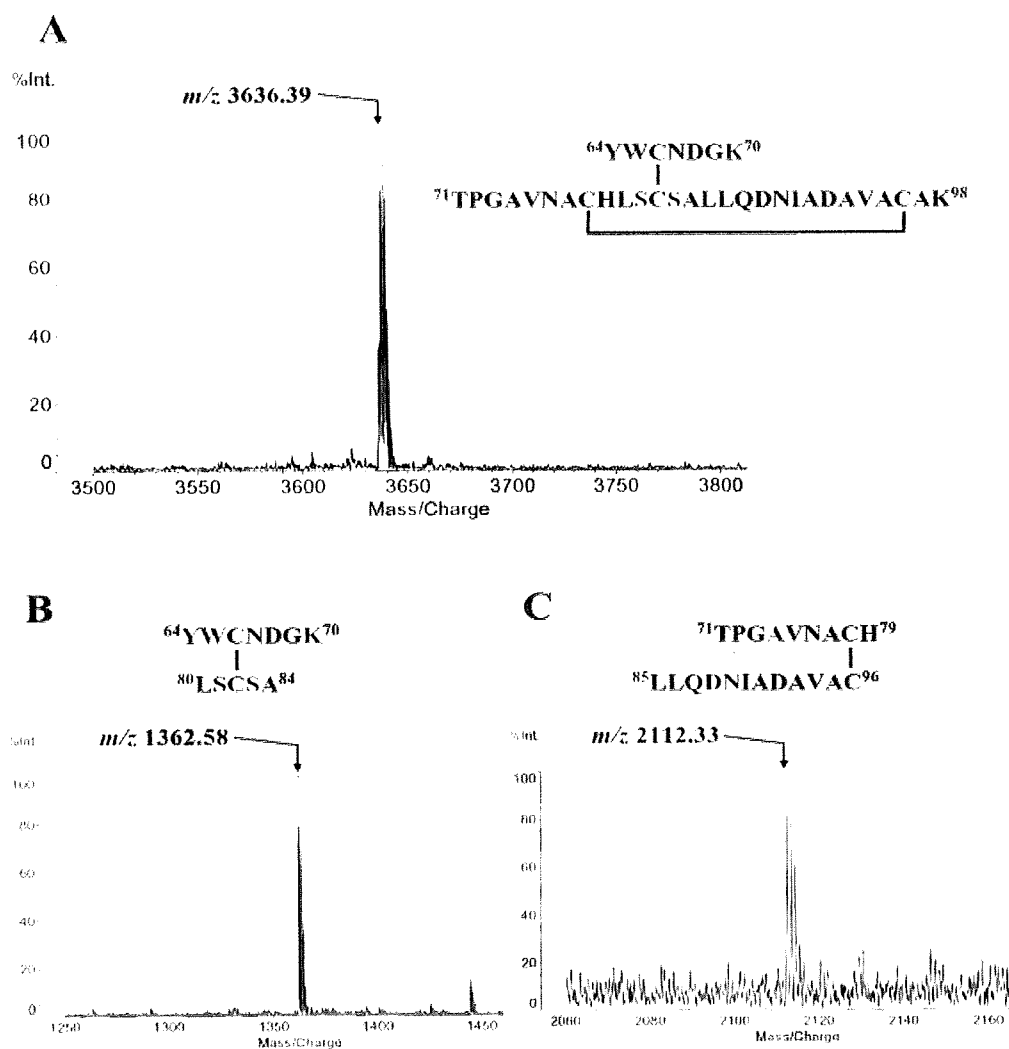


Figure 8. MALDI-mass spectra of thermolytic digests of peptide 8. The spectrum of peptide 8 (A). The acquired profiles were focused on mass ranges from 1250 to 1450 (B) and from 2060 to 2160 (C). Arrows and numerical values indicate specific peptide peaks and observed masses, respectively. Alphabet characters and lines show the corresponding amino acid sequences and disulfide linkages, respectively.

Table 6. Theoretical and observed monoisotopic mass values for tryptic digests of synthesized h-LYZ separated using reverse-phase HPLC

Peak ID	Position	Modification(s)	Theoretical mass value	Observed mass value
1	43–51		981.43	981.49
2	35–42		1012.45	1012.40
3	16–22		811.38	811.45
4	7–11	Disulfide bond:	3239.46	3239.14
	124–131	7–129		
5	109–114		788.42	788.04
6	52–63		1400.68	1400.34
7	23–34	Disulfide bond:	1823.89	1823.99
	117–120	31–117		
8	64–70	Disulfide bond:	3636.67	3636.39
	71–98	66–82, 78–96		

5 References

- [1] Ezure, T., Suzuki, T., Higashide, S., Shintani, E. *et al.*, Cell-free protein synthesis system prepared from insect cells by freeze-thawing. *Biotechnol. Prog.* 2006, 22, 1570–1577.
- [2] Suzuki, T., Ito, M., Ezure, T., Shikata, M. *et al.*, N-Terminal protein modifications in an insect cell-free protein synthesis system and their identification by mass spectrometry. *Proteomics* 2006, 6, 4486–4495.
- [3] Suzuki, T., Ito, M., Ezure, T., Shikata, M. *et al.*, Protein prenylation in an insect cell-free protein synthesis system and identification of products by mass spectrometry. *Proteomics* 2007, 7, 1942–1950.
- [4] Chen, H. Z., Zubay, G., Prokaryotic coupled transcription-translation. *Methods Enzymol.* 1983, 101, 674–690.
- [5] Erickson, A. H., Blobel, G., Cell-free translation of messenger RNA in a wheat germ system. *Methods Enzymol.* 1983, 96, 38–50.
- [6] Gasior, E., Herrera, F., Sadnik, I., McLaughlin, C. S. *et al.*, The preparation and characterization of a cell-free system from *Saccharomyces cerevisiae* that translates natural messenger ribonucleic acid. *J. Biol. Chem.* 1979, 254, 3965–3969.
- [7] Swerdel, M. R., Fallon, A. M., Cell-free translation in lysates from *Spodoptera frugiperda* (Lepidoptera: Noctuidae) cells. *Comp. Biochem. Physiol.* 1989, 93B, 803–806.
- [8] Jackson, R. J., Hunt, T., Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA. *Methods Enzymol.* 1983, 96, 50–74.
- [9] Ryabova, L. A., Desplancq, D., Spirin, A. S., Plückthun, A., Functional antibody production using cell-free translation: effects of protein disulfide isomerase and chaperones. *Nat. Biotechnol.* 1997, 15, 79–84.
- [10] Merk, H., Stiege, W., Tsumoto, K., Kumagai, I. *et al.*, Cell-free expression of two single-chain monoclonal antibodies against lysozyme: Effect of domain arrangement on the expression. *J. Biochem.* 1999, 125, 328–333.
- [11] Jiang, X. P., Ookubo, Y., Fujii, I., Nakano, H. *et al.*, Expression of Fab fragment of catalytic antibody 6D9 in an *Escherichia coli* *in vitro* coupled transcription/translation system. *FEBS Lett.* 2002, 514, 290–294.
- [12] Kim, D. M., Swartz, J. R., Efficient production of a bioactive, multiple disulfide-bonded protein using modified extracts of *Escherichia coli*. *Biotechnol. Bioeng.* 2004, 85, 122–129.
- [13] Oh, I. S., Kim, D. M., Kim, T. W., Park, C. G. *et al.*, Providing an oxidizing environment for the cell-free expression of disulfide-containing proteins by exhausting the reducing activity of *Escherichia coli* S30 extract. *Biotechnol. Prog.* 2006, 22, 1225–1228.
- [14] Kawasaki, T., Gouda, M. D., Sawasaki, T., Takai, K. *et al.*, Efficient synthesis of a disulfide-containing protein through a batch cell-free system from wheat germ. *Eur. J. Biochem.* 2003, 270, 4780–4786.
- [15] Bradshaw, R. A., Cancedda, F., Ericsson, L. H., Neumann, P. A. *et al.*, Amino acid sequence of *Escherichia coli* alkaline phosphatase. *Proc. Natl. Acad. Sci. USA* 1981, 78, 3473–3477.
- [16] Sone, M., Kishigami, S., Yoshihisa, T., Ito, K., Roles of disulfide bonds in bacterial alkaline phosphatase. *J. Biol. Chem.* 1997, 272, 6174–6178.
- [17] Kanaya, E., Kikuchi, M., Folding of human lysozyme *in vivo* by the formation of an alternative disulfide bond. *J. Biol. Chem.* 1992, 267, 15111–15115.
- [18] Kanaya, E., Ishihara, K., Tsunasawa, S., Nokihara, K. *et al.*, Indication of possible post-translational formation of disulfide bonds in the β -sheet domain of human lysozyme. *Biochem. J.* 1993, 292, 469–476.
- [19] Suzuki, T., Ito, M., Ezure, T., Kobayashi, S. *et al.*, Performance of expression vector, pTD1, in insect cell-free translation system. *J. Biosci. Bioeng.* 2006, 102, 69–71.
- [20] Yamaguchi, M., Nakazawa, T., Kuyama, H., Obama, T. *et al.*, High-throughput method for N-terminal sequencing of proteins by MALDI mass spectrometry. *Anal. Chem.* 2005, 77, 645–651.
- [21] Anfinsen, C. B., Haber, E., Sela, M., White, F. H., The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA* 1961, 47, 1309–1314.
- [22] Noiva, R., Lennarz, W. J., Protein disulfide isomerase. A multifunctional protein resident in the lumen of the endoplasmic reticulum. *J. Biol. Chem.* 1992, 267, 3553–3556.

総 説



安定同位体利用技術

プロテオーム解析のための安定同位体標識試薬[†]

九山浩樹*, 西村 紀*,**

*大阪大学蛋白質研究所

565-0871 大阪府吹田市山田丘 3-2

**株式会社 島津製作所 分析計測事業部

ライフサイエンスビジネスユニット ライフサイエンス研究所

604-8511 京都府京都市中京区西ノ京桑原町 1

Key Words : proteomics, quantification, differential analysis, 2-nitrobenzenesulfonyl chloride, stable isotope

1. はじめに

生命現象は遺伝子による精緻な制御のもとにある。そしてその直接的な担い手は主として蛋白質・ペプチドであると考えられてはいたが、生命現象の解明は遺伝子からのアプローチによって可能であるとされていた。このような潮流のもとで人的資源等が投入され生命の解明に大きく寄与してきた。しかしながら、mRNA レベルでみた遺伝子発現の変化が必ずしも細胞内蛋白質の量的変化とは一致しないこと、また多くの蛋白質が翻訳後に修飾を受けて機能するが、この翻訳後修飾も遺伝子情報からだけでは解明できないことなどから蛋白質レベルでの解析が極めて重要視されるようになった。著者(O.N)は製薬企業の基礎研究部門に長く籍を置き、創

薬研究に携わった経験からこのことを強く感じている。

このようなコンテキストにおける蛋白質の解析とはプロテオーム解析と呼ばれる。プロテオームという言葉は、gene-genome からの類推から生まれた言葉であり、プロテオーム解析とはある特定の細胞中で発現している蛋白質の時系列での網羅的定量解析と一般的に定義される。プロテオーム解析の方法論として従来から一般的なものは二次元電気泳動で分離・定量し、質量分析装置により蛋白質の同定を行うものである(2D-PAGE)。しかしながらゲルを用いる分離法には、ゲル板そのものの品質均一性、染色の再現性、極端な等電点・分子量を持つもの及び疎水性蛋白質などゲルでの分離に適用できない蛋白質の存在などの問題点があることが認識されており¹⁾、また、見かけ上は単一スポットであるが、co-elution によって複数の蛋白質が含まれる場合の蛍光検出での定量における問題点が安定同位体試薬による方法との比較で論じられている²⁾。このような問題点を解決する方法として開発されたのが安定同位体で標識された試薬と質量分析を組み合わせた方法論である。

本稿では、まずそのプロトコールについて述べ、次いで現在報告されている試薬について概

[†] "Applications of Stable Isotopes in Life Sciences". Stable Isotope Labeling Reagents for Proteome Analysis.

Hiroki KUYAMA* and Osamu NISHIMURA*,** :Institute for Protein Research, Osaka University, 3-2, Yamadaoka, Suita-shi, Osaka 565-0871, Japan, **Life Science Laboratory, Life Science Business Unit, Analytical & Measuring Instruments Division, Shimadzu Corporation, 1, Nishinokyo-Kuwabara-cho, Nakagyo-ku, Kyoto 604-8511, Japan.

観する。最後に筆者らが開発した 2-nitrobenzenesulfenyl chloride (NBSCl) を用いる方法³⁾について解説する。

2. 安定同位体試薬を用いるプロテオーム・ディファレンシャル解析のプロトコル⁴⁾⁻⁸⁾

安定同位体を用いた試薬の蛋白質（あるいはペプチド）への導入は生体組織あるいは体液から抽出された蛋白質・ペプチドサンプルに対して行う方法（*in vitro* ラベル化）と、培養細胞の培地に添加する方法（*in vivo* ラベル化）とに大別される。

2.1 *in vitro* ラベル化

生体サンプルから抽出・精製された蛋白質を安定同位体試薬でラベル化する方法である。2.2 で述べる *in vivo* ラベル化と比較すると、抽出された蛋白質を出発物質とするため、培地中で生育できない組織・細胞などにも適用でき、一般性がある方法論である。（相対）定量解析（ディファレンシャル解析）であるから比較の対象とするものが必要である。例えば、ある病気の場合に変動している蛋白質のプロテオーム解析を行う場合は、病態の組織あるいは体液から抽出された蛋白質と病態でない組織あるいは体液から抽出された蛋白質を用意する。ラベル化試薬にはその構成原子すなわち水素 (^1H)、炭素 (^{12}C)、窒素 (^{14}N) 等のうち単一、あるいは複数組み合わせてそれぞれの安定同位体 (D , ^{13}C , ^{15}N) と置換したものと、そうでないものの 2 種（重い試薬と軽い試薬）を用いる。これら 2 種類のラベル化試薬の質量は質量分析でのシグナルの同位体分布を考慮しペアで観測されるピークの重なりが相対定量に影響しないように 3 から 4 以上の差を持つように設計される。病態サンプルに対して、例えば重い試薬でラベル化し、もう一方の健常サンプルには軽い試薬と反応させる。このステップまでは二つのサンプルを独立して処理するが、この反応以降

は混合して進める。すなわち、二つを混ぜた反応溶液を脱塩精製、バッファー交換などを行った後、酵素消化を行い、次いでラベル化されたペプチド断片をアフィニティカラムで濃縮後、更に液体クロマトグラフィなどで分離精製し、質量分析を行う。図 1 にこの手順をまとめた。

2.2 *in vivo* ラベル化

この方法は生物体中での代謝によって安定同位体を含む化学種を取り込ませ、蛋白質をラベル化する方法である。従来からの放射性同位元素の取り込みによる方法論と類似している。細胞の培地中に安定同位体元素を含む化合物を加え、細胞の成長とターンオーバー（代謝回転）により細胞中の全ての蛋白質に取り込ませる。図 1 に示したように、培地中に加える安定同位体元素含有化合物が重いものと軽いものの 2 種あるため、細胞培養は別々に行う。2.1 で述べた *in vitro* ラベル化と実験プロセスを比較した場合、この方法は培地中で生育できないものには適用できないという点では一般性がないが、ラベル化は生体中でなされ、細胞の段階で混合されて以後の種々のプロセスを受けるところに有利な点がある。すなわち、混合後のプロセス、例えば蛋白質抽出、精製などは混合状態で進められ抽出効率・回収率などは 2 種のラベル化された蛋白質・ペプチドで差がないため、これら各ステップは後の質量分析装置での定量解析に影響を与えない（図 1）。

3. 発現蛋白質の相対定量比較（ディファレンシャル解析）に用いられる安定同位体標識試薬

ここでは、ラベル化のターゲットとなるアミノ酸残基及び官能基からみた試薬についていくつか紹介する（図 2, 図 3）。なお、トリプトファン残基に対する NBS 試薬によるラベル化の方法論については 4 で解説する。

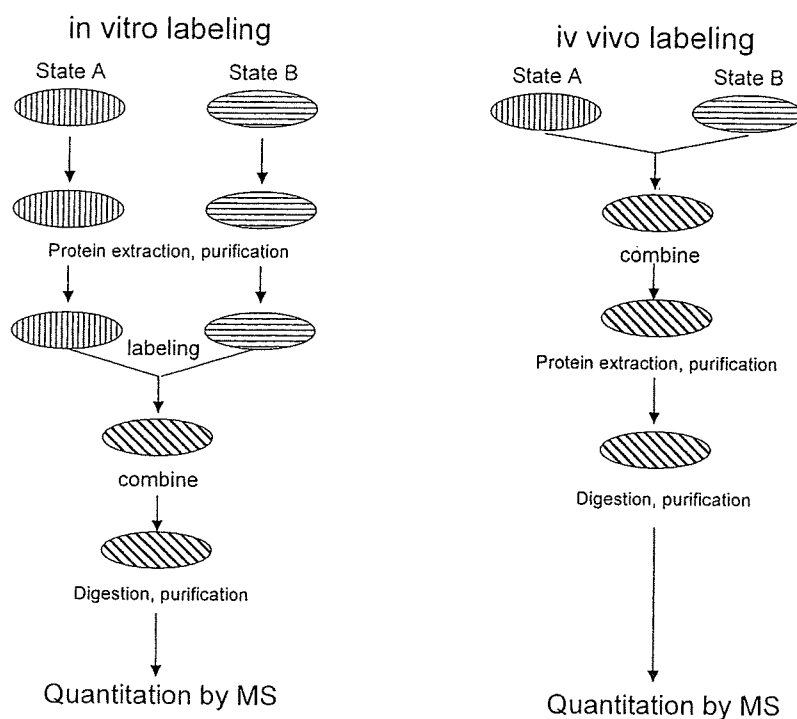


図1 解析プロトコール概略

3.1 システイン (スルフヒドリル基)

1999年に Aebersold らによって報告された isotope-coded affinity tag (ICAT) 法⁹⁾がこの種の方法論の最初である。この試薬はヨウ素化された反応部位、重水素化されたポリエーテルリンカー部、そしてビオチン部分からなる。リンカー部の水素が8個重水素に置換された構造であるので重い試薬と軽い試薬の質量数の差は8である。ビオチン部分はラベル化後の酵素消化断片の混合物からアビジンを用いるアフィニティカラムによってラベル化されたペプチド断片のみを濃縮・精製するために用いられる。重水素によって質量数の差を与える方法は液体クロマトグラフィで同位体効果がみられ溶出時間に差があるため、後に¹³Cによって質量差をあたえる構造の化合物が開発された (Cleavable ICAT 試薬^{10),11)}。この他、システイン (スルフヒドリル基) へのラベル化試薬として acrylamide^{12),13)} (D×3), 2-vinylpyridine¹⁴⁾ (D×4), iodoacetanilide¹⁵⁾ (D×5) などが報告されている (図2)。

3.2 リジン (アミノ基)

アミノ基一般をターゲットとする試薬については3.4で述べるが、リジン残基のアミノ基は O-methylisourea によって選択的にグアニジノ化されることを利用したものである^{16),17)}。重い試薬の化学構造はグアニジノ基の炭素原子、二つの窒素原子がそれぞれ¹³C, ¹⁵Nで置換されたものであり、軽い試薬との質量数の差は3である。リジン残基のアミノ基のグアニジノ基への変換は、すなわち残基としてはリジンのホモアルギニンへの変換である。MALDI-TOFMS (Matrix Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry) を用いた質量分析という観点から見れば、この試薬でラベル化された蛋白質はトリプシンによる (一般的な) 酵素消化を行った場合、得られるペプチド断片の全てのC末端にアルギニンあるいはホモアルギニンが位置することになる。一般的にC末端にアルギニンを有するペプチドは、MALDI-MS (Matrix Assisted Laser Desorption/Ionization Mass Spectrometry) で

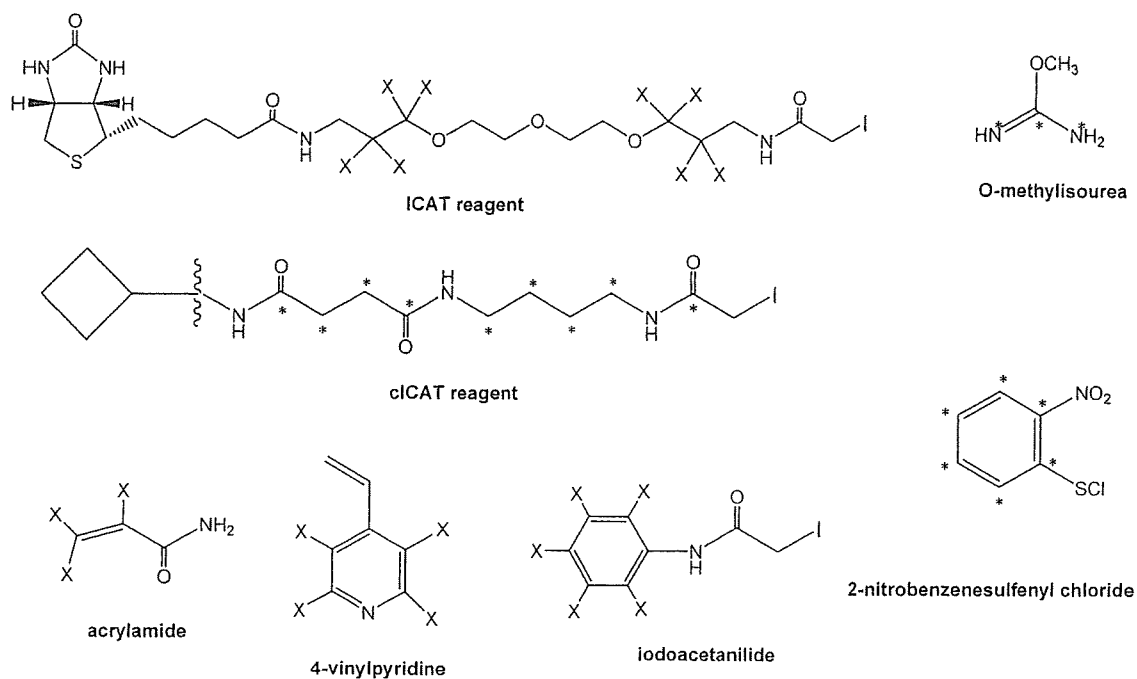


図2 安定同位体元素を含む試薬 (その1)

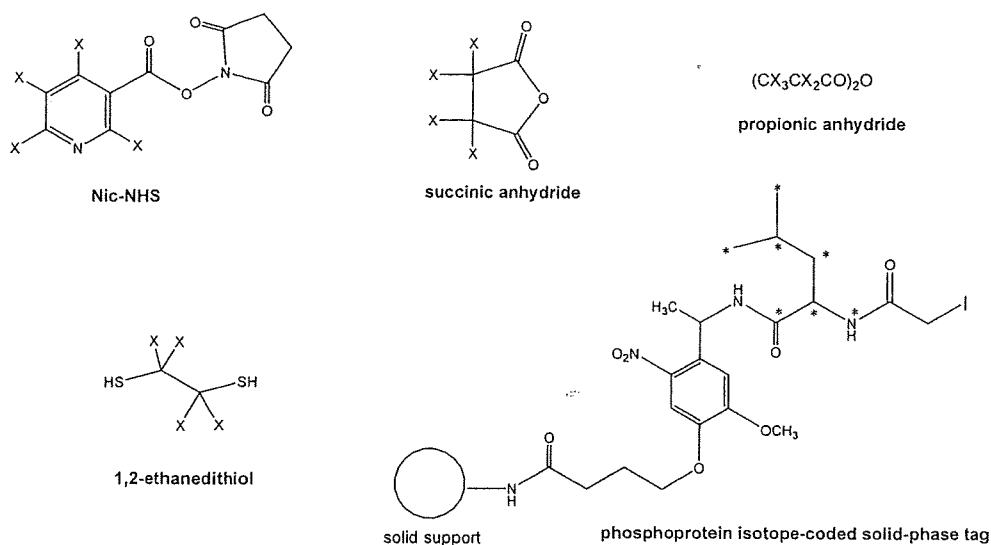
X: H or D, *: (^{12}C or ^{13}C) or (^{14}N or ^{15}N)

図3 安定同位体元素を含む試薬 (その2)

X: H or D, *: (^{12}C or ^{13}C) or (^{14}N or ^{15}N)

の感度が良いことが知られているので、トリプシン消化で生ずるペプチド断片の全てを感度よく観測することができる。また、タンデムマス測定時のフラグメントイオンが単純化されるので de novo 配列解析にも有効であることが報文¹⁸⁾中で述べられている (図2)。

3.3 リン酸化セリン・スレオニン

これは翻訳後修飾のひとつである蛋白質のリン酸化をターゲットとした解析法である。この方法はセリンあるいはスレオニンの水酸基とエステル結合したリン酸基が塩基性条件下でベータ脱離反応を起こすことを基礎としている。ベータ脱離が起こったセリン及びスレオニン残基

はデヒドロアラニン骨格となり, α, β -不飽和カルボニル系を有する残基へと変化する。生じた α, β -不飽和カルボニル系はマイケルアクセプタであるので, 適当な求核試薬を働かせてマイケル付加反応することにより安定同位体で標識された試薬でラベル化することができる。求核試薬・ラベル化試薬としては, 1,2エタンジチオール¹⁹⁾ ($D \times 4$) 等のチオール誘導体が用いられ, また, 光開裂性リンカーを有する化合物²⁰⁾ ($^{13}C \times 6, ^{15}N \times 1$) 等をマイケル付加反応後に生じる SH 基に反応させる 2段階での導入が報告されている。これらを用いると, 質量数の差がそれぞれ 4 及び 7 となる (図 3)。

3.4 アミノ基

従来から知られているアミノ基の修飾試薬に安定同位体元素を組み込んだ試薬がいくつか開発されている。ラベルするアミノ基は蛋白質の N 末端のみ, リジンの ϵ -アミノ基のみ (3.2), 酵素消化後のペプチドに対するラベル化など, いくつかの方法が報告されている。これらの方法と関連するが, 蛋白質中のアミノ基は数が多く, ラベル化後の酵素消化ペプチド混合物は極めて複雑であるので, 予め蛋白質の段階で 2D-PAGE, HPLC などで分離しておくか, リジンのグアニジノ化の場合もそうであるがラベル化—酵素消化後のペプチド混合物を HPLC で分離精製して質量分析を行う等, 複雑さを小さくするように実験系が工夫されるのが普通である。ラベル化試薬としては, コハク酸無水物²¹⁾ ($D \times 4$), ニコチン酸の N-ヒドロキシコハク酸イミドエステル^{22), 23)} (Nic-NHS: $D \times 4$), プロピオン酸無水物²⁴⁾ ($D \times 10$), ホルムアルデヒド²⁵⁾ ($D \times 2, ^{13}C \times 1$) などがあげられる (図 3)。

2D-HPLC 等の分離手法の発展と, 多検体の高速かつ高効率な解析を可能にする質量分析装置 (及び解析ソフトウェア) の進歩を背景にした, 蛋白質消化断片中のアミノ基を全てラベル化してプロテオーム解析を行うという方法も開発されている (iTRAQ 法^{4), 26)}。

3.5 その他

C 末端及び側鎖のカルボキシル基をエステル化によりラベル化する方法^{27), 28)}, セリン及びスレオニン残基の O-グリコシル化部位のラベル化による方法^{29) - 31)}, ^{18}O 水中でのトリプシン消化により各消化断片に C 末端に ^{18}O を導入する方法^{32), 33)} などが報告されている。

4. トリプトファンをターゲットとするプロテオーム解析法及びそのラベル化試薬

ここからは, 筆者らが開発したプロテオーム解析法³⁾について述べる。この方法は蛋白質構成アミノ酸残基中のトリプトファン残基をラベル化のターゲットとするもので, ラベル化にはハロゲン化スルフェニル試薬のひとつ, 塩化 2-ニトロベンゼンスルフェニル^{34), 35)} (2-nitrobenzenesulfonyl chloride: NBSCl) (別名: 2-nitrophenylsulfenyl chloride: NPSCl) を用いる。

トリプトファン残基は生体中で, 例えば触媒作用等, 様々な働きを担っており, 蛋白質の生物活性に重要な寄与をもつ残基として知られている。プロテオーム解析を目的とする場合, カバー率が重要なファクターであるが, トリプトファン残基の蛋白質中での分布は, マウスあるいはヒトの場合 90% 以上³⁶⁾であり, システイン残基の場合の 95% と比べても遜色が無いと考えられる。また, 個々の蛋白質での含有率は最も低いアミノ酸残基であるので最終的に得られるマススペクトルの複雑さがかなりの程度減少することが予測された。塩化ベンゼンスルフェニル (NBS) 試薬の反応性については, 報告^{34), 35)}がなされており, 選択性よくトリプトファン残基に反応することが知られている。ラベル化—酵素消化後の目的ペプチド断片の分離・濃縮に関して, トリプトファン含有ペプチドがセファデックス系の充填剤に親和性を示すことが報告^{37), 38)}されており, この方法では, トリプトファン残基の側鎖インドール環を疎水性の 2-ニトロベンゼンスルフェニル基で修飾するため, 疎水性が更に大きくなりセファデックス系の充

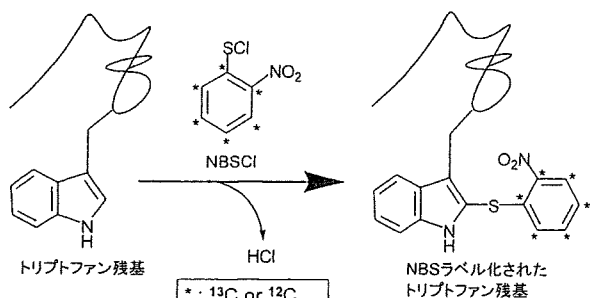


図4 トリプトファン残基とNBS試薬との反応

填剤に対してより強い親和性を示すと予想された。質量タグとしての2-ニトロベンゼンスルフェニル基は質量数が小さく (m/z : 153) 構造が比較的単純であるので、質量分析を行う際にタグ由来のフラグメンテーションによってデータベース検索に悪い影響を与えないと予測した。以上のようなストラテジーで本法の開発を進めた。

4.1 原理と実験プロセス

NBS試薬はトリプトファン残基のインドール環を酸性条件下で選択的に修飾し、重い試薬と軽い試薬が導入されたそれぞれのラベル化ペプチドの質量数の差は6Daである。図4にトリプトファン残基とNBS試薬の反応を示す。本法はNBSClのトリプトファン側鎖のインドール環に対する選択的な反応(求電子置換反応)に着目したところに新規性がある。一般的にハロゲン化アールスルフェニル試薬(2-nitrobenzenesulfenyl chloride^{34), 35), 4-nitrobenzenesulfenyl chloride^{39), 2,4-dinitrobenzenesulfenyl chloride^{40), 2-nitro-4-carboxybenzenesulfenyl chloride^{37), 41), 等)でこのような反応性が以前から知られている。これらのうちで2-nitrobenzenesulfenyl chlorideはトリプトファン、システイン残基以外のアミノ酸残基とは事実上反応しないことが報告³⁵⁾されていることから、本法でのラベル化試薬として採用した。反応の選択性に関しては、トリプトファン、システイン残基の双方に反応し、これら両者間での選択性はないと言ってよい。しかしながら、生}}}}

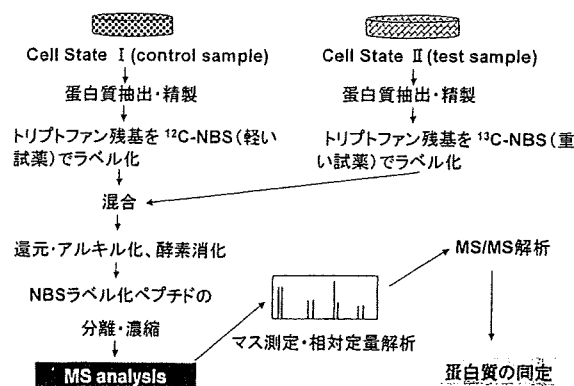


図5 NBS法のプロトコル概略

成する結合様式はトリプトファン残基とはスルフィド結合、システイン残基とはジスルフィド結合であるので、NBSタグが結合したシステイン残基は、以下にのべるように実験プロセス中の(蛋白質中のジスルフィド結合の)還元操作の際に還元的開裂を受けてスルフヒドリル基に戻り、次のヨードアセトアミドによるアルキル化でシステイン由来のスルフヒドリル基とともにアルキル化される³⁾。

さて、実験プロセスの概略は次の通りである。生体試料から抽出した蛋白質を、1) 変性させ、一方のサンプルを軽い試薬で、もう一方を重い試薬でラベル化する。2) これら二つのサンプルを混合し、過剰の試薬等をゲルろ過(セファデックス LH-20 等)により除去する。3) サンプル蛋白質中のジスルフィド結合を還元し次いでアルキル化する。4) 酵素消化を行い、ペプチド混合物とする。5) NBSラベル化ペプチドをフェニルセファロースカラムで分離・濃縮する。6) 必要ならば、更に逆相のHPLCで分離して質量分析を行う。上述のように、このプロセス全体で見ればNBSタグはトリプトファン残基のみに結合し選択的な反応が達成される(図5)。このプロセスは蛋白質変性剤及びラベル化ペプチド濃縮カラムの選択などで改良を加え⁴²⁾、またMALDI-TOFMSでの検出の際のマトリクスの探索を行って、NBSラベル化ペプチドを選択的に感度よく検出するマトリクス添加剤(3-hydroxy-4-nitrobenzoic acid)を見出している⁴³⁾(図6)。

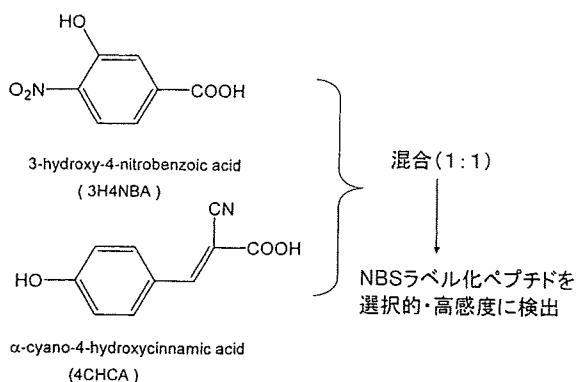


図6 NBSラベル化ペプチドを高感度・高選択的に検出するための混合マトリクス法

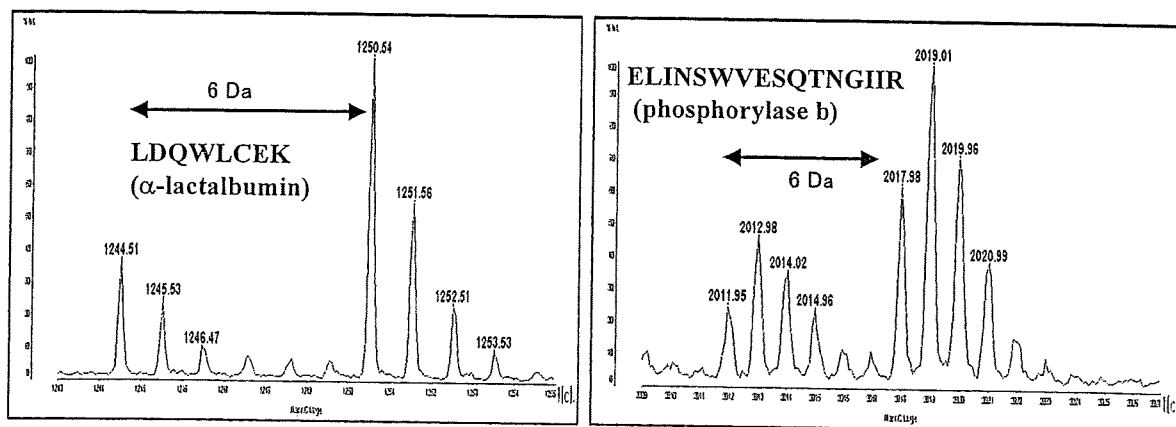
4.2 モデル蛋白質を用いた相対定量の実例³⁾

実際のマスペクトル上でどのようなシグナルが観測されるかをモデル蛋白質を用いた実験から例を示す。用いた蛋白質は glyceraldehyde-3-phosphate dehydrogenase (rabbit), phosphorylase b (rabbit), ovalbumin (chicken), α -lactalbumin (bovine) の4種であり、これらを用いて二つの状態を次のようにモデル化した: 状態A (4種を各々 12.5 μ g), 状態B (4種を各々 25 μ g)。これら二つの状態間での4種の蛋白質の比はいずれも 1:2である。このように調製した二つのサンプルを4.1で概略を述べたプロセスに従って処理し、目的とするNBSラベル化ペプチドの混合物を得た。この

混合物をMALDI-TOFMS (AXIMA-CFR: 島津製作所) で測定したスペクトルを図7に示す。ここでは例として4種の蛋白質のうち、 α -lactalbumin由来のLDQWLCEKと phosphorylase bからのELINSWVESQTNGIIRのシグナルを抽出した。いずれもペプチド断片中に一つのトリプトファン残基を有するため質量数6Da差のペアとして観測される。それぞれのペアピークの強度比は1:2であり、調製したモデル系での設定比とよい一致を示している。プロテオーム解析では質量分析装置での測定と同時に解析ソフトによる蛋白質同定も重要な位置を占める。図8にはデータベース検索による蛋白質同定の例を示す。これはlysozyme (chicken)を材料とした実験からのデータであるが、 m/z : 1198, 1204に観測される6Da差のペアピーク的一方、 m/z : 1204のピークに対してPSDを行ってマスコット検索した結果である。ニワトリ由来のリゾチームであることがペプチドの配列情報とともに一位にランクされている。

4.3 生体試料を用いた実例³⁾

実際の生体試料であるラット血清に対して筆者らの方法を適用して発現蛋白質の解析を試みた。用いた血清の由来は、Crj:Wister (正常



α -lactalbumin及び phosphorylase b由来のラベル化ペプチドを示す。それぞれのペアピークの強度比は1:2であり、モデルとして設定した比率と良い一致を示した。

図7 モデル蛋白質 (4種混合) での一例

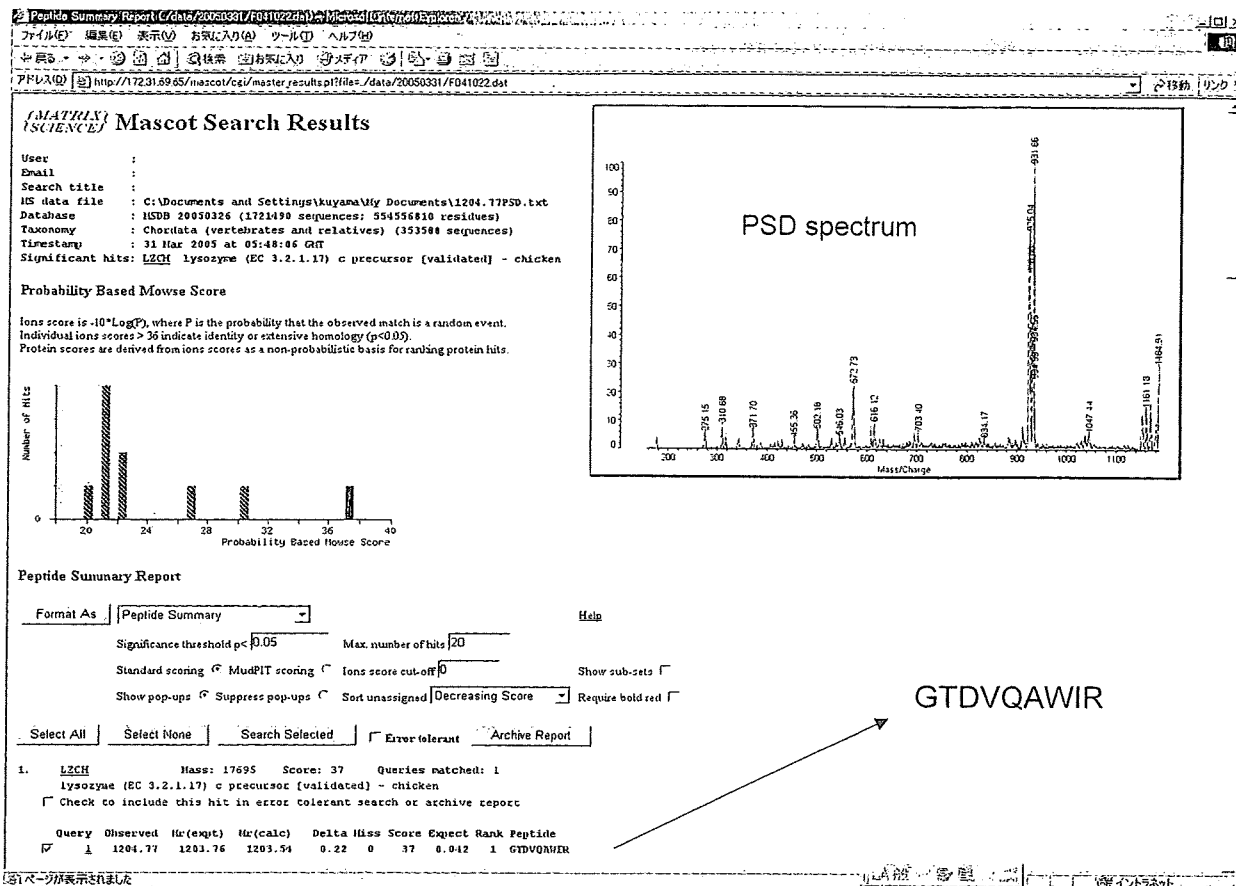


図8 蛋白質のデータベース検索例

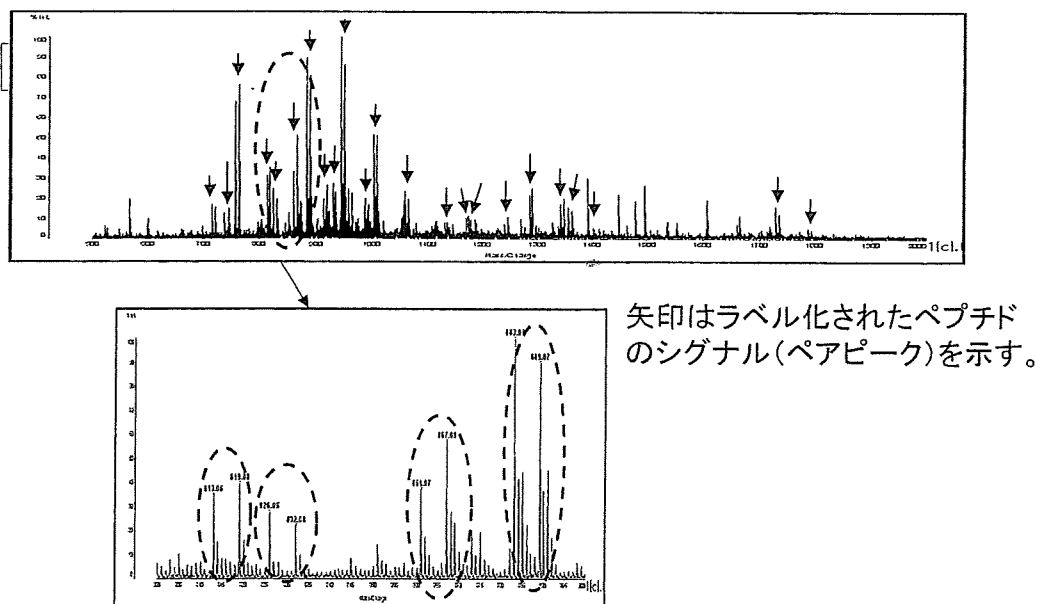


図9 ラット血清（正常及び高血糖）を用いた解析例

ラット) 及び Gk/Crj (高血糖ラット) である。各々 2 μ L の血清(総蛋白量 100 μ g 相当)を 4:1 で述べたプロセスに従って処理した(正常血清を軽い試薬で、高血糖血清を重い試薬でラベル

化した)。次いで得られたラベル化ペプチド混合物を MALDI-TOFMS で分析した。図9にはそのマスペクトルを示した。図中、矢印で示すようにラベル化されたペプチドのペアピー

クが多数観測された。蛋白質の同定等は現在進行中であるのでここでの記述は控えるが、HPLCと微量フラクションコレクター (AccuSpot: 島津製作所製) とを組み合わせたシステムによるラット及びマウス血清の分析例は報告済みである⁴⁴⁾。生体からのサンプルを用いたNBS法による研究 (バイオマーカー探索) に関しては、乳癌細胞を題材とした研究結果が既に報告²⁾されている。その他大腸癌^{45), 46)}, 肝臓癌^{47) - 49)}, 腎臓癌^{50), 51)}などを題材にして探索研究を進めており、マーカーの候補蛋白質が複数見つかった。

5. おわりに

この稿、前半部分では現在までに報告されている安定同位体標識試薬の中から、いくつかのものについてターゲットとなる残基あるいは官能基の観点から分類し解説した。後半では筆者らの開発したNBS法について実際のマススペクトル上のデータを示しながら解説し、現在進めている癌マーカー探索についても言及した。安定同位体試薬と質量分析装置を用いるプロテオーム解析は試薬の開発、装置側の進歩等により一大研究分野となっており、方法論も数多く報告されているが、マーカーとなる蛋白質を発見したという報告は現在のところみられない。このような現状では、種々提案されている方法論の中から自身の目的に合った、適切と考えられるものを選択して解析を行うのもひとつの方向であり、他とは違った新規な試薬・方法論を開発してターゲット蛋白質を探すのもまたひとつの方向である。いずれにせよ質量分析装置側の発展・進歩と試薬、解析ソフトウェアの進歩とが両輪となって進んでいくことがマーカー探索を前進させる原動力であるのは間違いなさであろう。

文 献

1) Gygi, S. P., Rist, B. and Aebersold, R., Measuring gene expression by quantitative proteome analy-

- sis, *Curr. Opin. Biotechnol.*, 11, 396-401 (2000)
- 2) Ou, K., Kesuma, D., Ganesan, K., Soon, S. Y., Lee, S. Y., Goh, X. P., Hooi, M., Chen, W., Jikuya, H., Ichikawa, T., Kuyama, H., Matsuo, E., Nishimura, O. and Tan, P., Quantitative profiling of drug-associated proteomic alterations by combined 2-nitrobenzenesulfonyl chloride (NBS) isotope labeling and 2DE/MS identification, *J. Proteome Res.*, 5, 2194-2206 (2006)
- 3) 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)
- 安定同位体試薬を用いたプロテオーム解析に関する総説を4から8に挙げる。
- 4) Leitner, A. and Lindner, W., Chemistry meets proteomics: The use of chemical tagging reactions for MS-based proteomics, *Proteomics*, 6, 5418-5434 (2006)
- 5) Brancia, F. L., Mass spectrometry based strategies in quantitative proteomics, *Curr. Anal. Chem.*, 2, 1-7 (2006)
- 6) Ong, S-E. and Mann, M., Mass spectrometry-based proteomics turns quantitative, *Nat. Chem. Biol.*, 5, 252-262 (2005)
- 7) Yan, W. and Chen, S. S., Mass spectrometry-based quantitative proteomic profiling, *Briefings in Functional Genomics and Proteomics*, 4, 1-12 (2005)
- 8) Zhang, H., Yan, W. and Aebersold, R., Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes, *Curr. Opin. Chem. Biol.*, 8, 66-75 (2004)
- 9) Gygi, S. P., Rist, B., Geber, S. A., Turecek, F., Gelb, M. H. and Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat. Biotechnol.*, 17, 994-998 (1999)
- 10) Hansen, K. C., Schmitt-Ulms, G., Chalkley, R. J., Hirsh, J. and Baldwin, M. A., Mass spectrometric analysis of protein mixtures at low levels using cleavable ¹³C-isotope-coded affinity tag and multidimensional chromatography, *Mol. Cell. Proteomics*, 2, 299-314 (2003)

- 11) Li, J., Steen, H., Gygi, S. P., Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents, *Mol. Cell. Proteomics*, **2**, 1198-1204 (2003)
- 12) Sechi, S. and Chait, B. T., Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification, *Anal. Chem.*, **70**, 5150-5158 (1998)
- 13) Faca, V., Coram, M., Phanstiel, D., Glukhova, V., Ahzng, Q., Fitzgibbon, M., McIntosh, M. and Hanash, S., Quantitative analysis of acrylamide labeled serum proteins by LC-MS/MS, *J. Proteome Res.*, **5**, 2009-2018 (2006)
- 14) Sebastiano, R., Citterio, A., Lapadula, M. and Rhigetti, P. G., A new deuterated alkylating agent for quantitative proteomics, *Rapid Commun. Mass Spectrom.*, **17**, 2380-2386 (2003)
- 15) Pasquarello, C., Sanchez, J-C., Hochstrasser, D. F. and Corthals, G. L., *N-t*-Butyliodoacetamide and iodoacetanilide : two new cycteine alkylating reagents for relative quantification of proteins, *Rapid Commun. Mass Spectrom.*, **18**, 117-127 (2004)
- 16) Brancia, F. L., Montgomery, H., Tanaka, K. and Kumashiro, S., Guanidino labeling derivatization strategy for global characterization of peptide mixtures by liquid chromatography matrix-assisted laser desorption/ionization mass spectrometry, *Anal. Chem.*, **76**, 2748-2755 (2004)
- 17) Warwood, S., Mohammed, S., Cristea, I. M., Evans, C., Whetton, A. D. and Gaskell, S. J., Guanidination chemistry for qualitative and quantitative proteomics, *Rapid Commun. Mass Spectrom.*, **20**, 3245-3256 (2006)
- 18) Brancia, F. L., Openshaw, M. E. and Kumashiro, S., Investigation of the electrospray response of lysine-, arginine-, and homoarginine-terminal peptide mixtures by liquid chromatography/mass spectrometry, *Rapid Commun. Mass Spectrom.*, **16**, 2255-2259 (2002)
- 19) Goshe, M. B., Conrads, T. P., Panisko, E. A., Angell, N. H., Veenstra, T. D. and Smith, R. D., Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analysis, *Anal. Chem.*, **73**, 2578-2586 (2001)
- 20) Qian, W-L., Goshe, M. B., Camp II, D. G., Yu, L-R., Tang, K. and Smith, R. D., Phosphoprotein isotope-coded solid-phase tag approach for enrichment and quantitative analysis of phosphopeptides from complex mixtures, *Anal. Chem.*, **75**, 5441-5450 (2003)
- 21) Wang, S. and Regnier, F. E., Proteomics based on selecting and quantifying cysteine containing peptides by covalent chromatography, *J. Chromatogr. A*, **924**, 345-357 (2001)
- 22) Munchbach, M., Quadroni, M., Miotto, G. and James, P., Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety, *Anal. Chem.*, **72**, 4047-4057 (2000)
- 23) Schmidt, A., Kellermann, J. and Lottspeich, F., A novel strategy for quantitative proteomics using isotope-coded protein labels, *Proteomics*, **5**, 4-15 (2005)
- 24) Zhang, X., Jin, Q. K., Carr, S. A. and Annan, R. S., N-terminal peptide labeling strategy for incorporation of isotopic tags : a method for the determination of site-specific absolute phosphorylation stoichiometry, *Rapid Commun. Mass Spectrom.*, **16**, 2325-2332 (2002)
- 25) Hsu, J-L., Huang, S-Y., Shiea, J-T., Huang, W-Y. and Chen, S-H., Stable-isotope dimethyl labeling for quantitative proteomics, *Anal. Chem.*, **75**, 6843-6852 (2003)
- 26) Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhász, P., Martin, S., Bartlett-Jones, M., He, F., Jacobson, A. and Pappin, D. J., Multiplexed protein quantitation in *saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents, *Mol. Cell. Proteomics*, **3**, 1154-1169 (2004)
- 27) Goodlett, D. R., Keller, A., Watts, J. D., Newitt, R., Yi, E. C., Purvine, S., Eng, J. K., von Haller, P., Aebersold, R. and Kolker, E., Differential stable isotope labeling of peptides for quantitation and *de novo* sequence derivatization, *Rapid Commun. Mass Spectrom.*, **15**, 1214-1221 (2001)
- 28) Šyka, J. E., Marto, J. A., Bai, D. L., Horning, S., Senko, M. W., Schwartz, J. C., Ueberheide, B., Garcia, B., Busby, S., Muratore, T., Shabanowitz,

- J. and Hunt, D. F., Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modification, *J. Proteome Res.*, **3**, 621-626 (2004)
- 29) Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J. and Hart, G. W., Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications, *Mol. Cell. Proteomics*, **1**, 791-804 (2002)
- 30) Amoresano, A., Marino, G., Cirulli, C. and Quemenet, E., Mapping phosphorylation sites: a new strategy based on the use of isotopically labeled DTT and mass spectrometry, *Eur. J. Mass Spectrom.*, **10**, 401-412 (2004)
- 31) Vosseller, K., Hansen, K. C., Chalkley, R. J., Trinidad, J. C., Wells, L., Hart, G. W. and Burlingame, A. L., Quantitative analysis of both protein expression and serine/threonine post-translational modifications through stable isotope labeling with dithiothreitol, *Proteomics*, **5**, 388-398 (2005)
- 32) Mirgorodskaya, O. A., Kozmin, Y. P., Titov, M. I., Koerner, R., Soenksen, C. P. and Poepstorff, P., Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using ^{18}O -labeled internal standards, *Rapid Commun. Mass Spectrom.*, **14**, 1226-1232 (2000)
- 33) Yao, X., Freas, A., Ramirez, J., Demirev, P. A. and Fenselau, C., Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus, *Anal. Chem.*, **73**, 2836-2842 (2001)
- 34) Scoffone, E., Fontana, A. and Rocchi, R., Selective modification of the tryptophan residue in peptides and proteins using sulfenyl halides, *Biochem. Biophys. Res. Commun.*, **25**, 170-174 (1966)
- 35) Scoffone, E., Fontana, A. and Rocchi, R., Sulfenyl halides as modifying reagents for polypeptides and proteins. I. Modification of tryptophan residues, *Biochemistry*, **7**, 971-979 (1968)
- 36) 渡辺 真, 西村 紀, 新規プロテオーム解析法の開発, バイオサイエンスとインダストリー, **62**, 49-52 (2004)
- 37) Tanaka, S., Mohri, N., Kihara, M. and Ohno, M., Selective separation of tryptophan-containing peptides via hydrophobic modification with 2-nitro-4-carboxy-phenylsulfenyl chloride, *J. Biochem. (Tokyo)*, **97**, 1377-1384 (1985)
- 38) Sasagawa, T., Titani, K. and Walsh, K. A., Selective isolation of tryptophan-containing peptides by hydrophobicity modulation, *Anal. Biochem.*, **134**, 224-229 (1983)
- 39) Zincke, Th. and Leinhardt, S., Über p-Nitrophenylschwefelchlorid und umwandlungsproducte, *Ann. Chem.*, **400**, 2-47 (1913)
- 40) Kharash, N., Gleason, G. I. and Buess, C. M., Derivatives of sulfenic acids. III. The catalytic chlorinolysis of 2, 4-dinitrophenyl disulfide, *J. Am. Chem. Soc.*, **72**, 1796-1798 (1950)
- 41) Havlic, A. J. and Kharash, N., Derivatives of sulfenic acids. XIX. Synthesis and characterization of 2-nitro-4-carboxybenzenesulfenyl chloride, *J. Am. Chem. Soc.*, **77**, 1150-1152 (1955)
- 42) Matsuo, E., Toda, C., Watanabe, M., Iida, T., Masuda, T., Minohata, T., Ando, E., Tsunawa, 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)
- 43) Matsuo, E., Toda, C., Watanabe, M., Ojima, N., Izumi, S., Tanaka, K., Tsunawa, 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)
- 44) Iida, T., Kuyama, H., Watanabe, M., Toda, C., Matsuo, E., Kido, A., Ando, E., Tsunawa, 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**, 333-341 (2006)
- 45) Watanabe, M., Takemasa, I., Matsubara, T., Nishimura, N., Matsuo, E., Masuda, T., Yoshioka, S., Miyake, M., Nagai, K., Monden, M., Tsunawa, S. and Nishimura, O., The proteomic profiling of human colorectal cancer tissue using NBS method, 54th ASMS Conference on Mass Spectrometry and Allied Topics, MP27-624 (2006)

- 46) Watanabe, M., Takemasa, I., Matsubara, T., Nishimura, N., Yoshioka, S., Miyake, M., Nagai, K., Monden, M., Tsunasawa, S. and Nishimura, O., The proteomic profiling of human colorectal cancer tissue using NBS method, 20th IUBMB International Congress of Biochemistry and Molecular Biology and 11th FAOBMB Congress, 4PA-382 (2006)
- 47) 竹政伊知朗, 渡辺 真, 吉岡慎一, 池田正孝, 山本浩文, 関本貢嗣, 西村 紀, 永井克也, 松原謙一, 門田守人, ポストゲノム時代における大腸癌の網羅的な遺伝子およびタンパク質の発現プロファイル解析, 第60回日本消化器外科学会学術総会 SY-2-6 (平成17年7月)
- 48) 竹政伊知朗, 永野浩昭, 吉岡慎一, 丸橋 繁, 宮本敦史, 武田 裕, 堂野恵三, 渡辺 真, 西村紀, 永井克也, 松原謙一, 門田守人, 肝癌の個別化治療の実践にむけて: ポストゲノム時代における網羅的な遺伝子及びタンパク質の発現プロファイル解析, 第64回日本癌学会学術総会 S16-2 (平成17年9月)
- 49) 竹政伊知朗, 渡辺 真, 吉岡慎一, 池田正孝, 山本浩文, 関本貢嗣, 西村 紀, 永井克也, 松原謙一, 門田守人, 大腸癌の発育・進展過程に関与する分子異常: 網羅的な遺伝子およびタンパク質の発現プロファイル解析, 第64回日本癌学会学術総会 W-011 (平成17年9月)
- 50) Masuda, Y., Okamura, N., Suganuma, K., Tanaka, H., Matsuo, E., Watanabe, M., Tsunasawa, S., Gotoh, A., Shirakawa, T., Terao, S., Okumura, K. and Nishimura, O., Construction of the proteomic profiles on human kidney cancer tissue using the NBS method, 54th ASMS Conference on Mass Spectrometry and Allied Topics, MP28-557 (2006)
- 51) Masuda, Y., Okamura, N., Suganuma, K., Tanaka, H., Gotoh, A., Shirakawa, T., Terao, S., Okumura, K. and Nishimura, O., Quantitative proteome analysis of renal cell carcinoma based on NBS method, 20th IUBMB International Congress of Biochemistry and Molecular Biology and 11th FAOBMB Congress, 4PA-365 (2006)

Gene expression of colorectal cancer: Preoperative genetic diagnosis using endoscopic biopsies

TAKAMICHI KOMORI¹, ICHIRO TAKEMASA¹, MAKOTO YAMASAKI¹, MASAOKI MOTOORI¹,
TAKESHI KATO², NOBUTERU KIKKAWA², NAOMASA KAWAGUCHI³,
MASATAKA IKEDA¹, HIROFUMI YAMAMOTO¹, MITSUGU SEKIMOTO¹,
KENICHI MATSUBARA⁴, NARIAKI MATSUURA³ and MORITO MONDEN¹

¹Department of Surgery, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871;

²Department of Surgery, Minoh City Hospital, 5-7-1 Sugano, Minoh, Osaka 562-0014; ³Department of Molecular Pathology, Graduate School of Medicine and Health Science, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871;

⁴DNA Chip Research Inc., 1-1-43 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan

Received July 5, 2007; Accepted August 30, 2007

Abstract. In colorectal cancer, to predict the response to chemo- and/or radio-therapy or the existence of lymph node metastasis preoperatively, a more competent diagnostic system is required, in addition to conventional diagnosis based on morphology and pathology. The application of gene expression profiling to preoperative cancer diagnosis using endoscopic biopsies could enable the selection of a more appropriate therapy for patients. In this study, we evaluated the feasibility of gene expression profiling using preoperative biopsies of colorectal tumors in a clinical setting, by investigating the influence of intra-tumor heterogeneity on the profiles and testing the prediction ability of tumor malignancy. Under endoscopic examination, two biopsies were sampled from each of 10 colorectal cancers and 10 adenomas, and their gene expression data were obtained using cDNA microarrays. The intra- and inter-tumor heterogeneities of the profiles were compared with unsupervised clustering analysis. Molecular prediction of tumor malignancy using biopsies was performed with the supervised classification algorithm. In clustering analysis, almost all paired biopsies from the same tumors joined each other. Pearson's correlation coefficients of the profiles between biopsies from the same tumors (mean, 0.83)

were significantly greater than those of the profiles between biopsies from other cancers (mean, 0.58) ($p < 0.0001$). In the supervised classification method, malignancy was correctly predicted in 39 out of 40 biopsies with 8-71 informative genes. Gene expression profiling using endoscopic biopsies of colorectal tumors revealed that the intra-tumor heterogeneity was smaller than the inter-tumor heterogeneity and tumor malignancy was correctly predicted. Our findings suggest that the technique of gene expression profiling accurately represents the biological properties of colorectal cancer and could help the preoperative diagnosis of this disease.

Introduction

The incidence of colorectal cancer (CRC) is increasing and it is one of the leading causes of cancer death in Japan (1). Conventional diagnosis based on morphology and pathology, such as Dukes' classification and the tumor-node-metastasis (TNM) staging system, has played an important role in the clinical decision-making and evaluation of prognosis for CRC (2-4). However, it is difficult to differentiate the response to chemo- and/or radio-therapy or the existence of lymph node metastasis preoperatively by conventional diagnosis. To predict such individual heterogeneous cancers' characteristics preoperatively, a more competent diagnostic system is required.

Comprehensive gene expression assay using microarray technology has provided insights into cancer pathogenesis and is expected to help to fulfil the clinical demands for individualized medicine (5). The discovery of a set of new molecular markers, which can classify cancers according to their properties using surgical specimens, has been reported in various cancers (6). In CRC, this technology has been used to elucidate the mechanisms involved in carcinogenesis (7-16) as well as to predict various clinicopathological aspects, such as recurrence after surgery (17-22). However, in clinical practice, this fruitful molecular prediction using surgical specimens will be limited to the selection of postoperative medicine such as adjuvant therapy and follow-up schedules. On the other hand, analysis using preoperative endoscopic biopsies, instead of

Correspondence to: Dr Ichiro Takemasa, Department of Surgery, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
E-mail: takemasa@surg2.med.osaka-u.ac.jp

Abbreviations: CRC, colorectal cancer; HCA, hierarchical clustering analysis; SNR, signal-to-noise ratio; IHC, immunohistochemical staining; pAb, polyclonal antibody; mAb, monoclonal antibody; FNAB, fine needle aspiration biopsy

Key words: colorectal cancer, gene expression profile, microarray, biopsy, tumor heterogeneity

Table I. Patients and tumor characteristics.

Case	Age (years)	Sex	Tumor location	Tumor size (cm)	Tumor type	TNM staging		
						T	N	Stage
Ca1	67	M	D	2.0	Well differentiated adenocarcinoma	T1	N0	I
Ca2	55	F	R	1.5	Well differentiated adenocarcinoma	T1	N0	I
Ca3	69	F	R	5.6	Well differentiated adenocarcinoma	T2	N0	I
Ca4	59	F	R	3.5	Moderately differentiated adenocarcinoma	T2	N0	I
Ca5	66	M	R	5.5	Moderately differentiated adenocarcinoma	T2	N0	I
Ca6	48	M	S	2.8	Moderately differentiated adenocarcinoma	T2	N0	I
Ca7	73	F	S	5.7	Well differentiated adenocarcinoma	T3	N0	II
Ca8	55	M	R	3.8	Moderately differentiated adenocarcinoma	T3	N0	II
Ca9	67	F	R	3.4	Moderately differentiated adenocarcinoma	T3	N0	II
Ca10	60	F	R	3.0	Well differentiated adenocarcinoma	T2	N1	III
Ad1	60	M	S	4.0	Tubulovillous adenoma			
Ad2	74	M	S	2.5	Tubular adenoma			
Ad3	77	M	A	2.5	Tubulovillous adenoma			
Ad4	68	M	T	1.1	Tubular adenoma			
Ad5	62	F	A	1.0	Tubulovillous adenoma			
Ad6	60	M	D	2.0	Tubulovillous adenoma			
Ad7	72	M	R	0.8	Tubulovillous adenoma			
Ad8	61	M	R	0.8	Tubular adenoma			
Ad9	64	F	S	1.2	Tubular adenoma			
Ad10	78	F	S	1.8	Tubular adenoma			

M, male; F, female; A, ascending colon; T, transverse colon; D, descending colon; S, sigmoid colon; R, rectum.

surgically resected samples, would widen the utility of microarray technology.

In rectal cancer, preoperative rather than postoperative chemo- and/or radio-therapy reduces local recurrence after surgery (23). Using preoperative biopsies under colonoscopic examination, prediction of the response to preoperative chemo- and/or radio-therapy would be useful for the selection of patients who would most benefit from preoperative therapies aimed at a better prognosis or improved chances of sphincter preservation (24,25). The prediction of lymph node metastasis could contribute to the avoidance of unnecessary surgery for early invasive CRC. This is clinically important since lymph node metastasis is found in only approximately 10% of early invasive CRCs (26); the remaining 90% without metastasis are more suited to undergo local excision such as colonoscopic resection or transanal endoscopic microsurgery.

To determine the clinical importance of diagnosis based on gene expression profiles using preoperative endoscopic biopsies in CRC, we investigated the data quality from low volume samples and the influence of intra-tumor heterogeneity on the profiles. Wide differences in profiles within a tumor would hinder the adoption of this technique. There are no reports that compared intra- and inter-tumor heterogeneity in expression profiles in colorectal tumors. In the present study, we sampled two biopsy specimens from each of 10 cancers and 10 adenomas obtained under colonoscopic examinations, and determined their gene expression profiles using cDNA

microarrays. By comparing their profiles, we assessed the intra-tumor heterogeneity in colorectal tumors. Moreover, as a first step for clinical applicability, by testing the molecular prediction ability of tumor malignancy, we investigated whether this technique, employing preoperative endoscopic biopsies, allows characterization of the biological properties of colorectal cancer.

Materials and methods

Patients and sample collection. From each of the 10 cancers and 10 adenomas, two biopsy specimens were obtained under colonoscopic examination. The clinicopathological data of patients and their tumors are summarized in Table I. Six bulky samples of the same cancers and 8 normal colorectal epithelium tissues were obtained from surgically resected specimens. None of the adenomas contained a cancerous component. Samples were stored in RNAlater (Qiagen, Valencia, CA) at -20°C after sampling until RNA extraction. This study was approved by the Institutional Review Board of Osaka University and Minoh City Hospital, and informed consent was obtained from all patients.

RNA extraction. Total RNA was extracted from biopsy samples using an RNeasy kit (Qiagen). The average volume of extracted total RNA from one biopsy was 21.1 µg in cancers and 22.6 µg in adenomas. From bulky samples, total RNA was extracted

using Trizol reagent (Invitrogen, Carlsbad, CA) and cleaned using the RNeasy kit. The quality of extracted total RNA was checked by 0.8% agarose-gel electrophoresis and a LabChip kit (Agilent Technologies, Palo Alto, CA), and the quantity was determined with a spectrophotometer. The total RNA solution was stored at -80°C until use.

Hybridization to cDNA microarray. We used cDNA microarrays containing 4608 clones which were derived from 30,000 clones expressed in CRC tissues (10,16,22). The Pearson's correlation coefficient of the data from replicated samples using our microarrays was 0.95. Gene expression data were obtained using a previously described method (10), with some improvements. As a standard normal control reference, a mixture of total RNA extracted from 40 normal colorectal epithelia was used. Labeled cDNA targets for hybridization were synthesized by reverse transcription from standard- and sample-total RNA respectively, with the indirect labeling method.

For each reverse transcription, 12 μg of total RNA was mixed with 1 μg of oligo-dT primer (Invitrogen) in a total volume of 15.5 μl , heated to 70°C for 10 min and cooled for 5 min. The mixture consisted of 6 μl of 5X first strand buffer, 3 μl of 0.1 M DTT, 3 μl of nucleotide cocktail (5 mM each dATP, dCTP, and dGTP, 3 mM dTTP and 2 mM aminoallyl-dUTP) and 0.5 μl of 40 U/ μl RNase inhibitor was added. After incubation for 2 min at 42°C , 2 μl of 200 U/ μl Superscript II reverse transcriptase (Invitrogen) was added. After incubation for 1 h at 42°C , 3.3 μl of 0.5 M EDTA was added and the RNA strand was degraded with 3.3 μl of 2 N NaOH and incubation at 70°C for 20 min. The mixture was neutralized with 3.3 μl of 2 N HCL and 60 μl of distilled water. This cDNA was purified using a QIAquick kit (Qiagen) and dried cDNA was dissolved in 9 μl of 0.2 M $\text{NaHCO}_3\text{-Na}_2\text{CO}_3$ (pH 9.0). Cy-dye solution (1 μl) (Amersham, Piscataway, NJ), Cy3 for the standard target, and Cy5 for the tested target, were added and incubated under shade at room temperature for 1 h, respectively. After purification using Micro Bio-Spin Columns P-30 Tris (Bio-Rad, Hercules, CA), the separately synthesized Cy3- and Cy5-labeled targets were combined and purified using Microcon YM-30 (Millipore, Billerica, MA), concentrated to a volume of 16.5 μl . To the concentrated target, 3.5 μl of human COT-1 DNA (Invitrogen), 7 μl of 20X SSC, 7 μl of 20X Denhart's solution, and 1 μl of 10% SDS were added. The 35 μl of target mixture was denatured by heating for 2 min at 95°C and cooled on ice. After the incubation at 50°C , the target was placed on the array. The array was incubated at 50°C for 14 h in a humid chamber. After hybridization, the slides were washed in 2X SSC with 0.1% SDS for 10 min, 0.1X SSC with 0.1% SDS for 10 min, and 0.1X SSC for 5 min at 30°C .

Scan and data processing. The array was scanned with ScanArray Lite (Perkin-Elmer, Wellesley, MA). The images were analyzed with QuantArray software (GSI Lumonics, Billerica, MA), converting signal intensities of each spot into numerical data. Data was processed after background subtraction, as described previously (10,16,22). Cy5/Cy3 ratios were log-transformed, and the global normalization was performed. Genes with >15% missing values in each

group of cancers and adenomas were excluded from further analysis.

Statistical analysis. First, gene expression profiles of 20 cancer biopsies and 8 normal epithelia were compared. Hierarchical clustering analysis (HCA) was performed with the software GeneMaths version 2.0 (Applied Maths, Inc., Austin, TX) using all 1966 genes after data processing. Pearson's correlation was used as the similarity coefficient and the unweighted pair group method using arithmetic average as the clustering algorithm. In HCA, similarity in the expression pattern of paired biopsies from the same cancer was compared with those from other cancers, to define the difference between intra- and inter-tumor heterogeneity of profiles in cancers. Statistical significance of the difference was determined using t-test. To verify the validity of the list of differentially expressed genes involved in colorectal carcinogenesis using biopsy samples, examined genes in cancer biopsies and normal epithelia were ranked according to the signal-to-noise ratio (SNR) (5) and they were compared with previous reports using tissue samples. The p-value of SNR was calculated by performing the random permutation test 1,000 times (5).

Next, using 20 cancer biopsies and 20 adenoma biopsies, the prediction of malignancy in colorectal tumors by their gene expression profiles was tested. Genes differentially expressed in cancer and adenoma biopsies were ranked according to SNR, and differential diagnosis was performed using the weighted-votes method (5). The prediction accuracy was determined with the leave-one-out cross validation (5). Positive prediction strength was judged to be a cancer and negative to be an adenoma. Using the gene set with the highest accuracy, HCA was performed to identify the difference between the intra- and the inter-tumor heterogeneity of the profiles in cancers and adenomas.

Immunohistochemical staining. Immunohistochemical staining (IHC) was performed to investigate the translation of mRNA of differentially expressed genes to each coding protein and to examine the intra-tumor heterogeneity of expression patterns in cancers at the protein level. Among differentially expressed genes between cancer biopsies and normal epithelia in our transcriptional analyses, 7 genes whose antibodies were commercially available were selected. Buffered formalin-fixed (10%), paraffin-embedded sections were prepared from 10 surgically resected cancers, whose biopsy specimens were included in the microarray analyses. The streptavidin-biotin immunoperoxidase complex method (27) was used for IHC. Primary antibodies used were as follows; polyclonal antibody (pAb) to human peroxiredoxin 1 (PRDX1, Alexis Biochemicals, Lausen, Switzerland), pAb to high-mobility group box 1 (HMGB1, Santa Cruz Biotechnology, Santa Cruz, CA), pAb to DEK oncogene (DEK, Santa Cruz Biotechnology), pAb to poly(A) binding protein, cytoplasmic 1 (PABPC1, Santa Cruz Biotechnology), monoclonal antibody (mAb) to heat shock 60-kDa protein 1 (HSPD1, Sigma Aldrich, St. Louis, MO), mAb to nucleolin (NCL, Santa Cruz Biotechnology) and mAb to carbonic anhydrase II (CA2, Rockland, Gilbertsville, PA). Sections for negative control were tested by using normal mouse serum instead of primary

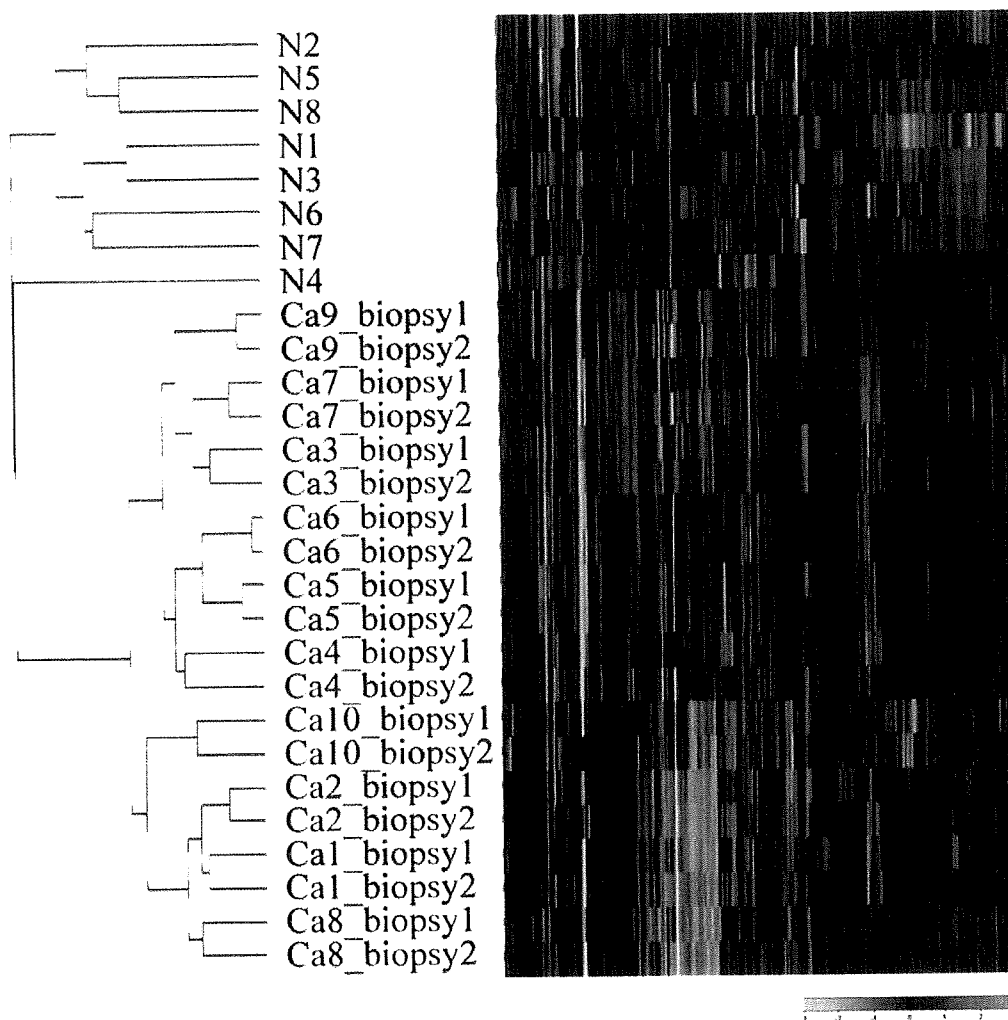


Figure 1. Hierarchical clustering analysis using full genes. Samples consisted of cancer biopsies (n=20; 10 pairs) and normal epithelia (n=8). Pearson's correlation was used as the similarity coefficient and the unweighted pair group method using the arithmetic average as the clustering algorithm. Red indicates overexpression, and green indicates underexpression. The respective paired biopsies from different areas of the same tumors are joined together.

antibody. Tissue sections of thyroid adenoma (for PRDX1), poorly differentiated gastric adenocarcinoma (for HMG1 and DEK), uterine cervical carcinoma (for HSPD1 and NCL), normal testis (for PABPC1) and normal colon (for CA2) were prepared as positive controls according to the recommendations of the manufacturer and previous publications on IHC. All slides were evaluated by a pathologist who was blinded to the microarray data. For each immunohistochemical analysis, the mean intensity in epithelial or tumor cells was evaluated in comparison with the positive controls as follows: weak, 1⁺; moderate, 2⁺; and strong, 3⁺.

Results

Intra-tumor heterogeneity in colorectal cancers. Comparison of gene expression profiles between cancers and normal epithelia by HCA showed clear separation of the 20 cancer biopsies from the 8 normal tissues using all genes (Fig. 1). Moreover, the respective paired biopsies from different areas of the same cancers were joined together. Pearson's correlation coefficients of paired biopsies from the same cancers (mean,

0.83; SD, 0.08) were significantly greater than those of unpaired biopsies from other cancers (mean, 0.58; SD, 0.12; $p < 0.0001$). Pearson's correlation coefficients between biopsies and their parent surgical bulky samples (mean, 0.60; SD, 0.15) were significantly greater than those between biopsies and surgical bulky samples from other cancers (mean, 0.37; SD, 0.16; $p < 0.0001$).

Differentially expressed genes in cancer biopsies. Differentially expressed genes in cancer biopsies and normal epithelia were ranked according to SNR. Among them, 692 up-regulated and 219 down-regulated genes in cancer biopsies had SNR ($p < 0.001$). The top 20 up-regulated and 20 down-regulated genes, excluding 4 down-regulated genes with no definition, are listed in Table II. PABPC1 and HSPD1 in the up-regulated genes and CA2, carboxylesterase 2 (CES2) and one EST in the down-regulated genes in cancer biopsies were also included in differentially expressed genes involved in colorectal carcinogenesis in our previous reports (10,16). Nine of the 40 differentially expressed genes were reported previously with respect to colorectal carcinogenesis with DNA microarray