

Table II. Functional gene classes demonstrating significant differential expression in each group.

Function name	Total	P (X2)	P (X1)	P (X)
<b>A group</b>				
Regulation of transcription	189	0.998	0.008	0.581
Regulation of transcription, DNA-dependent	184	0.997	0.011	0.582
Response to external stimuli	135	0.008	0.932	0.14
Mitotic cell cycle	48	0.628	1E-04	0.003
Ion transport	46	0.057	0.214	0.026
Regulation of cell cycle	46	0.901	0.001	0.055
Lipid metabolism	36	0.013	0.844	0.082
Protein folding	22	0.926	0.038	0.261
Alcohol metabolism	20	1	0.026	0.365
DNA replication and chromosome cycle	20	0.668	0.001	0.009
Response to endogenous stimulus	20	1	0.026	0.365
DNA replication	18	0.609	0.003	0.015
Nuclear organization and biogenesis	18	1	0.016	0.278
DNA repair	17	1	0.013	0.236
Main pathways of carbohydrate metabolism	14	0.808	0.029	0.126
Nucleocytoplasmic transport	14	1	0.029	0.297
DNA packaging	13	1	0.004	0.096
Carbohydrate catabolism	11	1	0.012	0.157
<b>B group</b>				
Transport	170	0.041	0.617	0.07
Response to external stimuli	135	0.019	0.219	0.01
Electron transport	56	0.026	0.65	0.045
Apoptosis	44	0.109	0.187	0.043
Energy pathways	37	0.003	0.497	0.005
Cell motility	36	0.455	0.024	0.06
Induction of programmed cell death	18	0.037	0.04	0.003
<b>C group</b>				
Protein metabolism	340	0.002	0.895	0.009
Biosynthesis	190	0.041	0.99	0.181
Response to stress	96	0.045	0.076	0.01
Defense response	85	0.642	0.047	0.296
Intracellular signaling cascade	82	0.022	0.04	0.003
Cell-cell signaling	36	0.081	0.319	0.05
Vehicle-mediated transport	36	0.424	0.024	0.101
Ribosome biogenesis and assembly	33	0.046	1	0.114
Protein kinase cascade	24	0.335	0.038	0.077

<sup>a</sup>Total number of genes represented in each group corresponding to a specific gene ontology term, <sup>b</sup>p-value in the X group (X=A, B or C).

We assessed the utility of these 119 genes to identify patients who would develop liver recurrence after surgery. The expression of these 119 genes was used to build a weighted vote-based classifier that could predict liver metastasis at the time of surgery. We applied this classifier to

the 28 independent samples. Eighteen and 10 patients were assigned to the localized signature and metastasized signature, respectively. Seventeen of the 18 (94.4%) patients assigned to the localized signature did not relapse at least 47 months after surgery, and 9 of the 10 (90%) patients assigned to the

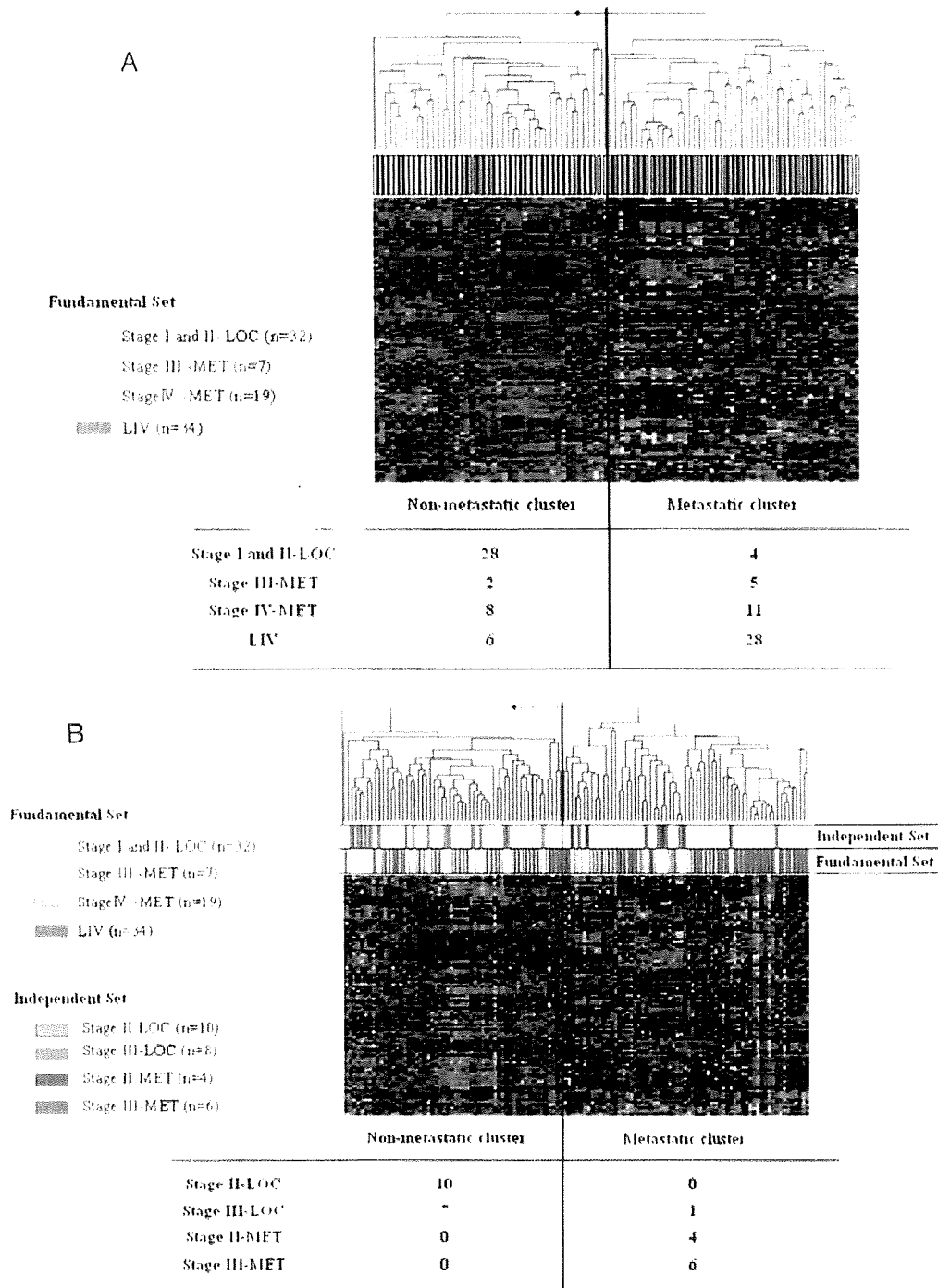


Figure 2. Hierarchical cluster analysis of metastasized and non-metastasized tumor signatures. The genes in group B are hierarchically clustered with respect to the 92 tumors in the fundamental set (A) and the 120 tumors adding these 92 to the 18 tumors in the independent set (B). Each row represents a gene and each column represents a sample. For the samples, the color bars below the dendrogram represent sample annotation in clinicopathology. The aqua bars represent localized primary tumors (LOC), the orange and pink bars represent stage III- and stage IV- ones (stage III-MET and stage IV-MET) and the red bars represent liver metastatic lesions (LIV) in the fundamental set. The blue and green bars represent LOC with stage II and III (stage II-LOC and stage III-LOC), and the brown and red bars represent MET with stage II and III (stage II-MET and stage III-MET), respectively, in the independent set. (A) Two major branches of the hierarchical tree are observed. The left and right branches are mainly comprised of non-metastasized (non-metastatic cluster) and metastasized tumors (metastatic cluster), respectively. These two clusters significantly correlated with the metastatic status of the tumors. (B) The clustering pattern was so robust against sample addition that the partitioning of the 92 fundamental tumors remained the same and showed little change. Out of the 28 independent tumors, all of the 10 metastasized tumors are in the metastatic cluster, and 17 of the 18 non-metastasized tumors are in the non-metastatic cluster.

metastasized signature recurred to liver metastasis after surgery (Table III). The successful prediction of liver metastasis could be shown in the independent set of tumors.

To assess the association between class assigned by gene expression profile and survival, we plotted Kaplan-Meier survival curves of 28 patients in the independent set (Fig. 3A

Table III. The predictive diagnosis of liver metastasis in the independent samples.

		Clinical diagnosis	
		Non-relapse	Relapse
Profiling diagnosis	LOC	17	1
	MET	1	9

and B). A significant difference was observed in the disease-free survival ( $p < 0.0001$ ) and overall survival ( $p = 0.0065$ ) of patients between the localized and metastasized signature.

## Discussion

The gene expression profile using DNA microarray is a promising approach to investigate the molecular complexity of cancer. Several studies have reported the usefulness of this technique such as the gene identification related to metastasis and possible prediction of metastasis in various cancers (16,17). However, these studies simply compared the gene expression profile of tumors between partial stages and did not directly address the metastatic process. To elucidate the phenomenon of metastasis more accurately, it is critical to analyze comprehensively the process of gradual change in cancer progression. Thus, we analyzed the gene expression profiles of 104 samples at four time points in cancer development; normal colon, localized and metastasized primary

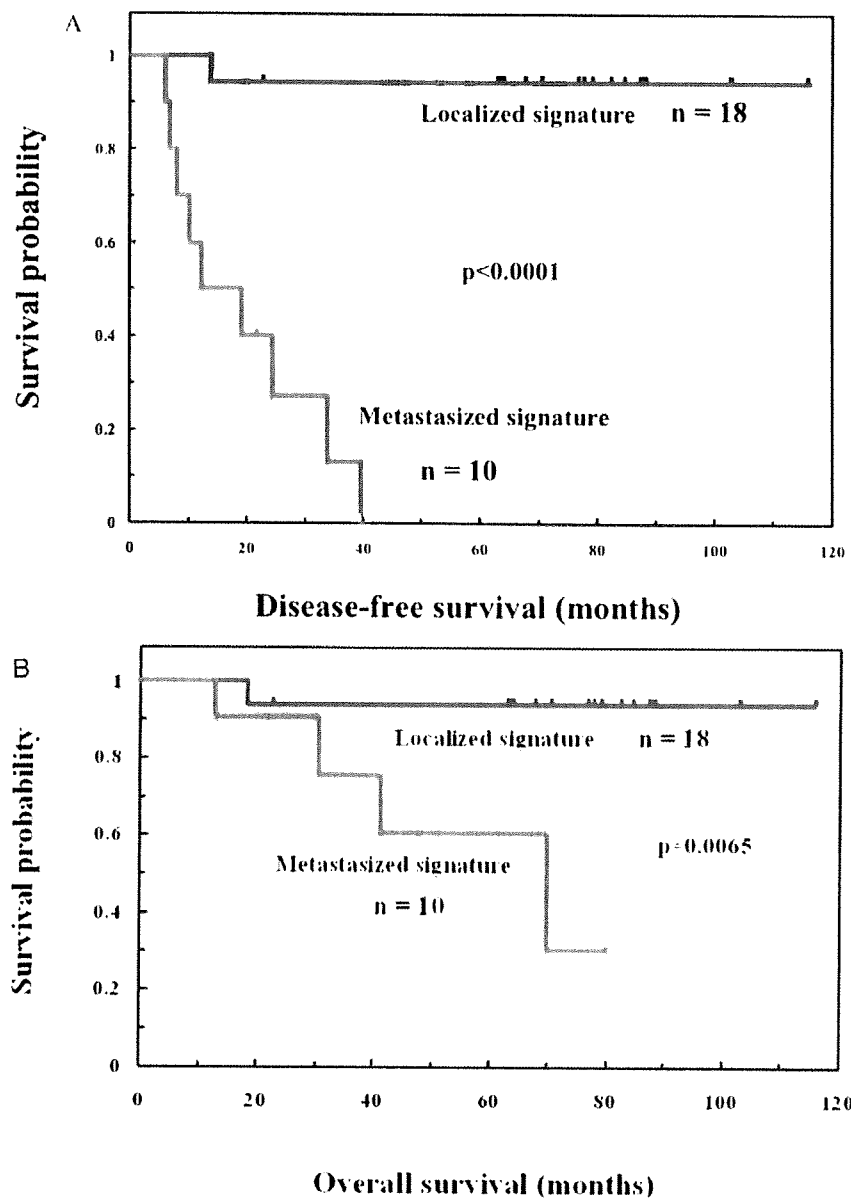


Figure 3. Kaplan-Meier plot of disease-free (A) and overall (B) survival of patients with the localized and metastasized signature. (A) Disease-free survival curves of patients with the localized signature were also higher than those of patients with the metastasized signature (log-rank test,  $p < 0.0001$ ). (B) Overall survival curves of patients with the localized signature were higher than those of patients with the metastasized signature (log-rank test,  $p = 0.0065$ ).

tumor and liver metastasis. These analyses indicate that each step of cancer progression can be characterized by a gene expression profile.

The expression pattern of group A is considered to be associated with the promotion of carcinogenesis in CRC. This group contained a significantly high proportion of genes with functions relevant to carcinogenesis such as cell proliferation and response to external and endogenous stimuli, and any imbalance in these functions initiates colorectal carcinogenesis. If these genes were expressed on the cell surface or secreted, this could be detected by easier methods such as immunohistochemical analysis and be applicable clinically as candidate diagnostic markers for the early detection of CRC (cellular-component, data not shown).

The expression of the genes represented in group B is significantly different in the primary tumor with or without metastasis, which directly reflects the event of metastasis itself. The loss or acquisition of this function in this group was closely relevant to the liver metastatic potential in colorectal cancer. This group contained a significantly high proportion of genes with functions relevant to the process of metastasis such as cell motility and apoptosis. The acquisition of cell motility is essential for the invasion of interstitial tissue and lymphatic or vascular vessel walls and the survival of cancer cells against immune and non-immune defenses while escape from apoptosis allows cancer cells to arrest a secondary organ apart from the primary lesion. Genes such as *DCC* and *GDP dissociation inhibitor 2 (GDI2)*, which are referred as relevant to metastasis, are included in group B. For example, *DCC* is a putative tumor suppressor gene, which has considerable homology to neural cell adhesion molecules and may be involved in the regulation of cell-to-cell or cell-to-substrate interactions with a potential functional role in the control of cell growth and the development of metastasis (22-24). Various studies reported that *DCC* is a predictor of distant metastasis and prognosis in colorectal cancer (24-26). Furthermore, *DCC* expression has been shown recently to predict clinical outcome to fluorouracil-based chemotherapy (27). Other studies have reported that the *DCC* gene product induced apoptosis through a caspase-dependent mechanism (28,29). In another study, *GDI2* was identified as a metastasis-suppressor gene (30). *GDI2* is an inhibitor of guanine nucleotide dissociation from RHO and RAC proteins, strongly inhibiting *GDP/GTP* exchange reaction. *GDI2*, RHO and RAC also modulate signal transduction such as the ERK, JNK and p38 MAPK pathways, three arms of which are involved in invasion, growth-factor-receptor signaling, the mitogen-activated protein kinase pathway, cell-cell communication and transcription, which are the main factors involved in the induction of liver metastasis (31,32).

The expression pattern of group C is closely associated with the colonization and maintenance of metastasis, which is the outgrowth of cancer cells after arrival at a secondary organ. Functional analysis indicated that metastatic colonization occurs as a result of the cancer cells' modification following interaction with the surrounding environment in the liver. Other investigators also reported similar results (16,33,34). Nishizuka *et al* (34) reported that cytokines produced by glial cells *in vivo* may contribute in a paracrine

manner to the development of brain metastases from breast cancer cells. Metastatic colonization differs from growth in the primary lesion by its requirement for angiogenesis, and changes in the local microenvironment in metastatic lesions and stressful conditions. The requirement for various growth factors and chemokines, the cell-cell or cell-extracellular matrix interaction and the activity of intracellular signaling pathways regulated by various stimuli and interactions could have considerable influence upon not only the initiation but also the maintenance of metastatic colonization (35,36). For example, *MAP3K4* and *ERK1*, which are differentially expressed in the metastatic tumor, stimulate the p38 and JNK MAPK pathway, and the MAPK-ERK pathway, respectively. Reduced JNK and p38 signals facilitate metastatic colonization by preventing apoptosis induced by the stress of a foreign environment (37). Numerous reports have linked the activation of *ERK1* with metastasis, motility and invasion, and angiogenesis (38,39). This result suggests that the cross-talk between signaling pathways, which were represented in the MAPK pathway, such as the ERK, p38 and JNK pathways, are linked closely within the cancer cells and interstitial tissue to adhesion, motility, proliferation, angiogenesis and apoptosis in metastatic lesions, which represents the final step in cancer development. These genes might be candidate markers of metastasis and could be suitable targets for the treatment of metastatic disease.

Chronological analysis allowed us to clarify each step of cancer progression, carcinogenesis to metastasis, at the molecular level. In particular, we were successful in separating metastasis into the acquisition of metastatic potential (group B) and the initiation and maintenance of metastatic colonization (group C).

Does the gene expression pattern correlate with the metastatic potential (such as that represented in group B) already present in the bulk of the primary tumor and detectable at the time of initial diagnosis? The conventional hypothesis about metastasis is that acquisition of the metastatic phenotype may occur late in tumor progression. Primary CRC cells arise from the normal colorectal epithelium through oncogenic mutations. The occasional cancer cells possessing metastatic ability in the primary tumor population can accumulate additional genetic changes that mediate metastasis to the liver. Thus, it has been hypothesized that primary tumors have gene expression profiles in which only a fraction of the metastatic potential exists. However, some groups have challenged this hypothesis (16,17). Ramaswamy *et al* (16) reported that the expression profile of primary tumors that had metastasized resembled that of metastatic tumors, and that a predictive diagnosis for metastasis was possible based on the analysis of the primary tumor profile. Van't Veer *et al* (17) also reported that patients with early-stage breast cancer could be classified according to their prognosis based on the gene expression signature obtained from the primary tumors, and specific gene expression profiles were predictive for distant metastases. Thus, the new hypothesis indicates that the acquisition of metastatic ability does not begin in just a few cells in the primary tumor but instead most cells in such tumors are inherently capable of metastasis.

Our hierarchical clustering with the expression profile of the genes closely relevant to the metastatic potential shows

that the profile of the metastasized primary tumor resembled one of a metastatic lesion apart from a primary lesion rather than one of a non-metastasized primary tumor. This means that the tumors are segregated according to their final metastatic status rather than their clinical stages. This suggests that the metastatic potential of CRC is not acquired in proportion to cancer progression but has already been encoded in the primary tumor. Thus, our findings support the above new hypothesis rather than the conventional one.

We also succeeded in the verification of this hypothesis using independent primary tumors. Clustering analysis with these genes on CRC samples diagnosed as localized or regional disease allowed the robust classification of metastasized tumors vs non-metastasized tumors. Supervised classification analysis enabled us to predict the postoperative relapse risk of the patients who were diagnosed as localized or regional disease at the time of surgery. The classification appropriately reflected the outcome of CRC patients. Our microarray data provide important information on the metastatic potential in CRC and indicate that the expression profiles in the bulk of primary tumors reflect the character of individual tumors and are possible and valid enough to predict metastasis.

The clinical outcome of patients with CRC is governed to some extent, if not fully, by the inherent character of tumor cells. The programmed expression pattern is detectable in the bulk of primary tumors. The characterization of cancer based on gene expression profile promises to shift the paradigm of cancer diagnosis from general taxonomy by clinicohistopathological parameters to individualized risk assessment.

Currently, because it is clinically difficult to accurately predict metachronous liver metastasis at the time of initial diagnosis, a lot of time and effort is directed toward the detection of asymptomatic recurrences using postoperative monitoring such as CT and the CEA test. Therefore, a more accurate prediction of liver metastasis and clinical outcome should benefit the optimization of cancer treatment, including adjuvant therapy and postoperative surveillance, for individual patients, and at the same time avoid over- or unnecessary treatment and reduce overall cost. Chronological analysis of the gene expression profile should enable a better understanding of the metastatic process. The identified 'interesting' genes and profiles could help design new strategies for the treatment and diagnosis of metastasis.

## References

1. Fearon ER, Hamilton SR and Vogelstein B: Clonal analysis of human colorectal tumors. *Science* 238: 193-197, 1987.
2. Hanahan D and Weinberg RA: The hallmarks of cancer. *Cell* 100: 57-70, 2000.
3. Fidler IJ: Critical factors in the biology of human cancer metastasis: twenty-eighth G.H.A. Clowes memorial award lecture. *Cancer Res* 50: 6130-6138, 1990.
4. Gutman M and Fidler IJ: Biology of human colon cancer metastasis. *World J Surg* 19: 226-234, 1995.
5. Woodhouse EC, Chuaiqui RF and Liotta LA: General mechanisms of metastasis. *Cancer* 80: 1529-1537, 1997.
6. Chambers AF, Naumov GN, Varghese HJ, Nadkarni KV, MacDonald IC and Groom AC: Critical steps in hematogenous metastasis: an overview. *Surg Oncol Clin N Am* 10: 243-255, 2001.
7. Fidler IJ: Orthotopic implantation of human colon carcinomas into nude mice provides a valuable model for the biology and therapy of metastasis. *Cancer Metastasis Rev* 10: 229-243, 1991.
8. Welch DR: Technical considerations for studying cancer metastasis *in vivo*. *Clin Exp Metastasis* 15: 272-306, 1997.
9. Mohr S, Leikauf GD, Keith G and Rihn BH: Microarrays as cancer keys: an array of possibilities. *J Clin Oncol* 20: 3165-3175, 2002.
10. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JJ, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO and Staudt LM: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403: 503-511, 2000.
11. Notterman DA, Alon U, Sierk AJ and Levine AJ: Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res* 61: 3124-3130, 2001.
12. Bertucci F, Salas S, Eysteris S, Nasser V, Finetti P, Ginestier C, Charafe-Jauffret E, Liorid B, Bachelart L, Montfort J, Victorero G, Viret F, Ollendorff V, Fert V, Giovaninni M, Delpero JR, Nguyen C, Viens P, Monges G, Birnbaum D and Houlgatte R: Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene* 23: 1377-1391, 2004.
13. Safi F and Beyer HG: The value of follow-up after curative surgery of colorectal carcinoma. *Cancer Detect Prev* 17: 417-424, 1993.
14. Compton CC: Colorectal carcinoma: diagnostic, prognostic, and molecular features. *Mod Pathol* 16: 376-388, 2003.
15. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, Ray TS, Koval MA, Last KW, Norton A, Lister TA, Mesirov J, Neuberg DS, Lander ES, Aster JC and Golub TR: Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 8: 68-74, 2002.
16. Ramaswamy S, Ross KN, Lander ES and Golub TR: A molecular signature of metastasis in primary solid tumors. *Nat Genet* 33: 49-54, 2003.
17. Van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R and Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415: 530-536, 2002.
18. Takemasa I, Higuchi H, Yamamoto H, Sekimoto M, Tomita N, Nakamori S, Matoba R, Monden M and Matsubara K: Construction of preferential cDNA microarray specialized for human colorectal carcinoma: molecular sketch of colorectal cancer. *Biochem Biophys Res Commun* 285: 1244-1249, 2001.
19. Komori T, Takemasa I, Higuchi H, Yamasaki M, Ikeda M, Yamamoto H, Ohue M, Nakamori S, Sekimoto M, Matsubara K and Monden M: Identification of differentially expressed genes involved in colorectal carcinogenesis using a cDNA microarray. *J Exp Clin Cancer Res* 23: 521-527, 2004.
20. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286: 531-537, 1999.
21. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC and Weinstein JN: GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol* 4: R28, 2003.
22. Fearon ER, Cho KR, Nigro JM, Kern SE, Simons JW, Ruppert JM, Hamilton SR, Preisinger AC, Thomas G and Kinzler KW: Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* 247: 49-56, 1990.
23. Cho KR and Fearon ER: DCC: linking tumour suppressor genes and altered cell surface interactions in cancer? *Eur J Cancer* 31A: 1055-1060, 1995.
24. Shibata D, Reale MA, Lavin P, Silverman M, Fearon ER, Steele G Jr, Jessup JM, Loda M and Summerhayes IC: The DCC protein and prognosis in colorectal cancer. *N Engl J Med* 335: 1727-1732, 1996.
25. Raymond MA, Dworak O, Remke S, Hohenberger W, Kirchner T and Kockerling F: DCC protein as a predictor of distant metastases after curative surgery for rectal cancer. *Dis Colon Rectum* 41: 755-760, 1998.

26. Saito M, Yamaguchi A, Goi T, Tsuchiyama T, Nakagawara G, Urano T, Shiku H and Furukawa K: Expression of DCC protein in colorectal tumors and its relationship to tumor progression and metastasis. *Oncology* 56: 134-141, 1999.
27. Aschele C, Debernardis D, Lonardi S, Bandelloni R, Casazza S, Monfardini S and Gallo L: Deleted in colon cancer protein expression in colorectal cancer metastases: a major predictor of survival in patients with unresectable metastatic disease receiving palliative fluorouracil-based chemotherapy. *J Clin Oncol* 22: 3758-3765, 2004.
28. Chen YQ, Hsieh JT, Yao F, Fang B, Pong RC, Cipriano SC and Krepulat F: Induction of apoptosis and G2/M cell cycle arrest by DCC. *Oncogene* 18: 2747-2754, 1999.
29. Mehlen P, Rabizadeh S, Snipas SJ, Assa-Munt N, Salvesen GS and Bredeben DE: The DCC gene product induces apoptosis by a mechanism requiring receptor proteolysis. *Nature* 395: 801-804, 1998.
30. Gildea JJ, Seraj MJ, Oxford G, Harding MA, Hampton GM, Moskaluk CA, Frierson HF, Conaway MR and Theodorescu D: RhoGDI2 is an invasion and metastasis suppressor gene in human cancer. *Cancer Res* 62: 6418-6423, 2002.
31. Jaffe AB and Hall A: Rho GTPases in transformation and metastasis. *Adv Cancer Res* 84: 57-80, 2002.
32. Arozarena I, Matalanas D and Crespo P: Maintenance of CDC42 GDP-bound state by Rho-GDI inhibits MAP kinase activation by the exchange factor Ras-GRF. Evidence for Ras-GRF function being inhibited by Cdc42-GDP but unaffected by CDC42-GTP. *J Biol Chem* 276: 21878-21884, 2001.
33. LaTulippe E, Satagopan J, Smith A, Scher H, Scardino P, Reuter V and Gerald WL: Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Res* 62: 4499-4506, 2002.
34. Nishizuka I, Ishikawa T, Hamaguchi Y, Kamiyama M, Ichikawa Y, Kadota K, Miki R, Tomaru Y, Mizuno Y, Tominaga N, Yano R, Goto H, Nitanda H, Togo S, Okazaki Y, Hayashizaki Y and Shimada H: Analysis of gene expression involved in brain metastasis from breast cancer using cDNA microarray. *Breast Cancer* 9: 26-32, 2002.
35. Liotta LA and Kohn EC: The microenvironment of the tumour-host interface. *Nature* 411: 375-379, 2001.
36. Chambers AF, Groom AC and MacDonald IC: Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer* 2: 563-572, 2002.
37. Takekawa M, Posas F and Saito H: A human homolog of the yeast Ssk2/Ssk22 MAP kinase kinase kinases, MTK1, mediates stress-induced activation of the p38 and JNK pathways. *EMBO J* 16: 4973-4982, 1997.
38. Webb CP, Van Aelst L, Wigler MH and Woude GF: Signaling pathways in Ras-mediated tumorigenicity and metastasis. *Proc Natl Acad Sci USA* 95: 8773-8778, 1998.
39. Simon C, Hicks MJ, Nemecek AJ, Mehta R, O'Malley BW Jr, Goepfert H, Flaitz CM and Boyd D: PD 098059, an inhibitor of ERK1 activation, attenuates the *in vivo* invasiveness of head and neck squamous cell carcinoma. *Br J Cancer* 80: 1412-1419, 1999.

## 遺伝子発現プロファイルによる 消化器癌転移推測の可能性を探る

竹政伊知朗\* 池田正孝\* 山本浩文\* 関本貢嗣\* 門田守人\*

### KEY WORDS

遺伝子発現プロファイル, 消化器癌, 転移予測, DNA マイクロアレイ

### SUMMARY

進行消化器癌治療における主たる目的は癌の再発・転移の制御にあり, 精度の高い診断技術とそれに対応した治療方針が重要である. 従来の確率論的な医療から脱却を図り, 癌治療成績を改善するためには, 分子生物学的な手法を積極的に臨床応用し, 癌腫, 宿主の生物学的な特性を反映した個別化診断・治療戦略の確立が急務である. ヒト全遺伝子の発現情報が1枚のDNA マイクロアレイで解析可能な時代となり, 遺伝子発現プロファイル解析は, 癌の発育・進展の分子生物学的な解明において大きな役割を果たすようになった. 今後, 癌転移・再発のメカニズムの解明を進め, 悪性度診断や治療方針の個別化に役立てることが期待されている.

### はじめに

癌の死亡者数は年約2%ずつ増加し, 2005年には32万5885人となり, 消化器癌はその54%を占めている. 消化器癌の致死率に最も影響を与える因子は転移であり, その制圧を目的とした新しい診断・治療体系の確立が急務である. 癌細胞は無秩序に多数の臓器に転移するのではなく, ある一定の臓器選択性をもって転移する特徴がある. その機構は, 原発巣から遊離した癌細胞が血液動態に伴い転移巣を形成するという血液動態説と, 癌細胞と転移標的臓器との密接な相互作用による“seed and soil”説によって説明されてきた. その後, 腫瘍には組織不均一性があり, 癌の発育・進展過程において比較的后期に, ごくわずかな細胞が遺伝子レベルで変異を起こすことで転移能を有し, それらの細胞が原発巣から離脱することで遠隔転移のステップが開始されるとする多段階発癌説が提唱され, 一般的な癌転移モデルとして解釈されている. しかし, 実際には同一病期内でも症例により再発の有無, 予後にはばらつきがある. このような症例による違いを従来の多段階発癌モデルだけで説明することは困

\* TAKEMASA Ichiro, IKEDA Masataka, YAMAMOTO Hirofumi, SEKIMOTO Mitsugu, MONDEN Morito/  
大阪大学大学院医学系研究科外科学講座消化器外科

難であるため、癌は臓器による違いだけでなく、同種の癌でもそれぞれのケースにおける宿主、癌腫両面の遺伝的表現型の違いにより種々の個性をもち、さまざまな臨床経過をたどると考えられるようになった<sup>1)</sup>。

2003年、ヒトゲノムの解読完了が宣言され、癌医療はこれまでの生命科学研究の集大成を土台に大きな革新を迎えている。なかでも網羅的な遺伝子発現解析は、癌の発育・進展の分子生物学的な解明において大きな役割を果たすようになった。最近の遺伝子発現プロファイルの研究で、癌の転移能力は発育段階の早期にすでに確立されており、転移に関与する特有の遺伝子発現プロファイルが原発性腫瘍に存在するとの報告があいつぎ、癌の転移メカニズムに関する新たな知見と、それらを応用した転移・再発予測診断の可能性がさかんに議論されるようになった<sup>2)</sup>。

本稿では、われわれの教室の取り組みを含め、消化器癌の転移予測診断における遺伝子発現プロファイルの研究の現状を概要する。

## 癌の遠隔転移について

正常の細胞が増殖異常を起こして悪性化した後、転移巣を形成するまでには、細胞間接着能の低下による癌細胞の主病巣からの離脱にはじまり、基底膜への接着と浸潤、組織内移動、血管内への侵入と移動、遠隔臓器の血管内皮細胞への接着、血管外への脱出、遠隔臓器での浸潤、血管新生を伴う増殖といった数多くの複雑なステップが存在する<sup>3)</sup>。また癌細胞は癌周囲の間質細胞と互いに影響を及ぼしあい直接的、間接的に転移を促進させるとともに、それぞれの段階で異なる遺伝子群が特有の変化を起こし、転移過程が推移すると考えられている。よって、局所的に単一分子を解析するよりも包括的に腫瘍の遺伝子変化を解析することことで、はじめて転移過程の分子生物学的な全体像をとらえることが可能となると考えられる<sup>4)</sup>。

## 癌の遺伝子発現プロファイル解析研究の現状とは？

遺伝子発現プロファイル解析とは、網羅的な遺伝子発現情報と生物学的な特徴の関係を比較検討し、有意義な

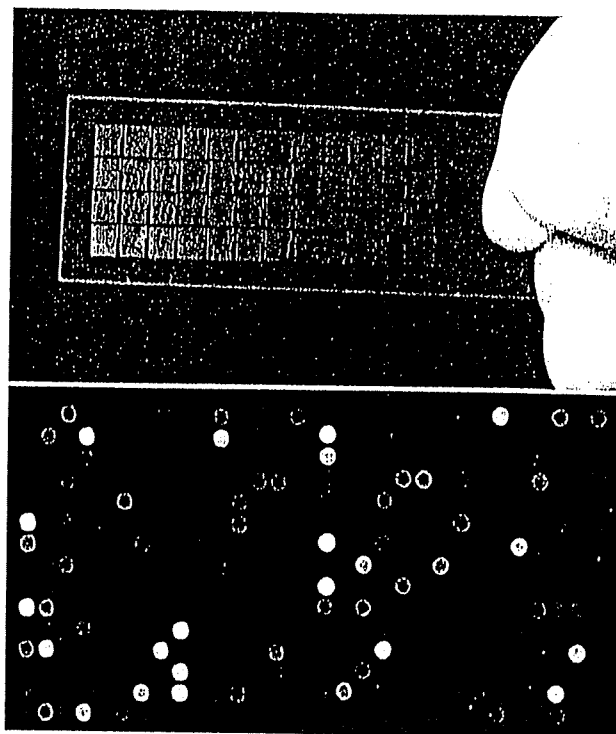


図1 ヒト全遺伝子型 DNA マイクロアレイ (AceGene: DNA チップ研究所) とハイブリダイゼーションイメージ  
50 mer のオリゴヌクレオチド 30000 種あまりが搭載された DNA マイクロアレイ。

情報を検索することである<sup>5)</sup>。現在では、ヒト全遺伝子の発現情報が1枚のDNAマイクロアレイで解析可能となり、遺伝子発現プロファイル研究に拍車がかかっている(図1)。癌研究でも遺伝子発現プロファイル解析がさかんに用いられ、すべての癌腫で報告がみられるようになった。当初は癌細胞と対照細胞を比較する単純なものが中心であったが、技術が進むにつれ、多数の癌の生物学的特徴と発現プロファイルによる分類の対応をつけようという総合的な試みがされるようになり、癌の個性化診断・治療への応用に関する報告も多数みられるようになった。なかでも2002年頃より癌の遠隔転移・再発に関係する重要な論文があいついで報告されるようになった。

van't Veer ら<sup>6)</sup>は117例の若年性乳癌を対象として遠隔転移を特有の遺伝子発現の刻印により予測できると報告した。引き続き van de Vijver ら<sup>7)</sup>は stage I または II の若年性乳癌295例を70種の遺伝子発現パターンによって、腋窩リンパ節転移など従来の臨床病理予後指標



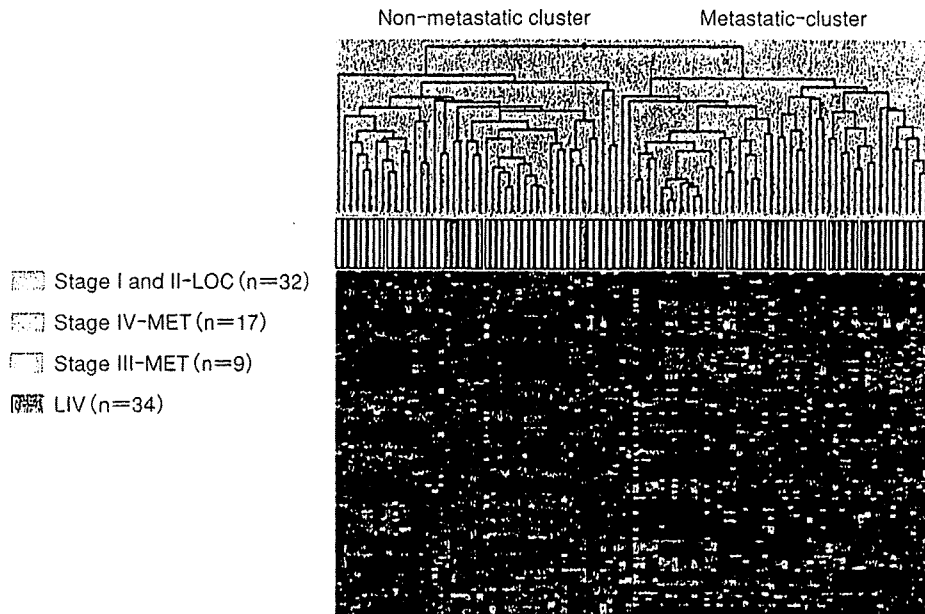


図2 大腸癌原発巣と肝転移巣の遺伝子発現プロファイル解析  
 119種類の大腸癌の発育進展に関する遺伝子群の発現パターンでは、肝転移のある原発巣 (MET) は肝転移のない原発巣 (LOC) よりもむしろ肝転移巣 (LIV) の発現プロファイルに近い。  
 (Yamasaki M *et al.*, 2007<sup>19)</sup>より引用)

の有無とは独立して予後良好群と不良群に判別することができ、さらに予後不良群を示す発現パターンは遠隔転移に深く関連していることを示した。これらの結果より、遠隔転移能は比較的早期よりすでに獲得されており、腫瘍内に本質的に備わっているという仮説がたてられるようになった。

Ramaswamy ら<sup>9)</sup>は、数種類の腺癌原発巣 64 例と転移巣 12 例の遺伝子発現プロファイルを比較し、原発巣、転移巣それぞれで転移に関与深い 64 種ずつの遺伝子群による刻印を同定することによって、この刻印を有する原発巣はさまざまな癌腫において有意に予後不良であることを示した。これは、原発巣内にはすでに転移能をもった細胞が十分量存在し、その個体の予後を規定する因子となっていることを示している。また Ye ら<sup>9)</sup>は、肝細胞癌の原発巣と転移巣では遺伝子発現プロファイルは類似するが、転移の有無によってプロファイルは大きく異なることより、原発性肝細胞癌の段階で転移する可能性はすでに決まっていると、90%の確率で転移予測できることを示した。

Kang ら<sup>10)</sup>は種々の乳癌の cell line を用いて、骨、肺、肝に転移する遺伝子発現プロファイルを比較し、原発腫

瘍内には予後不良な遺伝子発現の刻印が存在するだけでなく、遠隔転移巣別に組織特有の遺伝子発現パターンが内在することを示した。

これらの報告は遺伝子発現プロファイル解析が、遠隔転移・再発のメカニズムの解明の一助となり、転移予測に大きく貢献することを証明したものであるとともに、従来の癌転移モデルとは異なる解釈の可能性を示したものである。

## 大腸癌の転移予測診断の可能性を探る

大腸癌でも遺伝子プロファイルによるリンパ節転移、遠隔転移、再発の予測診断の可能性について多数報告されている。Kwon ら<sup>11)</sup>はリンパ節転移に関与する遺伝子群を同定し、その発現プロファイルによる転移予測の可能性を報告した。Croner ら<sup>12)</sup>はリンパ節転移の予測判断には従来の臨床病理因子より遺伝子発現プロファイルが優れているとした。Wang ら<sup>13)</sup>は 23 種の遺伝子発現パターンによって、Dukes' B 大腸癌の術後 3 年以内の再発が 13 倍もの高危険度となる群を予測できるとし、Eschrich ら<sup>14)</sup>は 43 種の遺伝子の発現プロファイルによる分子 staging が Dukes' 病期より予後予測に優れている

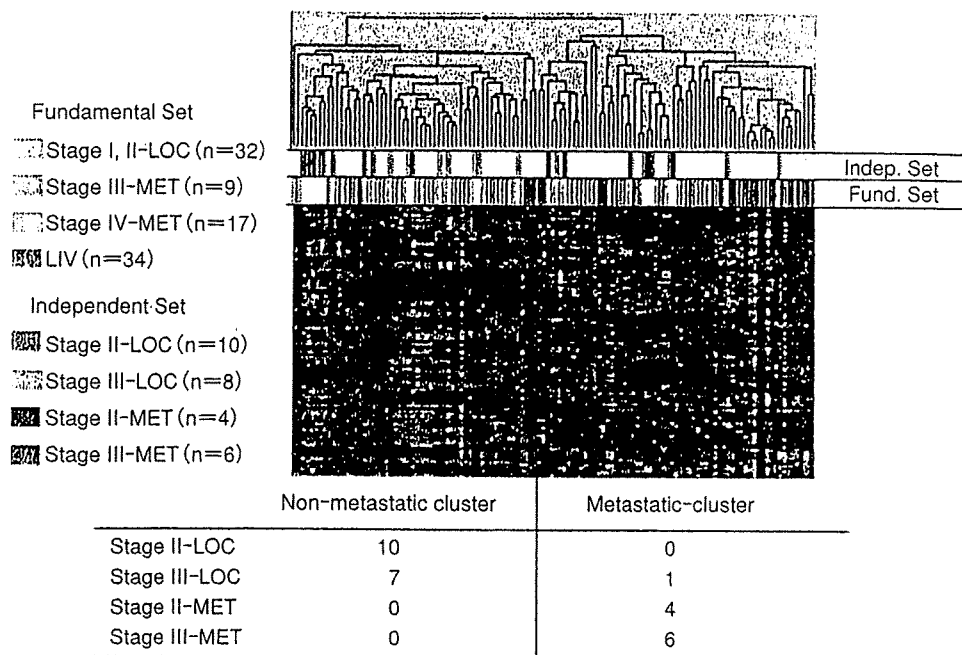


図4 肝転移の有無による大腸癌の遺伝子発現プロファイル解析  
64種類の転移関連遺伝子群の発現パターンは転移のある症例(MET)と肝転移のない症例(LOC)ではまったく異なっている。(Yamasaki M *et al.*, 2007<sup>19)</sup>より引用)

と報告した。Barrierら<sup>15)16)</sup>は、stage II大腸癌の異時性転移の予測診断の可能性について、さらには切除された癌周囲の大腸粘膜からも同様に再発予後予測が可能であることを示した。Stage II結腸癌は一般的に予後良好で、術後補助療法の有用性は検証されていないが、個々の症例の分子生物学的特性を反映した予後予測診断が可能となれば、一部の高リスク症例群に対して効率的な補助化学療法など個別化医療の実践が期待できる点で臨床的意義が大きいと考えられる。また Arangoら<sup>17)</sup>は Dukes'C大腸癌の予後予測に従来の分子マーカーである TP-53 や K-RAS の遺伝子変異、18q染色体のゲノム不安定性よりも遺伝子プロファイルが優れていたと報告した。

われわれ<sup>18)</sup>の教室では大腸癌解析用の cDNA マクロアレイを用いて大腸癌の肝転移予測診断に取り組んでいる。肝転移を起こした原発巣は肝転移を伴わない原発巣よりも肝転移巣の発現プロファイルに近いこと(図2)、また Dukes 病期によらず、原発巣の解析により肝転移の有無によりその発現パターンはまったく異なることがわかってきた(図3)。これら結果は大腸癌の肝転移に関しても、原発巣に転移刻印が内在することを示唆しており、

刻印の違いによって disease free survival, overall survival とともに有意差があることを示した(図4, 5)<sup>19)</sup>。

### 遺伝子発現プロファイル研究の今後の展望とは?

今後の消化器癌の転移研究は DNA マイクロアレイ技術による遺伝子発現プロファイルをはじめ、各種の解析手段を組み合わせることで DNA, RNA, 蛋白質での相互作用を包括的に解析する方向へ発展すると思われる。しかし基礎的研究で得られた成果を実際の臨床に応用するためにはまだ多くの課題と改善点が残されている。

転移という現象をより明確にとらえるためには、解析サンプルの対象を、laser captured microdissection (LCM)法によって純粋な癌細胞だけとするのか、周囲細胞との interaction も考慮にいれて whole tissue とするのか議論が続けられている<sup>20)</sup>。また全身麻酔下手術も遺伝子発現に影響があり研究デザインには十分な注意が必要である<sup>21)</sup>。

解析デザインに関しては、解析遺伝子数に対する症例数が少ないことや、同一データを対象としても用いる解

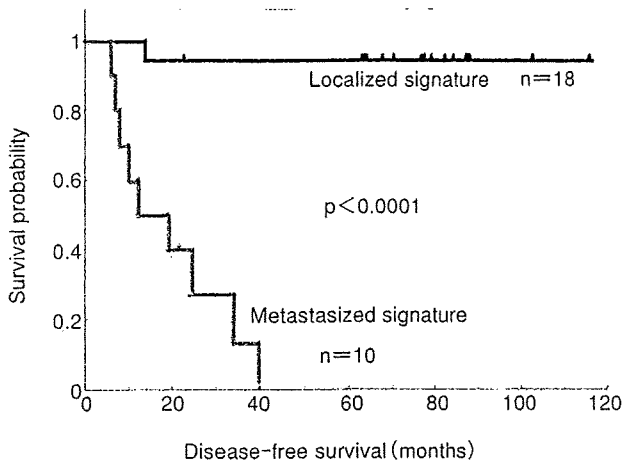


図4 刻印による disease-free survival (Yamasaki M *et al*, 2007<sup>19)</sup>より引用)

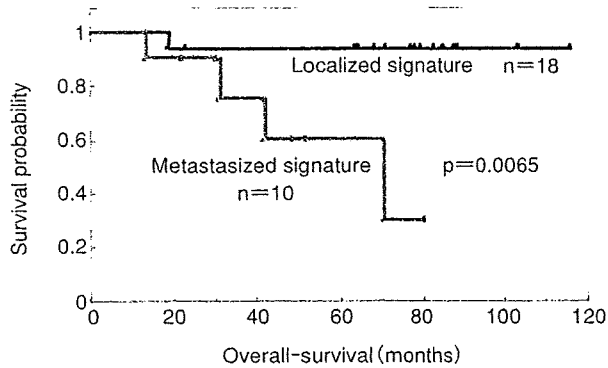


図5 刻印による overall-survival (Yamasaki M *et al*, 2007<sup>19)</sup>より引用)

析手法により結果が異なるといった問題点があげられる。実際に研究単位により報告されている転移関連遺伝子の相関性が少ないことや、判別信頼性が不足していることが指摘されている<sup>22)</sup>。その他DNAマイクロアレイのplatformの種類、対照サンプルの設定でも予測精度が変動することが知られるようになり注意深い検討が必要である。

今後はより多数症例の解析による validation と再現性を確保し、さらにプロスペクティブな臨床試験によって、予後予測の臨床応用のエビデンスを証明することが必要である<sup>23)</sup>。

## おわりに

遺伝子発現プロファイル解析による癌の転移研究と新たな知見について概説した。今後、網羅的な分子生物学

的研究により、どのような癌細胞集団が、どのような転移様式をとって転移巣を形成するのか、そのメカニズムの解明がさらに進むと考えられる。近い将来、従来困難であった個々の患者の転移・再発状況に即した予後を予測することが可能になり、消化器癌の個性化医療の実践が期待される。



## 文献

- 1) 竹政伊知朗, 門田守人: ヒト癌の遺伝子発現プロファイル解析の現況と展望: 松原謙一編. ゲノム機能, 発現プロファイルとトランスクリプトーム中山書店, 東京, 2000, pp. 83-101
- 2) Ntzani EE, Ioannidis JP: Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 362: 1439-1444, 2003
- 3) Liotta LA, Tryggvason K, Garbisa S *et al*: Metastatic potential correlates with enzymatic degradation of basement membrane collagen. *Nature* 284: 67-68, 1980
- 4) 竹政伊知朗, 門田守人, 松原謙一: 癌の悪性度とマイクロアレイ; 遺伝子発現プロファイル解析. 消化器外科 24: 1733-1739, 2001
- 5) 竹政伊知朗, 山崎誠, 門田守人: DNAチップを用いた癌研究の現状と展望. *Molecular Medicine* (臨時増刊号 癌ゲノム学) 39: 54-65, 2002
- 6) van't Veer LJ, Dai H, van de Vijver MJ *et al*: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415: 530-536, 2002
- 7) van de Vijver MJ, He YD, van't Veer LJ *et al*: A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347: 1999-2009, 2002
- 8) Ramaswamy S, Ross KN, Lander ES *et al*: A molecular signature of metastasis in primary solid tumors. *Nat Genet* 33: 49-54, 2003
- 9) Ye QH, Qin LX, Forques M *et al*: Predicting hepatitis B virus positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat Med* 9: 416-423, 2003
- 10) Kang Y, Siegel PM, Shu W *et al*: A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 3: 537-549, 2003
- 11) Kwon HC, Kim SH, Roh MS *et al*: Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer. *Dis Colon Rectum* 47: 141-152,

2004

- 12) Croner RS, Peters A, Brueckl WM *et al* : Microarray versus conventional prediction of lymph node metastasis in colorectal carcinoma. *Cancer* 104 : 395-404, 2005
- 13) Wang Y, Jatkoe T, Zhang Y *et al* : Gene expression profiles and molecular markers to predict recurrence of Dukes'B colon cancer. *J Clin Oncol* 22 : 1564-1571, 2004
- 14) Eschrich S, Yang I, Bloom G *et al* : Gene expression profiles and molecular markers to predict recurrence of Dukes'B colon cancer. J Clin Oncol Molecular staging for survival prediction of colorectal cancer patients. *J Clin Oncol* 23 : 3526-3535, 2005
- 15) Barrier A, Boelle PY, Roser F *et al* : Stage II colon cancer prognosis prediction by tumor gene expression profiling. *J Clin Oncol* 24 : 4685-4691, 2006
- 16) Barrier A, Roser F, Boelle PY *et al* : Prognosis of stage II colon cancer by non-neoplastic mucosa gene expression profiling. *Oncogene*, Oct 9, 2006
- 17) Arango D, Laiho P, Kokko A *et al* : Gene-expression profiling predicts recurrence in Dukes'C colorectal cancer. *Gastroenterology* 129 : 874-884, 2005
- 18) Takemasa, I. Higuchi H, Yamamoto H *et al* : Construction of preferential cDNA microarray specialized for human colorectal carcinoma : molecular sketch of colorectal cancer. *Biochem Biophys Res Commun* 285 : 1244-1249, 2001
- 19) Yamasaki M, Takemasa I, Komori T *et al* : The gene expression profile represents the molecular nature of liver metastasis in colorectal cancer. *Int J Oncol* 30 : 129-138, 2007
- 20) Hynes RO : Metastatic potential : generic predisposition of the primary tumor or rare, metastatic variants-or both? *Cell* 113 : 821-823, 2003
- 21) Lin DW, Coleman IM, Hawley S *et al* : Influence of surgical manipulation on prostate gene expression : implications for molecular correlates of treatment effects and disease prognosis. *J Clin Oncol* 23 : 3763-3770, 2006
- 22) Michiels S, Koscielny S, Hill C : Prediction of cancer outcome with microarrays : a multiple random validation strategy. *Lancet* 365 : 488-492, 2005
- 23) 竹政伊知朗, 池田正孝, 山本浩文ほか : 網羅的遺伝子・タンパク解析, DNA マイクロアレイ, 癌転移もメカニズムがよくわかる, 肝転移のすべて, 門田守人, 松浦成昭編, 永井書店, 東京, 2006, pp. 139-146

たけまさ・いちろう

竹政伊知朗 大阪大学大学院医学系研究科外科学講座消化器外科助手

専門は, 消化器外科学, 消化器内視鏡学.  
研究テーマは, 消化器癌における OMICS 解析の臨床応用.

## RESEARCH ARTICLE

# A simple and highly successful C-terminal sequence analysis of proteins by mass spectrometry

Hiroki Kuyama<sup>1\*</sup>, Keisuke Shima<sup>2\*</sup>, Kazuhiro Sonomura<sup>2</sup>, Minoru Yamaguchi<sup>2</sup>, Eiji Ando<sup>2</sup>, Osamu Nishimura<sup>1, 2</sup> and Susumu Tsunasawa<sup>1, 2</sup>

<sup>1</sup> Institute for Protein Research, Osaka University, Suita, Japan

<sup>2</sup> Life Science Laboratory, SHIMADZU CORPORATION, Kyoto, Japan

A simple and efficient method for C-terminal sequencing of proteins has long been pursued because it would provide substantial information for identifying the covalent structure, including post-translational modifications. However, there are still significant impediments to both direct sequencing from C termini of proteins and specific isolation of C-terminal peptides from proteins. We describe here a highly successful, *de novo* C-terminal sequencing method of proteins by employing succinimidylloxycarbonylmethyl tris (2,4,6-trimethoxyphenyl) phosphonium bromide and mass spectrometry.

Received: November 12, 2007

Revised: December 11, 2007

Accepted: December 19, 2007

**Keywords:**

C-terminal peptide / Isolation / Mass spectrometry / Post-translational modification / Sequencing

## 1 Introduction

Determining amino acid sequences of proteins, including their PTM, is important in life sciences because the covalent structures themselves are responsible for the functions of individual proteins in living systems. MS has emerged as a key technology for determining the covalent structures of proteins, and two approaches based on MS, categorized as “top-down” and “bottom-up”, have been developed [1]. The former approach is more attractive [2, 3] because a particular series of fragment ions from the N and C terminus of an intact protein is obtained with high sensitivity and accuracy, by which its amino acid sequence, including PTM, is directly read out. However, it is still difficult to effectively dissociate large proteins into a series of fragment ions for se-

quencing, although sophisticated analyzers such as FT-ICR and orbitrap mass spectrometers have been developed. Therefore, the latter approach is generally used to determine the covalent structures of proteins [4]. It involves two steps: (i) determination of the molecular weight of the target protein as accurately as possible, (ii) structural analyses by PMF and/or MS/MS, utilizing information from genome and protein databases. These analyses are very useful, but they often fail to identify post-translationally modified residues, or to detect sequence polymorphisms observed in various protein isoforms. This is because sequence coverage by PMF is usually less than 60% and protein databases are still insufficiently comprehensive to effectively search for any protein from the various creatures on the Earth. Therefore, it is necessary to determine the complete covalent structures of proteins by orthogonal methodologies. The addition of information from N- and C-terminal sequencing of proteins into databases is practically significant and is expected to maximize the chance of identifying sequence variants [5]. In line with this expectation, several methods for N-terminal [6–8] and C-terminal [9, 10] sequence analyses by MS have been reported, and we have proposed that this scientific category be called “terminal proteomics” [11].

\* These authors contributed equally to this work.

**Correspondence:** Dr. Hiroki Kuyama, Institute for Protein Research, Osaka University, Suita 565-0871, Japan

**E-mail:** kuyama@protein.osaka-u.ac.jp

**Fax:** +81-6-6879-4320

**Abbreviations:** DITC, *p*-phenylenediisothiocyanate; LysC, lysyl endopeptidase; Pfu, *Pyrococcus furiosus*; RPA, replication protein A; TCEP, tris (2-carboxyethyl) phosphine hydrochloride; TMPP-Ac-OSu, succinimidylloxycarbonylmethyl tris (2,4,6-trimethoxyphenyl) phosphonium bromide

The development of a simple and highly successful method for C-terminal sequencing by MS has particularly been a growing need for a long time.

This paper describes selective recovery and *de novo* sequencing of C-terminal fragments of proteins of interest. The method is based on highly specific modification of the  $\alpha$ -amino group of peptides for attaching a positively charged group in a pool of lysyl endopeptidase (LysC) digest of proteins, and scavenging peptides having a free amino group for isolating the desired C-terminal fragment that does not contain a free amino group.

## 2 Materials and methods

### 2.1 Materials

Chicken ovalbumin, chicken lysozyme, bovine  $\alpha$ -casein, horse cytochrome *c*, and iodoacetamide were obtained from Sigma (St. Louis, MO). Tris (2-carboxyethyl) phosphine hydrochloride (TCEP) and succinimidylloxycarbonylmethyl tris (2,4,6-trimethoxyphenyl) phosphonium bromide (TMPP-Ac-OSu) were obtained from Fluka (Switzerland). LysC, sodium hydrogen carbonate (NaHCO<sub>3</sub>), ACN, 2-propanol and TFA were purchased from Wako Pure Chemical Industries, (Osaka, Japan). CHCA (high-purity mass-spectrometric grade) was obtained from Shimadzu GLC (Tokyo, Japan). Four peptides (RKRSAE, AAKIQASFRGMHARKK, TRDIYETDYYRK, and MHRQETVDCLK-NH<sub>2</sub>) were from AnaSpec (San Jose, CA). A peptide (GFETVPEG-NH<sub>2</sub>) was synthesized in house using Shimadzu peptide synthesizer PSSM-8 by the Fmoc strategy. A recombinant of replication protein A (RPA) complex from the hyperthermophilic archaeon, *Pyrococcus furiosus* (Pfu) (PfuRPA; 1.54 mg/mL solution containing 50 mM Tris-HCl, 0.4 M NaCl, 0.5 mM DTT, 0.1 mM EDTA, and 10% glycerol) was provided by Professor Ishino (Kyushu University). *p*-Phenylenediisothiocyanate (DITC) resin used in this study was obtained from Shimadzu Corporation (Kyoto, Japan) as an item packaged in an ORFinder-NB Mass Sequencing Kit. Water used in all experiments was purified using a Milli Q water Purification System. All other chemicals were of analytical reagent grade and were used without further purification.

### 2.2 Methods

#### 2.2.1 General protocol

One hundred micrograms of sample protein was dissolved in 23  $\mu$ L of 8 M urea-50 mM NaHCO<sub>3</sub>. To the solution was added TCEP solution (1  $\mu$ L; 0.1  $\mu$ mol/ $\mu$ L) and incubated for 1 h at 37°C, followed by addition of iodoacetamide solution (1  $\mu$ L; 0.3  $\mu$ mol/ $\mu$ L, 1-h incubation at room temperature). To this solution was added LysC solution (10  $\mu$ g in 250  $\mu$ L of ACN-50 mM NaHCO<sub>3</sub>; 1:9), and this was left standing for 15 h at 37°C. An aliquot corresponding to 10 pmol of protein

was taken from the digest solution and diluted to 10  $\mu$ L with 50 mM NaHCO<sub>3</sub>. To this diluted solution was added 3  $\mu$ L of 1 mM TMPP-Ac-OSu [12] solution (ACN-water; 1:4), followed by sonication in a water bath for 30 min. DITC resin (5 mg) was pre-washed with 20% ACN solution (60  $\mu$ L; three times) and 50 mM NaHCO<sub>3</sub> (60  $\mu$ L; three times). To the washed resin the TMPP modified solution was added, which was allowed to stand for 2 h in a water bath at 60°C. The extraction was carried out using ACN-50 mM NaHCO<sub>3</sub> (1:9, 60  $\mu$ L; twice) and 2-propanol-acetonitrile-0.1%TFA (1:1:2, 60  $\mu$ L; three times). The extracts were combined and dried in a vacuum centrifuge.

#### 2.2.2 General protocol for proteins on gel slices

A sample protein (30 pmol) was separated by SDS-PAGE in a 12.5% acrylamide gel. The Coomassie-stained protein band was excised and washed with 50% v/v ACN in 100 mM NaHCO<sub>3</sub>. The washed gel piece was dehydrated with ACN and dried in a vacuum centrifuge. To the dried gel 100  $\mu$ L of 10 mM aqueous TCEP solution was added to reduce disulfide bonds. This solution was incubated for 30 min at 37°C. S-alkylation was accomplished by replacing the TCEP solution with 55 mM iodoacetamide in 100 mM NaHCO<sub>3</sub>. After a 45-min incubation at room temperature in the dark, the gel piece was washed with 100  $\mu$ L of 50 mM NaHCO<sub>3</sub>, shrunk by dehydration in ACN, and dried in a vacuum centrifuge. The gel piece was then rehydrated with 2  $\mu$ L of ACN-50 mM NaHCO<sub>3</sub> (1:9) containing 200 ng of LysC. After 5 min, 50 mM of NaHCO<sub>3</sub> solution (15  $\mu$ L) was added to keep the gel piece moist during digestion (37°C, overnight). To extract the resulting peptides, 30  $\mu$ L of 50% ACN containing 0.05% TFA was added to the digestion mixture, and the gel piece was sonicated in a water bath for 10 min, after which the supernatant was collected. This extraction procedure was repeated three times. The extract was combined and lyophilized. The resulting powder was dissolved with 10  $\mu$ L of ACN-50 mM NaHCO<sub>3</sub> (1:9). To this solution 1  $\mu$ L of TMPP-Ac-OSu solution (10 mM in ACN-water, 1:4) was added and the mixture was sonicated in a water bath for 30 min. The TMPP-modified solution was added to the pre-washed DITC resin (described above), and this was allowed to stand for 2 h in a water bath at 60°C. The extraction was done using ACN-50 mM NaHCO<sub>3</sub> (1:9, 60  $\mu$ L; twice) and 2-propanol-ACN-0.1%TFA (1:1:2, 60  $\mu$ L; three times). The extracts were combined and dried in a vacuum centrifuge.

#### 2.2.3 MALDI-TOF MS

MALDI mass spectra were recorded on AXIMA CFR-plus and AXIMA TOF<sup>2</sup> (SHIMADZU/KRATOS, Manchester, UK) reflectron TOF mass spectrometers equipped with a nitrogen laser (337 nm, 3-ns pulse width). All measurements were performed in positive-ion reflectron mode. The ion acceleration voltage was set to 20 kV, and the reflectron detector was operated at 24 kV. The flight path in the reflectron mode is

about 240 cm for both types of instruments. For MS/MS experiments, CID was carried out using helium at a pressure of *ca.*  $5 \times 10^{-6}$  mbar in the collision cell.

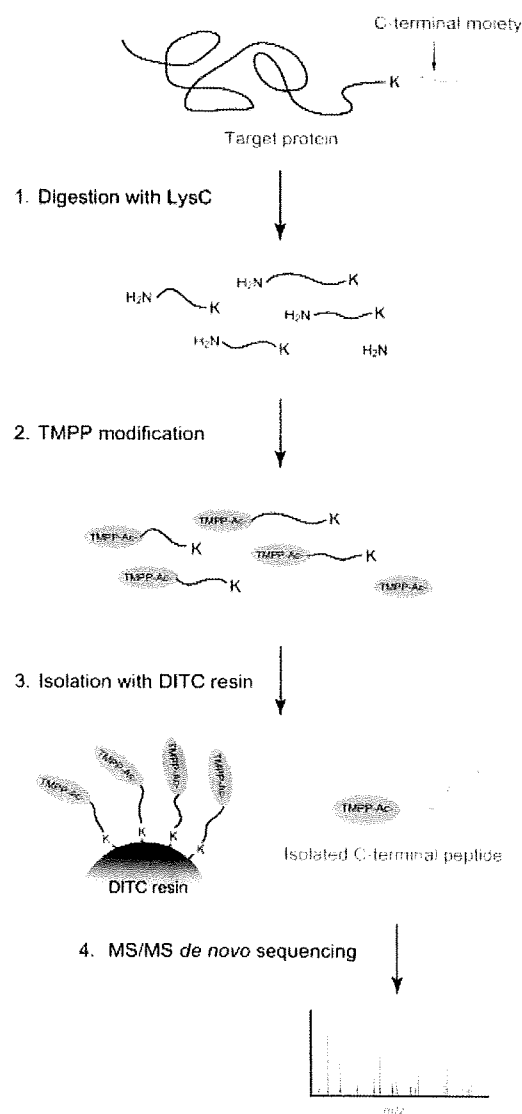
CHCA was used as a matrix, which was dissolved to saturation in 50% aqueous ACN containing 0.05% TFA. A portion (0.5  $\mu$ L) of sample solution was mixed with an equivalent volume of matrix solution on the MALDI target plate and analyzed after drying. A double layer protocol [12] was used for TMPP-Ac peptides. The *m/z* values in spectra were externally calibrated with angiotensin II (human) and ACTH fragment 18–39 (human) using CHCA as a matrix.

### 3 Results and discussion

Our project for terminal sequencing of protein has set the goal on both of isolation and *de novo* sequencing of N- [8, 13] and C-terminal peptides [14, 15] by MS. In the course of the study, we found that TMPP-Ac-OSu [12] exhibited selectivity exclusively for the  $\alpha$ -amino group, which was more than expected. Furthermore, because side reactions were described but were not actually seen in mass analysis, we thought of using this selective tagging agent for LysC digest of protein, and amino-reactive group immobilized material. Hence, we tried DITC resin as a scavenger of peptides having free amino group(s), which in turn afforded good recovery of peptides devoid of amino group(s).

A technique using the combination of LysC and DITC resin for isolating C-terminal peptide was reported almost two decades ago [16]. In the method, the sequence determination of an isolated C-terminal peptide was carried out by Edman method (not by MS) after cleavage of the peptide from the resin with use of hazardous conc. TFA. Therefore, compared with the present method, some drawbacks in the technique can be pointed out as follows: (i) sequence analysis by Edman reaction remains ambiguous for identifying several amino acids as Ser, Thr etc. and post-translationally modified residues, (ii) low yield of PTH-amino acid adjacent to the C terminus leads to difficulty for complete sequence analysis, (iii) mixture sample can not be analyzed without further fractionation of isolated C-terminal peptide, and (iv) C-terminal determination is not possible if the peptide consists of only one amino acid. With this in mind, we set out to the investigation of selective isolation of C-terminal peptides and their *de novo* sequencing with combinational use of an  $\alpha$ -amino group selective modifying reagent (TMPP-Ac-OSu) and MS.

Before describing the results, we will present the outline of our method in Fig. 1. First, the protein of interest is digested with LysC to yield peptides that have amino groups at both ends, but a C-terminal peptide has only the  $\alpha$ -amino group in it. If a protein is N-terminally blocked by any PTM, the N-terminal peptide has only the  $\epsilon$ -amino group. The next step is TMPP-Ac derivatization with TMPP-Ac-OSu to produce TMPP-Ac peptides. These peptides are selectively modified at the  $\alpha$ -amino group (not at the  $\epsilon$ -amino group on



**Figure 1.** Schematic overview of the method. A protein is first digested with LysC to yield peptides incorporating amino groups at both ends but the C-terminal peptide contains only an  $\alpha$ -amino group. This is followed by selective attachment of the TMPP-Ac moiety to  $\alpha$ -amino groups, and isolation of the C-terminal peptide using a DITC resin.

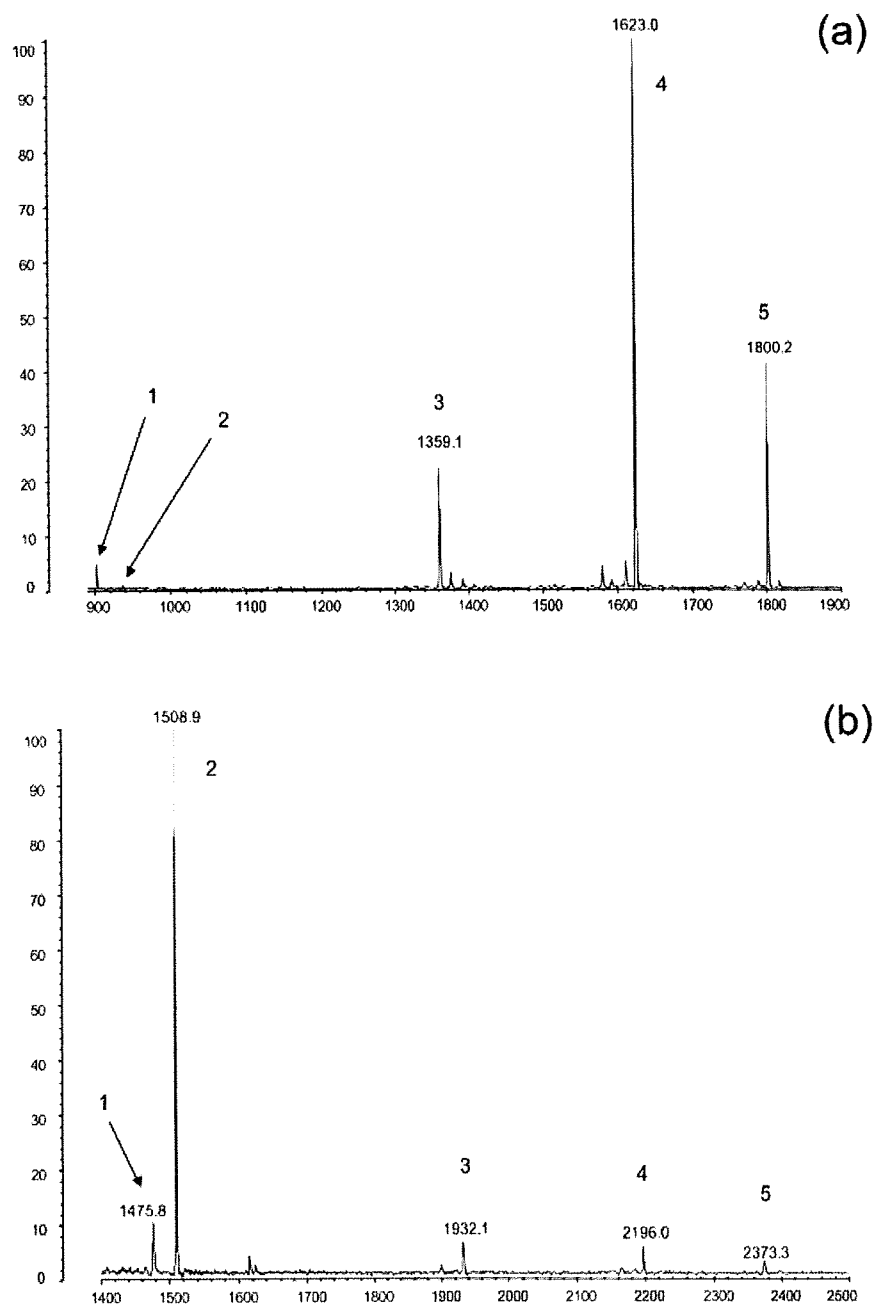
the lysine residue). Hence, the resultant modified peptides have free amino groups at their lysine residue, but the C-terminal peptide itself has no free amino group left. In the final step, peptides with free amino groups are attached covalently to the DITC resin, and the C-terminal peptide thus recovered in the supernatant solution is subjected to *de novo* sequencing analysis by MALDI-MS.

The results obtained are described using (i) model peptides, (ii) model proteins, and (iii) a recombinant protein (PfuRPA).

### 3.1 Model peptides

To demonstrate the usefulness of the method, we tried this strategy using model peptides. Five peptides were used for a model peptide mixture (1) RKRSRAE ( $m/z$  901.5), (2) GFETVPETG-NH<sub>2</sub> ( $m/z$  934.4), (3) MHRQETVDCLK-NH<sub>2</sub> ( $m/z$  1357.7), (4) TRDIYETDYYRK ( $m/z$  1621.8), and (5) AAKIQASFRGHMARKK ( $m/z$  1799.0). Peptide (2) is devoid of lysine residue and thus has only one amino group at its N

terminus. It was expected that peptide (2) would be isolated after treatment with DITC resin. Fifty microliters of equimolar solution of these five peptides (2 pmol/ $\mu$ L each) was prepared, and to this solution was added TMPP-Ac-OSu solution (5  $\mu$ L of 1 mM solution in ACN-water, 1:4). The solution was then sonicated in a water bath for 30 min. An aliquot (4  $\mu$ L) was taken and treated with DITC resin for 2 h at 60°C. Figure 2 presents the MALDI-MS spectra of the equimolar mixture of the five peptides before and after TMPP



**Figure 2.** MALDI-TOF mass spectra before and after TMPP-Ac modification of the mixture of five model peptides. The five peptides are as follows: (1) RKRSRAE ( $m/z$  901.5); (2) GFETVPETG-NH<sub>2</sub> ( $m/z$  934.4); (3) MHRQETVDCLK-NH<sub>2</sub> ( $m/z$  1357.7); (4) TRDIYETDYYRK ( $m/z$  1621.8); (5) AAKIQASFRGHMARKK ( $m/z$  1799.0). Panel (a) shows the spectrum before TMPP-Ac modification, while (b) shows the signals of TMPP modified peptides after TMPP-Ac modification. The x-axis and y-axis represent  $m/z$  and % intensity, respectively, for all mass spectra.



modification. Side reactions at the potential reactive sites on the residues of Tyr, Lys, Thr, and Ser were not actually observed (data not shown). In Fig. 2a peptide (2) has the least intense peak before TMPP-Ac modification. In Fig. 2b, however, the modified peptide (2) has the most intense peak after the modification. If a peptide devoid of Arg and/or Lys residue in its sequence is modified with a TMPP-Ac reagent, the signal intensity in MALDI-MS can be enhanced. TMPP-Ac peptide (2) was clearly recovered from the mixture of TMPP-Ac peptides, as depicted in

Fig. 3a. The CID spectrum was acquired (Fig. 3b) for sequencing the TMPP-Ac peptide. Relatively simple fragmentation was observed, which was enough for *de novo* sequencing [17–19]. In addition to the readout of the sequence using a-type ions, several characteristic fragmentations were observed, including  $d_3/d_7$  (-58),  $d_4/d_8$  (-16), and  $d_5$  (-14), which can be very helpful for assigning the residues [20]. The peak of  $m/z$  573 is derived from one of the common reagent-related background ions [TMPP-CH=CO]<sup>+</sup> [12].

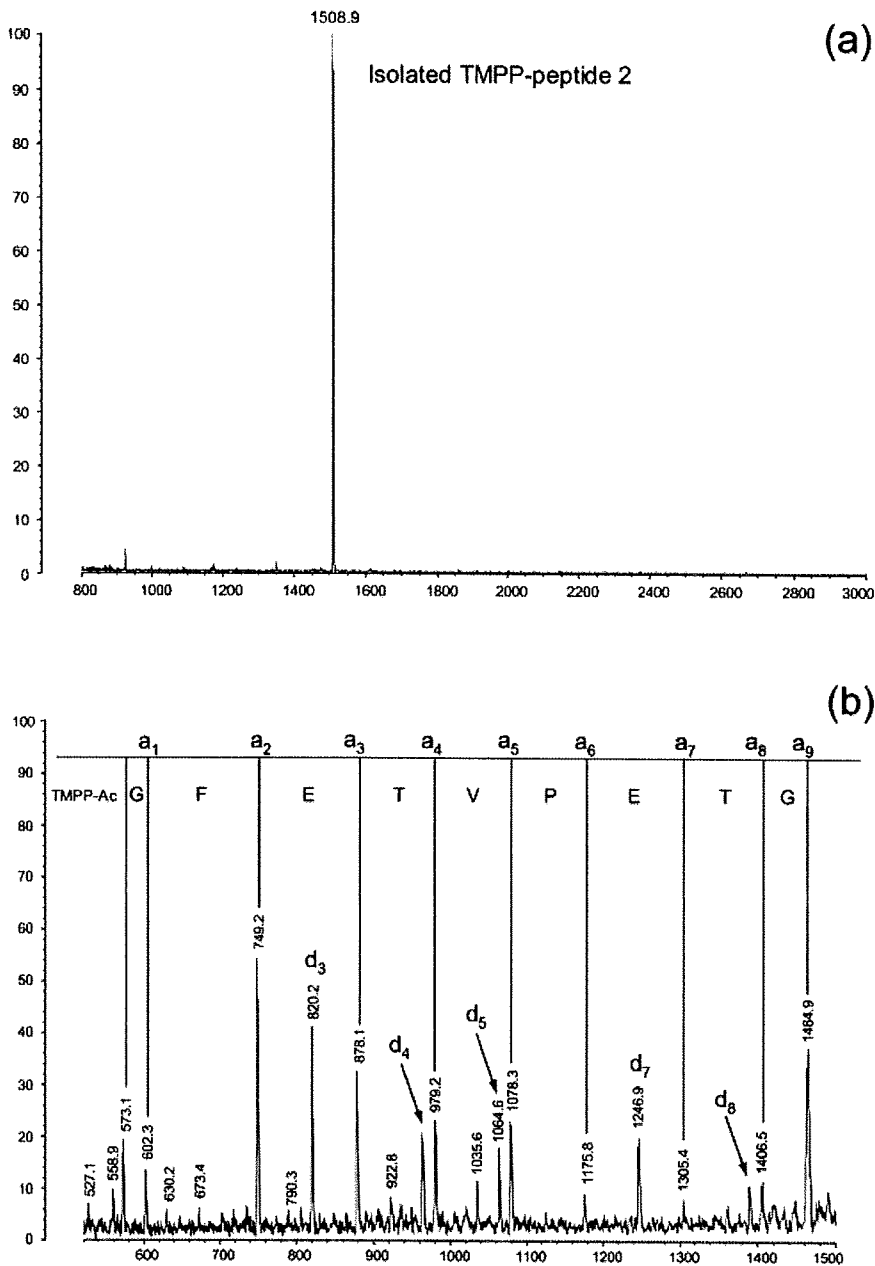


Figure 3. MALDI-TOF mass spectrum of the isolated peptide (2) after DITC resin treatment (panel a) and the CID spectrum (panel b).

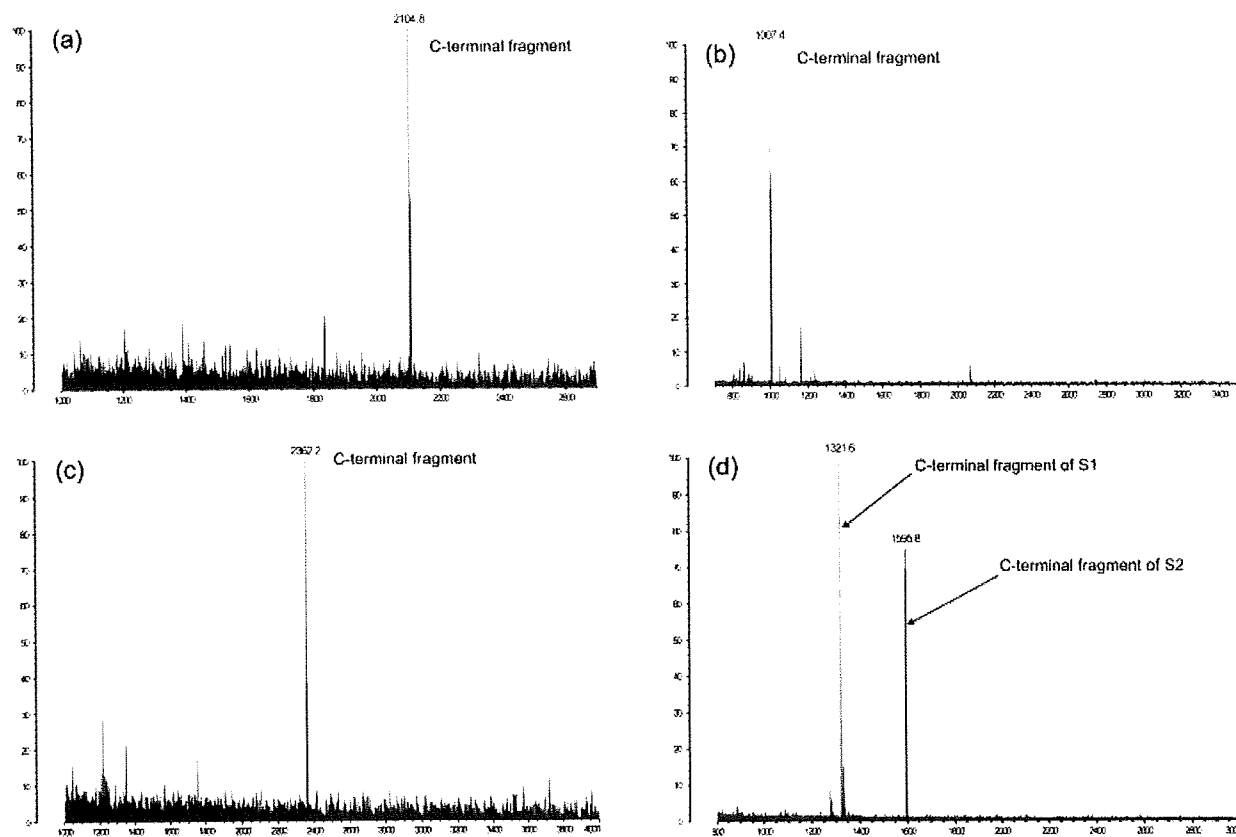
### 3.2 Model proteins

The model proteins used were lysozyme (chicken), ovalbumin (chicken), cytochrome *c* (horse), and  $\alpha$ -casein (bovine). Both the lyophilized powder and the excised gel strip obtained after purification by SDS-PAGE of each protein were used as starting material. For the solution phase experiment, 10 pmol of each LysC digest was used for the TMPP-Ac modification, followed by DITC resin treatment for the selective recovery of the C-terminal peptide. To prepare 1-D-separated gel samples, 30 pmol was applied to the gel, and each stained band was excised after separation. The resulting gel pieces were treated with LysC, TMPP-Ac reagent and DITC resin as described in Section 2.2.1. In this protocol, amino group-containing reagents were avoided because they may consume the NCS group on the resin. Figure 4 presents the MALDI-TOF mass spectra of their isolated C-terminal fragments. Figures 4a and b depict gel-started results, and the other two figures depict solution-started results. Isolation was successful regardless of whether gel pieces or solutions were used as the starting material. Their C-terminal peptides are GTDVQAWIRGC\*RL (lysozyme,  $m/z$  1530.8: \*S-alkylated with iodoacetamide), HIATN-

AVLFFGRC\*VSP (ovalbumin,  $m/z$  1787.9: \*S-alkylated with iodoacetamide), ANTE (cytochrome *c*,  $m/z$  433.2), TTMLPW ( $\alpha$ -casein S1,  $m/z$  747.4), and VIPYVRYL ( $\alpha$ -casein S2,  $m/z$  1021.6). The C-terminal peptides of lysozyme, cytochrome *c* and ovalbumin were sequenced in CID mode using a MALDI-TOF mass spectrometer. The spectra are plotted in Fig. 5. Sequencing was performed using mainly a-type ions and using the structural information from characteristic d-type ions that occurred due to the elimination of the side chain from certain amino acid residues. For example, Fig. 5a plotted the CID spectrum of the C-terminal fragment of lysozyme, in which d-type ions are formed on Thr (-16), Asp (-44), Val (-14), Gln (-57), Ile (-28), Arg (-85), Cys (-104: S-alkylated with iodoacetamide), and Leu (-42). The d-type ions are very helpful for *de novo* sequencing [20].

Two peaks were observed for  $\alpha$ -casein (Fig. 4d) because it consists of two components (S1 and S2), which indicates that protein mixture (*e.g.* hetero-oligomeric complex consisting of multiple subunits) can be analyzed with this strategy.

The isolation of C-terminal peptide and its *de novo* sequencing were successfully carried out using the four model proteins. Hence, we next used a recombinant protein consisting of three subunits to analyze their C-terminal peptides.



**Figure 4.** MALDI-TOF mass spectra of the isolated C-terminal peptides from four model proteins. (a) The isolated peptide from lysozyme. (b) The isolated peptide from cytochrome *c*. (c) The isolated peptide from ovalbumin. (d) The isolated peptide from  $\alpha$ -casein.

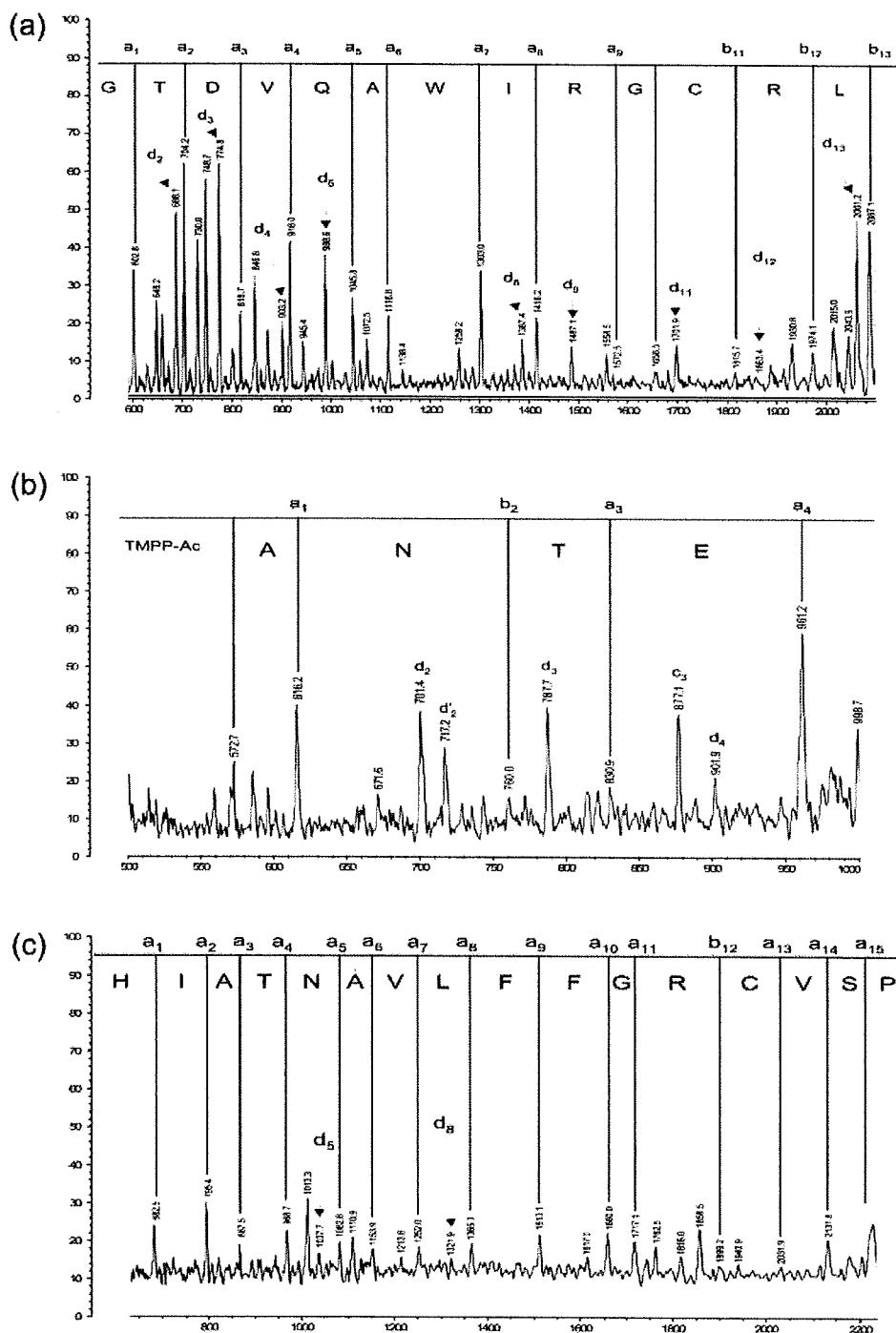


Figure 5. CID spectra of the isolated C-terminal peptides from lysozyme (panel a), cytochrome c (panel b), ovalbumin (panel c).

### 3.3 Recombinant protein

We then examined the method using a recombinant protein (PfuRPA) [21], consisting of three subunits, RPA14, RPA32,

and RPA41. Figures in the abbreviations indicate their molecular weights.

The recombinant protein was first analyzed without separation into subunits. The whole protein (100 µg) was

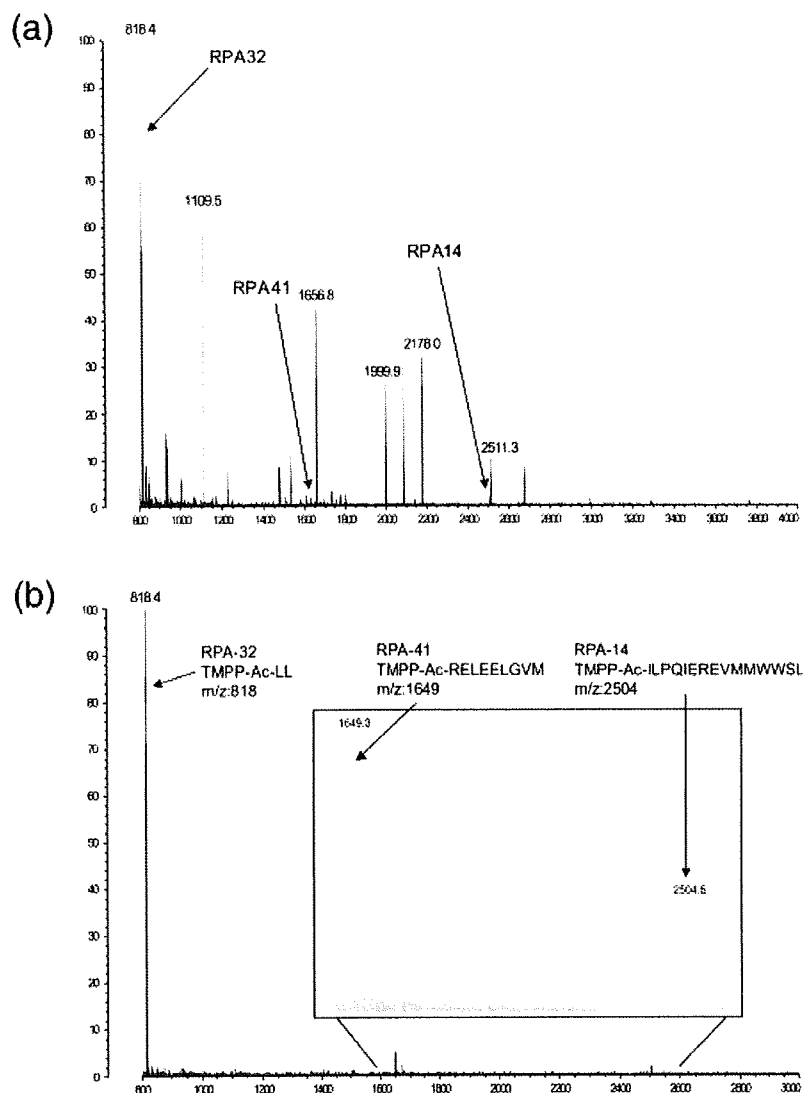
processed through reduction and S-alkylation, followed by digestion with LysC (10 µg) in 250 µL of ACN-50 mM NaHCO<sub>3</sub> (1:9). An aliquot (50 µL) was then treated with TMPP reagent, and one-tenth of the resulting mixture was added to 5 mg of DITC resin.

The results of TMPP-Ac modification and selective recovery of the three C-terminal peptides are depicted in Fig. 6. Among the many TMPP-Ac peptides that appeared after TMPP-Ac modification (Fig. 6a), the three objective C-terminal peptides were selectively recovered with DITC resin (Fig. 6b).

We next used three subunits of the recombinant protein separated on gel: 25 µg of the protein was applied for the separation on gel. After separation, the excised gel pieces were processed through reduction/S-alkylation, digestion (LysC), TMPP-Ac modification, and isolation with DITC

resin. Figure 7 presents MALDI-MS spectra at the stages of TMPP-Ac modification (Figs. 7a, b, c) and isolation (Figs. 7d, e, f). These experiments demonstrated the successful isolation of the C-terminal peptides, regardless of the methods used for sample preparation. The three C-terminal peptides were sequenced by MALDI-TOF-MS. Their CID spectra are shown in Fig. 8. The sequence readouts of the C-terminal peptides derived from RPA14 (Fig. 8a) and RPA32 (Fig. 8b) are consistent with their reported C-terminal sequences. The sequencing was performed using a-type and d-type ion signals (Ile, Leu, Gln, Glu, Arg, Val, Met, etc.). The signals of the d-type ions were quite helpful in the sequencing [20].

In the case of the C-terminal peptide of RPA41, no significant fragmentation peaks were observed in the CID spectrum (data not shown), hence its sequencing failed. This seemed attributable to the highest basicity of the arginine



**Figure 6.** MALDI-TOF mass spectra for C-terminal peptides isolated from PfuRPA protein complex. MALDI-TOF mass spectrum (a) after TMPP-Ac modification of the LysC digest, (b) after isolation of three C-terminal peptides using DITC resin. Arrows indicate signals of the C-terminal peptides derived from the three subunits, RPA14, RPA32, and RPA41 of PfuRPA.