Table 2. Breast Cancer Biomarker Candidates Were Cross-Validated by 2-DE/MS and TMA Analyses Using Human Breast Tissue Samples[a]

	2-DE on Cell Lines					Subcellular Location	
candidates	tumor	normal	TMA results	optimal conditions[b]	primary antibody suppliers	predicted	exp. and lit.
ANX1	↓ **(MCF7, HCC38)**	↑	88% of tumor tissues show negative staining	Pressure cook 1:500	BD Transduction Laboratories 610066	Cyto 52% Nucl 22%	Plasma mem and cyto
CRAB	↓ **(MCF7)**	↑	96% of tumor tissues show negative staining	Dk9 1:400	Stressgen #SPA-222	Cyto 35% Nucl 30%	Cyto and nucl
6PGL	↑ **(MCF7)**	↓	98% of tumor tissues show stronger positive staining	Tris-EDTA 1:50	Customized (BioGenes)	Cyto 35% Mito 22%	Cyto
CAZ2	↑ **(HCC38)**	↓	69% of tumor tissues show stronger positive staining	Dk6 1:400	Customized (BioGenes)	Cyto 48% Nucl 26%	Cyto
K2C7	↑ **(HCC38)**	ND	Both tumor and normal tissues show positive staining	In-house protocol 1:80	DakoCytomation M7018	Mito 61% Nucl 26%	Intermediate filament
LAM1	↑ **(MCF7, HCC38)**	ND	Both tumor and normal tissues show positive staining	DK6 1:50	Santa Cruz sc-20682	Nucl 52% Plasma mem 17%	Inner nucl membrane
TKT	↑ **(MCF7)**	↓	Both tumor and normal tissues show positive staining	Dk9 1:50	Customized (BioGenes)	Cyto 70% 22%	Cyto > nucl
PSD2	↑ **(HCC38)**	ND	Both tumor and normal tissues show positive staining	Dk9 1:200	Customized (BioGenes)	Nucl 74% Cyto 13%	Cyto and nucl
PAK2	↑ **(MCF7)**	ND	Both tumor and normal tissues show negative staining	Dk9 1:100	Zymed Laboratories 51–3000	Endopla 33% Mito 22%	Cyto and nucl

[a] Optimal conditions include the method for deparaffinization, primary antibody titers, and antigen subcellular localization. Customized polyclonal antibodies were raised against the selected polypeptides from the corresponding proteins (BioGenes GmbH, Berlin, Germany). [b] Optimal conditions include methods for deparaffinization and concentrations of primary antibody. ND: no detection/identification.

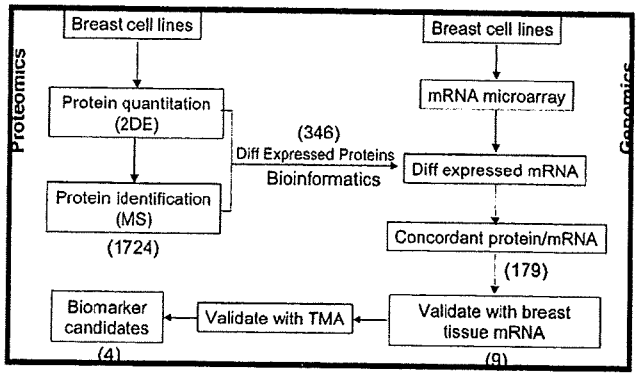

Figure 1. Schematic of an integrated proteo-transcriptomic approach for breast cancer biomarker discovery. Numbers in parentheses represent the number of proteins or mRNAs after each filtering step.

expressed spots and their protein identities is presented in the Supporting Information.

Biomarker Prioritization by Proteo-Transcriptomic Mapping. We then integrated the proteomic information with gene expression data. First, we focused on the 78 overexpressed and 32 underexpressed proteins in MCF-7 relative to CCD-1059Sk (Supplementary Table S2 in Supporting Information). Using Affymetrix U133_plus chip annotations and matching Swiss-Prot IDs, we mapped these differentially expressed proteins to their corresponding mRNA counterparts and identified two sets of microarray probes: one set ('MCF-7 upregulated') corresponding to MCF-7 up-regulated proteins (198 probes), and a second set ('MCF-7 downregulated') corresponding to MCF-7 down-regulated proteins (84 probes). We then ranked all ~50K probes on the microarray according to their strength of differential gene expression between MCF-7 and CCD-1059Sk, and used Gene Set Enrichment Analysis (GSEA) to examine the distribution of the two 'MCF-7 upregulated' and 'MCF-7-downregulated' probe sets against the ranked list. GSEA is a recently described powerful method for testing the coordinated over or underexpression of sets of genes using the Kolmogorov–Smirnov test over a weighted summation.[25] We found that the "MCF-7 upregulated" probes were significantly enriched in genes exhibiting elevated gene expression levels in MCF-7 compared to CCD-1059Sk cells ($p < 0.001$), indicating a good concordance between the proteomic and transcriptomic data. Likewise, the "MCF-7 downregulated" probe set was significantly enriched in genes exhibiting decreased gene expression in MCF-7 compared to CCD-1059Sk cells ($p < 0.001$) (Figure 3A). We then performed a similar analysis for the "HCC-38 upregulated" and "HCC-38 downregulated" probe sets, corresponding to the 81 overexpressed and 31 underexpressed proteins in HCC-38 cells compared to controls (Supplementary Table S3 in Supporting Information). The "HCC-38 upregulated" probe set (195 probes) was significantly enriched in genes highly expressed in HCC-38 cells ($p < 0.001$), while the "HCC-downregulated" probe set (71 probes) was significantly enriched in genes exhibiting decreased expression in HCC-38 cells ($p < 0.001$) (Figure 3A). These results indicate that a significant proportion of differentially expressed proteins (79–82%) identified in the 2DE/MS analysis are also associated with concomitant alterations in mRNA expression. By focusing on these concordantly expressed proteins, we were able to further shortlist the list of putative biomarker candidates to 87 proteins for MCF-7 versus CCD-1059Sk and 92 proteins for HCC-38 versus CCD-1059Sk, respectively. These proteins were then combined into a single list, representing 30 underexpressed and 105 overexpressed proteins differentially expressed between cancer and control cells.

To move beyond the *in vitro* cell line setting, we then prioritized the biomarker candidates by identifying proteins whose cognate genes were differentially expressed at the gene expression level between primary human breast tumors and normal breast samples. We constructed a training mRNA data set of 39 primary tissue samples (7 normal and 32 tumors), and used a combination of a Support Vector Machine (SVM) machine learning algorithm, coupled with 100 cross-valiation runs, to identify biomakers that could robustly discriminate between breast tumors and normal tissues. The SVM identified nine biomarkers for further evaluation, including Annexin I (ANX1_HUMAN), Alpha basic-Crystallin (CRAB_HUMAN), 6-phosphogluconolactonase (6PGL_HUMAN), F-actin capping protein alpha-2 subunit (CAZ2_HUMAN), type II cytoskeletal 7 (K2C7_HUMAN), Lamin B1 (LAM1_HUMAN), Transketolase (TKT_HUMAN), Serine/threonine-protein kinase PAK 2 (PAK2_HUMAN), and 26S proteasome non-ATPase regulatory subunit 2 (PSD2_HUMAN). Of these nine candidates, ANX1 and CRAB were down-regulated, while the rest were up-regulated in breast cancer cells (Table 2). Importantly, the gene expression patterns of the top nine biomarkers were concordant between the cell lines and human tumors. To further validate these nine biomarkers, we then assessed their ability to further classify an independent test set of 36 tissue samples (31 tumors and 5 normals) (Figure 3B). As shown in Figure 3B, the nine biomarkers could unambiguously differentiate all the tumor and normal samples in the test data set, indicating that they are relatively robust.

Validation of Potential Biomarkers by Tissue Microarray Analysis. Our decision to select the top nine biomarkers out of 179 was motivated by the empirical need to select a small set of candidate biomarkers for subsequent immunohistochemical validation. To validate the potential biomarkers at the protein level, we performed an immunohistochemical (IHC) analysis of the biomarkers on an independent breast cancer tissue microarray (TMA) of approximately 100 breast tumors and 100 normal breast tissues. Of the nine biomarker candidates, we purchased commercial antibodies for five (ANX1, CRAB, K2C7, LAM1, PAK2) and generated custom antibodies for the remaining four (6PGL, CAZ2, TKT, PSD2). Prior to IHC, the antibodies were optimized against paraffin-embedded cell line blocks, inferring the likely subcellular localization of the biomarker from either vendor-provided data (for commercial antibodies), prediction using bioinformatic analysis,[23] or Western blotting experiments using subcellularly fractionated cell line lysates (see Materials and Methods). Of the nine biomarker candidates, four (ANX1, CRAB, 6PGL, and CAZ2) were successfully validated on the TMA as bona fide differentially expressed proteins between normal breasts and breast tumors (Table 2). With the exception of ANX1 where a previous association with breast cancer has been demonstrated,[26-29] the other three proteins (CRAB, 6PGL, and CAZ2) represent novel breast cancer biomarkers.

The TMA analysis for ANX1 revealed that it was strongly expressed in normal breast tissues (10%), while 88% of the breast tumors exhibited positive staining only in fibroblast cells but not in tumor cells (Figure 4A,B, $p = 4 \times 10^{-30}$, Fisher test). These findings confirm previous reports that only the stromal cells in tumors, but not the epithelial cells, are reactive against

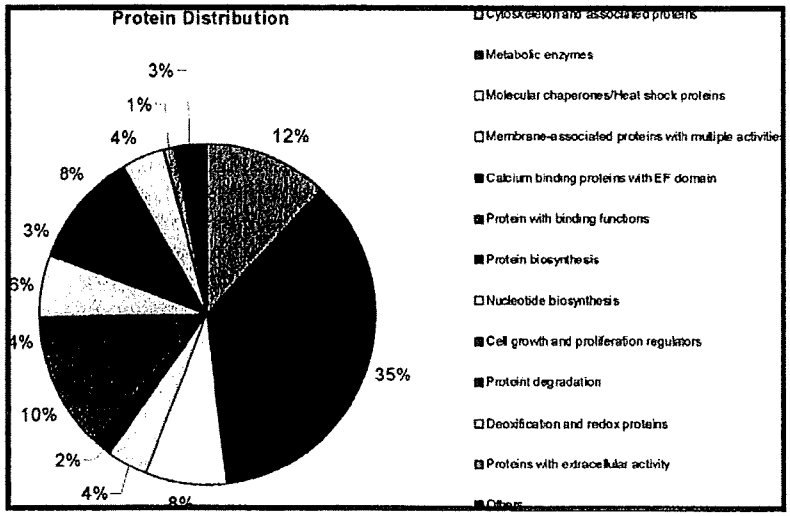

Figure 2. A total of 484 different protein species were identified from the 3 breast cell lines (MCF-7, HCC-38, and CCD-1059Sk). Their functional distributions are grouped into 13 categories. These categorizations should be treated as general divisions, as one major function to each protein. However, some proteins can have multiple functions and may belong to multiple functional categories.

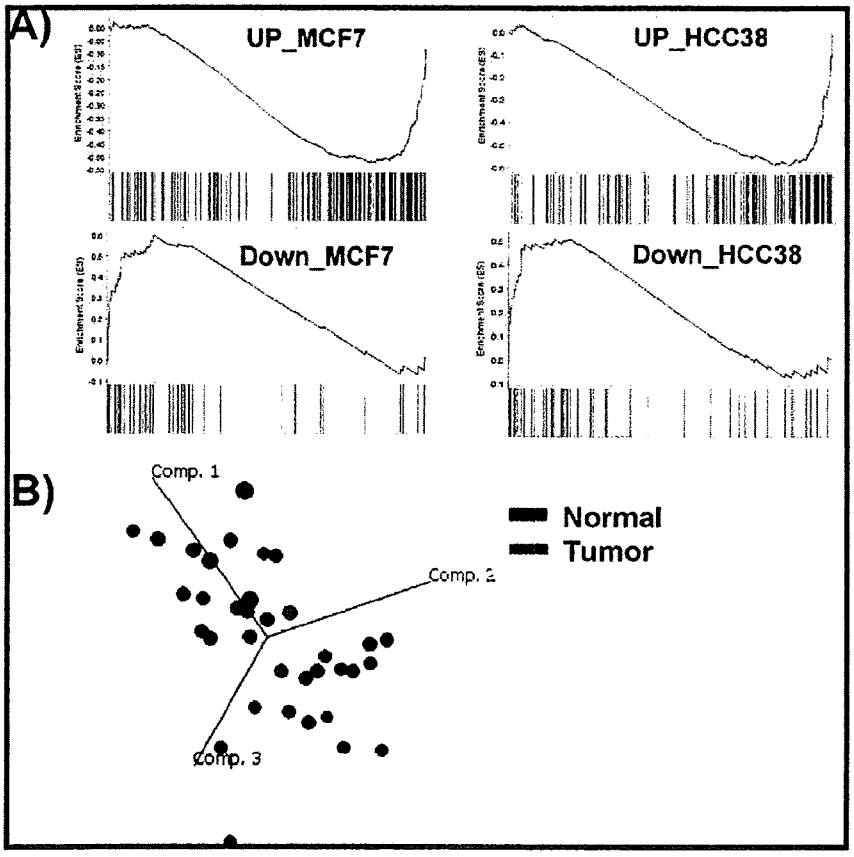

Figure 3. (A) Gene set enrichment analysis of differentially expressed proteins in mRNA expression data. (Left panel) Proteins identified by comparison between MCF-7 and CCD-1059Sk; (Right panel) Proteins identified by comparison between HCC-38 and CCD-1059Sk. The genes are ranked based on the ratio of their differential gene expression between control and cancer cell lines. The significantly differentially expressed protein spots were mapped to their corresponding mRNAs and used as test groups against the ranked gene lists by GSEA (see Materials and Methods). (B) Principal component analysis (PCA) of mRNA expression of nine biomarkers in an independent test set, including 5 normal and 31 tumors. The brown spots represent normal samples, while blue spots are tumor samples. Using nine biomarkers, normals are clearly distinguished from tumors.

the ANX1 antibody,[30] and are consistent with proposals that ANX1 might be an endogenous suppressor of cancer development.[27] Similarly, CRAB was expressed in only 5 out of 98 breast tumor tissue cores, while all 98 normal tissues exhibited strong CRAB expression in the myoepithelial cells of the normal ducts (Figure 4C,D, $p = 3 \times 10^{-50}$). Interestingly, the five breast tumor tissues exhibiting positive CRAB staining were all ER negative samples, a class of tumors proposed to have a myoepithelial

cancers and normal tissues, and validated three other novel candidates, CRAB, 6PGL, and CAZ2, as bona fide differentially expressed proteins by immunohistochemistry on breast tissue microarrays. In total, close to half (4/9) of our protein biomarker candidates were successfully validated. It is worth noting that the bulk of the analytical pipeline prior to the final validation step did not require focusing on individual proteins. We thus believe that our approach is likely to be highly competitive compared to other biomarker identification strategies in terms of cost, throughput, and efficiency.

The initial step in our approach involved the delineation of a large set of differentially expressed proteins that were subsequently prioritized and refined. We elected to use cancer cell lines for this purpose, so as to minimize contributions from nontumor cellular compartments, and to maximize the ability of our candidate list to represent tumor-centric proteins. Indeed, the 2DE/MS analysis reported in this study represents the most comprehensive proteomic investigation of breast cancer cell lines ever reported using a 2DE/MS platform, dramatically exceeding the closest previous study where 162 proteins (including isoforms and variants) were identified.[24] Notably, when the differentially expressed cell line proteins were treated as a collective set, they did not exhibit a significant concordance with primary tumor gene expression data when analyzed by GSEA. This lack of global concordance, which persisted even after the cell lines proteins were filtered by cell line mRNA data, is likely due to the many differences between *in vitro* cultured cell lines and primary tumors. In our strategy, we thus deliberately introduced an extra filtering step where candidate biomarkers were further stratified by their concordance to primary tumor gene expression (see Figure 1). This resulted in nine biomarkers being finally chosen for final validation on the basis of their concordant gene expression patterns between cell lines and primary tumors. One potential criticism of our study might be its reliance of 2DE/MS, which has acknowledged limitations in detecting proteins that are lowly expressed (sensitivity) or associated with extreme p*I* and MW.[38] However, there is no conceptual reason why our approach should not be transportable to other newer generation proteomic technologies, such as shotgun-based proteomics[39] and stable isotope labeling proteomics,[17,40] so long as the candidate proteins are identified before the mRNA integration step. It is also important to note that, while we elected to use a particular set of analytical algorithms in our strategy, other methodologies can also be applied. For example, the gene ranking step in our study was performed using a Support Vector Machine, as this algorithm has been used in many other gene expression studies.[41] However, this does not rule out the use of simpler methods such as a Venn diagram analysis to perform the ranking. Future work should be focused on defining the most accurate and efficient set of analytical methods for this approach.

Of the nine biomarker candidates, four proteins including ANX1, CRAB, 6PGL, and CAZ2 successfully passed the TMA validation. ANX1 is a member of the annexin family of calcium and phospholipid-binding proteins, which are involved in diverse biological processes including signal transduction, mediation of apoptosis, and immunosuppression.[27] Interestingly, the exact expression status of ANX1 in breast cancer is controversial. While some groups have reported that ANX1 expression is undetectable in breast cancer cells, Ahn et al. recently reported that ANX1 was generally expressed in various types of breast cancers, including noninvasive ductal carcinoma *in situ* (DCIS), invasive and metastatic breast tumors.[29] The proteomic and gene expression results of our current study strongly suggests that ANX1 is down-regulated in breast cancer cells and only detected in the stromal cells of tumors, consistent with proposed models where ANX1 may play a critical role in maintaining normal breast biology.[42] Similar to ANX1, we also identified CRAB as a down-regulated protein in tumors. CRAB belongs to the small heat shock protein HSP20 family and is a well-known structural component of the eye lens.[43] A potential role for CRAB in cancer development has been suggested as it was previously found to be down-regulated in anaplastic thyroid carcinomas.[44] We found that CRAB was strongly expressed in normal breast tissues line as compared to the majority of breast tumors. Interestingly, we detected CRAB expression in five breast tumors that were ER negative. Recently, CRAB expression was reported to associate with the basal-like subtype of human breast tumor[31,45] which are typically ER negative. The potential role of CRAB in the development of ER negative breast cancer deserves to be further studied.

We identified 6PGL and CAZ2 as two proteins up-regulated in tumors. 6PGL is a well-known enzyme involved in glucose metabolism. Although there is no current evidence directly relating 6PGL to cancer, its high expression in tumors may contribute to the 'Warburg effect' where tumors undergo a metabolic shift from fermenting glucose to lactate even in the presence of oxygen (aerobic glycolysis).[46] 6PGL expression has been associated with pentose phosphate pathway (PPP) activity in several species,[47,48] and the PPP regulates glucose conversion to ribose for nucleic acid synthesis and glucose degradation to lactate. Interestingly, besides being simply a metabolic end-product, lactate has been proposed to directly mediate tumor invasion through the promotion of angiogenesis, the proteolytic cleavage of matrix proteins, and inhibition of the immune responses.[49] Finally, CAZ2 binds to the fast growing ends of actin filaments (barbed end) in the presence of Ca^{2+}, thereby blocking the exchange of subunits at these ends. CAZ2 has been reported to be amplified in glioblastomas,[50] and F-actin capping protein has been identified as one of the potential tumor associated antigens (TAA) in human renal cell carcinoma (RCC).[51] It is possible that CAZ2 may play a role in regulating tumor-specific aspects of cell motility.

It is also worth revisiting the five biomarker candidates that failed to pass the TMA validation. Four of these, including K2C7, LAM1, TKT, and PSD2, were initially identified as differentially expressed proteins in the proteo-transcriptomic pipeline but ultimately showed equally positive IHC staining on both tumor and normal tissues. There could be several reasons for this discrepancy. For example, although TKT was selected by proteo-transcriptomic analysis as an up-regulated caner biomarker, the TMA analysis revealed that its expression was equally strong in both tumor and normal cells. This could be due to differences between *in vitro* and *in vivo* growth. Alternatively, since breast tumors typically contain higher relative proportions of epithelial cells than normal controls that comprise predominantly of stromal tissue,[52] this difference in relative proportions might have resulted in relatively greater amounts of TKT protein/mRNA being extracted from tumor cells. Finally, while PAK2 protein was detected in cell lines, we did not identify PAK2 in a preliminary 2DE analysis of breast cancer tissues (data not shown), which is consistent with the negative staining results in the breast TMA. Thus, PAK2

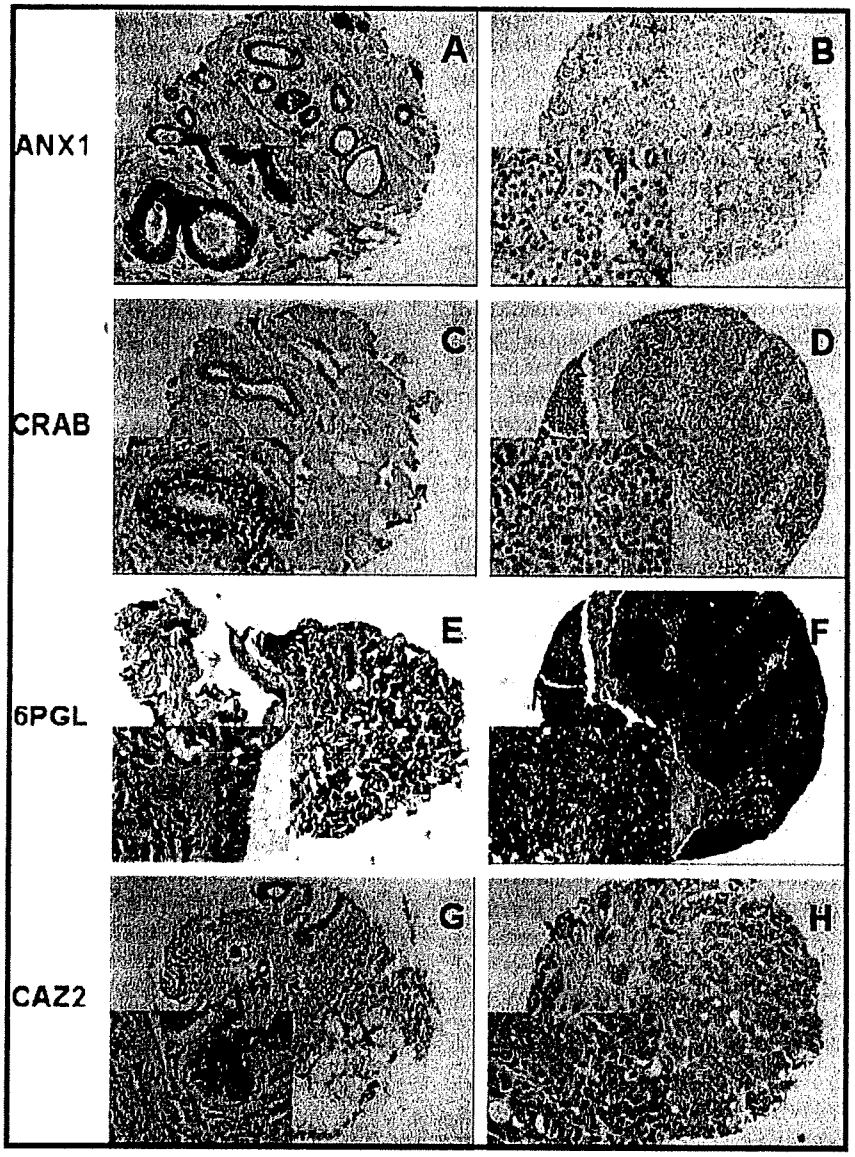

Figure 4. TMA analysis of breast cancer biomarker candidate expression. (A) Strong positive ANX1 staining in the myoepithelial cells of normal breast tissues, while (B) 88% of the breast cancer tissues show negative ANX1staining. (C) Strong positive CRAB staining in the myoepithelial cells of normal breast tissues, while (D) 96% of the breast cancer tissues show negative CRAB staining. (E) All of the normal breast tissues showed a score of 1 or 2 for 6PGL staining, while (F) 98% of the breast cancer tissues show strong (2 or 3) positive 6PGL staining. (G) Majority of the normal breast tissues showed negative CAZ2 staining, while (H) 69% of the breast cancer tissues show strong positive CAZ2 staining. Insets are zoom-in images highlighting the respective protein expression patterns in breast tissue cells.

origin.[31] Conversely, 6PGL was lowly expressed in all the normal breast tissues (intensity scores 1–2), while 98% of the breast cancer tissues exhibited strong positive 6PGL staining (scores 2–3) ($p = 1.7 \times 10^{-54}$). In the cancer cells, 6PGL was largely expressed in the cytoplasm, consistent with bioinformatic predictions. Finally, CAZ2 was expressed in 69% of the tumor tissue cores (58 out of 84) also in a predominantly cytoplasmic location. Specifically, 15.5% of the tumor cores showed score 3 staining, 16.7% showed score 2 staining, and 36.9% showed score 1 staining. In contrast, the majority of normal tissue cores showed negative staining, with only occasional samples showing score 1 positive staining (Figure 4G,H, $p = 2.6 \times 10^{-18}$).

Discussion

In this report, we used a proteo-transcriptomic integrative strategy for the rapid discovery of new cancer biomarkers. Mirroring the clinical reality that most currently approved biomarkers are likely to be protein-based, a major strength of our approach lies in the combining the direct visualization of differentially expressed proteins with the high-throughput scale of gene expression profiling. Although several studies have previously attempted to combine proteomic data with gene expression from bacteria,[32,33] yeast,[34] cell lines,[16,35] animals,[33] and humans,[36] comparatively fewer studies have attempted to utilize such integration efforts for the purposes of biomarker candidate prioritisation and discovery.[37] Using breast cancer as a case example, we combined 2DE/MS proteomic maps of cancer (MCF-7 and HCC-38) and control (CCD-1059Sk) cell lines with gene expression databases of cancer cell lines and primary breast tumors to identify nine selected biomarker candidates. From these, we reidentified ANX1, a protein previously reported to be differentially expressed in breast

expression may be induced during *in vitro* growth and may only be present in certain breast cell lines.

In conclusion, we have in this report used a novel integrated proteo-transcriptomic analytical pipeline to successfully reconfirm one biomarker and to identify three novel potential biomarkers for breast cancer. We acknowledge there are some limitations in the current work. First, although our 67% protein identification rate is generally acceptable, several highly ranked protein spots were excluded from the analysis because they were not identified. Thus, improving the protein identification rate by, for example, implementing MS/MS peptide sequencing, would definitely enhance the process of biomarker selection. Second, only three cell lines were used in this proof-of-concept study, and a larger number of cell lines will undoubtedly be required to accurately capture the molecular heterogeneity of breast cancer. We believe that all are readily solvable deficiencies. Thus, with fine-tuning, this versatile strategy may also prove valuable for biomarker identification in other cancer types for either molecular diagnostics or therapeutic applications.

Abbreviations: ER, estrogen receptor; GSEA, Gene Set Enrichment Analysis; SVM, Support Vector Machine; TMA, tissue microarrays; IHC, immunohistochemical analysis.

Acknowledgment. The authors thank Dr. Yonghui Wu for advice on biostatistical analysis. This work was funded by a grant to P.T. from Agenica Research, and grants from the Singapore Cancer Syndicate to M.S-.T. (MN005 and MN-077).

Supporting Information Available: Supporting Information Figure 1, 2DE gel mage analysis of MCF-7, HCC-38 and CCD-1059Sk cells (pH 4–7) using the PDQuest 7.3 software. Three replicates were made. Supporting Information Figure 2, 2DE gel mage analysis of MCF-7, HCC-38 and CCD-1059Sk cells (pH 6–9) using the PDQuest 7.3 software. Four replicates were made. Supporting Information Figure 3, annotated 2DE map of MCF-7 pH4–7. Supporting Information Figure 4, annotated 2DE map of MCF-7 pH6–9. Supporting Information Figure 5, annotated 2DE map of HCC-38 pH4–7. Supporting Information Figure 6, annotated 2DE map of HCC-38 pH6–9. Supporting Information Figure 7, annotated 2DE map of CCD-1059Sk pH4–7. Supporting Information Figure 8, annotated 2DE map of CCD-1059Sk pH6–9. Supporting Information Table 1, a complete list of proteins identified from 3 breast cell lines (MCF-7, HCC-38 and CCD-1059Sk), which are grouped into 13 catalogues based on their functions. Supporting Information Table 2, the list of 110 unique proteins short-listed from CCD-1059Sk vs MCF-7 comparison. Supporting Information Table 3, the list of 112 unique proteins short-listed from CCD-1059Sk vs HCC-38 comparison. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Chia, K. S.; Du, W. B.; Sankaranarayanan, R.; Sankila, R.; Wang, H.; Lee, J.; Seow, A.; Lee, H. P. *Int. J. Cancer* **2004**, *108*, 761–765.
(2) Parkin, D. M. *Lancet Oncol.* **2001**, *2*, 533–543.
(3) Gerrero, M. R.; Weber, B. L. *Curr. Opin. Oncol.* **2001**, *13*, 415–419.
(4) Pole, J. C.; Gold, L. I.; Orton, T.; Huby, R.; Carmichael, P. L. *Toxicology* **2005**, *206*, 91–109.
(5) Wulfkuhle, J. D.; Liotta, L. A.; Petricoin, E. F. *Nat. Rev. Cancer* **2003**, *3*, 267–275.
(6) Rai, A. J.; Zhang, Z.; Rosenzweig, J.; Shih, I.; Pham, T.; Fung, E. T.; Sokoll, L. J.; Chan, D. W. *Arch. Pathol. Lab. Med.* **2002**, *126*, 1518–1526.
(7) Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I.; Williams, K. L.; Hochstrasser, D. F. *Biotechnology (N.Y.)* **1996**, *14*, 61–65.
(8) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198–207.
(9) Verma, M.; Wright, G. L., Jr.; Hanash, S. M.; Gopal-Srivastava, R.; Srivastava, S. *Ann. N.Y. Acad. Sci.* **2001**, *945*, 103–115.
(10) Hondermarck, H. *Mol. Cell. Proteomics* **2003**, *2*, 281–291.
(11) Yu, K.; Lee, C. H.; Tan, P. H.; Hong, G. S.; Wee, S. B.; Wong, C. Y.; Tan, P. *Cancer Res.* **2004**, *64*, 2962–2968.
(12) Yu, K.; Lee, C. H.; Tan, P. H.; Tan, P. *Clin. Cancer Res.* **2004**, *10*, 5508–5517.
(13) Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R. *Mol. Cell. Biol.* **1999**, *19*, 1720–1730.
(14) Chen, G.; Gharib, T. G.; Huang, C. C.; Taylor, J. M.; Misek, D. E.; Kardia, S. L.; Giordano, T. J.; Iannettoni, M. D.; Orringer, M. B.; Hanash, S. M.; Beer, D. G. *Mol. Cell. Proteomics* **2002**, *1*, 304–313.
(15) Pradet-Balade, B.; Boulme, F.; Beug, H.; Mullner, E. W.; Garcia-Sanz, J. A. *Trends Biochem. Sci.* **2001**, *26*, 225–229.
(16) Tian, Q.; Stepaniants, S. B.; Mao, M.; Weng, L.; Feetham, M. C.; Doyle, M. J.; Yi, E. C.; Dai, H.; Thorsson, V.; Eng, J.; Goodlett, D.; Berger, J. P.; Gunter, B.; Linseley, P. S.; Stoughton, R. B.; Aebersold, R.; Collins, S. J.; Hanlon, W. A.; Hood, L. E. *Mol.Cell. Proteomics* **2004**, *3*, 960–969.
(17) Ou, K.; Kesuma, D.; Ganesan, K.; Yu, K.; Soon, S. Y.; Lee, S. Y.; Goh, X. P.; Hooi, M.; Chen, W.; Jikuya, H.; Ichikawa, T.; Kuyama, H.; Matsuo, E.; Nishimura, O.; Tan, P. *J. Proteome Res.* **2006**, *5*, 2194–2206.
(18) Jiang, Z.; Gentleman, R. *Bioinformatics* **2007**, *23*, 306–313.
(19) Agranoff, D.; Fernandez-Reyes, D.; Papadopoulos, M. C.; Rojas, S. A.; Herbster, M.; Loosemore, A.; Tarelli, E.; Sheldon, J.; Schwenk, A.; Pollok, R.; Rayner, C. F.; Krishna, S. *Lancet* **2006**, *368*, 1012–1021.
(20) Salto-Tellez, M.; Lee, S. C.; Chiu, L. L.; Lee, C. K.; Yong, M. C.; Koay, E. S. *Clin. Chem.* **2004**, *50*, 1082–1086.
(21) Zhang, D.; Salto-Tellez, M.; Putti, T. C.; Do, E.; Koay, E. S. *Mod. Pathol.* **2003**, *16*, 79–84.
(22) Chen, W.; Salto-Tellez, M.; Palanisamy, N.; Ganesan, K.; Hou, Q.; Tan, L. K.; Sii, L. H.; Ito, K.; Tan, B.; Wu, J.; Tay, A.; Tan, K. C.; Ang, E.; Tan, B. K.; Tan, P. H.; Ito, Y.; Tan, P. *Genes, Chromosomes Cancer* **2007**, *46*, 288–301.
(23) Nakai, K.; Horton, P. *Trends Biochem. Sci.* **1999**, *24*, 34–36.
(24) Pucci-Minafra, I.; Cancemi, P.; Fontana, S.; Minafra, L.; Feo, S.; Becchi, M.; Freyria, A. M.; Minafra, S. *Proteomics* **2006**, *6*, 2609–2625.
(25) Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15545–15550.
(26) Shen, D.; Chang, H. R.; Chen, Z.; He, J.; Lonsberry, V.; Elshimali, Y.; Chia, D.; Seligson, D.; Goodglick, L.; Nelson, S. F.; Gornbein, J. A. *Biochem. Biophys. Res. Commun.* **2005**, *326*, 218–227.
(27) Vishwanatha, J. K.; Salazar, E.; Gopalakrishnan, V. K. *BMC Cancer* **2004**, *4*, 8.
(28) Pencil, S. D.; Toth, M. *Clin. Exp. Metastasis* **1998**, *16*, 113–121.
(29) Ahn, S. H.; Sawada, H.; Ro, J. Y.; Nicolson, G. L. *Clin. Exp. Metastasis* **1997**, *15*, 151–156.
(30) Schwartz-Albiez, R.; Koretz, K.; Moller, P.; Wirl, G. *Differentiation* **1993**, *52*, 229–237.
(31) Moyano, J. V.; Evans, J. R.; Chen, F.; Lu, M.; Werner, M. E.; Yehiely, F.; Diaz, L. K.; Turbin, D.; Karaca, G.; Wiley, E.; Nielsen, T. O.; Perou, C. M.; Cryns, V. L. *J. Clin. Invest.* **2006**, *116*, 261–270.
(32) Ou, K.; Ong, C.; Koh, S. Y.; Rodrigues, F.; Sim, S. H.; Wong, D.; Ooi, C. H.; Ng, K. C.; Jikuya, H.; Yau, C. C.; Soon, S. Y.; Kesuma, D.; Lee, M. A.; Tan, P. *J. Bacteriol.* **2005**, *187*, 4276–4285.
(33) Naranjo, V.; Villar, M.; Martin-Hernando, M. P.; Vidal, D.; Hofle, U.; Gortazar, C.; Kocan, K. M.; Vazquez, J.; de la, F. J. *Proteomics* **2007**, *7*, 220–231.
(34) Washburn, M. P.; Koller, A.; Oshiro, G.; Ulaszek, R. R.; Plouffe, D.; Deciu, C.; Winzeler, E.; Yates, J. R., III *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3107–3112.
(35) Chen, Y. R.; Juan, H. F.; Huang, H. C.; Huang, H. H.; Lee, Y. J.; Liao, M. Y.; Tseng, C. W.; Lin, L. L.; Chen, J. Y.; Wang, M. J.; Chen, J. H.; Chen, Y. J. *J. Proteome Res.* **2006**, *5*, 2727–2742.
(36) Seshi, B. *Proteomics* **2006**, *6*, 5169–5182.
(37) Varambally, S.; Yu, J.; Laxman, B.; Rhodes, D. R.; Mehra, R.; Tomlins, S. A.; Shah, R. B.; Chandran, U.; Monzon, F. A.; Becich, M. J.; Wei, J. T.; Pienta, K. J.; Ghosh, D.; Rubin, M. A.; Chinnaiyan, A. M. *Cancer Cell* **2005**, *8*, 393–406.
(38) Young, R. A. *Cell* **2000**, *102*, 9–15.

(39) Malmstrom, J.; Lee, H.; Aebersold, R. *Curr. Opin. Biotechnol.* **2007**, *18*, 378–384.
(40) Smolka, M.; Zhou, H.; Aebersold, R. *Mol. Cell. Proteomics* **2002**, *1*, 19–29.
(41) Ramaswamy, S.; Tamayo, P.; Rifkin, R.; Mukherjee, S.; Yeang, C. H.; Angelo, M.; Ladd, C.; Reich, M.; Latulippe, E.; Mesirov, J. P.; Poggio, T.; Gerald, W.; Loda, M.; Lander, E. S.; Golub, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 15149–54.
(42) Shen, D.; Chang, H. R.; Chen, Z.; He, J.; Lonsberry, V.; Elshimali, Y.; Chia, D.; Seligson, D.; Goodglick, L.; Nelson, S. F.; Gornbein, J. A. *Biochem. Biophys. Res. Commun.* **2005**, *326*, 218–227.
(43) Horwitz, J. *Exp. Eye Res.* **2003**, *76*, 145–153.
(44) Mineva, I.; Gartner, W.; Hauser, P.; Kainz, A.; Loffler, M.; Wolf, G.; Oberbauer, R.; Weissel, M.; Wagner, L. *Cell Stress Chaperones* **2005**, *10*, 171–184.
(45) Gruvberger-Saal, S. K.; Parsons, R. *J. Clin. Invest.* **2006**, *116*, 30–32.
(46) Stern, R.; Shuster, S.; Neudecker, B. A.; Formby, B. *Exp. Cell Res.* **2002**, *276*, 24–31.
(47) Zimenkov, D.; Gulevich, A.; Skorokhodova, A.; Biriukova, I.; Kozlov, Y.; Mashko, S. *FEMS Microbiol. Lett.* **2005**, *244*, 275–280.
(48) Clarke, J. L.; Scopes, D. A.; Sodeinde, O.; Mason, P. J. *Eur. J. Biochem.* **2001**, *268*, 2013–2019.
(49) Gatenby, R. A.; Gawlinski, E. T. *Cancer Res.* **2003**, *63*, 3847–3854.
(50) Mueller, H. W.; Michel, A.; Heckel, D.; Fischer, U.; Tonnes, M.; Tsui, L. C.; Scherer, S.; Zang, K. D.; Meese, E. *Hum. Genet.* **1997**, *101*, 190–197.
(51) Kellner, R.; Lichtenfels, R.; Atkins, D.; Bukur, J.; Ackermann, A.; Beck, J.; Brenner, W.; Melchior, S.; Seliger, B. *Proteomics* **2002**, *2*, 1743–1751.
(52) Hawes, D.; Downey, S.; Pearce, C. L.; Bartow, S.; Wan, P.; Pike, M. C.; Wu, A. H. *Breast Cancer Res.* **2006**, *8*, R24.

PR700820G

Analytical Biochemistry 380 (2008) 291–296

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

A method for N-terminal *de novo* sequence analysis of proteins by matrix-assisted laser desorption/ionization mass spectrometry

Hiroki Kuyama [a,*], Kazuhiro Sonomura [b], Osamu Nishimura [a,b], Susumu Tsunasawa [a]

[a] *Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan*
[b] *Life Science Laboratory, Shimadzu Corporation, Kyoto 604-8511, Japan*

ARTICLE INFO

Article history:
Received 16 April 2008
Available online 7 June 2008

Keywords:
N-terminal sequence analysis
Mass spectrometry
N-terminal modification
Succinimidyloxycarbonylmethyl tris(2,4,6-trimethoxyphenyl)phosphonium bromide

ABSTRACT

A novel method for isolation and *de novo* sequencing of N-terminal peptides from proteins is described. The method presented here combines selective chemical tagging using succinimidyloxycarbonylmethyl tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP-Ac-OSu) at the N^{α}-amino group of peptides after digestion by metalloendopeptidase (from *Grifola frondosa*) and selective capture procedures using *p*-phenylenediisothiocyanate resin, by which the N-terminal peptide can be isolated, whether or not it is N-terminally blocked. The isolated N-terminal peptide modified N-terminally with TMPP-Ac-OSu reagent produces a simple fragmentation pattern under tandem mass spectrometric analysis to significantly facilitate sequencing.

Analyzing amino acid sequences of proteins is highly important in the life sciences because the covalent structures are responsible for the biological functions of individual proteins. A "bottom-up" (or "shotgun") approach is generally used to determine covalent structures of proteins [1]. This usually involves peptide mass fingerprinting (PMF)[1] [2–5] or MS/MS ion search [6,7]. However, a mature protein formed under particular physiological conditions in cells or organs is usually altered from that expected from the DNA sequence alone because of processes such as splicing and shuffling of the gene or mRNA, and/or the various posttranslational modifications (PTMs) that occur in newly expressed proteins [8]. Therefore, widely used methods such as PMF and MS/MS ion search often fail to identify posttranslationally modified proteins or to detect sequence polymorphisms observed in various protein isoforms because the database is insufficiently comprehensive for an effective search for any type of protein.

The superiority of *de novo* sequencing using N- and C-terminal peptides ("terminal proteomics") has been stressed because of the global applicability of sequencing peptides to structural determination and/or protein identification [9]. One of the most advantageous features of terminal sequencing is that as few as four residues from an N-terminal amino acid sequence have proven sufficient for specifying 43 to 83% of proteins, and those from the C-terminal counterpart raise the success rate to 74 to 97% of proteins [10]. Hence, *de novo* sequencing of terminal parts of a protein affords much higher fidelity and throughput.

Recently we developed a method for specific isolation of C-terminal peptides from proteins and their *de novo* sequencing [11]. This method is as follows: (1) proteolytic digestion of a protein with lysyl endopeptidase (LysC) to produce peptides having α- and ε-amino groups except for the C-terminal peptide, which has only an α-amino group at its N-terminus; (2) modification of the N^{α}-amino groups of resulting peptides with TMPP-Ac-OSu [12,13] to yield TMPP-Ac peptides in which the C-terminal peptide incorporates no free amino group; (3) binding of the resulting peptides with a free ε-amino group to *p*-phenylenediisocyanate resin (DITC resin) to produce an unreacted C-terminal peptide that remains intact in the supernatant of the reaction mixture. The C-terminal peptide obtained is then sequenced *de novo* by tandem mass spectrometry.

While studying the C-terminal sequencing, we noticed that the use of *Grifola frondosa* metalloendopeptidase (LysN) [14–16] instead of LysC enables N-terminal-specific isolation, regardless of whether the N-terminal amino group is blocked or not. The TMPP-Ac tag is reported to facilitate sequence analysis by simplifying fragmentation under MS/MS analysis [12,13]. Therefore, *de novo* sequencing of the isolated N-terminal peptide can be successfully performed for a protein with a nonblocked N-terminus. As a

* Corresponding author. Fax: +81 6 6879 4320.
E-mail address: kuyama@protein.osaka-u.ac.jp (H. Kuyama).

[1] *Abbreviations used:* PMF, peptide mass fingerprinting; PTM, posttranslational modification; LysC, lysyl endopeptidase; LysN, metalloendopeptidase; TMPP-Ac-OSu, succinimidyloxycarbonylmethyl tris(2,4,6-trimethoxyphenyl)phosphonium bromide; DITC, *p*-phenylenediisothiocyanate; TCEP, tris(2-carboxyethyl)phosphine hydrochloride; TFA, trifluoroacetic acid; CHCA, α-cyano-4-hydroxycinnamic acid; MDPNA, methanediphosphonic acid; CID, collision-induced dissociation; MALDI, matrix-assisted laser desorption/ionization; ToF, time of flight; MS/MS, tandem mass spectrometry.

0003-2697/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2008.05.053

whole, this method can facilitate N-terminal sequence analysis by tandem mass spectrometry.

In this report, we describe a new method for isolation and *de novo* sequence analysis of N-terminal peptides from proteins.

Materials and methods

Materials

Bovine α-lactalbumin, bovine carbonic anhydrase II, human hemoglobin, and iodoacetamide were obtained from Sigma (St. Louis, MO, USA). Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and succinimidyloxycarbonylmethyl tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP-Ac-OSu) were obtained from Fluka (Switzerland). Metalloendopeptidase (from *Grifola frondosa*: LysN) was purchased from Seikagaku Corporation (Tokyo, Japan). Sodium hydrogen carbonate ($NaHCO_3$), acetonitrile, 2-propanol, and trifluoroacetic acid (TFA) were purchased from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). α-Cyano-4-hydroxycinnamic acid (CHCA: high-purity mass spectrometric grade) was obtained from Shimadzu GLC (Tokyo, Japan). *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. Methanediphosphonic acid (MDPNA) was obtained from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan). Water used in all experiments was purified using a MilliQ water purification system. All other chemicals were of analytical reagent grade and were used without further purification.

Methods

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) acetonitrile in 100 mM $NaHCO_3$. The washed gel piece was dehydrated with acetonitrile and dried in a vacuum centrifuge. To the dried gel was added 100 μL of 10 mM aqueous TCEP solution to reduce disulfide bonds. This solution was incubated for 30 min at 37 °C. S-alkylation was performed by replacing the TCEP solution with 55 mM iodoacetamide in 100 mM $NaHCO_3$. After a 45-min incubation at room temperature in the dark, the gel piece was washed with 100 μL of 50 mM $NaHCO_3$, dehydrated to shrink in acetonitrile, and dried in a vacuum centrifuge. The gel piece was then rehydrated with 2 μL of acetonitrile–50 mM $NaHCO_3$ (1:9) containing 10 ng of LysN. After 5 min, 50 mM $NaHCO_3$ solution (15 μL) was added to keep the gel piece moist during digestion (room temperature, overnight). To extract the resulting peptides, 30 μL of 50% acetonitrile 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 μL of acetonitrile–50 mM $NaHCO_3$ (1:9). To this solution was added 1 μL of TMPP-Ac-OSu solution (10 mM in acetonitrile–water = 1:4), and the mixture was sonicated in a water bath for 30 min. The TMPP-Ac-modified peptide solution was added to the prewashed DITC resin (5 mg), and this was allowed to stand for 2 h in a water bath at 60 °C. The extraction was done using acetonitrile–50 mM $NaHCO_3$ (1:9, 60 μL; twice) and 2-propanol–acetonitrile–0.1%TFA (1:1:2, 60 μL; three times). The extracts were combined and dried in a vacuum centrifuge.

MALDI-ToF MS

MALDI mass spectra were recorded on Axima CFR-plus or Axima TOF^2 (Shimadzu/Kratos, Manchester UK) reflectron time-of-flight 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 at 20 kV, and the reflectron detector was operated at 24 kV. The flight path in the reflectron mode is about 240 cm for both instruments. For MS/MS experiments, collision-induced dissociation (CID) was carried out using helium at a pressure of ca. 5×10^{-6} mbar in the collision cell.

CHCA was used as a matrix, which was dissolved to saturation in 50% aqueous acetonitrile containing 0.05% TFA. We used MDPNA, which has been proven to be useful for MALDI analysis of salt-containing samples, as a matrix additive [17]. MDPNA was used as a 1 to 2% aqueous solution. An aliquot (0.4 μL) of sample solution was mixed with an equivalent volume of matrix solution and matrix additive solution on the MALDI target plate and analyzed after drying.

m/z values in the spectra were externally calibrated with angiotensin II (human) and ACTH fragment 18–39 (human) using CHCA as a matrix.

Results and discussion

Several methods for selectively isolating the N-terminal peptide followed by sequencing by tandem mass spectrometry to determine N-terminal amino acid sequences of proteins with high throughput and high fidelity have been reported [18–26]. Our group has been developing methodologies for "terminal proteomics" [9], and we recently developed a method for sequencing N-terminal peptides that is based on a concept different from that of the method described in this report and will be published elsewhere.

In our previous report of C-terminal isolation/tandem mass analysis [11], we used tagging into an N^{α}-amino group (with TMPP-Ac-OSu reagent) and depletion of the reactive amino group-containing peptides (using DITC resin or glass). The selectivity of the TMPP-Ac-tagging reaction to the N^{α}-amino group is essential to the method. This is achieved by adjusting the reaction pH to 8.2 [13]. In the method for N-terminal specific isolation followed by MS/MS analysis described in this article, we employed a similar procedure except for digestion with LysN instead of LysC. Fig. 1 outlines the procedure. A protein sample is first digested with LysN to yield peptides incorporating a lysine residue at their N-termini except for the N-terminal peptide, which possesses no lysine residue. If a protein is N-terminally blocked by any PTM, the N-terminal peptide has no amino group in its sequence. The next step is to modify their N^{α}-amino groups with TMPP-Ac-OSu to produce TMPP-Ac peptides. The resulting modified peptides have one more active amino group (ε-amino group) in the lysine residue, except for the N-terminal peptide, which has no active amino group after TMPP-Ac modification. In the final step, peptides with free amino groups are depleted using DITC resin (or DITC glass [27]), and the N-terminal peptide, whether N-terminally blocked or not, is thus recovered in the supernatant solution, which is then subjected to *de novo* sequencing analysis by tandem mass spectrometry.

The results obtained are described using three model proteins (bovine α-lactalbumin, human hemoglobin, and bovine carbonic anhydrase II). For each protein, we used excised gel pieces after purification by SDS–PAGE.

Fig. 2 presents MALDI-ToF mass spectra of the LysN digest after TMPP-Ac modification (Figs. 2a, c, and e) and the isolated N-terminal peptides (Figs. 2b, d, and f) of the three proteins used. Figs. 2a and b illustrate the results for bovine α-lactalbumin. The N-terminal peptide (TMPP-Ac-EQLT) was singly recovered after depletion of undesired amino group-containing peptides using DITC resin

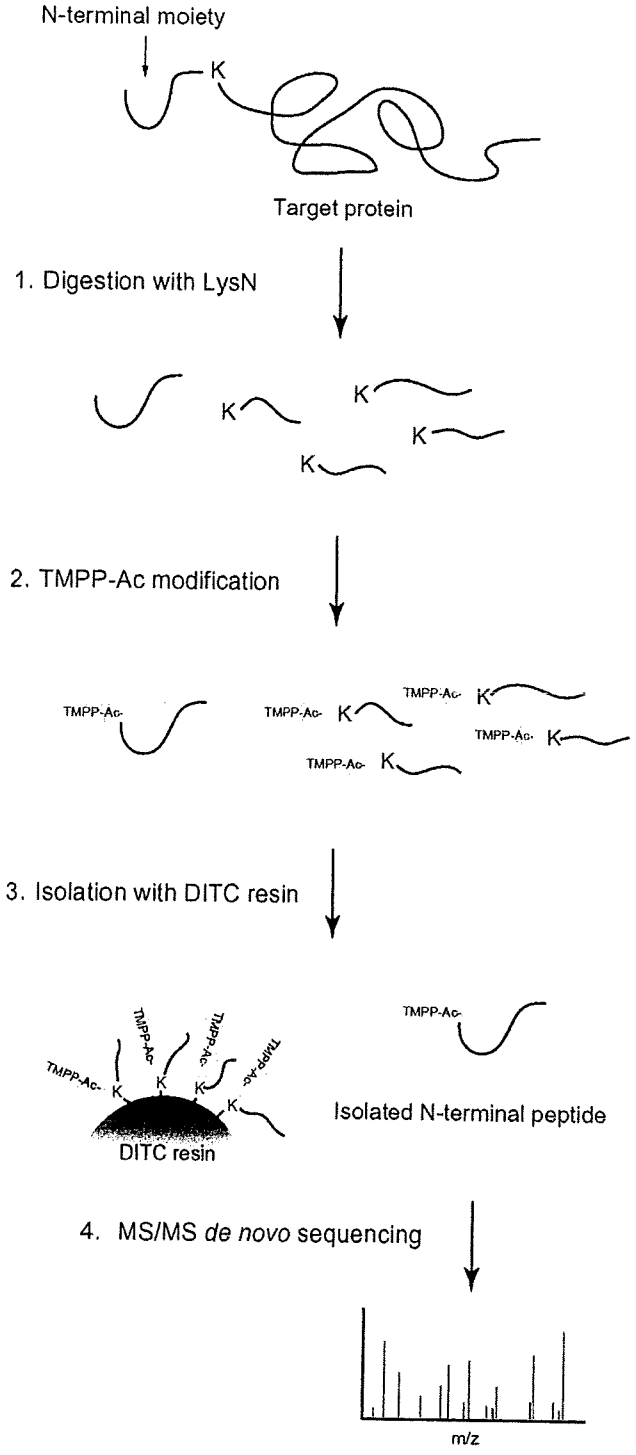

Fig. 1. Outline of procedure. A protein is first digested with LysN to yield peptide fragments incorporating α- and ε-amino groups except for the N-terminal peptide having only an α-amino group. This is followed by selective modification of α-amino groups with TMPP-Ac-OSu and isolation of the N-terminal peptide using a DITC resin.

(Fig. 2b). Figs. 2c and d illustrate the results for human hemoglobin. Human hemoglobin comprises two subunits (α and β). Therefore, two N-terminal peptides (α: TMPP-Ac-VLSPAD,β: TMPP-Ac-VHLTPEE) were cleanly recovered (Fig. 2d). Figs. 2e and f show the results for bovine carbonic anhydrase II. The N-terminal peptide incorporating an acetyl group on its N-terminal amino group (Ac-SHHWGYG) was isolated after depletion of the amino group-containing peptides with DITC resin (Fig. 2f). In this case, an N-terminally blocked peptide does not have the benefits of the TMPP-Ac tag. For blocked proteins, a simpler method has been shown to work [26].

TMPP-Ac-modified peptides give simple fragmentation patterns biasing strongly toward producing a-type ions [12,13]. Fig. 3a is the CID spectrum of the N-terminal peptide of bovine α-lactalbumin. A simplified pattern was obtained to show all *a*-type ions (a_1, a_2, a_3, and a_4). In addition, a few *d*-type ions (d_1, d_2, d_3, and d_4) are clearly discernible and helpful for sequencing. Figs. 3b and c depict CID spectra of N-terminal peptides from α and β subunits of human hemoglobin. Both spectra display all *a*-type ions and some *d*-type ions (d_2, d_3, and d_6 for the N-terminal peptide from the α subunit of hemoglobin; d_1, d_3, d_4, and d_6 for the N-terminal peptide from the β subunit of hemoglobin).

Mass spectrometry sometimes cannot discriminate between leucine and isoleucine. However, in a TMPP-Ac-modified peptide, loss of an isopropyl group (−42) is often observed for leucine residues, whereas loss of an ethyl group (−28) is observed for isoleucine residues [12,13]. Each of the three N-terminal peptides from α-lactalbumin and hemoglobin contains one leucine residue, which was differentiated from its isobaric counterpart using the *d*-type ion indicating loss of the isopropyl group (−42) (d_3 in Fig. 3a, d_2 in Fig. 3b, and d_3 in Fig. 3c). Thus d_i ions help to differentiate isobaric amino acids on the peptide chain.

Glutamine and lysine residues have very close masses (monoisotopic mass: 128.058 for glutamine residue and 128.094 for lysine residue). Therefore, differentiation by mass spectrometry is generally difficult. However, this is not problematic because this method does not incorporate lysine residues into the isolated N-terminal peptide.

LysN from *Grifora frondosa* is known to cleave specifically peptidyl-lysine bonds, but certain X-Lys bonds (where X is an acidic residue or proline) are difficult to cleave [15]. In this study, however, we observed cleavage of such X-Lys bonds in the two N-terminal peptides from human hemoglobin: TMPP-Ac-VLSPAD(K) from the α subunit and TMPP-Ac-VHLTPEE(K) from the β subunit.

The amino group-containing peptides were depleted using DITC resin. In principle, depletion can be performed with any kind of amine scavenger. In this study, we tested a couple of commercially available polymer-bound isothiocyanates as well as DITC glass prepared in accordance with the study by Wachter et al. [27]. For in-house preparation of DITC glass, aminopropyl glass (pore size 17 nm, 200–400 mesh, amine content: 162 μmol/g) was purchased from Sigma. Details on the DITC resin were not disclosed by Shimadzu. The DITC resin (Shimadzu) and the in-house-prepared DITC glass gave better results in that background peaks were not discernible over the *m/z* 600 range.

Although lysine residues are equally distributed across all positions of proteins [26], we sometimes have the problem that an isolated N-terminal peptide fragment is too big for sequence analysis by mass spectrometry. One solution to such a problem is subdigestion of the large fragment with other proteases (such as GluC, trypsin, and chymotrypsin), which may yield an appropriate fragment length for mass analysis. Alternatively, the sequence of the N-terminal part of a protein can easily be estimated using the observed *m/z* value of the isolated N-terminal peptide because the full-length sequence of the protein can be retrieved by database searches such as MS/MS ion search or PMF.

In a protein sequence, the lysine residue can be N-terminally positioned and its ε-amino group is free. In this case, every peptide fragment, including N-terminal peptide, is bound to DITC resin and totally depleted from the supernatant solution, which eliminates peptide-derived signals in the mass spectrum. Therefore, when

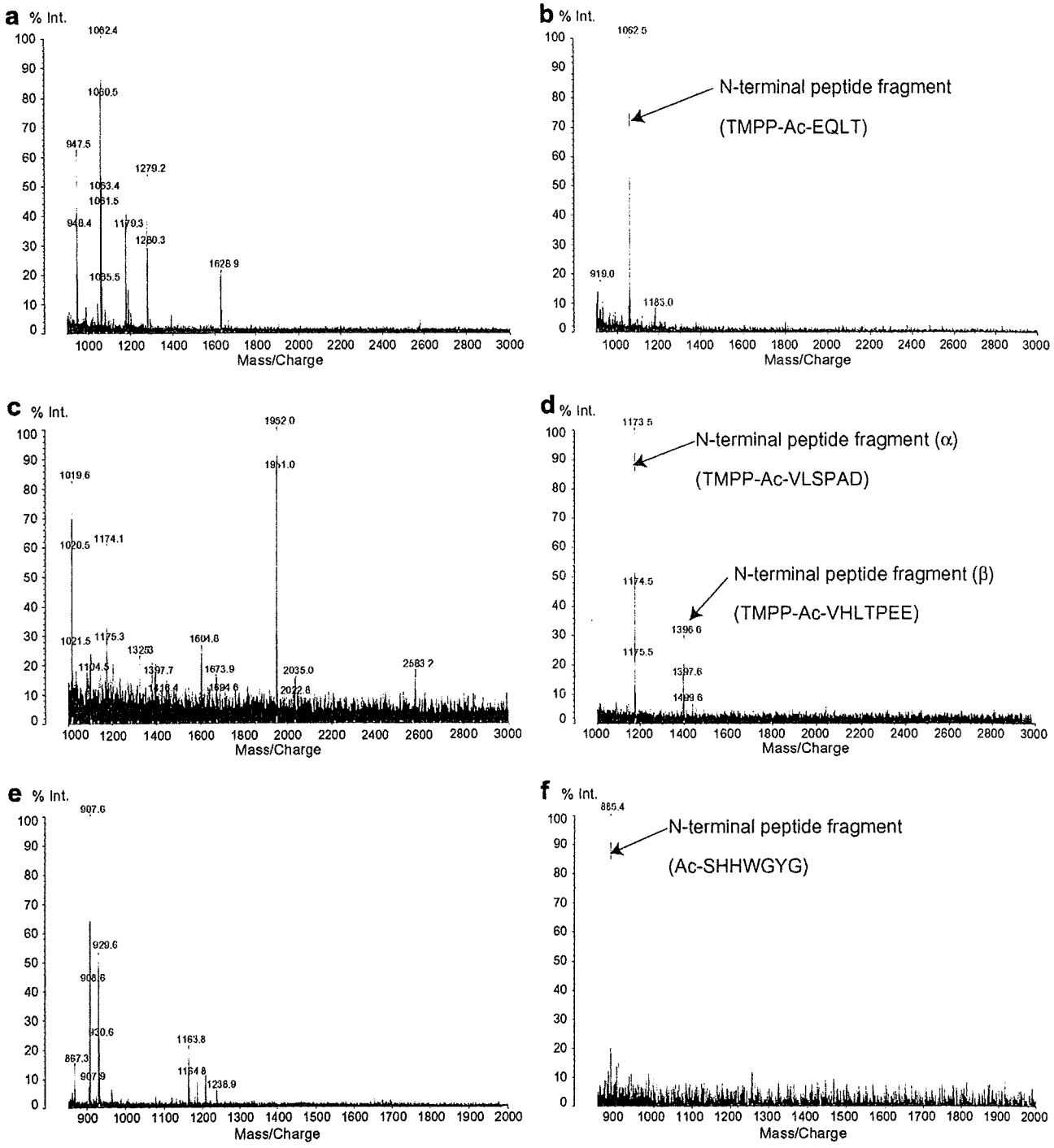

Fig. 2. MALDI-ToF mass spectra of LysN digest after TMPP-Ac modification (a, c, e) and the isolated N-terminal peptides (b, d, f) of the three proteins used: bovine α-lactalbumin (a and b), human hemoglobin (c and d), bovine carbonic anhydrase II (e and f). Each arrowed signal corresponds to the mass calculated from the already reported N-terminal sequence with TMPP-Ac modification.

any peptide signal is not observed in a mass spectrum, it can be assumed that the protein has a lysine residue at its N-terminus (and its ε-amino group is free), and the N-terminal sequence can be estimated from a readily obtained full-length sequence of the protein as described above.

As a pretreatment of sample solution for mass analysis, a micro-desalting approach is generally employed using ZipTip-type pipet tips. However, we have sometimes experienced nonrecovery of TMPP-Ac peptides using this type of tip. On the other hand, alkali metal adduct signals were effectively eliminated from mass spectra when MDPNA was used as a matrix additive in the sensitive detection of phosphopeptides [17]. Hence, we used MDPNA, instead of a micro-desalting approach, to obtain better-quality mass spectra.

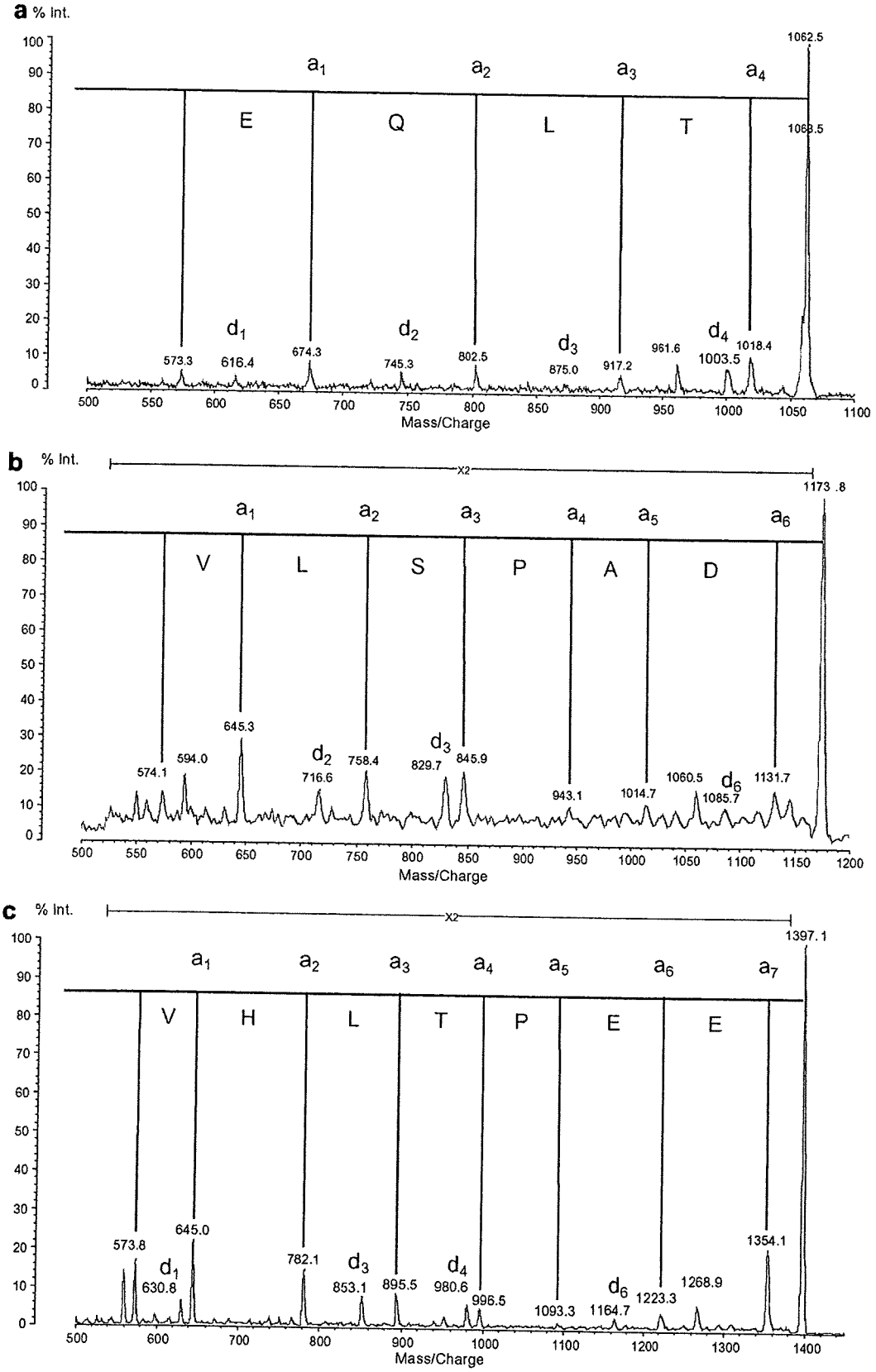

Fig. 3. CID spectra of the isolated N-terminal peptide fragments from bovine α-lactalbumin (a) and human hemoglobin (b: α subunit, c: β subunit).

In summary, we have described a simple alternative method for the isolation and *de novo* sequencing of N-terminal peptides from proteins.

References

[1] B.T. Chait, Mass spectrometry: bottom-up or top-down?, Science 314 (2006) 65–66

[2] W.J. Henzel, T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, C. Watanabe, Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases, Proc. Natl. Acad. Sci. USA 90 (1993) 5011–5015.

[3] M. Mann, P. Hojrup, P. Roepstorff, Use of mass spectrometric molecular weight information to identify proteins in sequence databases, Biol. Mass Spectrom. 22 (1993) 338–345.

[4] P. James, M. Quadroni, E. Carafoli, G. Gonnet, Protein identification by mass profile fingerprinting, Biochem. Biophys. Res. Commun. 195 (1993) 58–64.

[5] J.R. Yates, S. Speicher, P.R. Griffin, T. Hunkapiller, Peptide mass maps: a highly informative approach to protein identification, Anal. Biochem. 214 (1993) 397–408.

[6] J.K. Eng, A.L. McCormack, J.R. Yates, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, J. Am. Soc. Mass Spectrom. 5 (1994) 976–989.

[7] M. Mann, M. Wilm, Error-tolerant identification of peptides in sequence databases by peptide sequence tags, Anal. Chem. 66 (1994) 4390–4399.

[8] The FANTOM Consortium and RIKEN Genome Exploration Research Group and Genome Science Group, The transcriptional landscape of the mammalian genome, Science 309 (2005) 1559–1563.

[9] T. Nakazawa, M. Yamaguchi, T. Okamura, E. Ando, O. Nishimura, S. Tsunasawa, Terminal proteomics: N- and C-terminal analyses for high-fidelity identification of proteins using MS, Proteomics 8 (2008) 673–685.

[10] M.R. Wilkins, E. Gasteiger, L. Tonella, K. Ou, M. Tyler, J.C. Sanchez, A.A. Gooley, B.J. Walsh, A. Bairoch, R.D. Appel, K.L. Williams, D.F. Hochstrasser, Protein identification with N and C-terminal sequence tags in proteome projects, J. Mol. Biol. 278 (1998) 599–608.

[11] H. Kuyama, K. Shima, K. Sonomura, M. Yamaguchi, E. Ando, O. Nishimura, S. Tsunasawa, A simple and highly successful C-terminal sequence analysis of proteins by mass spectrometry, Proteomics 8 (2008) 1539–1550.

[12] Z.H. Huang, J. Wu, K.D.W. Roth, Y. Yang, D.A. Gage, J.T. Watson, A picomole-scale method for charge derivatization of peptides for sequence analysis by mass spectrometry, Anal. Chem. 69 (1997) 137–144.

[13] Z.H. Huang, T. Shen, J. Wu, D.A. Gage, J.T. Watson, Protein sequencing by matrix-assisted laser desorption ionization–postsource decay–mass spectrometry analysis of the N-tris(2,4,6-trimethoxyphenyl) phosphine-acetylated tryptic digests, Anal. Biochem. 268 (1999) 305–317.

[14] T. Nonaka, H. Ishikawa, Y. Tsumuraya, Y. Hashimoto, N. Dohmae, K. Takio, Characterization of a thermostable lysine-specific metalloendopeptidase from the fruiting bodies of a basidiomycete, *Grifola frondosa*, J. Biochem. 118 (1995) 1014–1020.

[15] T. Nonaka, Y. Hashimoto, K. Takio, Kinetic characterization of lysine-specific metalloendopeptidases from *Grifola frondosa* and *Pleurotus ostreatus* fruiting bodies, J. Biochem. 124 (1998) 157–162.

[16] T. Hori, T. Kumasaka, M. Yamamoto, N. Nonaka, N. Tanaka, Y. Hashimoto, T. Ueki, K. Takio, Structure of a new 'aspzincin' metalloendopeptidase from *Grifola frondosa*: implications for the catalytic mechanism and substrate specificity based on several different crystal forms, Acta Crystallogr. D 57 (2001) 361–368.

[17] H. Kuyama, K. Sonomura, O. Nishimura, Sensitive detection of phosphopeptides by matrix-assisted laser desorption/ionization mass spectrometry: use of alkylphosphonic acids as matrix additives, Rapid Commun. Mass Spectrom. 22 (2008) 1109–1116.

[18] T.H. Akiyama, T. Sasagawa, M. Suzuki, K. Titani, A method for selective isolation of the amino-terminal peptide from α-amino-blocked proteins, Anal. Biochem. 222 (1994) 210–216.

[19] K. Gevaert, M. Goethals, L. Martens, J. Van Damme, A. Staes, G.R. Thomas, J. Vandekerckhove, Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides, Nat. Biotechnol. 21 (2003) 566–569.

[20] K. Kuhn, A. Thompson, T. Prinz, J. Muller, C. Baumann, G. Schmidt, T. Neumann, C. Hamon, Isolation of N-terminal protein sequence tags from cyanogen bromide cleaved proteins as a novel approach to investigate hydrophobic proteins, J. Proteome Res. 2 (2003) 598–609.

[21] L. McDonald, D.H.L. Robertson, J.L. Hurst, R.J. Beynon, Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides, Nat. Methods 2 (2005) 955–957.

[22] M. Yamaguchi, T. Nakazawa, H. Kuyama, T. Obama, E. Ando, T. Okamura, N. Ueyama, S. Norioka, High-throughput method for N-terminal sequencing of proteins by MALDI mass spectrometry, Anal. Chem. 77 (2005) 645–651.

[23] C. Zhou, Y. Zhang, P. Qin, X. Liu, L. Zhao, S. Yang, Y. Cai, X. Qian, A method for rapidly confirming protein N-terminal sequences by matrix-assisted laser-desorption/ionization mass spectrometry, Rapid Commun. Mass Spectrom. 20 (2006) 2878–2884.

[24] M. Yamaguchi, T. Obama, H. Kuyama, D. Nakayama, E. Ando, T. Okamura, N. Ueyama, T. Nakazawa, S. Norioka, O. Nishimura, S. Tsunasawa, Specific isolation of N-terminal fragments from proteins and their high-fidelity *de novo* sequencing, Rapid Commun. Mass Spectrom. 21 (2007) 3329–3336.

[25] T. Mikami, T. Takao, Selective isolation of N-blocked peptides by isocyanate-coupled resin, Anal. Chem. 79 (2007) 7910–7915.

[26] G. Coussot, D.H. Hawke, A. Mularz, J.M. Koomen, R. Kobayashi, A method for the isolation of blocked N-terminal peptides, Anal. Biochem. 361 (2007) 302–304.

[27] E. Wachter, W. Machleidt, H. Hofner, J. Otto, Aminopropyl glass and its *p* -phenylene diisothiocyanate derivative, a new support in solid-phase Edman degradation of peptides and proteins, FEBS Lett. 35 (1973) 97–102.

RAPID COMMUNICATIONS IN MASS SPECTROMETRY
Rapid Commun. Mass Spectrom. 2008; **22**: 2063–2072
Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/rcm.3587

RCM

An improved method for *de novo* sequencing of arginine-containing, N^{α}-tris(2,4,6-trimethoxyphenyl)-phosphonium-acetylated peptides

Hiroki Kuyama[1]*, Kazuhiro Sonomura[2], Keisuke Shima[2], Osamu Nishimura[1,2] and Susumu Tsunasawa[1,2]

[1]Institute for Protein Research, Osaka University, Suita 565-0871, Japan
[2]Life Science Laboratory, Shimadzu Corporation, Kyoto 604-8511, Japan

Received 19 March 2008; Revised 14 April 2008; Accepted 20 April 2008

An improved method for *de novo* sequencing of arginine-containing peptides modified with succinimidyloxycarbonylmethyl tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP-Ac-OSu) is reported. A tagging reagent, TMPP-Ac-OSu, was introduced to improve the sequence analysis of peptides owing to the simplified fragmentation pattern. However, peptides containing arginine residues did not fragment efficiently even after TMPP-Ac modification at their N-termini. This report describes how fragmentation efficiency of TMPP-Ac-modified arginine-containing peptides was significantly improved by modifying the guanidino group on the side chain of arginine with acetylacetone. Copyright © 2008 John Wiley & Sons, Ltd.

Peptide sequencing by tandem mass spectrometry (MS/MS) is an established methodology in proteomic analysis.[1] This is normally performed by comparing fragmentation data with protein and/or genomic databases. Although this approach is very effective, it often fails to identify post-translationally modified residues or to detect sequence polymorphisms observed in a variety of protein isoforms. In these cases, *de novo* sequencing, where peptide sequences are interpreted directly from the MS/MS spectra, is the only practical approach by MS/MS.

Peptide *de novo* sequencing implies determining the peptide covalent structure, which is done from a given MS/MS spectrum, not with any information from an existing protein or DNA database. This is an alternative approach for analyzing peptide sequences and is especially effective for analytes with sequences that are difficult to analyze via the normally employed, database searching approach.

However, this approach by MS/MS is limited by the complex fragmentation patterns of various types of fragmentation.[2] The complexity of the MS/MS spectrum drastically reduces the possibility of *de novo* sequencing. Hence, fragmentation that is simple enough for manual interpretation (i.e., generation of a particular type of fragment ions) has been demanded.

Several simplification strategies have been reported using matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS).[3–11] *De novo* sequencing becomes easier and is facilitated with simple fragmentation.

Succinimidyloxycarbonylmethyl tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP-Ac-OSu)[6–9] is a reagent that has already been reported for simplification. The tagging reagent reacts selectively with the α-amino group at the N-terminus. The resulting peptide incorporating a fixed positive charge at its N-terminus biases the fragment ion intensities strongly toward a-type ions, leading to an easy and clear sequence determination, compared with its unmodified counterpart. However, arginine-containing peptides did not produce fragment ions well even after modification with TMPP-Ac reagent.[12]

It has already been reported that arginine-containing peptides exhibit a low degree of structurally informative fragmentation when investigated by commonly used collision-induced dissociation (CID) analysis. The arginine residue has the most basic property and easily protonates, which has deleterious effects on structural analysis by MS/MS.[13] For example, protonation of the basic groups results in charge localization, which subsequently results in suppression of random cleavage of backbone bonds. Several modification methods have been reported for enhancing the fragmentation of arginine-containing peptides.[14–18] Of these, derivatization with acetylacetone has been developed by reducing the basic nature of the guanidino group.[14]

Few studies, however, have been performed on the influence of arginine on the fragmentation behavior of TMPP-Ac-modified peptides. Recently, Chen and coworkers reported that enzymatic elimination of arginine residue from TMPP-Ac-modified peptides improved fragmentation.[12] However, this method has limited use for peptides containing arginine at their C-termini.

In this paper, we applied acetylacetone derivatization to TMPP-Ac-modified arginine-containing peptides and evaluated the fragmentation efficiency of derivatized peptides in comparison with the fragmentation efficiency for nonderivatized counterparts. This study covers four types of

*Correspondence to: H. Kuyama, Institute for Protein Research, Osaka University, Suita 565-0871, Japan.
E-mail: kuyama@protein.osaka-u.ac.jp

Table 1. Fragmentation efficiency of TMPP-Ac-modified peptides with and without arginine derivatization in MALDI-PSD and MALDI-CID

	Fragmentation efficiency[a] PSD		Fragmentation efficiency[a] CID	
TMPP-Ac-modified peptides	non-derivatized	Arg derivatized	non-derivatized	Arg derivatized
N-terminal arginine				
RGDS	0.25	1.8	0.84	2.2
RVYIHPF	0.53	5.5	1.3	14
RAHYNIVTF	—[b]	4.1	1.4	12
Internal arginine				
APDTRPAPG	5.3	7.8	11	14
TPRVTGGGAM	3.2	4.7	6.6	9.6
DAEFRHDSGYE	4.4	5.2	7.5	15
C-terminal arginine				
SFLLR	4.4	4.4	9.0	7.8
YIGSR-NH_2	6.0	5.6	8.3	9.6
multiple arginines[c]				
RLRFH	—[b]	4.3 (1.4)	—[b]	6.5 (3.5)
RPGFSPFR	—[b]	4.4 (3.4)	1.8	11 (8.1)
APRLRFYSL	1.9	6.5 (2.1)	5.0	14 (9.0)

[a] "Fragmentation efficiency" is calculated as the sum of a_n ion absolute abundances divided by the absolute abundance of the $[M+H]^+$ ion signal.
[b] Not detected.
[c] The values of multiple arginine-containing peptides indicate those for doubly derivatized peptides. The values in parentheses indicate those for singly derivatized peptides.

peptides, arginine residue at the N-terminus, at the C-terminus, in the middle of the sequence, and peptides with two arginine residues.

EXPERIMENTAL

Materials

Pig albumin and iodoacetamide were purchased from Sigma (St. Louis, MO, USA). Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and succinimidyloxycarbonylmethyl tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP-Ac-OSu) were obtained from Fluka (Switzerland). Lysyl endopeptidase (LysC), sodium hydrogen carbonate ($NaHCO_3$), acetonitrile, 2-propanol, and trifluoroacetic acid (TFA) were purchased from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). α-Cyano-4-hydroxycinnamic acid (CHCA; high-purity mass-spectrometric grade) was obtained from Shimadzu GLC (Tokyo, Japan). Two peptides (RGDS and RPGFSPFR) were purchased from Peptide Institute, Inc. (Osaka, Japan). Other peptides used in this study (listed in Table 1) were purchased from AnaSpec (San Jose, CA, USA). Methanediphosphonic acid (MDPNA) was obtained from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan).

Derivatization of peptides

Model peptides

Each peptide was dissolved in 100 mM $NaHCO_3$ containing 10% acetonitrile at a concentration of 10 nM. A volume of 5 μL of 10 mM TMPP-Ac-OSu solution (in water/acetonitrile 4:1) was added to 45 μL of peptide solution. The mixture was sonicated in a water bath for 5 min and incubated at 50°C for 30 min. Then 4 μL of Na_2CO_3 and 6 μL of acetylacetone were added to this reaction solution, followed by incubation at 80°C for 3 h. The reaction solution was used directly for MALDI-TOF MS analysis without any purification.

Isolated C-terminal peptide

The C-terminal peptide of pig albumin was isolated according to our method.[19] The solution containing the C-terminal peptide isolated from separated proteins in an excised gel strip (loaded amount: 40 pmol) was lyophilized and redissolved in 18 μL of 100 mM $NaHCO_3/Na_2CO_3$ buffer (pH 9.0). Then acetylacetone (2 μL) was added to this solution, followed by incubation at 37°C for 12 h. The reaction solution was used directly for MALDI-TOF MS analysis without any purification.

MALDI-MS

MALDI mass spectra were recorded on AXIMA TOF^2 (SHIMADZU/KRATOS, Manchester, UK) reflectron time-of-flight (TOF) mass spectrometers equipped with a nitrogen laser (337 nm, 3ns 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, collision-induced dissociation (CID) was carried out using helium at a pressure of 5×10^{-6} mbar in the collision cell.

α-Cyano-4-hydroxycinnamic acid (CHCA) was used as a matrix, which was dissolved to saturation in 50% aqueous acetonitrile containing 0.05% TFA. We used MDPNA,[20] which has been proven to be useful for MALDI analysis of salt-containing samples, as a matrix additive. MDPNA was used as 1–2% aqueous solution. An aliquot (0.4 μL) of sample solution was mixed with an equivalent volume of matrix

solution and matrix additive solution on the MALDI target plate and analyzed after drying.

We set the laser fluence identically for the MS/MS analysis of each TMPP-Ac-modified peptide before and after arginine modification. All spectra were obtained by accumulating data from 200 laser shots.

RESULTS AND DISCUSSION

We have recently developed a simple method for C-terminal-specific isolation and *de novo* sequencing of proteins by MALDI-MS using the TMPP-Ac reagent.[19] In this method, protein was digested with LysC; thus an isolated C-terminal peptide may contain arginine residue(s) at any position(s) in its sequence. Therefore, proteolytic elimination of an arginine residue for enhancing fragmentation was beyond the scope for the peptides isolated by this method. Dikler and co-workers described the enhancing effect of fragmentation by MS/MS using derivatization of a guanidino group in peptides with acetylacetone yielding the N^5-(4,6-dimethyl-2-pyrimidinyl)ornithine (Pyo) residue.[14] This modification reduces the basicity of the guanidino group in the arginine residue. It has been reported that fragmentation of protonated peptides is suppressed due to the high basicity of the guanidino group. The MS signal itself has also been reported to be enhanced by introducing the pyrimidine ring.[21] In addition, the modification can be performed in an aqueous media as in the TMPP-Ac modification. Therefore, we adopted the acetylacetone derivatization procedure to TMPP-Ac-modified peptides.

In one C-terminal peptide in our report,[19] one arginine residue was positioned at its N-terminus and gave practically no fragmentation signals in the MS/MS experiment. We applied acetylacetone derivatization to the peptide and obtained significant enhancement of structurally informative fragmentation in the MS/MS experiment.

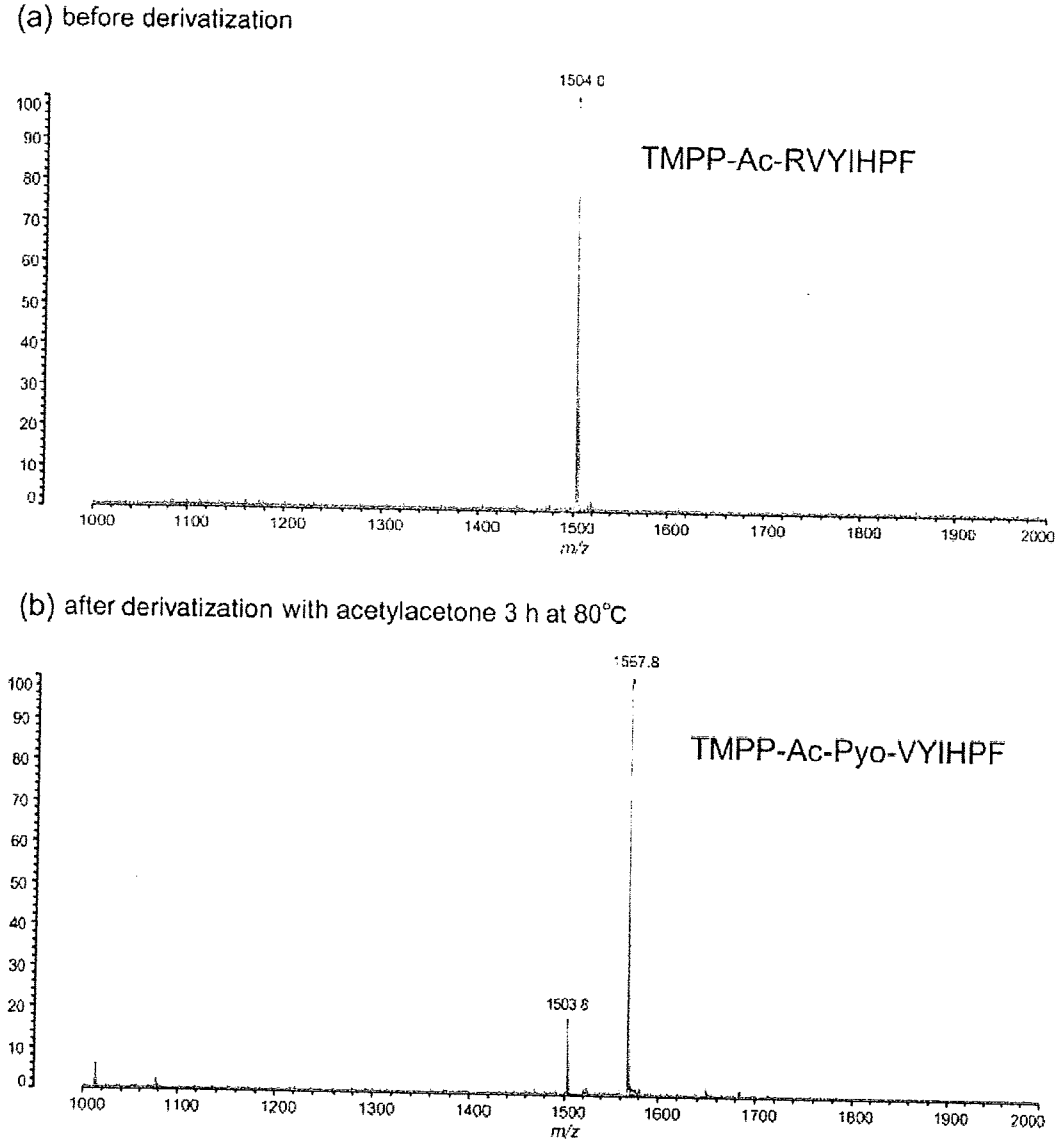

Figure 1. Derivatization with acetylacetone at elevated temperature. The reaction was accelerated at 80°C even with the peptide with an arginine residue at its N-terminus (TMPP-Ac-RVYIHPF). Under this condition, no side reaction is observed (b).

DOI: 10.1002/rcm

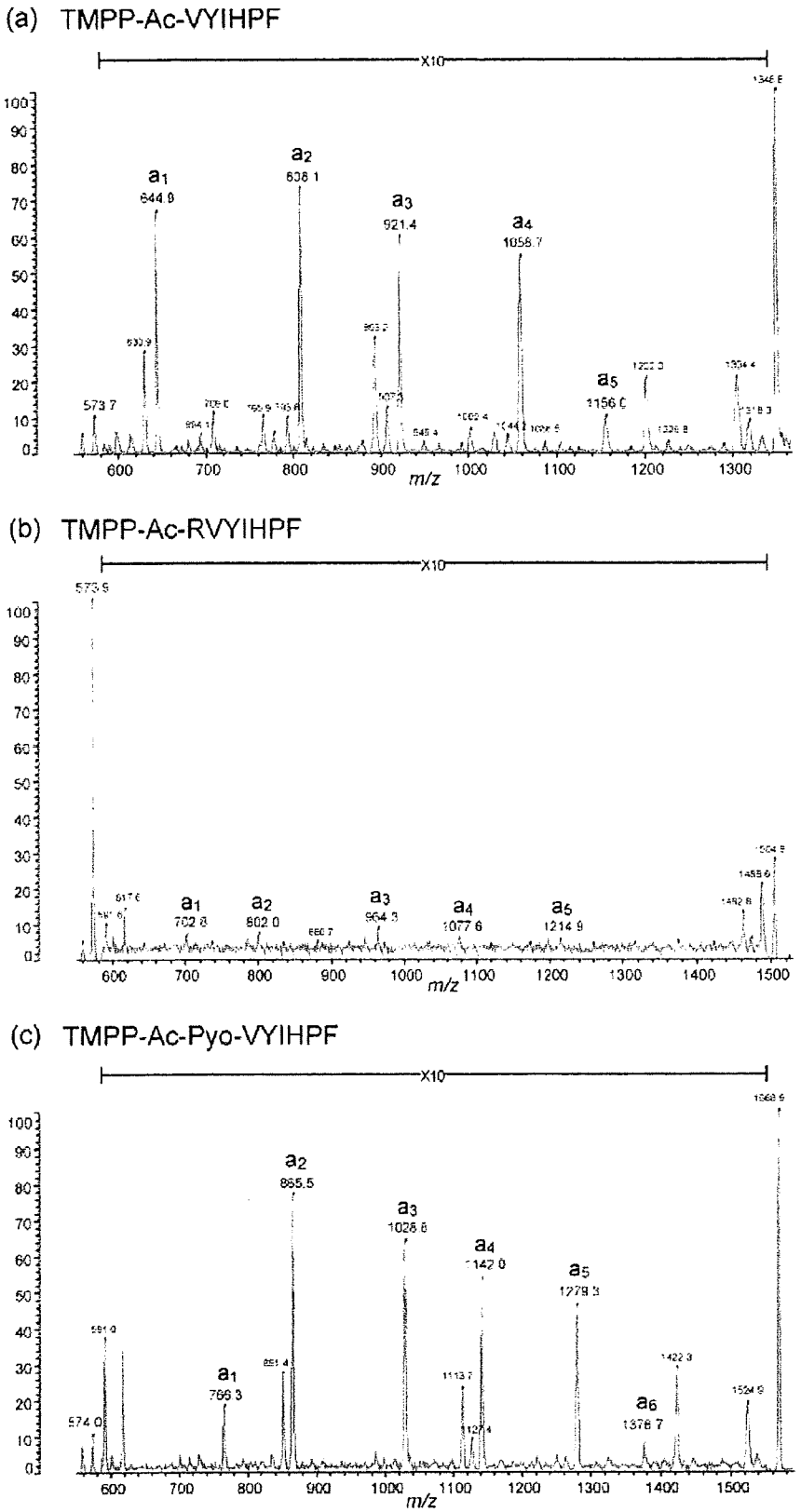

Figure 2. Enhancing effect of arginine derivatization using a peptide incorporating an arginine residue at its N-terminus (TMPP-Ac-RVYIHPF). (b) A non-derivatized peptide and (c) TMPP-Ac peptide after derivatization of Arg with acetylacetone. The influence of the arginine residue on fragmentation is also depicted using a peptide without an arginine residue (a: TMPP-Ac-VYIHPF) and with an arginine residue at its N-terminus (b: TMPP-Ac-RVYIHPF).

DOI: 10.1002/rcm

Hence, we began a detailed investigation of fragmentation efficiency by derivatizing arginine residue(s) of TMPP-Ac-modified peptides.

Reaction

We first investigated the reaction conditions of TMPP-Ac modification. It has been reported that peptides with an N-terminal arginine residue and peptides containing more than two arginine residues are resistant to TMPP-Ac modification.[8] In our experiments, peptides with N-terminal arginine did not react with TMPP-Ac-OSu under the original conditions.[8] We found that it is possible to derivatize arginine residues in both types of peptides by heating at 50°C. No by-products were detected in the MS spectrum (data not shown).

We next investigated the reaction conditions of arginine derivatization. One experimental problem of the procedure is that the reaction is very slow, and a long reaction time (3 days) is required for modification. This problem was solved by heating at 80°C. The reaction was checked by MS, and the modification was achieved with more than 80% completion after 3-h reaction at 80°C (Figs. 1(a) and 1(b)). The reaction efficiency was simply evaluated by using signal intensities of derivatized and non-derivatized peptides (Fig. 1(b)). However, the improved procedure of arginine modification was not directly applicable to the peptides isolated by our C-terminal isolation and sequencing method.[19] This may be attributable to the impurities in the isolated peptide solution. Therefore, a detailed examination of reaction conditions is required to improve the reaction efficiency. During the present experiment, we obtained

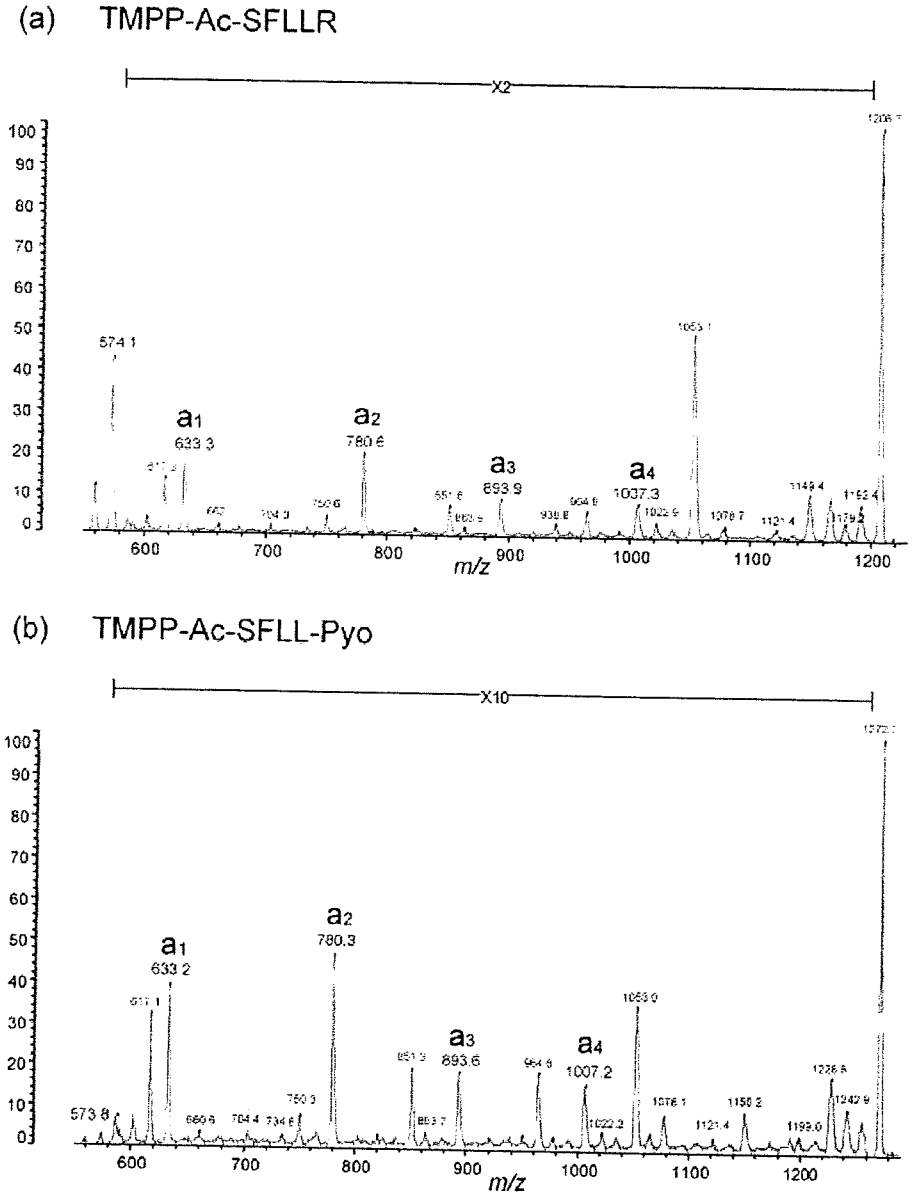

Figure 3. Enhancing effect of arginine derivatization using a peptide with a C-terminal arginine residue (TMPP-Ac-SFLLR): (a) non-derivatized peptide and (b) TMPP-Ac peptide after derivatization of Arg with acetylacetone.

Rapid Commun. Mass Spectrom. 2008; 22: 2063–2072
DOI: 10.1002/rcm

sufficient spectra of these derivatization-reluctant peptides for sequence determination for reaction conditions of 12 h at 37°C.

Improvement of fragmentation efficiency of model peptides

To confirm whether an arginine residue influences fragmentation behavior of TMPP-modified peptides, we compared MALDI-CID spectra of two model peptides (TMPP-Ac-VYIHPF and TMPP-Ac-RVYIHPF). In the spectrum of TMPP-Ac-VYIHPF (Fig. 2(a)), all the a-type ions were detected, and it was possible to directly interpret the amino acid sequence of the peptide, while in the spectrum of TMPP-Ac-RVYIHPF (Fig. 2(b)), all the a-type ions were detected, but signals were rather low and masked by noise peaks. The only difference between these two peptides is the presence of N-terminal arginine. This indicates that the arginine residue greatly influences the fragmentation pattern of TMPP-Ac-modified peptides.

Figure 2(c) depicts the MALDI-CID spectra of the peptide TMPP-Ac-RVYIHPF after arginine derivatization with acetylacetone. In the spectrum, all the a-type ions were clearly detected as in the spectrum of TMPP-Ac-VYIHPF, and the sequence was easily determined.

To study how the effectiveness of arginine modification depends on the position and the number of arginine residues, four groups of model peptides (peptides incorporating an arginine residue at their N-termini, C-termini, an internal position, and peptides containing two arginines) were derivatized with acetylacetone and subjected to MALDI-PSD (post-source decay) and MALDI-CID analysis. To compare the fragmentation of the derivatized and non-derivatized peptides, we defined a 'fragmentation efficiency' as in the report by Dikler *et al.*,[14] which is calculated as the sum of the absolute abundances of the a_n ion divided by the absolute

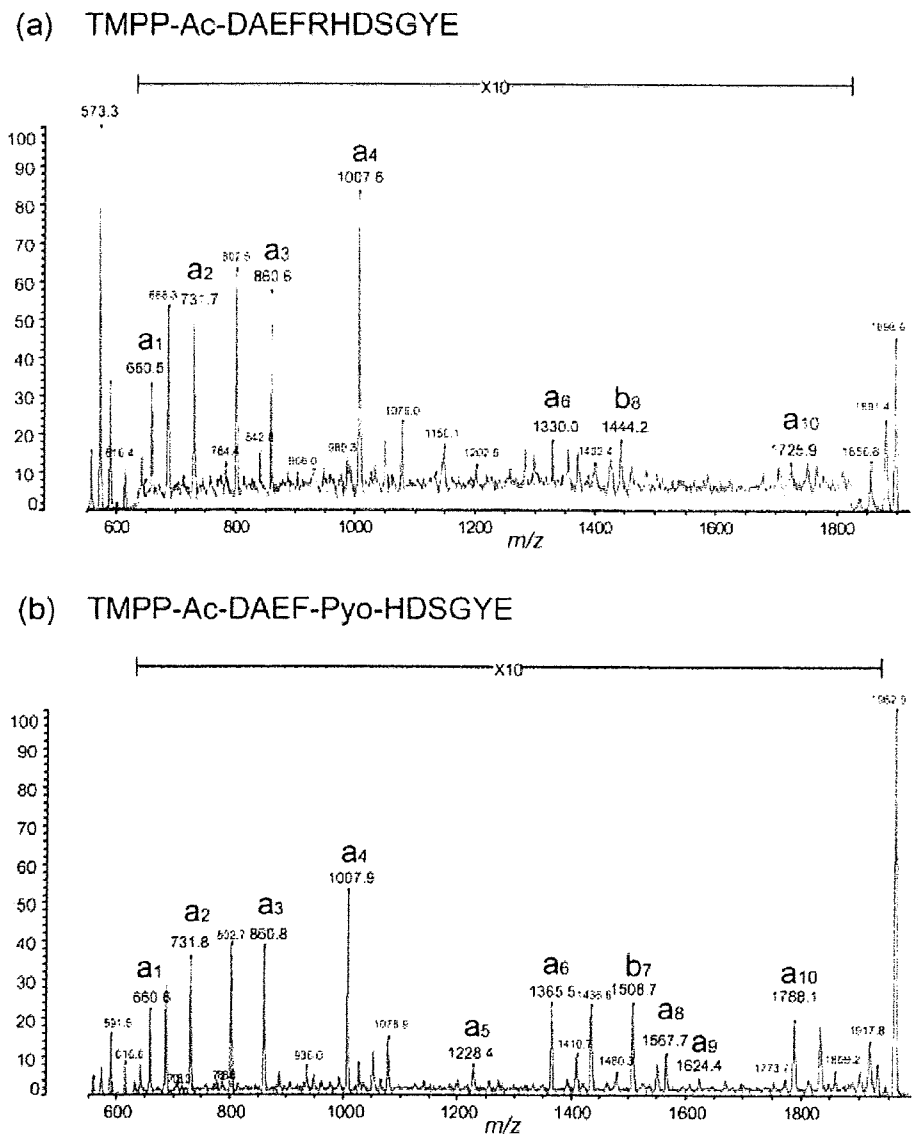

Figure 4. Enhancing effect of arginine derivatization using a peptide incorporating an arginine residue at an internal position (TMPP-Ac-DAEFRHDSGYE): (a) non-derivatized peptide and (b) TMPP-Ac peptide after derivatization of Arg with acetylacetone.

DOI: 10.1002/rcm

abundance of the $[M+H]^+$ ion signal. Amino acid sequences and fragmentation efficiencies of the peptides are listed in Table 1. The values were averaged from the data of two independent experiments.

1. **Peptides incorporating an arginine residue at their N-termini**
   Of the four groups of peptides, the fragmentation efficiency of peptides incorporating an arginine residue at their N-termini (TMPP-Ac-RGDS, TMPP-Ac-RVYIHPF and TMPP-Ac-RAHYNIVTF) was most improved by a factor of 7.2 to 10.4 in PSD and 2.7 to 10.8 in CID (Table 1). In the spectrum of one representative peptide (TMPP-Ac-RVYIHPF) before arginine modification, a-type ions were weak (a_1, a_2, a_3, a_4 and a_5) or not detectable (a_6) (Fig. 2(b)). After modification, all the a-type ions were clearly detected and sequences were successfully determined (Fig. 2(c)).
2. **Peptides incorporating an arginine residue at their C-termini**
   Fragmentation of C-terminal arginine-containing peptides (TMPP-Ac-SFLLR and TMPP-Ac-YIGSR-NH_2) was little affected by derivatization of arginine residues. Fragmentation efficiencies of these peptides were similar before and after arginine derivatization in both PSD and CID analysis (Table 1). Figures 3(a) and 3(b) present spectra before and after the arginine derivatization (Pyo) of a peptide (TMPP-Ac-SFLLR), respectively.
3. **Peptides incorporating an arginine residue at an internal position**
   In the MS/MS analysis of peptides containing an arginine residue at an internal position (TMPP-Ac-APDTRPAPG, TMPP-Ac-TPRVT and TMPP-Ac-DAEFRHDSGYE), a-type ions with an arginine residue were weak or non-detectable before arginine derivatization, but a-type ions without an arginine residue were detected. Figure 4(a) depicts one representative example of such peptides. In the spectrum, a_1–a_4 ions were strongly detected, but the other a-type ions were weak (a_6 and a_{10}) or not detectable (a_5, a_7, a_8, and a_9). After the derivatization of arginine, a_5-a_{10} ions became detectable, and it was possible to determine the amino acid sequence (Fig. 4(b)). In these peptides, fragmentation efficiency was improved by a factor of 1.2 to 1.5 in PSD and 1.3 to 2.0 in CID (Table 1). Figure 7 shows CID spectra of the C-terminal peptide[19] isolated from albumin (pig) before (Fig. 7(a)) and after (Fig. 7(b)) arginine derivatization. Before the derivatization, a-type ions containing an arginine residue (a_6, a_7) were not detected, though other a-type ions were clearly discernible (Fig. 7(a)). As with the peptide in Fig. 3, missing ions a_6 and a_7 were detected after the derivatization with acetylacetone (Fig. 7(b)).
4. **Peptides incorporating two arginine residues**
   For this experiment, three peptides were used (TMPP-Ac-RLRFH, TMPP-Ac-RPGFSPFR and TMPP-Ac-APRLRFYSL). To present the results graphically, we selected one peptide (TMPP-Ac-APRLRFYSL). This peptide incorporated two arginine residues, one each at the third and fifth positions from the N-terminus. The reactivity of these two arginines exhibited slight differences to derivatization. The reaction time course is presented in Fig. 5. The derivatization reaction at 37°C for 1 day mainly yielded a singly derivatized product, though a signal of a non-derivatized peptide was still at the highest intensity (Fig. 5(b)). The derivatization reaction occurred first at the third position from its N-terminus, as determined by the CID analysis. The CID spectrum is depicted in Fig. 6(b). The enhancing effect was not enough to produce all a-type ions; a_5 and a_8 ions were still missing. After a prolonged reaction for 3 days, the signal of the

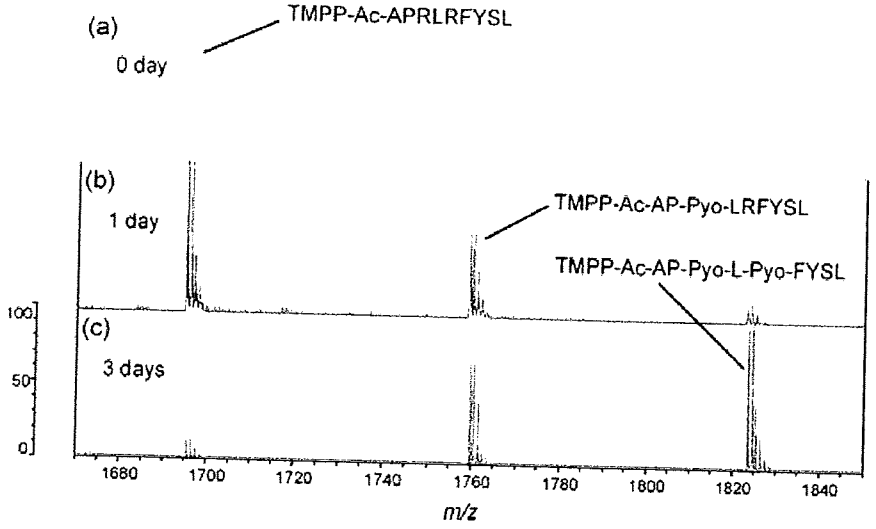

Figure 5. Reaction time course of a peptide incorporating two arginine residues (TMPP-Ac-APRLRFYSL) for derivatization with acetylacetone. After 1 day, the singly derivatized peptide emerged as the main product, though the non-derivatized peptide signal is strongest in intensity (b). After 3 days, the starting peptide almost disappeared and primarily the double derivatization was observed (c).

DOI: 10.1002/rcm