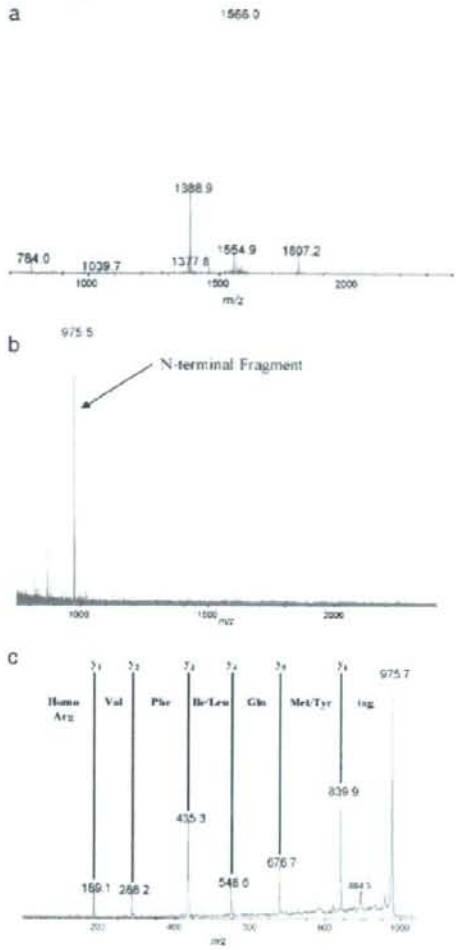

Figure 7. MALDI spectra at each step during N-terminal sequencing of spot No. 34 after 2D SDS-PAGE of the human epidermoid carcinoma cell lysate. (a) MALDI-TOF MS spectrum after trypsin digestion. (b) MALDI-TOF MS spectrum of the N-terminal peptides specifically isolated from the trypsin digests. (c)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

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

CONCLUSIONS

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

REFERENCES

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

DOI: 10.1002/rcm

Proteomics 2008, *8*, 673–685 DOI 10.1002/pmic.200700084

REVIEW

Terminal proteomics: N- and C-terminal analyses for high-fidelity identification of proteins using MS

*Takashi Nakazawa[1], Minoru Yamaguchi[2], Taka-aki Okamura[3], Eiji Ando[2], Osamu Nishimura[2, 4] and Susumu Tsunasawa[2, 4]**

[1] Department of Chemistry, Nara Women's University, Nara, Japan
[2] Life Science Laboratory, Shimadzu Corporation, Kyoto, Japan
[3] Department of Macromolecular Science, Graduate School of Science, Osaka University, Toyonaka, Osaka, Japan
[4] Institute for Protein Research, Osaka University, Suita, Osaka, Japan

In proteomics, MS plays an essential role in identifying and quantifying proteins. To characterize mature target proteins from living cells, candidate proteins are often analyzed with PMF and MS/MS ion search methods in combination with computational search routines based on bioinformatics. In contrast to shotgun proteomics, which is widely used to identify proteins, proteomics based on the analysis of N- and C-terminal amino acid sequences (terminal proteomics) should render higher fidelity results because of the high information content of terminal sequence and potentially high throughput of the method not requiring very high sequence coverage to be achieved by extensive sequencing. In line with this expectation, we review recent advances in methods for N- and C-terminal amino acid sequencing of proteins. This review focuses mainly on the methods of N- and C-terminal analyses based on MALDI-TOF MS for its easy accessibility, with several complementary approaches using LC/MS/MS. We also describe problems associated with MS and possible remedies, including chemical and enzymatic procedures to enhance the fidelity of these methods.

Received: January 30, 2007
Revised: August 20, 2007
Accepted: August 29, 2007

Keywords:
Chemically assisted fragmentation / C-terminal sequence analysis / Ladder sequencing / MALDI-TOF MS, MALDI-PSD-MS / N-terminal sequence analysis

1 Introduction

1.1 The value and information content of the N- and C-termini of proteins

The native folded structure of a protein is dictated solely by its covalent structure, that is, its amino acid sequence and PTMs. Because the characteristic three-dimensional structure is vital for the biological function of the protein, determining the primary amino acid sequences and PTMs of proteins has been an important topic in the life sciences as a means of understanding the complicated chemical reactions performed within living cells. The amino acid sequence of each protein is encoded by its corresponding gene, providing proteomics researchers with one of the most reliable clues for the identification of a protein of interest. However, it is usually altered from that expected from the DNA sequence alone, owing to splicing, shuffling, and processing of the gene or mRNA [1]. Protein sequence databases with additional information about the N-terminal peptide of proteins

Correspondence: Dr. Takashi Nakazawa, Department of Chemistry, Nara Women's University, Nara 630 8506, Japan
E-mail: t.nakazawa@cc.nara-wu.ac.jp
Fax: +81-742-20-3396

Abbreviations: COFRADIC, combined fractional diagonal chromatography; **CPase,** carboxypeptidase; **DITC,** *p*-phenylene diisothiocyanate; **ISD,** in-source decay; **sulfo-NHS-SS-biotin,** sulfosuccimidyl 3-(2-(biotinamido)ethyldisulfanyl)propionate

* Additional corresponding author: Dr. Susumu Tsunasawa, E-mail: s-tsunas@shimadzu.co.jp

allow efficient identification of protein isoforms involving single nucleotide polymorphisms in the coding regions [2]. The datasets obtained by large-scale identification of N-terminal peptides in the Halophilic archea, *Halobacterium salinarum* and *Natronomonas pharaonis*, led to both the validation/improvement of start codon assignments and the unraveling of N-terminal protein maturation in these archaea [3, 4].

After mRNA processing, the newly expressed proteins may undergo a variety of PTMs catalyzed by, for example, processing enzymes, acyl transferases, glycosylases, phosphatases, and kinases, which modify strictly specific sites (Fig. 1). Additional PTMs include N-terminal methionine excision [5], ubiquitination [6], membrane attachment of proteins modified at C-terminal carboxyl groups with glucosyl phosphatidylinositol [7], and truncation of the C-terminal part of a protein [8]. These PTMs are responsible for the final maturation of proteins and are regarded as typical mechanisms for regulating protein functions and influencing stability and activity. The stability of mature proteins is often associated with N-terminal ubiquitination, which occurs as a signal for initiating the degradation. Notably, approximately one third of the nuclear-encoded proteins starting with unprocessed and unprotected methionine are at least one order of magnitude less stable than proteins as a whole in higher eukaryotes, as they are rapidly degraded after N-terminal ubiquitination [9]. Considering the main pathway of protein maturation including N-terminal methionine excision, acetylation, and leader peptide removal, we could expect that analysis of protein termini is very important for the identification of target proteins and to fully understand protein functions in a given cell under a particular stage of physiological process. Furthermore, N-terminal sequence may be closely associated with subcellular localization of proteins [10]. Using N-terminal sequence information only, a neural network-based tool discriminated between proteins destined for distinct subcellular locations with a success rate of about 85% [11]. Such a high score of prediction in terms of genomics should, however, be checked by minutely comparing with the N-terminal sequence of mature proteins so that proteomics can appreciate the implication of the relationship between N-terminal amino acid sequences and subcellular distribution of proteins.

1.2 MS for the identification of proteins

Recent developments in MS and massive accumulation of genomic databases have brought revolutionary changes to the field of protein analysis with respect to both sensitivity and throughput. Protein analysis with MS less suffers from N- and C-terminal modifications, owing to its low dependency on the chemical nature of terminal groups. The methods for the identification of proteins include "top-down" and "bottom-up" (or "shotgun") proteomics, using ESI- or MALDI-MS to provide robust database search programs with a complex set of mass data [12].

In a majority of bottom-up approaches, a sample protein is purified with 2-D PAGE or HPLC, and is digested with a highly specific enzyme such as trypsin, before or after MS analysis, with which the sequence information is acquired. There are two distinctive methods for the identification of sample proteins. First, in PMF, candidate proteins are identified on the basis of sufficiently high score of match between the masses of the peptides produced by protease (*i.e.* trypsin)

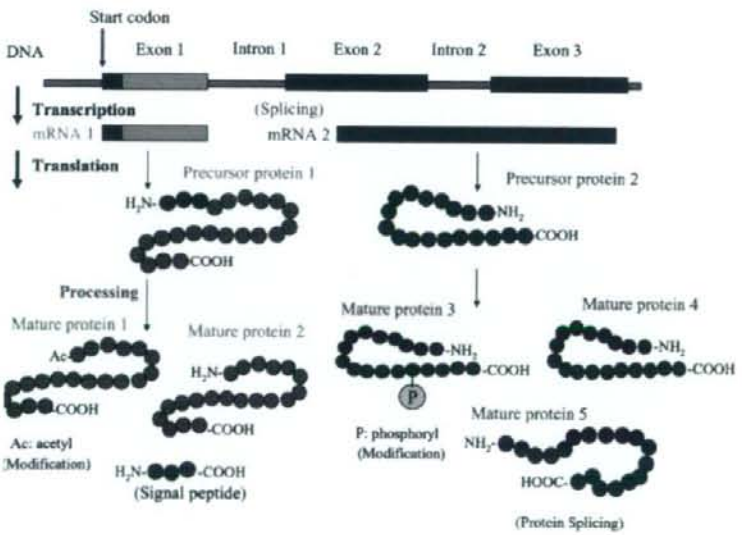

Figure 1. A scheme for protein processing *via* particular coding genes. A small number of proteins are expressed *via* different types of transcriptional mechanisms and then post-translationally processed and/or modified. A large number of mature proteins are thus made even from a relatively small number of coding genes. Figure shows typical examples of protein processing.

digestion and the mass list predicted from proteins in genome or protein databases, using an appropriate search algorithm incorporated in programs such as MASCOT and SEQUEST. Because the highly reliable identification could be achieved through the closest match between experimental and theoretical masses of peptides, any contamination of sample protein with peptides should be avoided. Second, in MS/MS database search for a protein, MS/MS spectra of a series of enzyme-digested peptides are compared with those of product ions predicted from protein sequence databases [13]. For the identification of sample protein in high fidelity, these methods require an extensive coverage of experimental data over the whole sequence of the protein to compensate for the difficulty in discerning the N- and C-terminal sequences of the sample protein. More than 200 types of PTMs impose a further limitation in the fidelity of protein identification because of our inherent inability to predict the occurrence of PTMs from the genome (EST) and protein sequence databases. The suggested remedies include a positional proteomics strategy, which greatly reduces the complexity of analytes by precisely locating a peptide at a particular position in a protein [14].

The top-down approach is attractive for N- and C-terminal amino acid sequencing of proteins. To be successful, a particular series of fragment ions from the N- or C-terminus of an intact protein must be obtained by FT-ICR MS instruments, which are used because of their high sensitivity and accuracy [15]. However, it is still difficult to effectively dissociate intact large proteins into a series of fragment ions for sequencing by further fragmentation. The accessibility to top-down proteomics has been improved by the development of an efficient mass analyzer, the orbitrap, which is less expensive than the FT-ICR analyzer without degrading accuracy, resolution, and sensitivity [16]. Suckau and Resemann [17] suggested an alternative top-down strategy based on the information about fragment ions effectively produced by in-source decay (ISD) in MALDI-TOF/TOF MS instruments. Despite the limitation of applicability to intact purified peptides and proteins, MALDI-ISD permits the analysis of much longer N-terminal amino acid sequence than does the MALDI-PSD MS without being affected by PTMs (for a review, see Hardouin [18]).

Here, we review recent methods developed for N- and C-terminal amino acid sequence analyses of proteins using MALDI-TOF MS as a key means for the identification of primary structures of mature proteins formed under particular physiological conditions in cells and organs. One of the most advantageous features of terminal sequencing is that as few as four residues from the N-terminal amino acid sequence (sequence tag) have proved sufficient for specifying between 43 and 83% of proteins, and those from the C-terminal counterpart further raise the success rate to 74–97% of proteins [19]. This illuminates a sharp contrast to shotgun proteomics in which high sequence coverage throughout the whole protein is necessary for the convincing identification. Therefore, positional proteomics [14], or more specifically, terminal proteomics, which deals with protein identification based on N- and C-terminal amino acid sequencing, can allow the analysis to be achieved in high-throughput. However, it is inevitable that the advantage of high-throughput in protein sequencing is often compromised by fidelity. Sequence data deduced from a fully instrumented and, consequently, high-throughput method should ultimately be confirmed by more reliable chemical and enzymatic approaches, especially when it has proved to involve covalent modification(s). The requisite manual analysis impedes throughput to some extent. Accordingly, we also include alternative methods based on ESI-MS in combination with LC to allow LC-MS/MS analysis of a series of purified peptide or protein samples.

2 Strategies for terminal analysis of peptides and proteins

In the first report of protein sequencing, Sanger used partial acid hydrolysis of bovine insulin, followed by labeling (or tagging) of the amino groups with 1-fluoro-2,4-dinitrobenzene to determine the primary structure [20]. Soon afterwards, however, the method of Sanger involving the partial acid hydrolysis of proteins was succeeded by the Edman chemistry [21], of which elaboration enabled high fidelity N-terminal sequential degradation of a peptide chain. Interestingly, the partial acid hydrolysis and the Edman degradation can still be useful to some extent in current proteomics research, even though the needs of high-throughput and sensitivity for large-scale analysis take first priority.

For the C-terminal version of Edman degradation, oxazolone chemistry was developed because an oxazolone only arises at the C-terminus of a peptide chain, discriminating against all side-chain carboxyl groups. Oxazolones are usually converted to a hydantoin, which is removed from the rest of the peptide chain, providing the next cycle of reactions with the new terminal carboxyl group of the penultimate amino acid residue. The repetition of this reaction cycle allows stepwise degradation of a peptide chain from the C-terminus, leading to an automated C-terminal sequencing method. It is now possible to sequence more than ten amino acid residues at high pmol sample levels [22]. While this method has been highly recommended for standard use in laboratory manuals and textbooks [23, 24], very few proteomics researchers appreciate its potential utility for the unambiguous determination of amino acid sequences. One of the most charitable assessments of this method stated that it was of limited use when 'you have nothing to lose' and if 'you don't get anywhere with your sample' ([24], pp. 184). Nevertheless, some courageous chemists have taken on the challenge of improving the method to enhance reaction yield and sensitivity. Their efforts to modify the simpler thiohydantoin chemistry into the more effective alkylation chemistry of thiohydantoin [25, 26] now allow the method to be performed at low picomole amounts on a protein sequencer [27].

2.1 Ladder sequencing

2.1.1 Controlled acid hydrolysis and chemical degradation

Probably the first example of exploiting MS in this context is "ladder sequencing" of proteins [28]. This method employs MALDI-MS for the analysis of a mixture of peptides produced after several cycles of Edman degradation with phenylisothiocyanate (95%) doped with phenylisocyanate (5%). This creates a ladder composed of a series of sequentially truncated peptides by partially terminating the degradation at each cycle of the reaction. The N-terminal amino acid sequence is thus elucidated from the ladder in a manner that is later adopted in MS/MS analysis. This greatly diminished the difficulties of the original Edman method, including the decreasing solubility of protein samples during the process and extensive sample losses for shorter hydrophobic peptides during washing of the excess reagents [29–31]. Nonetheless, many N-terminally blocked proteins remain to be analyzed. Aside from other modifications, N-terminal acetylation has been estimated to occur in 30–80% of eukaryotic proteins [9, 32], heavily impeding the approaches requiring the free N-terminal α-amino group. It is therefore necessary to perform the proteomic analysis by taking the precaution that more than 50% of N-termini in mature proteins may be blocked by some kind of modifications.

Consequently, the controlled acid hydrolysis primitively used by Sanger has been re-examined as a possible alternative for generating a ladder for MS analysis. This method is less likely to be adversely affected by the blockage of the N- or C-terminus. It is also possible to identify PTM amino acid residues, modifiers, and their attachment sites, if such modifications are not susceptible to acid hydrolysis. Li and coworkers observed cleavage of a single peptide bond by microwave irradiation in acidic solution of a protein, creating a pair of the N- and C-terminal fragments, thus resulting in a full set of peptides constituting a ladder for MALDI-MS analysis [33]. This method would be more attractive if the N- or C-terminus of a protein were tagged appropriately so that the peptides arising from either one of these termini could be distinguished from the others. For such a purpose, mass or charge tags (Section 2.2) employed in MS/MS analysis might be useful. Note, however, that the fidelity of the methods based on ladder sequencing of proteins falls as the molecular weight increases above 15 kDa. This is due to the difficulty in identifying the particular ladder corresponding to, say, the N-terminal sequence among all the others.

Under highly controlled conditions for partial acid hydrolysis, the free C-terminal carboxyl group assists the cleavage of the adjacent peptide bond, leading to sequential degradation adoptable to ladder sequencing [34]. Undesirable degradation associated with acid hydrolysis of peptide bonds was suppressed by carrying out the reaction in a vapor consisting of pentafluoropropionic or hexafluorobutyric acid and acetic anhydride [34]. The procedure has been improved so that proteins can be truncated with acetic anhydride in gel [35]. Although these methods have potentially wider applicability, the appeal of simply treating samples chemically is somewhat reduced by the complexity in interpreting the resulting ladder, in which careful allowances need to be made for mass shifts due to *N*- and *O*-acetylation along with accidental dehydration.

To identify a protein more effectively through the C-terminal sequence, it is possible to label all the peptide fragments except that at the C-terminus with ^{18}O by hydrolysis in $H_2{}^{18}O$. If a protein were subjected to proteolysis in $H_2{}^{18}O$, the resulting C-terminal peptide could be distinguished from others in the digest by mass spectra because it does not incorporate ^{18}O [36–38]. This isotope-labeling technique has rarely been used for C-terminal sequencing of whole proteins because of the general lack of C-terminal basic amino acids (Lys and Arg) known to assist the detection by MALDI-MS in a high sensitivity. Nevertheless, considering the requirements of extensive sequence coverage in bottom-up approaches, we can expect the potential utility of this technique for discriminating between the internal and C-terminal peptides in terminal proteomics.

2.1.2 The use of proteolytic enzymes

One of the most specific and reliable ways to generate a C-terminal ladder is to use a carboxypeptidase (CPase) such as CPase Y [39]. To avoid missing a particular step in the ladder, a mixture of CPases Y and P is often used for its wider substrate specificity [40, 41]. With this enzymatic ladder-creating method, up to ten or more amino acid residues have been sequenced at the pmol range by using MALDI-MS. The main obstacle to the method has been the difficulty of distinguishing between the C-terminal peptide and other peptide fragments when the method is applied to the proteolytic digest of a large protein. Suggested remedies include combining this method with cyanogen bromide cleavage of a protein at Met residues, discriminating the genuine C-terminal peptide from all the other peptide fragments, which are not digested by CPases owing to the blockage of the carboxyl group of the C-terminal homoserine by the lactone structure [42]. The limitation of this method is that it requires Met at an appropriate position near the C-terminus.

2.1.3 Gas-phase fragmentation

A ladder can also be produced in gas-phase fragmentation in MS/MS. Sequencing with MS/MS or MS^n takes the place of conventional chemical methods for creating the ladder through the use of state-of-the-art ion-trap facilities and sophisticated fragmentation techniques such as CID, electron transfer dissociation, or electron capture dissociation. These techniques automatically generate the ladder from a given ion selected in the mass spectrum. However, there still remains the problem of clearly distinguishing a ladder belonging to, at best, either the b- or y-ion series. It is difficult

to distinguish the peak of the N- or C-terminal peptide for the subsequent fragmentation, unless either one or both termini is appropriately tagged chemically or enzymatically as described in Section 2.2.

2.2 Tagging of proteins at N- and C-termini for selective isolation and identification

Tagging a protein at either terminus limits our target to relatively short peptides. This avoids confusion arising from complex mass spectra with a large number of peaks, and thus enhances the fidelity of sequencing. For *de novo* protein sequencing, tags are usually introduced into the N-terminus of a peptide with a reagent that enhances the sensitivity of MS. It is desirable that the tag facilitates the interpretation of mass spectra by providing the peak of interest with a readily discernible feature.

2.2.1 Methods for N-terminal tagging

There are two approaches to sample preparation in N-terminal tagging: (i) preferential guanidination of lysine ε-amino groups with *O*-methylisourea, leaving the terminal α-amino group unprotected for the subsequent tagging (disulfide bonds are reduced and *S*-alkylated prior to the modification) [29]; (ii) direct labeling of both α- and ε-amino groups of the sample protein with the tagging reagent, followed by digestion with a proteolytic enzyme, such as trypsin and chymotrypsin. In method (i), the N-terminal peptide can be easily found and analyzed in the presence of other peptides, owing to the specific tag thus introduced. In contrast, method (ii) needs either PMF or *de novo* sequencing with MS/MS analyses of target peptides to confirm the identity of the resultant peptides. Sometimes, it is desirable to avoid labeling of the ε-amino groups of lysine residues. Towards that end, removal of lysine residues by carboxypeptidase B and use of Lys-N as a proteolytic enzyme have been examined. For these strategies, an ESI-MS instrument is effective because each tagged peptide can be separated by LC before MS and MS/MS analyses. The following compounds have been employed: partially deuterated nicotinic acid [43], 5-bromonicotinic acid [44], 4-sulfophenyl isothiocyanate with a strongly acidic group [45], either a tertiary or quaternary ammonium group bearing a fixed positive charge like the iTRAQ™ reagent [46], a mixture of acetic and deuterioacetic anhydride [47], and bis(terpyridine)ruthenium(II) reagent ([(tpy)Ru(II)(tpy-C_6H_4COONSu)]X_2) (where tpy: terpyridine; NSu: *N*-succinimidyl; X: Cl, or PF_6) [48]. The chemical structure of bis(terpyridine) ruthenium(II) reagent and its use in N-terminal sequence analysis of bovine ubiquitin are shown in Fig. 2. These reagents might facilitate the sequencing of naturally blocked N-terminal peptide by modifying all the non-N-terminal peptides to be removed after tryptic digestion. 2,4,6-Trinitrobenzenesulfonic acid is also used for the modification of the non-terminal amino groups to enable the separation of the N-terminal peptide in a method termed combined fractional diagonal chromatography (COFRADIC) [49]. Apparently, these N-terminal tags, especially isotopic mass tags, can also be exploited for quantification of proteins in the manner of iTRAQ [46].

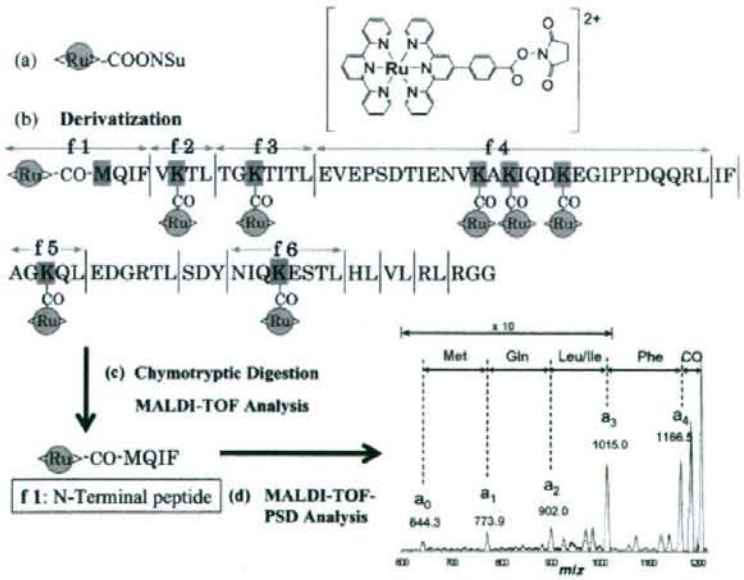

Figure 2. N-Terminal amino acid sequencing of ubiquitin tagged with [(tpy)Ru(II)(tpy-C_6H_4COONSu)]X_2 (<Ru>-COONSu). The N-Terminal amino acid sequence of bovine ubiquitin could be elucidated by analyzing its <Ru>-CO-tagged chymotryptic digest with MALDI-TOF/MALDI-TOF-PSD MS. The chemical structure of the reagent (<Ru>-COONSu) is shown in (a). Six peaks (b) having characteristic isotopic patterns depending on Ru(I) ions, which were assumed to be produced at one-half from Ru(II) by reduction, were detected in a MALDI-TOF mass spectrum of the chymotryptic digest (c). The amino acid sequence of a peptide corresponding to the N-terminal part was determined by MALDI-TOF-PSD MS (d).

2.2.2 Specific isolation of the N-terminal peptide and its *de novo* sequencing

To determine N-terminal amino acid sequences of proteins with high-throughput and high fidelity (regardless of protein mass), several methods for selective isolation of the N-terminal peptide followed by *de novo* sequencing have been reported [14, 50–53]. As a typical example, our approach [53] is illustrated in Fig. 3. The method consists of the following five steps: (i) reduction of disulfide bonds, followed by *S*-alkylation, and guanidination of the ε-amino groups; (ii) acylation of the α-amino groups with sulfosuccimidyl 3-(2-(biotinamido)ethyldisulfanyl)propionate (sulfo-NHS-SS-biotin) to introduce a removable biotin-tag; (iii) digestion of modified proteins by trypsin; (iv) specific isolation of N-terminal fragments of proteins by affinity capture using a biotin-avidin system; (v) *de novo* sequence analysis of peptides by MALDI-TOF/MALDI-TOF-PSD MS with effective utilization of the chemically assisted fragmentation method developed by Keough *et al.* [54]. Note that biotinyl cysteic acid used in the previous method [50] is replaced with sulfo-NHS-SS-biotin, from which the sulfonyl group is generated at the N-termini by performic acid oxidation [53]. This permits specific capture of N-terminal peptides from tryptic digests of proteins by utilizing a biotin-avidin system, followed by their release by oxidative cleavage of the disulfide bond of sulfo-S-S-biotin tag. These processes proceed sequentially with better yields for samples at low pmol levels. Amino acid sequences of the N-terminal peptides thus isolated are easily determined *de novo* by exclusive observation of y-series fragment ions in the MALDI-PSD spectrum (Fig. 4). The efficacy of the method is such that 34 proteins in an *Escherichia coli* extract were successfully identified by amino acid sequencing after separation by 2-D SDS-PAGE, followed by excision of each protein spot [50].

The throughput can be much enhanced by carrying out the reaction for several samples in parallel. However, for sequencing of N-terminally blocked proteins, this method needs further elaboration. To address this issue, we have developed a new technique in which two types of columns bearing *p*-phenylene diisothiocyanate (DITC) and avidin are used successively [55]. In this way, N-terminally blocked peptides are recovered from the first column in the unadsorbed fraction, and biotinylated N-terminal peptides are recovered in the adsorbed fraction of the avidin column. If we use a mixture of $H_2{}^{16}O$ and $H_2{}^{18}O$ instead of natural light ($H_2{}^{16}O$) water for tryptic digestion, the C-termini of N-blocked peptides are doubly labeled with ^{16}O and ^{18}O, facilitating their sequencing (to be reported elsewhere). A similar strategy uses acetylation of all the amino groups of a protein, followed by biotinylation of the α-amino groups of internal peptides released by tryptic digestion. This permits one to isolate the acetylated N-terminal peptides after removal of N-biotinylated internal peptides on an avidin column [14]. Similar to the COFRADIC approach, the advantages of this method are that it is easy to perform by obviating the guanidination and that it can be applied to identification of naturally N^{α}-acetylated proteins if deuterated ($[^2H_6]$) acetic anhydride is used appropriately to distinguish them from the N-terminal free proteins. Recently, the method has been improved by replacing the biotin-streptavidin system with an amine-reactive (NHS-activated) Sepharose column for the removal of internal peptides [56]. The improvement in method for the isolation or enrichment of the N-terminal

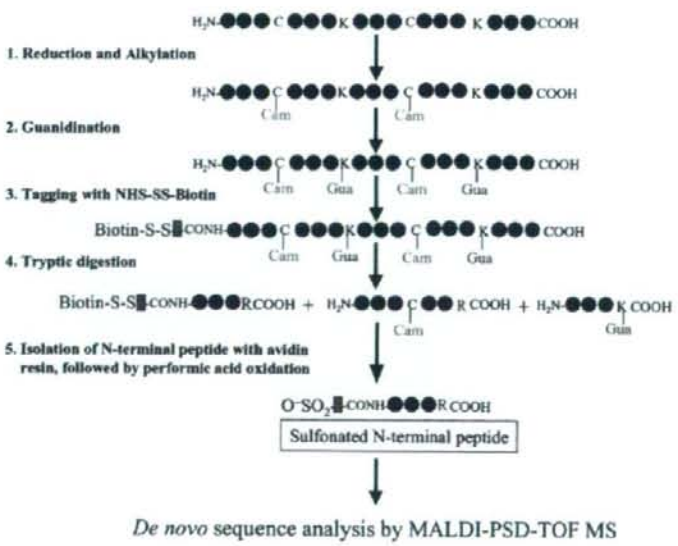

Figure 3. Protocol for selective isolation of N-terminal peptides from proteins and their *de novo* sequencing. The method consisting of five steps is shown. For the isolated N-terminal fragment, its sequence analysis is performed *de novo* with MALDI-TOF/MALDI-TOF-PSD MS by utilization of the chemically assisted fragmentation method [53].

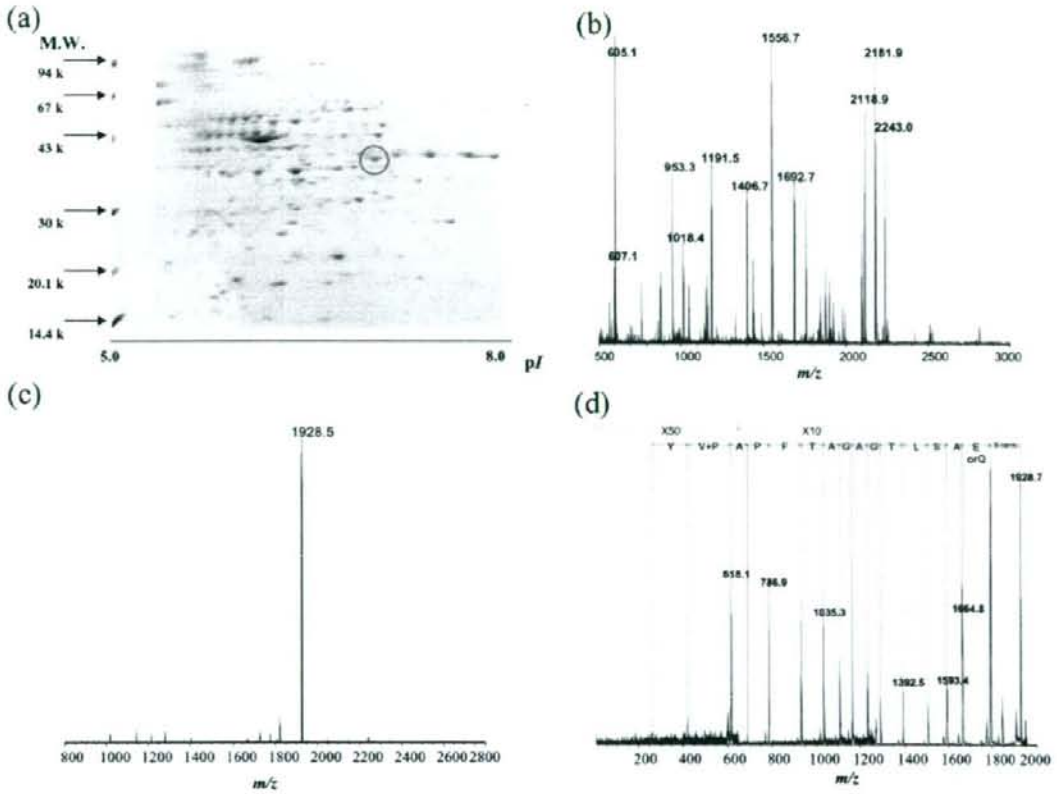

Figure 4. N-Terminal sequencing of a protein separated from *Escherichia coli* extracts according to the protocol shown in Fig. 3. *Escherichia coli* strain K-12 crude extracts were prepared using ProteoPrep™ Sample Extraction Kit (Sigma, St. Louis, MO). Samples containing 500 μg of protein were separated first by IEF on an Immobiline dry strip gel of pH 5.0–8.0 (Amersham Pharmacia Biotech, Bucks, UK) after dissolving in rehydration buffer containing 0.2% carrier ampholyte (Amersham Pharmacia Biotech). (a) 2-D PAGE image of proteins separated and stained by CBB. (b) MALDI-TOF MS spectrum of tryptic digest of a protein (approximately 5–10 pmol), excised from the darkest spot found in a blue circle in (a) and modified according to the protocol shown in Fig. 3. (c) MALDI-TOF MS spectrum of the N-terminal peptides specifically isolated from the trypsin digests using the avidin column. (d) MALDI-PSD spectrum of the precursor ion at *m/z* 1928.

peptide greatly reduces the analyte complexity, facilitating interpretation of the mass spectra, and thus allowing to achieve a high-fidelity analysis due to the drastically limited database search focused on the terminal part of cDNA or genomic data.

2.2.3 Methods for C-terminal tagging

For identification of a protein by database search, the C-terminal amino acid sequence offers more information than the N-terminal counterpart [19]. Moreover, C-terminal sequencing is less likely to suffer from PTMs than is N-terminal sequencing. In spite of these advantages, severe difficulties in chemically discriminating between main-chain and side-chain carboxyl groups have prevented C-terminal sequencing from performing as effectively as N-terminal analysis using Edman chemistry. The suggested remedies include the formation of an oxazolone, which occurs solely at the C-termini of peptides and proteins, while keeping all the other carboxyl groups intact. Because the oxazolone group is an active form of the carboxyl group, it enables the site-specific modification of the C-terminal carboxyl group (Fig. 5).

Generally, when a ladder is obtained by non-specific degradation of a peptide, it is impossible to determine the terminus from which it originated, with the exception of a ladder generated by CPase-digestion. A possible solution is chemical labeling of either the N- or C-terminus of a peptide. In MS/MS, the most promising tag for C-terminal sequencing should accentuate the responses of C-terminal fragments (x-, y-, and z-type ions) and suppress those of the N-terminal

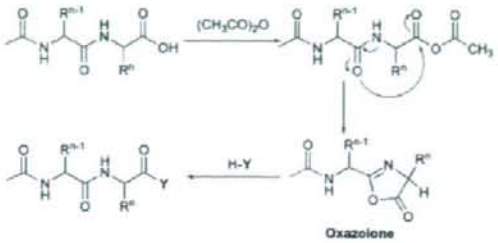

Figure 5. Formation and reaction of oxazolone at the C-terminus of a protein. The C-terminal carboxyl group of a protein reacts with acid anhydride (acetic anhydride, for example) to form an oxazolone *via* a mixed anhydride. The oxazolone can be utilized for subsequent derivatization by the reaction with a nucleophilic reagent (Y-H).

fragments (a-, b- and c-type ions) (see refs cited in 2.2.4). However, conventional methods used to derivatize the amino groups are not applicable to introduce a tag into the C-terminus of a peptide or protein, unless the C-terminal peptide is successfully isolated for sequencing. To achieve the requisite tagging for characterization and identification of the protein C-terminus, it is indispensable to modify the C-terminal carboxyl group. As described in the section on stepwise C-terminal degradation, the chemistry of oxazolone could provide us with a variety of methods for the derivatization [57]. However, the relatively low yield of oxazolone formation and the elaborate protocols for chemical manipulations, which are often unfriendly for proteomics researchers, need to be improved. Such hurdles inherent in oxazolone chemistry could be partially overcome by using MS for the detection of products with high sensitivity and throughput.

To take advantage of MS for selectively detecting and identifying peptides bearing a fixed positive charge, an oxazolone was exploited for coupling with a derivative of arginine [58]. Although the peptides attached with arginine or its ester exhibited peaks with much enhanced intensities in MALDI-MS, the reaction carried out in acetic anhydride at 115°C for oxazolone formation was so severe that several side reactions could not be suppressed. The improved protocol involves *in situ* conversion of oxazolone (formed in a mixture of acetic anhydride and formic acid) into the more stable active ester, with which a nucleophile such as arginine or 2-hydrazino-2-imidazoline bearing a positively charged group is coupled [59]. With this improvement, it could be used for C-terminal charge tagging of proteins. The tagged protein was digested with chymotrypsin to yield a mixture of peptides, among which the MALDI peak of the C-terminal peptide fragment could specifically be enhanced (Fig. 6). Note that in this setting, trypsin is not appropriate for use because it leaves basic amino acids at the C-termini of all the peptide fragments after the digestion. The remaining practical problems, which include the difficulties of discriminating between genuine and false-positive peaks in MALDI-MS, might be solved by developing an improved reagent for tagging so that modified C-terminal peptides are more clearly discerned and sequenced by MS/MS including LC-ESI-MS/MS. Note that the method is potentially extended for the isolation of C-terminal peptides in the manner described in the Section 2.2.2, provided that a specific functional group reactive to a solid support is introduced in the tag.

2.2.4 Specific isolation of C-terminal peptides and their sequencing

When there arises a need to sequence a protein from the C-terminus using MS, one might attempt to obtain the C-terminal peptide fragment from the enzymatic digest rather than deal with the whole protein. This would be advantageous because a small peptide is generally easier to handle, well within the accuracy limits of MS. Furthermore, a sequence of 5–10 amino acid residues is sufficient to characterize the protein with a genomic database [19]. To avoid being confused by contamination, a specific stratagem for isolating the genuine C-terminal from a mixture of peptide fragments has been suggested. Many authors prefer to use trypsin, or more specifically lysyl endopeptidase because these proteases leave basic amino acid residues at the C-terminus of every peptide except for that at the C-terminus. This permits specific isolation of C-terminal peptides using a column of immobilized anhydrotrypsin, which removes peptides containing lysine or arginine residues at their C-termini [60]. The eluted solution containing the uniquely unbound C-terminal peptide may be subjected to MS for sequencing. DITC attached to glass beads might serve as an alternative to immobilized anhydrotrypsin because all the side-chain amino groups of lysine residues occurring at the C-termini of lysyl endopeptidase digests are covalently bound to the glass beads, while the C-terminal peptide that most probably lacks a lysine residue is released from the solid support after cleavage of the N-terminal amino acid residue [61].

The isolated C-terminal peptide can be subjected to ordinary techniques of N-terminal sequencing if its length is modest enough to determine the entire covalent structure. This circumvents the risk of trying the still immature method of C-terminal sequencing. A reasonably short (but not too short) peptide is eligible for any kind of sequencing using MS. Therefore, the efficacy of these methods based on the isolation of C-terminal peptides depends on the amino acid sequence of a protein from which a medium-sized peptide is to be released. Except for an unfortunate case where the protein has arginine or lysine at the C-terminus, it is claimed that the method has proved successful in sequencing a protein at the 20 pmol level [62].

2.3 Chemical strategies for terminal access

Oxazolone chemistry associated with C-terminal modification of proteins is less effective than the established Edman chemistry for N-terminal sequencing. Nevertheless, it pro-

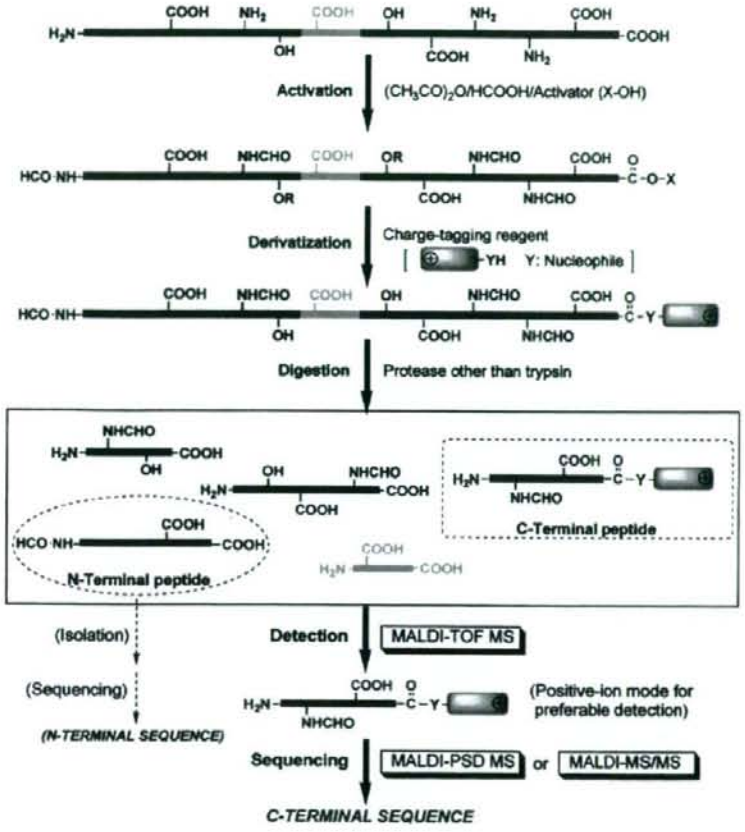

Figure 6. A scheme for C-terminal sequencing of peptides and proteins [59]. The method involves C-terminal activation, derivatization for charge tagging, and protease digestion, followed by successive analyses with MALDI-MS and MALDI-PSD-MS (or MALDI-MS/MS). In the activation process, some of the hydroxyl groups may be formylated (R = CHO or H), but the subsequent reaction with a nucleophile can deacylate. Although this scheme shows an example of charge tagging, it is also possible to label the peptide with an isotopically labeled reagent as a mass tag or with a bifunctional reagent that allows the isolation of the tagged peptide. After the step of protease digestion, the N-terminal peptide could possibly be isolated as suggested in literature [14].

vides us with the sole means to deal with C-terminal carboxyl groups, strictly discriminating against the same group on side-chains. This is the most advantageous feature of oxazolone chemistry compared to conventional approaches. For example, carbodiimides activate every carboxyl group in a protein for subsequent modification. A major breakthrough in C-terminal modification of proteins could therefore be achieved through implementing oxazolone chemistry.

2.4 Difficulties of protein identification and prospects of future developments

Protein sequencing was one of the most challenging subjects for biochemists until the middle 1970's, prior to the era of genomics. It required skills, profound knowledge of protein chemistry, and a great deal of patience with laborious experiments, which often lasted for years for each protein. Although the throughput of the sequencing was understandably low even after the automation of Edman degradation, the fidelity of the resulting sequence data was incredibly high. Today, the majority of proteomics researchers who are sequencing proteins are provided with laboratory manuals instructing them how to handle proteins with a kit of enzymes and reagents. Except for preliminary separation and manipulation of proteins, researchers focus on MS and a search for candidate proteins in a genome database. This approach requires the use of sensitive and high-throughput methods. However, proteomics intrinsically deals with thousands of proteins from a particular cell type or tissue in a given physiological condition. The enormously wide distribution of molecular weights and concentrations of proteins to be analyzed makes it prohibitively difficult to assess the throughput of different methods [63]. Nonetheless, there arises a rational relationship that the performance of protein sequencing is a compromise between throughput and reliability (Table 1). Apparently the rate-determining step in the proteomics analysis for identifying proteins should be placed on the development of methods, which deal with difficult proteins by means of special approaches. Considering the successful automation of Edman degradation and DNA

Table 1. Throughput and fidelity in the performance of methods using MS.

Method	Throughput	Fidelity	Comments (Instrument(s) required)	References
Top-down	Excellent	Modest	• Very high accuracy and sensitivity [FT ICR or orbitrap mass analyzer]	[12, 15, 16]
			• Easy to perform [MALDI-ISD-TOF/TOF]	[17, 18]
			• Limitations in higher mass range	
Bottom-up (PMF)	High	Modest[a)]	• High sequence coverage and high sample purity required.	[13, 65]
			• Easy to perform [MALDI]	
Bottom-up (MS/MS ion search)	High	Modest[a)]	• High sequence coverage required	[15, 66]
			• Robust search programs required for high throughput and fidelity	
Ladder sequencing (Acid hydrolysis)	Modest– High	Modest	• Easy to perform	[32]
			• Difficult to control the reaction	
			• High sample purity required	
Ladder sequencing (Chemical degradation)	Low	Modest– High	• Useful for terminal proteomics	[28, 33, 34, 39–41]
			• High sample purity required	
Ladder sequencing (Combination of Chemical & Enzymatic)	Modest– Low	Modest– High	• Probably most reliable for C-terminal sequencing	[42]
			• Met residue required for cleavage	
Isolation (enrichment) of N-terminal peptide	Modest– Low	Modest– High	• Modest technical skills required	[14, 50–53, 55, 56]
			• Adoptable to proteins with PTMs	
Isolation (enrichment) of C-terminal peptide	Modest– Low	Modest– High	• Modest/high technical skills required	[60–62]
			• Least likely affected by PTMs	
N-Terminal tagging	High– Low	Modest– High	• Modest technical skills required	[29, 43–49, 54]
			• Applicable to protein mixtures	
C-Terminal tagging	Modest– Low	Modest– High	• Advanced technical skills required	[58, 59]
			• Optimization required	
			• High fidelity expected	

a) Fidelity of identification resulting from these approaches depends strongly on the quality of peptide mass data and database search programs used [65, 66].

sequencing [64], we expect that some of the routine protocols of terminal sequencing may be replaced by robotic manipulation to increase the throughput of the method. The virtue of a high-throughput method is that it minimizes the routinely (often tedious) task of sequencing simple proteins, thus allowing researchers to concentrate on elucidating the structure of more complex proteins, which might be processed by a myriad of PTMs.

3 Concluding remarks

In this review, we referred to some chemical or enzymatic approaches, which may be still immature for practical use, along with standard methods for protein sequencing. This is because we believe that the tools of proteomics would not be complete without implementing new strategies for analyzing proteins not accessible by ordinal approaches. Since proteomics focuses on 'comprehensive analysis of whole proteins in a cell, tissue, or organism', the analytical tools of the subject must also be comprehensive. Probably, the most promising breakthrough arises through a rational combination of MS with highly site-specific chemical or enzymatic approaches to make it easier to interpret the mass spectra. A rational strategy of terminal proteomics might address the issues by choosing (i) the target at a specific site of a protein for tagging, isolation, or enrichment; (ii) reagent(s) for precisely reacting with the target functional group, discriminating (if necessary) between the groups of the same kind but of different site; (iii) an enzyme (or chemical) to generate either or both of terminal peptide. Most of the examples listed in the respective entries of Table 1 have such a structure. It is further desired that any method for proteomics study should also be able to deal with a complex mixture of proteins with a reasonably high sensitivity.

Terminal sequencing techniques using MS should minimize possible confusion caused by unpredictable or unexpected mass shifts due to the attachment of non-amino acid components to proteins by PTMs, because the probability is very low that both N- and C-termini of a protein are modified. Once the protein of interest is thus identified, the more laborious but more inspiring work would follow to characterize the protein with such modifiers. Thus, the completion of terminal sequencing is just the first step in proteomics research.

Our terminal proteomics project was partly supported by Grant-in-Aid for Scientific Research A-15201043 for Dr. Shigemi Norioka (Osaka University) and Taka-aki Okamura, and B-18310146 for Takashi Nakazawa, from the Ministry of Education, Culture, Sports, Science and Technology of Japan. We thank Dr. Hiroki Kuyama (Osaka University) for his valuable advices. We also sincerely thank Mr. Shigehiko Hattori (President, Shimadzu Co.) and Mr. Akira Nakamoto (Executive Director, Shimadzu Co.) for financial support.

The authors have declared no conflict of interest.

4 References

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

[2] Schandorff, S., Olsen, J. V., Bunkenborg, J., Blagoev, B. *et al.*, A mass spectrometry-friendly database for cSNP identification. *Nat. Methods.* 2007, *4*, 465–466.

[3] Falb, M., Aivaliotis, M., Garcia-Rizo, C., Bisle, B. *et al.*, Archaeal N-terminal protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey. *J. Mol. Biol.* 2006, *362*, 915–924.

[4] Aivaliotis, M., Gevaert, K., Falb, M., Tebbe, A. *et al.*, Large-scale identification of N-terminal peptides in the halophilic archaea *Halobacterium salinarum* and *Natronomonas pharaonis*. *J. Proteome Res.* 2007, *6*, 2195–2204.

[5] Sherman, F., Stewart, J.W., Tsunasawa, S., Methionine or not methionine at the beginning of a protein. *BioEssays* 1985, *3*, 27–31.

[6] Cantin, G. T., Yates III, J. R., Strategies for shotgun identification of post-translational modifications by mass spectrometry. *J. Chromatogr. A* 2004, *1053*, 7–14.

[7] Omaetxebarria, M. J., Hagglund, P., Elortza, F., Hopper, N. M. *et al.*, Isolation and characterization of glycosylphosphatidylinositol-anchored peptides by hydrophilic interaction chromatography and MALDI tandem mass spectrometry. *Anal. Chem.* 2006, *78*, 3335–3341.

[8] Selinheimo, E., Saloheimo, M., Ahola, E., Westerholm-Parvinen, A. *et al.*, Production and characterization of a secreted, C-terminally processed tyrosinase from the filamentous fungus *Trichoderma reesei*. *FEBS J.* 2006, *273*, 4322–4335.

[9] Meinnel, T., Peynot, P., Giglione, C., Processed N-termini of mature proteins in higher eukaryotes and their major contribution to dynamic proteomics. *Biochimie* 2005, *87*, 701–712.

[10] Höglund, A., Dönnes, P., Blum, T., Adolph, H.-W. *et al.*, MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics* 2006, *22*, 1158–1165.

[11] Emanuelsson, O., Nielsen, H., Brunak, S., von Heine, G., Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 2000, *300*, 1005–1016.

[12] Chait, B. T., Chemistry. Mass spectrometry: bottom-up or top-down? *Science* 2006, *314*, 65–66.

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

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

[15] Han, X., Jin, M., Breuker, K., McLafferty, F. W., Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science* 2006, *314*, 109–112.

[16] Macek, B., Waanders, L. F., Olsen, J. V., Mann, M., Top-down protein sequencing and MS^3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. *Mol. Cell. Proteomics* 2006, *5*, 949–958.

[17] Suckau, D., Resemann, A., T3-sequencing: targeted characterization of the N- and C-termini of undigested proteins by mass spectrometry. *Anal. Chem.* 2003, *75*, 5817–5824.

[18] Hardouin, J., Protein sequence information by matrix-assisted laser desorption/ionization in-source decay mass spectrometry. *Mass Spectrom. Rev.* 2007, *26*, 672–682.

[19] Wilkins, M. R., Gasteiger, E., Tonella, L., Ou, K. *et al.*, Protein identification with N and C-terminal sequence tags in proteome projects. *J. Mol. Biol.* 1998, *278*, 599–608.

[20] Sanger, F., Tuppy, H., The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *Biochem. J.* 1951, *49*, 463–481.

[21] Edman, P., On the mechanism of the phenyl isothiocyanate degradation of peptides. *Acta Chem. Scand.* 1956, *10*, 761–768.

[22] Bergman, T., Cederlund, E., Jörnvall, H., Chemical C-terminal protein sequence analysis: improved sensitivity, length of degradation, proline passage, and combination with Edman degradation. *Anal. Biochem.* 2001, *290*, 74–82.

[23] Simpson, R. J., *Proteins and Proteomics – A Laboratory Manual*, 2003, Cold Spring Harbor Laboratory Press, New York , pp. 311–317.

[24] Rehm, H., *Protein Biochemistry and Proteomics*, Elsevier Academic Press, Amsterdam 2006.

[25] Boyd, V. L., Bozzini, M., Zon, G., Noble, R. L. *et al.*, Sequencing of peptides and proteins from the carboxy terminus. *Anal. Biochem.* 1992, *206*, 344–352.

[26] Bailey, J. M., Chemical methods of protein sequence analysis. *J. Chromatogr. A*, 1995, *705*, 47–65.

[27] Samyn, B., Hardeman, K., Van der Eycken, J., Van Beeumen, J., Applicability of the alkylation chemistry for chemical C-terminal protein sequence analysis. *Anal. Chem.* 2000, *72*, 1389–1399.

[28] Chait, T., Wang, R., Beavis, R. C., Kent, S. B., Protein ladder sequencing. *Science* 1993, *262*, 89–92.

[29] Bartlet-Jones, M., Jeffery, W. A., Hansen, H. F., Pappin, D. J., Peptide ladder sequencing by mass spectrometry using a novel, volatile degradation reagent. *Rapid Commun. Mass Spectrom.* 1994, *8*, 737–742.

[30] Gu, Q. M., Prestwich, G. D., Efficient peptide ladder sequencing by MALDI-TOF mass spectrometry using allyl isothiocyanate. *J. Pept. Res.* 1997, *49*, 484–491.

[31] Hoving, S., Münchbach, M., Schmid, H., Signor, L. *et al.*, A method for the chemical generation of N-terminal peptide

sequence tags for rapid protein identification. *Anal. Chem.* 2000, *72*, 1006–1014.

[32] Brown, J. L., Roberts, W. K., Evidence that approximately eighty per cent of the soluble proteins from Ehrlich ascites cells are N^{α}-acetylated. *J. Biol. Chem.* 1976, *251*, 1009–1014.

[33] Zhong, H., Zhang, Y., Wen, Z., Li, L., Protein Sequencing by Mass Analysis of Polypeptide-Ladders after Controlled Protein Hydrolysis. *Nat. Biotech.* 2004, *10*, 1281–1296.

[34] Tsugita, A., Takamoto, K., Kamo, M., Iwadate, H., C-Terminal sequencing of protein: a novel partial acid hydrolysis and analysis by mass spectrometry. *Eur. J. Biochem.* 1992, *206*, 691–699.

[35] Miyazaki, K., Tsugita, A., C-Terminal sequencing method for proteins in polyacrylamide gel by the reaction of acetic anhydride. *Proteomics* 2006, *6*, 2026–2033.

[36] Rose, K., Savoy, L.-A., Simona, M. G, Offord, R. E. *et al.*, C-Terminal peptide identification by fast atom bombardment mass spectrometry. *Biochem. J.* 1988, *250*, 253–259.

[37] Shevchenko, A., Chernushevich, I., Ens, W., Standing, K. G. *et al.*, Rapid *'de novo'* peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* 1997, *11*, 1015–1024.

[38] Schnölzer, M., Jedrzejewski, P., Lehmann, W. D., Protease-catalyzed incorporation of ^{18}O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* 1996, *17*, 945–953.

[39] Patterson, D. H., Tarr, G. E., Regnier, F. E., Martin, S. A., C-Terminal ladder sequencing *via* matrix-assisted laser-desorption mass-spectrometry coupled with carboxypeptidase-Y time-dependent and concentration-dependent digestions. *Anal. Chem.* 1995, *67*, 3981–3987.

[40] Thiede, B., Wittmann-Liebold, B., Biernert, M., Krause, E., MALDI-MS for C-terminal sequence determination of peptides and proteins degraded by carboxypeptidase Y and P. *FEBS Lett.* 1995, *357*, 65–69.

[41] Samyn, B., Sergeant, K., Castanheira, P., Faro, C. *et al.*, A new method for C-terminal sequence analysis in the proteomic era. *Nat. Methods.* 2005, *2*, 193–200.

[42] Samyn, B., Sergeant, K., Van Beeumen, J., A method for C-terminal sequence analysis in the proteomic era (proteins cleaved with cyanogen bromide). *Nat. Protoc.* 2006, *1*, 318–323.

[43] Münchbach, M., Quadroni, M., Miotto, M., James, P., Quantitation and facilitated *de novo* sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety. *Anal. Chem.* 2000, *72*, 4047–4057.

[44] Miyagi, M., Nakao, M., Nakazawa, T., Kato, I. *et al.*, A novel derivatization method with 5-bromonicotinic acid N-hydroxysuccinimide for determination of the amino acid sequences of peptides. *Rapid Commun. Mass Spectrom.* 1998, *12*, 603–608.

[45] Wang, D., Kalb, S. R., Cotter, R. J., Improved procedures for N-terminal sulfonation of peptides for matrix-assisted laser desorption/ionization post-source decay peptide sequencing. *Rapid Commun. Mass Spectrom.* 2004, *18*, 96–102.

[46] Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B. *et al.*, Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cellular Proteomics* 2004, *3*, 1154–1169.

[47] Noga, M. J., Lewandowski, J. J., Suder, P., Silberring, J., An enhanced method for peptides sequencing by N-terminal derivatization and MS. *Proteomics* 2005, *5*, 4367–4375.

[48] Okamura, T., Iwamura, T., Ito, A., Kaneko, M. *et al.*, Application of bis(terpyridine)ruthenium(II) to N-terminal amino acid sequencing. *Chem. Lett. (Tokyo)* 2005, *34*, 332–333.

[49] Gevaert, K., Goethals, M., Martens, G. R., Van Damme, J. *et al.*, Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat. Biotech.* 2003, *21*, 566–569.

[50] Yamaghchi, M., Nakazawa, T., Kuyama, H., Obama, T. *et al.*, High-throughput method for N-terminal sequencing of proteins by MALDI mass spectrometry. *Anal. Chem.* 2005, *77*, 645–651.

[51] Kuhn, K., Thompson, A., Prinz, T., Müller, J. *et al.*, Isolation of N-terminal protein sequence tags from cyanogen bromide cleaved proteins as a novel approach to investigate hydrophobic proteins. *J. Proteome Res.* 2003, *2*, 598–609.

[52] Zhou, C., Zhang, Y., Qin, P., Liu, X. *et al.*, A method for rapidly confirming protein N-terminal sequences by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 2006, *20*, 2878–2884.

[53] Yamaguchi, M., Obama, T., Kuyama, Nakayama, D. *et al.*, Specific isolation of N-terminal fragments from proteins and their high-fidelity *de novo* sequencing. *Rapid Commun. Mass Spectrom.* 2007, *21*, 3329–3336.

[54] Keough, T., Lacey, M. P., Youngquist, R. S., Derivatization procedures to facilitate *de novo* sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 2000, *14*, 2348–2356.

[55] Nakayama, D., Yamaguchi, M., Kuyama, H., Ando, E. *et al.*, Development of a universal method for selective isolation of N-terminal peptides from proteins and their *de novo* sequencing. *Proc. 55th ASMS Conf. Mass Spectrometry and Allied Topics* 2007, Indianapolis, IN, June 3–7.

[56] McDonald, L., Beynon, R. J., Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. *Nat. Protoc.* 2006, *1*, 1790–1798.

[57] Nakazawa, T., Chemistry in proteomics: an interplay between classical methods in chemical modification of proteins and mass spectrometry at the cutting edge. *Current Proteomics*, 2006, *3*, 33–54.

[58] Nakazawa, T., Yamaguchi, M., Nishida, K., Kuyama, H. *et al.*, Enhanced responses in matrix-assisted laser desorption/ionization mass spectrometry of peptides derivatized with arginine *via* a C-terminal oxazolone. *Rapid Commun. Mass Spectrom.* 2004, *18*, 799–807.

[59] Yamaguchi, M., Oka, M., Nishida, K., Ishida, M. *et al.*, Enhancement of MALDI-MS spectra of C-terminal peptides by the modification of proteins *via* an active ester generated *in situ* from an oxazolone. *Anal. Chem.* 2006, *78*, 7861–7869.

[60] Sechi, S., Chait, B. T., A method to define the carboxyl terminal of proteins. *Anal. Chem.* 2000, *72*, 3374–3378.

[61] Miyazaki, K., Tsugita, A., C-Terminal sequencing method for peptides and proteins by the reaction with a vapor of perfluoric acid in acetic anhydride. *Proteomics* 2004, *4*, 11–19.

[62] Kondo, J., Ohuchi, C., Method for purifying and isolating carboxyl-terminal peptides. *US Patent 5 104 973*, April 14, 1992.

[63] Rabilloud, T., in: Veenstra, T. D., and Yates III, J. R. (Eds.) *Proteomics for Biological Discovery*, John Wiley & Sons, New Jersey 2006, pp. 111–133.

[64] Zimmermann, J., Voss, H., Schwager, C., Stegemann, J. *et al.*, Automated Sanger dideoxy sequencing reaction protocol. *FEBS Lett.* 1988, *233*, 432–436.

[65] Hjernø, K., Protein identification by peptide mass fingerprinting. *Methods Mol. Biol.* 2007, *367*, 61–75.

[66] Nesvizhskii, A. I., Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol. Biol.* 2007, *367*, 87–119.

Note added in proof

The authors have recently developed a simple and efficient method for C-terminal sequencing of proteins with MS by employing succinimidyloxycarbonylmethyl tris- (2,4,6-trimethoxyphenyl) phosphonium bromide and DITC [Kuyama, H., Shima, K., Sonomura, K., Yamaguchi, M. *et al.*, A simple and highly successful C-terminal sequence analysis of proteins by mass spectrometry. *Proteomics* 2008 (in press)].

C-terminal Sequence Analysis of Protein by MS

Efficient Isolation and *de novo* Sequencing

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 structures of target proteins, 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 succinimidyloxycarbonylmethyl tris (2,4,6-trimethoxyphenyl) phosphonium bromide (TMPP-Ac-OSu) and mass spectrometry (MS).

From left: Prof. Susumu Tsunasawa, Associate Prof. Hiroki Kuyama, Prof. Osamu Nishimura, Institute for Protein Research, Osaka University

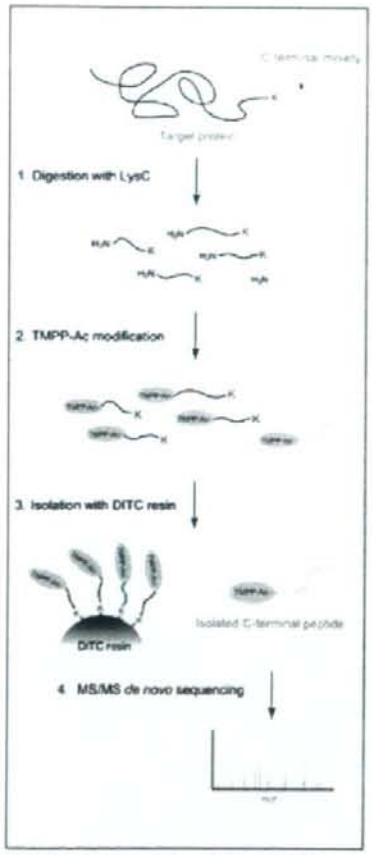

Fig. 1: Schematic overview of the method
A protein is first digested with LysC to yield peptides incorporating amino groups at both ends but the C-terminal peptide contains only an α-amino group. This is followed by selective attachment of the TMPP-Ac moiety to α-amino groups, and isolation of the C-terminal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

Why C-terminal peptide? Why *de novo* Sequencing?

Determining amino acid sequences of proteins, including their post-translational modifications (PTMs), is highly important in life sciences because the covalent structures are directly responsible for the biological functions of individual proteins. A "bottom-up" (or "shotgun") approach by MS has been generally used to determine covalent structures of proteins [1]. This usually involves peptide mass fingerprinting (PMF) or MS/MS ion search by mass spectrometry.

However, the amino acid sequence of a mature protein is usually altered from that expected from the DNA sequence due to processing such as editing and splicing of RNA before translation, and/or various PTMs that occur to newly expressed proteins. Therefore, widely used methods such as PMF or MS/MS ion search often fail to identify post-translationally modified proteins, or to detect sequence polymorphisms observed in various protein isoforms because the database is not sufficiently comprehensive for an effective search for any type of protein. In these cases, *de novo* sequencing, where peptide sequences are interpreted directly from the MS/MS spectra, is the only practical approach by MS.

The advantage of *de novo* sequencing using N- and C-terminal peptide ("Terminal proteomics") has been well acknowledged because it can be applied globally to peptide sequencing for structural determination and/or identification of protein [2]. One of the most advantageous features of terminal sequencing is that as few as four residues from N-terminal amino acid sequence have proven sufficient for specifying 43 to 83% of proteins, and those from the C-terminal counterpart further raise the success rate to 74 to 97% of proteins [3]. Hence, *de novo* sequencing of terminal parts of a protein affords much higher reliability of search results. Of both terminal analyses, several practical methods for N-terminal sequencing have been reported [2], but any satisfactory methods for C-terminal sequencing have not been developed. Therefore, 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.

Here in this article, simple and highly successful method of isolation of C-terminal peptide from protein followed by its *de novo* sequencing by MS is described [4].

Outline of the Method

The outline of this method is illustrated in figure1. First, a protein of interest is digested with lysylendopeptidase (LysC) to yield peptides that have two amino groups at their termini, except the C-terminal peptide which has only one amino group at the N-terminus. In case a protein of interest is N-terminally blocked by any PTM, the N-terminal peptide has one amino group on the side chain of lysine residue. TMPP-Ac derivatisation with TMPP-Ac-OSu to produce TMPP-Ac peptides is then carried out in the next step [5]. These peptides are selectively modified at the α-amino group (not at the ε-amino group on the lysine residue). Hence, the resulting 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 p-phenylenediisothiocyanate (DITC) resin (or glass). The C-terminal peptide thus recovered in the elution solution is subjected to *de novo* sequencing analysis by MALDI-MS.

An Example: Isolation of C-terminal Peptide from BSA and its *de novo* Sequencing

Here we show the result of C-terminal sequencing analysis of bovine serum albumin (BSA). Thirty pmol of BSA was dissolved in 5 µl of urea solution (8M). To the solution was added 3 µl of Tris (2-carboxyethyl) phosphine hydrochloride solution (1 nmol/µl) and reacted for 1 h at 37°C, followed by the addition of iodoacetamide solution (3 µl of 3 nmol/µl solution). After the S-reduction and alkylation, the reaction solution was diluted with 24 µl of 50 mM $NaHCO_3$ solution and 5 µl of acetonitrile followed by the addition of 9 µl of LysC (5 ng/µl) for digestion for 16 h at 37°C. Then 2 µl of TMPP-Ac-OSu solution (10 nmol/µl) was added to the digest and incubated for 30 min in a sonication bath. To the pre-washed DITC glass (5 mg) was added the TMPP modified solution, which was allowed to stand for 2 h in a water bath at 60°C. The extraction was carried out using acetonitrile-50 mM $NaHCO_3$ (1:9, 20 µl; twice) and 2-propanol-acetonitrile-0.1%TFA (1:1:2, 20 µl; three times). The extracts were combined and dried in a vacuum centrifuge. MALDI-TOF mass spectra were recorded on an AXIMA TOF² (Shimadzu/Kratos) mass spectrometer. Figure 2 presents the MALDI-TOF mass spectra of the TMPP-Ac-modified peptide mixture (fig. 2a) and an isolated C-terminal fragment (fig. 2b). The isolated C-terminal peptide was then sequenced by tandem mass spectrometry (CID mode). The spectrum is plotted in figure 3. Sequencing was performed using mainly a-type ions (a_1-a_{10}) as well as using the helpful structural information from characteristic d-type ions (d_1-d_7, and d_9) that occurred due to the elimination of the side chain from certain amino acid residues [6]. The C-terminal amino acid sequence of BSA was thus determined to be LVVSTQTALA. For the results of other proteins, refer to the report [4].

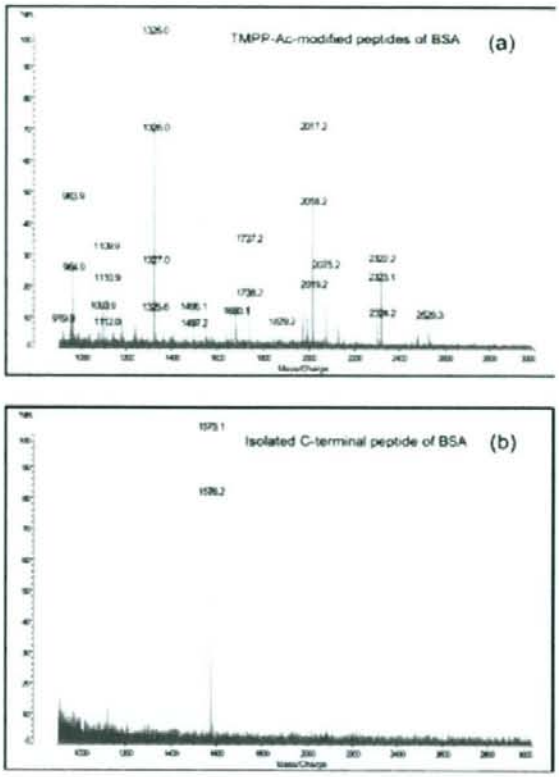

Fig. 2: MALDI-TOF mass spectra of (a) TMPP-Ac-modified peptides after LysC digestion, (b) Isolated TMPP-Ac-modified C-terminal peptide after DITC glass treatment

Discussion

In case the isolated C-terminal peptide incorporates arginine residue(s), *de novo* sequence analysis of the peptide may fail because arginine-containing peptide does not fragment well in a structurally informative manner. We applied a derivatisation protocol [7] to TMPP-Ac-peptide incorporating arginine residue(s), after which the problem was solved and *de novo* sequence analysis was successfully performed [8].

Proteolytic fragmentation was carried out with LysC. If an isolated C-terminal peptide is larger than m/z 3000, though such cases are not many, the *de novo* sequence analysis may be problematic. One solution to such a problem is sub-digesting the large fragment with other proteases (such as GluC, trypsin, and chymotrypsin) followed by TMPP modification, which may yield an appropriate fragment length for *de novo* sequencing by mass analysis. Alternatively, the sequence of

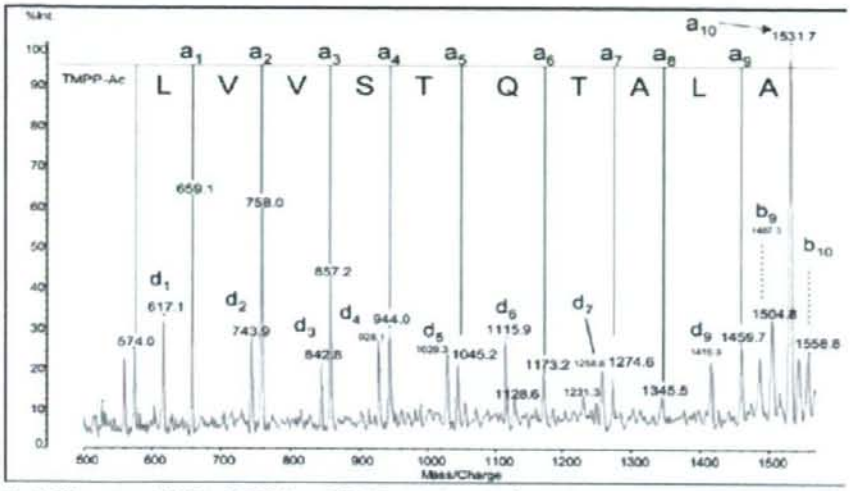

Fig. 3: CID spectrum of isolated TMPP-Ac-modified C-terminal peptide from BSA

the C-terminal moiety of a protein can easily be estimated using the observed m/z value of the isolated C-terminal peptide, since the predicted, full-length sequence of the protein can be obtained by searching a database such as an MS/MS ion search or PMF.

The concept of this method was successfully applied to the sequence analysis of N-terminal peptide [9], which indicates that this methodology may have practical applicability for isolation/separation of peptides.

Conclusion

We have developed a simple and highly successful method to perform C-terminal sequencing of proteins, using a combination of chemical/enzymatic procedures (LysC digestion, modification of α-amino groups with TMPP-Ac-OSu and isolation of the C-terminal peptide using DITC resin or glass) and *de novo* sequence analysis by MALDI-TOF-MS. We believe the method described here is widely applicable to high-throughput C-terminal analysis of proteins.

References

[1] Chait B.T.: Science 314, 65–66 (2006)

[2] Nakazawa T. *et al.*: Proteomics 8, 673–685 (2008)

[3] Wilkins M.R. *et al.*: J. Mol. Biol. 278, 599–608 (1998)

[4] Kuyama H. *et al.*: Proteomics 8, 1539–1550 (2008)

[5] Huang Z-H. *et al.*: Anal. Biochem. 268, 305–317 (1999)

[6] Huang Z-H. *et al.*: Anal. Chem. 69, 137–144 (1997)

[7] Dikler S. *et al.*: J. Mass Spectrom. 32, 1337–1349 (1997)

[8] Kuyama H. *et al.*: Rapid Commun. Mass Spectrom. 22, 2063–2072 (2008)

[9] Kuyama H. *et al.*: Anal. Biochem. 380, 291–296 (2008)

► www.eMagazineBIOforum.com

CONTACT

Hiroki Kuyama
Associate Professor
Division of Disease Proteomics
Institute for Protein Research
Osaka University, Japan
Tel. and Fax: +81 6 6879 4320
kuyama@protein.osaka-u.ac.jp

The 12th Annual European Conference on Micro & Nanoscale Technologies for the Biosciences

17–19 November 2008, Eden Palace Hotel Montreux, Switzerland:

This unique conference at the frontier of micronanotech and bio applications will feature high level invited speakers as well as selected presentations on current topics related to micro- and nanotechnologies in the biosciences. Participants from over 15 countries in Europe, America and Asia meet in a variety of contexts ideal for creative exchange of ideas and new project developments.

The popular NanoTech Short Course will be given 15–16 November 2008. This short course gives newcomers to the field an introduction to micro- and nanotechnologies.

The Exhibition provides an excellent exposure for startup companies.

NanoTech ~Montreux

www.nanotech-montreux.com

RAPID COMMUNICATIONS IN MASS SPECTROMETRY
Rapid Commun. Mass Spectrom. 2009; **23**: 611–618
Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/rcm.3920

RCM

The specific isolation of C-terminal peptides of proteins through a transamination reaction and its advantage for introducing functional groups into the peptide

Kazuhiro Sonomura[1,2], Hiroki Kuyama[1], Ei-ichi Matsuo[1,2], Susumu Tsunasawa[1] and Osamu Nishimura[1]*

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

Received 29 September 2008; Revised 11 December 2008; Accepted 15 December 2008

A novel method for isolating C-terminal peptides from proteolytic digests of proteins was developed. Proteins were digested with lysyl endopeptidase (LysC) and applied to metal-ion-catalyzed transamination reactions. This reaction enabled the selective conversion of an N^{α}-amino group to a carbonyl group. Subsequent incubation with *p*-phenylenediisothiocyanate (DITC) glass effectively scavenged the lysine-containing N-terminus and internal peptides. The obtained C-terminal peptide is open to modification with reagents having virtually any type of functionality owing to the reactive α-ketocarbonyl group. In this report, 2,4-dinitrophenylhydrazine (DNPH) was used as an example of a nucleophile to the carbonyl group. The isolated C-terminal peptide was modified with DNPH, which exhibited signal enhancement, and was sequenced by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). Copyright © 2009 John Wiley & Sons, Ltd.

The C-terminus of proteins acts as an important determinant of cellular functions such as protein trafficking, degradation, and complex formation.[1] Therefore, the significance in developing C-terminal analytical methodology has been widely acknowledged, and efficient and practical analytical techniques have long been pursued. Mass spectrometry (MS) has been the central detection method in analyzing proteins and peptides, and various methodologies have been developed.

Mass-based methods for protein and peptide analysis can be categorized as 'top-down' and 'bottom-up'. Direct sequencing ('top-down') is a promising technique for analyzing the C-terminal structure,[2,3] but there are still significant impediments to its practical use. In contrast, several 'bottom-up' methodologies, which require pinning down the signal of the C-terminal peptide in the mass spectrum or isolating the target peptide from the peptide pool, have been reported.[4–7] As with the top-down technique, the bottom-up method also harbors significant impediments for efficient and practical C-terminal analysis. For C-terminal analysis using the bottom-up method, isolating the target peptide is beneficial, particularly for downstream operations such as chemical transformation. Therefore, a practical isolation technique has been sought.

We have been engaged in developing terminal sequencing methods by MS. Recently, we developed N- and C-terminal sequencing methods of proteins using tris(2,4,6-trimethoxyphenyl)phosphoniumacetyl (TMPP-Ac) tagging.[7,8] In these methods, differentiating the reactivities of N^{α}- and N^{ε}-amino groups is highly important for isolating these target peptides. TMPP-Ac selective tagging proceeded successfully, and facilitated the subsequent isolation and *de novo* sequencing by MS. However, the selective tagging reaction of the N^{α}-amino group is generally difficult. Hence, we have also been seeking orthogonal reagents or reactions that lead to the selective recovery and sequence analyses of terminal peptides. One of the authors (O.N.) has used chemical transamination[9] which selectively converts N^{α}-NH_2 into a carbonyl group. This prompted us to investigate its feasibility for C-terminal-specific isolation and modification.

The transamination reaction was originally developed in the 1950s and improved by Dixon.[10,11] An advantage of the reaction is its high selectivity for the N^{α}-amino group. The selectivity is achieved by the direct participation of the adjacent carbonyl oxygen. Moreover, this technique introduces a reactive α-diketone moiety, thus enabling further modification with a variety of nucleophiles, such as hydrazine compounds. A very wide variety of applications has been reported, e.g. removal of N-terminal residues from recombinant proteins,[9] N-terminal specific labeling of proteins and peptides,[12–16] and applications to peptide synthesis.[17]

Here we report a new approach for isolating C-terminal peptides of proteins and sequencing utilizing the transamination reaction.

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

RCM

EXPERIMENTAL

Materials

Bovine cathepsin B, pig albumin, bovine α-casein, 1,4-phenylenediisothiocyanate, and iodoacetamide were purchased from Sigma (St. Louis, MO, USA). Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was obtained from Fluka (Switzerland). Pyridine was purchased from Alfa Aesar (UK). Lysyl endopeptidase (LysC), sodium hydrogen carbonate ($NaHCO_3$), sodium acetate, acetic acid, glyoxylic acid monohydrate, copper sulfate pentahydrate, acetonitrile, 2-propanol, dimethylformamide (DMF), methanol and trifluoroacetic acid (TFA) were all obtained from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). α-Cyano-4-hydroxycinnamic acid (CHCA: high-purity mass-spectrometric grade) and 2,5-dihydroxybenzoic acid (DHBA: high-purity mass-spectrometric grade) were purchased from Shimadzu GLC (Tokyo, Japan). Three peptides (MHRQETVDCLK-NH_2, AAKIQASFRGHMARKK, and SFLLR) were obtained from AnaSpec (San Jose, CA, USA). A peptide (VIYHPF) was purchased from Peptide Institute, Inc. (Osaka, Japan). Methanediphosphonic acid (MDPNA) and 2,4-dinitrophenylhydrazine hydrochloride (DNPH) were obtained from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan). DITC glass was prepared according to the procedure of Wachter *et al.*[18] Briefly, 500 mg of aminopropyl glass (Sigma, St. Louis, MO, USA) was added in small portions to 6 mL of 1,4-phenylenediisothiocyanate solution (83 mg/mL, in DMF). The reaction mixture was stirred for 2 h at room temperature. The modified glass was washed to remove excess reagent with DMF and methanol and dried under vacuum.

In-gel digestion with LysC

A sample protein (30 pmol) was separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) in a 12.5% acrylamide gel. The Coomassie-stained protein band was excised, washed with 50% (v/v) acetonitrile in 100 mM $NaHCO_3$, 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 100 μL of iodoacetamide solution (55 mM, in 100 mM $NaHCO_3$). After 45 min incubation at room temperature in the dark, the gel piece was washed with 100 μL of 50 mM $NaHCO_3$, shrunk by dehydration 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 100 ng of LysC. After 5 min, 50 mM of $NaHCO_3$ solution (18 μL) was added to keep the gel piece moist during digestion (37°C, overnight). To extract the resulting peptides, 40 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, followed by collection of the supernatant. This extraction procedure was repeated three times. The extract was combined and lyophilized. The resulting dried material was applied to the transamination reaction.

Transamination of peptides

Transamination was performed with a freshly prepared aqueous solution of 10% pyridine containing 0.2 M glyoxylic acid and 6 mM $CuSO_4$ for 30 min at room temperature. The reaction was also investigated in a sodium acetate buffer (1 M sodium acetate, 5 mM $CuSO_4$, 0.2 M acetic acid, and 0.2 M glyoxylic acid). The transamination reaction was successfully performed under both conditions. The reaction mixture was purified using ZipTip μ-C18 (Millipore, Billerica, MA, USA) according to the manufacturer's protocol. The peptides were eluted using 5 μL of 50% acetonitrile containing 0.05% TFA. An aliquot of the eluted sample was analyzed by MALDI-MS and the rest was dried in a vacuum centrifuge and applied to the following C-terminal isolation step.

Isolation of C-terminal peptides with DITC glass

DITC glass was washed before use with acetonitrile, then with 50 mM $NaHCO_3$ buffer containing 10% acetonitrile. The dried material after transamination was dissolved in 10 μL of 50 mM $NaHCO_3$ buffer containing 10% acetonitrile, applied to 5 mg of DITC glass and incubated for 2 h at 60°C. To extract peptides, 40 μL of acetonitrile/2-propanol/0.1% TFA (1:1:2) was added to the suspension of DITC glass and the supernatant was collected. This extraction procedure was repeated twice, and the recovered solution was dried in a vacuum centrifuge. The dried sample was dissolved in 0.1% TFA and purified with ZipTip μ-C18. An aliquot of the eluted sample was analyzed by MALDI-TOF MS, and the rest was applied to the following modification step.

Modification with 2,4-dinitrophenylhydrazine

Isolated C-terminal peptides were modified with DNPH reagent. The eluted sample (5 μL, in 50% acetonitrile containing 0.05% TFA) was mixed with DNPH solution (5 μL, 1 mg/mL in 50% acetonitrile containing 0.05% TFA) and incubated at 37°C for 1 h. The reaction mixture was directly analyzed by MALDI-TOF MS.

MALDI-TOF MS

A matrix solution was prepared by dissolving 5 mg of CHCA or DHBA in 0.5 mL of 50% acetonitrile containing 0.05% TFA. DHBA was used only for detecting the C-terminal peptide from α_{s1}-casein. We used MDPNA as a matrix additive, which was found to be an effective suppressor of alkali metal adduct formation.[19] MDPNA was dissolved in water to give 2% (w/w) solution. The sample solution, matrix solution and MDPNA solution (0.5 μL each) were mixed on the target plate and air dried. MALDI-TOF mass spectra and tandem mass spectra (post-source decay (PSD) mode) were recorded on an AXIMA-CFR-Plus mass spectrometer or an AXIMA-TOF^2 (Shimadzu, Kyoto, Japan; Kratos, Manchester, UK). The m/z values found in the mass spectra are observed monoisotopic values of single positively charged ions, whereas those in the tandem mass spectrum are averaged. Theoretical monoisotopic masses for peptides and their derivatives are in the text and legends.

DOI: 10.1002/rcm

RESULTS

Outline of the method

Figure 1 illustrates the general workflow for the method. The protocol consists of the following five steps: (1) digestion of proteins with lysyl endopeptidase (LysC); (2) transamination of digested fragments; (3) treatment with DITC glass; (4) modification through hydrazone formation; and (5) MALDI-TOF MS and MALDI-PSD analysis of the modified C-terminal peptide.

LysC cleaves the peptide bond at the C-terminal side of the lysine residue. Peptide fragments in LysC digests contain lysine residues, with the exception of the C-terminal peptide. Transamination converts the N^{α}-amino group into the carbonyl group (the N^{ε}-amino group is not affected). After transamination, peptides have a free amino group in the lysine side chain, but the C-terminal peptide has no free amino group. DITC glass scavenges peptides having free amino groups to leave the target C-terminal peptide in the eluate.

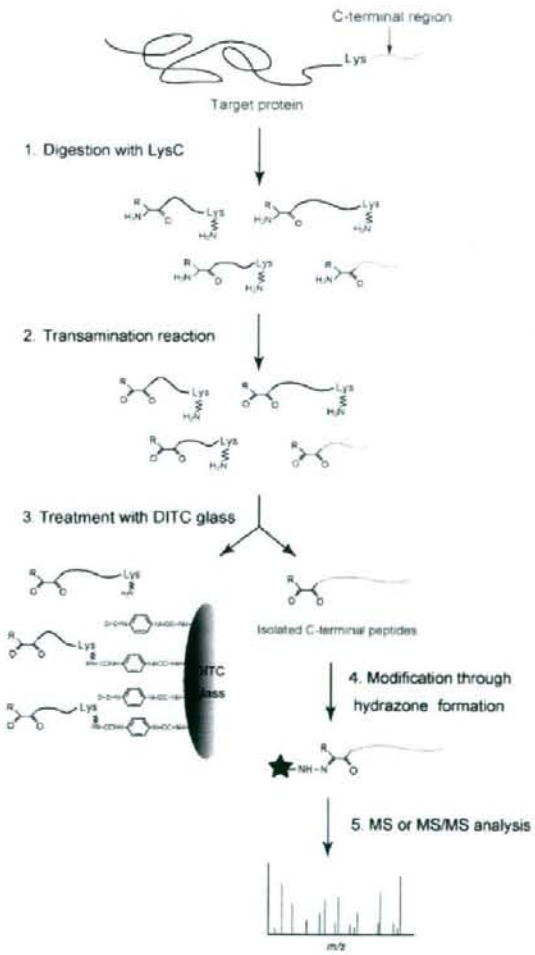

Figure 1. Schematic representation of the protocol for isolating and modifying C-terminal peptides.

The isolated C-terminal peptide has an α-ketocarbonyl group that reacts readily with hydrazine (hydroxylamine) to form hydrazone (oxime). In this paper, we used DNPH reagent to modify the transaminated C-terminal peptide. DNPH is well known as a derivatizing reagent used for determining carbonyl compounds in chromatography methods and has been used to modify 4-hydroxy-2-nonenal-modified peptides for sensitive detection by MALDI-MS.[20]

Estimation of the transamination conversion yield

The transaminated peptide was quantified via 2,4-dinitrophenylhydrazone formation according to the method reported by Fields and Dixon.[21] A model peptide (AAKIQASFRGHMARKK) was used for this purpose. The peptide (1 nmol) was treated in a pyridine solution containing glyoxylic acid and $CuSO_4$ for 30 min at room temperature and purified with ZipTip C18. After evaporation of the eluate from ZipTip C18 to dryness using a vacuum centrifuge, the dried material was dissolved in 4 μL of DNPH solution (2.5 mM, in 1 M HCl) and incubated for 30 min at 37°C. The reaction mixture was then diluted to 50 μL and the absorbance at 370 nm was determined. The amount of the transaminated peptides was calculated from the absorptivity of pyruvoylglycine (ε: $1.2 \times 10^4\ M^{-1}\ cm^{-1}$).[11]

Following this procedure, the conversion yield was calculated to be 70%. This value is the total yield of the three steps (transamination reaction, ZipTip treatment, DNPH modification).

Model experiment using peptide mixture

To investigate the effectiveness of the method, we first applied this strategy to a peptide mixture. The mixture contained three peptides **1** VYIHPF ($[M+H]^+$: 775.4), **2** MHRQETVDCLK-NH_2 ($[M+H]^+$: 1358.7), and **3** AAKIQASFRGHMARKK ($[M+H]^+$: 1800.0). We used an equimolar mixture of these three peptides (100 pmol each), which constituted a model for a LysC digest of proteins. LysC digestion produces peptide fragments having a lysine residue, except for the C-terminal peptide. Hence, peptide **1** corresponds to the C-terminal peptide, while the other two are N-terminal or internal peptides. It was expected that peptide **1** would be isolated after transamination and treatment with DITC glass because only peptide **1** is transformed to a peptide devoid of a free amino group that would covalently attach to isothiocyanate groups on DITC glass beads, while the other two peptides still contain ε-NH_2 that survived the transamination reaction and therefore would bind to the glass beads. The mass difference between the peptides before and after the transamination reaction is 1 Da, which can be detected by MS (Fig. 2).

The three peptides (100 pmol each) were dissolved in 10 μL of a pyridine solution containing glyoxylic acid and $CuSO_4$ and reacted for 30 min at room temperature. The reaction mixture was purified with ZipTip μ-C18 and applied to DITC glass. Figure 3(a) depicts the MALDI-TOF mass spectra of the peptide mixture before the transamination reaction. After transamination, the three peptides were detected with a mass decrease of 1 Da (Fig. 3(b)). As

DOI: 10.1002/rcm