Journal of Chromatography B, 874 (2008) 101–110

Contents lists available at ScienceDirect

Journal of Chromatography B

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

Detection of C-terminal peptide of proteins using isotope coding strategies

Samir Julka*, Demetrius Dielman, Scott A. Young

The Dow Chemical Company, Analytical Sciences, 1897 Building, Midland, MI 48667, USA

ARTICLE INFO

Article history:
Received 23 July 2008
Accepted 15 September 2008
Available online 19 September 2008

Keywords:
C-terminal detection
Isotope coding
O^{18}
Acylation
Esterification
LC–ESI–MS

ABSTRACT

The determination of C-terminal peptide sequence is critical since the C-terminal peptide contains biologically relevant information and often undergoes post-translational processing. Another important application is in estimating purity of the biopharmaceuticals, especially for determining the presence of ragged processed ends and for N-terminally blocked polypeptides and proteins. In this paper, different isotope coding strategies in combination with reversed phase chromatography (RPC) coupled with electrospray ionization–mass spectrometry (ESI–MS) were evaluated to detect the C-terminal peptide from proteolytic digests. These were (i) O^{18} (ii) acylation and (iii) esterification based isotope coding strategies. Using reversed phase chromatography, the C-terminal peptide was resolved from other internal peptides. The isotope coding approaches specifically rendered a characteristic MS signature to the C-terminal peptide, thereby facilitating its detection. The unique MS signature, along with accurate mass data for the C-terminal peptide was found to be sufficient for its detection and identification. The advantages and limitations of the three approaches will be discussed.

1. Introduction

Analysis of N- and C-terminal peptide is important in estimating the purity of proteins for therapeutic product registration [1]. The detection of correct N- and C-terminal sequence assists in ascertaining that the protein has been correctly translated based upon the open reading frames of corresponding mRNA [2]. Moreover, during the past decade, federal regulatory agencies have come to depend heavily on peptide mapping by reversed phase chromatography (RPC) and RPC–MS to confirm primary structure and regulate the production of biopharmaceuticals [3]. Detection of the correct C-terminal peptide is critical since it (i) contains biologically relevant information [4–6], (ii) often has ragged processed ends and undergoes post-translational processing [1,7] and (iii) in cases where the N-terminal peptide is blocked [8]. Although, a number of methods have been developed to detect N-terminal peptides, only limited literature precedence is available for detection of C-terminal peptides.

One of the popular strategies to determine the C-terminal peptide of a protein involves C-terminal processing of the protein by carboxypeptidase followed by comparative peptide mapping of the intact and the carboxypeptidase-excised protein [8–10]. The above comparative method is tedious, sample intensive, and requires careful optimization of carboxypeptidase digestion. A few oxazolone based approaches to detect the C-terminal peptide have also been reported [2,7,11,12]. In a recent approach, oxazolone derivatives at the C-terminal end of the protein were formed followed by further derivatization to selectively facilitate the signal enhancement of the C-terminal peptide [12]. Although successful, the strategy requires multiple derivatizations and the hydrazide conversion for signal enhancement typically has a low yield.

During the last decade, isotope coding-based strategies have been routinely utilized for expression based bottom-up proteomics applications, and in facilitating MS/MS interpretation [13]. Of the different isotope based strategies, global coding approaches directed either at the C-terminus or at the N-terminus of peptides have the potential for identification of the C-terminal peptide of a protein of interest. Some of these approaches were previously utilized for detection of the C-terminal peptide of proteins [14–17]. The strategies are based on the premise that by applying different derivatization based isotope coding approaches followed by reversed phase chromatography–mass spectrometry analyses, the C-terminal peptide can be specifically identified by its characteristic MS signature in comparison to the internal peptides.

The main intent of this study was to develop an approach for detection of C-terminal peptide of purified protein/s in context with biopharmaceutical applications mentioned earlier. In this paper, the feasibility of three different isotope coding strategies to identify the C-terminal peptide of a protein were compared. Also, an acylation-based isotope coding-based approach to detect the C-terminal peptide is described. The three strategies that were investigated were: (i) O^{18}, (ii) acylation and (iii) esterification.

* Corresponding author. Tel.: +1 989 636 1533; fax: +1 989 638 6443.
E-mail address: SJulka@dow.com (S. Julka).

1570-0232/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2008.09.012

The general schematic of the approaches is described in Fig. 1. Although the objective is to detect the C-terminal peptide of the protein of interest, the three approaches have unique properties. The O^{18} approach specifically targets the carboxylate group at the C-terminus of only the internal peptides. The esterification based approach targets all carboxylate centers of all peptides (C-terminus carboxylate group, side chains of glutamic acid and aspartic acid) while the acylation based approach targets primary amine centers of all peptides (N-terminus, side chain of lysine). Moreover, in the O^{18} approach, coding is performed during proteolysis. In contrast, coding is performed after proteolysis in both esterification and acylation-based approaches. The following criteria were utilized in order to develop a suitable C-terminal peptide detection approach. The successful approach for C-terminal peptide detection needs to have (i) broad applicability, (ii) ease of use, (iii) requires minimum derivatization steps, (iv) sensitive/not sample intensive and (v) yield no/minimum side reactions.

2. Experimental

2.1. Materials

The proteins, ribonuclease A (RNAase A; bovine), cytochrome *c* (bovine), lysozyme (chicken), myoglobin (equine) and lactoglobulin (bovine), were purchased from Sigma (St. Louis, MO). The following reagents were purchased from the specified companies: acetic anhydride (d_0) and acetic anhydride (d_6); acetoyl chloride, methanol (d_3), methanol (d_0), ammonium bicarbonate, poly-DL-alanine, Leucine Enkephalin and iodoacetamide (IAM) were purchased from Sigma (St. Louis, MO). Tris was purchased from Bio-Rad (Hercules, CA). Guanidine hydrochloride (Gu:HCl) and dithiothretiol (DTT) were purchased from Pierce Biotechnology (Rockford, IL). Trypsin, Glu-C and Lys-C were purchased from Roche Diagnostics (Indianapolis, IN). Formic acid (FA) 96%, and ammonium formate were purchased from Fisher Scientific (Pittsburg, PA). HPLC-grade acetonitrile (AcN) was purchased from J.T. Baker (Phillipsburg, NJ). For all analyses, Milli-Q (Millipore, Billerica, MA) deionized water was used.

2.2. Proteolysis of cytochrome c, myoglobin, lysozyme, lactoglobulin and RNAase A

Approximately 720 μL of protein dissolution buffer (6 M guanidine hydrochloride/0.4 M ammonium bicarbonate, pH 7.8) was added separately to 1 mg of RNAase A, lactoglobulin, myoglobin, cytochrome *c* and lysozyme. This was followed by addition of 20 μL of 100 mM DTT (reducing reagent) solution. Microtubes were sealed, vortexed, and incubated at 65 °C for 40 min in an Eppendorf thermomixer R (Brinkmann, Westbury, NY) at 1100 rpm. Microtubes were then cooled to room temperature, centrifuged at 6000 rpm for 30 s, and 40 μL of 200 mM IAM (alkylating reagent) solution was added to each microtube. Samples were incubated in the dark at room temperature for 1 h. Approximately 80 μL of DTT solution was added to consume the unreacted IAM and the microtubes were allowed to stand for 20 min at room temperature. The total reaction volume was approximately 300 μL. The reduced/alkylated protein samples were desalted using NAP-5 gravity cartridges (Sephadex G-25) as per the manufacturer's procedure (GE Healthcare, Piscataway, NJ). NAP-5 cartridges were pre-equilibrated with 100 mM Tris buffer, pH 8.4, and protein elution was performed with the same buffer (final volume 1 mL).

2.3. Enzymatic digestion/O^{18} coding

Before O^{18} coding, the desalted protein was completely evaporated to dryness. To the dried protein sample was added 150 μL of H_2O^{18} and 100 μL of Tris buffer and 50 μL of protease solution (prepared in 100 mM Tris buffer (H_2O^{16})). This ensured that the H_2O^{16} and H_2O^{18} were present in a 1:1 ratio. The protein: enzyme concentration was maintained at approximately 20:1 ratio. The

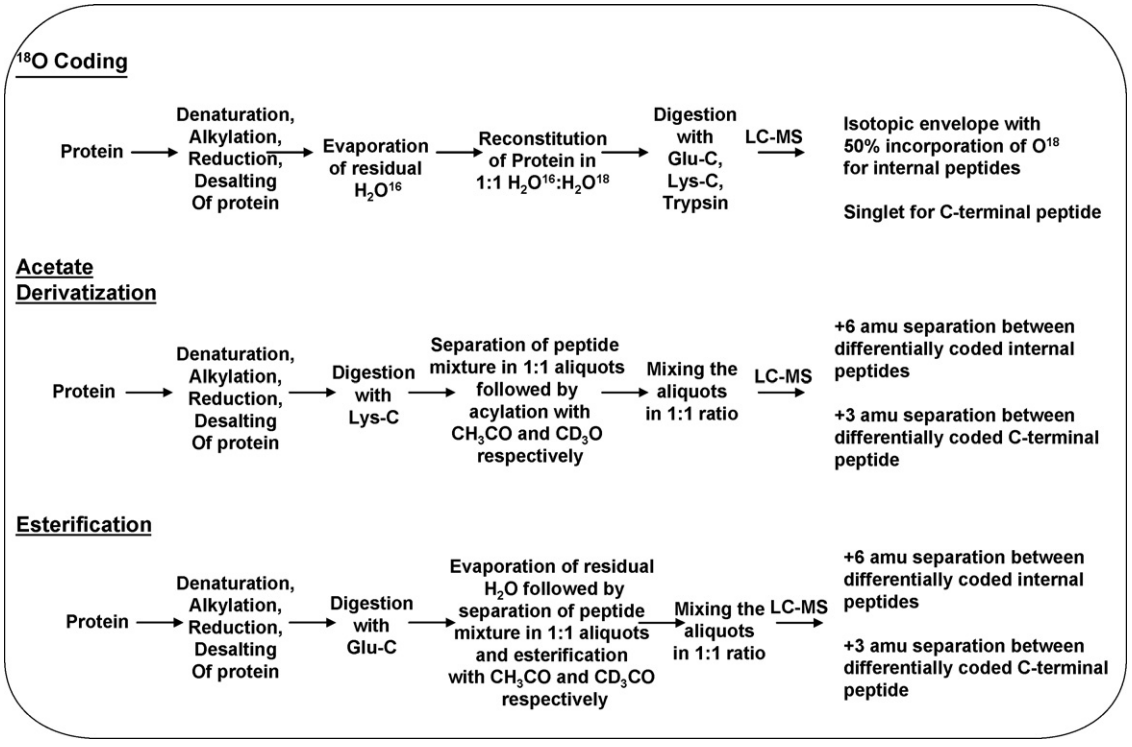

Fig. 1. Schematic presentation of the strategies for selectively coding the C-terminal peptide of proteins.

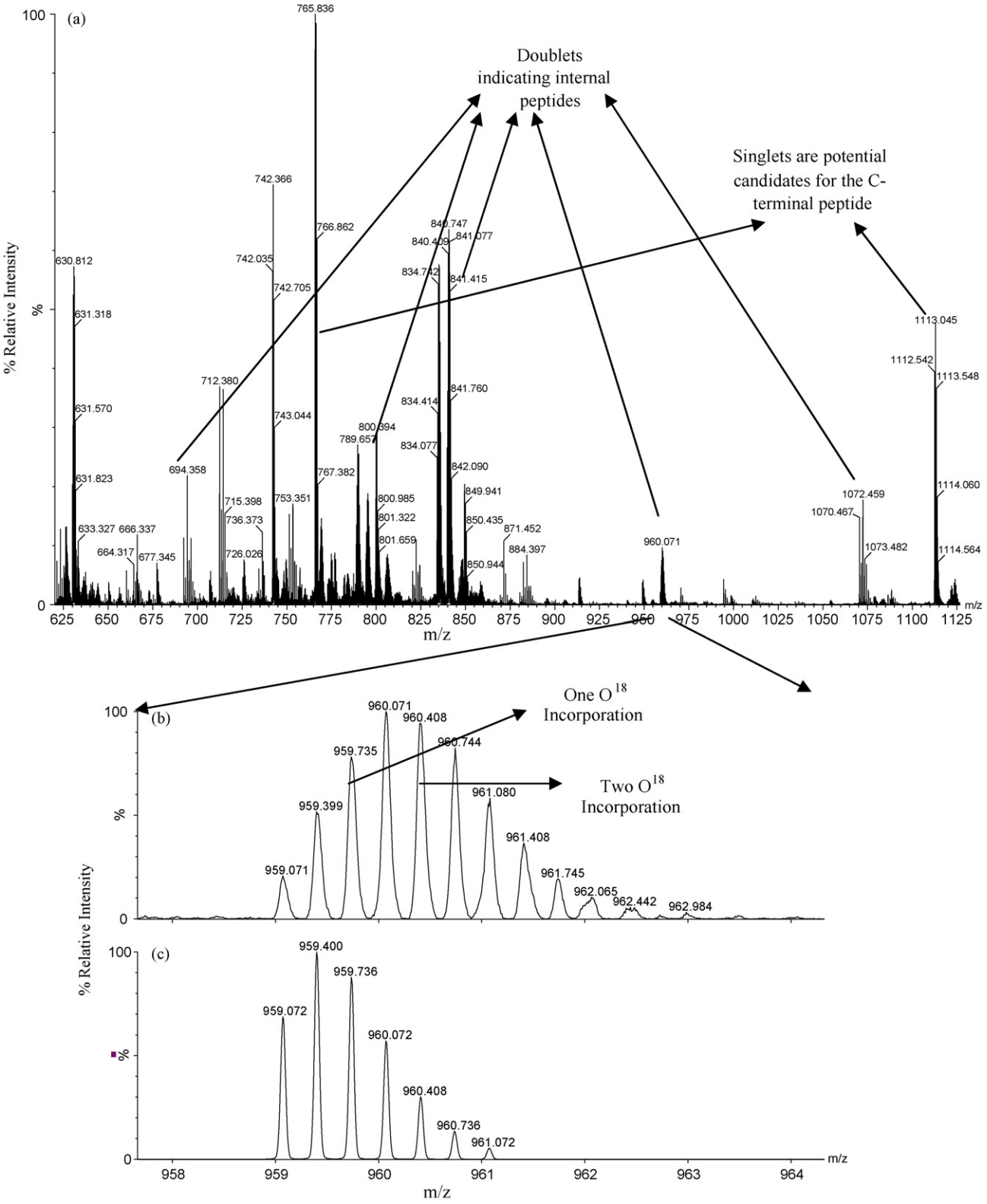

Fig. 2. (a) Mass spectrum of a tryptic digest of RNAase A in H_2O^{16}:H_2O^{18} Tris buffer. The doublets (due to differential coding in H_2O^{16}:H_2O^{18} water) are conclusive evidence for internal peptides. The peptides with singlets (no evidence of incorporation of the O^{18} tag) are potential candidates for the C-terminal peptide of protein, (b) inset of observed isotope envelope for molecular ion $[M+H]^{3+}$ of an internal peptide doublet, NGQTNCYQSYSTMSITDCRETGSSK, depicting incorporation of O^{18} tag and (c) theoretical isotope envelope with no O^{18} incorporation.

digestion was carried out overnight at 37 °C for respective proteolytic digestions with trypsin and Lys-C. The digestion was carried out overnight at 25 °C for Glu-C digestion.

2.4. Acylation

Acylation was carried out only with proteolytic Lys-C digests. The respective proteolytic Lys-C digests were divided into two equal aliquots. The derivatization protocol was followed as described previously [18]. A 50-fold molar excess of derivatization reagent; i.e., d_0 and d_6 acetic anhydride, was added individually to the peptide mixtures. Acylation was allowed to proceed for 1 h at room temperature. This was followed by mixing the two aliquots and heating the final solution at 100 °C for approximately 3 h to remove the partial O-acylation [19]. This was followed by subsequent RPC–MS analysis.

2.5. Glu-C digestion/esterification

In order to achieve proteolytic cleavage at both glutamic acid and aspartic acid residues, digestion with Glu-C protease was carried out for approximately 24 h at room temperature. The protein:enzyme concentration was maintained at approximately a 10:1 ratio. The desalted Glu-C proteolytic digest was divided into two equal aliquots. This was followed by evaporating the respective aliquots to dryness. The two aliquots were then each reconstituted with freshly prepared methanolic HCl as described previously [20]. The derivatization was carried out for 2 h, and was stopped by evaporating the solution to dryness. The two aliquots were reconstituted in 100 μL of Mobile Phase A and mixed in equal volumes for subsequent LC–MS analysis.

2.6. Reversed phase chromatography of peptides

The peptides were separated by linear gradient elution from a ACQUITY BEH C18 column (2.1 mm × 150 mm; C18; 1.7 μm; 135 Å) on an ACQUITY UPLC® system (Waters, Framingham, MA). The C18 column was equilibrated using 95% mobile phase A (0.1% FA in ddI H_2O (double distilled water)) at a flow rate of 100 μL/min for 5 min. The peptide mixtures (2 nmol) were injected and eluted at a flow rate of 100 μL/min in a linear gradient ranging over 40 min from 95% mobile phase A to 45% mobile phase B (0.1% FA in AcN). At the end of this period, a second linear gradient was applied in 5 min from 45% B to 90% B and held at 90% B for 5 min at the same flow rate. The column was then re-equilibrated with 95% A for 14 min before injection of next peptide mixture. Throughout the analysis, an online photo-diode array detector was used to monitor separation of the peptide mixtures.

2.7. LC–MS and MS/MS analysis

Electrospray ionization (ESI) was performed using a Z spray source equipped with a lockspray interface was used. The Time of Flight (ToF) analyzer was calibrated prior to analysis using a 20 μM solution (0.1% formic acid in 98% acetonitrile) of PolyAlanine at 5–7 μL/min flow rate. Data acquisition was performed with a cycle time of 1 scan/s (scan acquisition time: 0.88 s; interscan delay: 0.1 s) in the MS mode. The lockspray reference ion used was the singly charged Leu Enkephalin ion at 556.2771 amu. The acquired mass spectrometry data was manually processed. The scans for each individual chromatographic peak were summed, centroided and mass measured. The experimental m/z was then compared to the theoretical m/z obtained by *in-silico* digest of the protein. In MS/MS mode, default collision energy was set to 23 V, and the mass scan range was set to 100–1900 amu. Tandem MS data acquisition used user-specified m/z analysis mode (where the precursor ion to be chosen was specified in a file). Collision energies were also specified for each of the precursor ions.

3. Results and discussion

3.1. O^{18} coding strategy

The premise of this strategy is that when enzymatic digestion of proteins is carried out in H_2O^{18} containing buffer (pH 7.5–8.5), all specifically cleaved internal peptides will exchange the O^{18} tag except for the peptide originating from the carboxy-terminus of the protein [14–16,21]. In the O^{18} strategy for detection of C-terminal peptide, the protein is first denatured, reduced and alkylated. This is followed by removal of salts using the Sephadex columns. The resulting solution is then evaporated to dryness to eliminate any residual H_2O^{16}. After complete evaporation, the protein is reconstituted in a 50% H_2O^{18} and 50% 100 mM Tris buffer (H_2O^{16}) and digested with a specific protease as described in Section 2.

Four model proteins, RNAase A, lactoglobulin, cytochrome *c* and myoglobin were selected for studying the feasibility of O^{18} strategy with trypsin as a protease. Since the number of peptides in the proteolytic digests is increased by a factor of two due to coding of each peptide with light and heavy isoforms of reagent, it is critical to separate the components before mass spectrometric analyses. After O^{18} derivatization, differentially coded proteolytic peptide pairs were separated by reversed phase chromatography followed by detection using electrospray ionization–mass spectrometry. Fig. 2a represents the broad band mass spectra of tryptic digest of RNAase A with internal peptides depicting incorporation of O^{18} tag. The singlets (due to lack of the O^{18} tag) are potential candidates for the C-terminal peptide of the protein. Fig. 2b and c show the inset for experimentally measured isotope envelope for molecular ion $[M+H]^{3+}$ of peptide doublet, NGQTNCYQSYSTMSITDCRETGSSK at 959.0 Da and its theoretically calculated isotope envelope with no O^{18} incorporation, respectively. Based on their respective isotope envelopes, it can be easily inferred that the peptide was an internal tryptic peptide. Table 1 provides a summary of the peptide mass fingerprinting of RNAase A coupled with O^{18} strategy. Based on scan-to-scan analyses of tryptic digest of RNAase A, a characteristic MS signature was observed for all internal tryptic peptides. As observed in the table, two peptides which distinctly did not incorporate the O^{18} tag were potential candidates for the C-terminal peptide of RNAase A. Based on the accurate mass analysis, it was determined that the two peptides originated from the C-terminal peptide of RNAase A. The molecular ion $[M+H]^{2+}$ at 1112.542 Da was the actual complete C-terminal peptide, HIIVACEGNPYVPVHFDASV. The molecular ion $[M+H]^{2+}$ at 765.86 Da (EGNPYVPVHFDASV), resulted from partial non-specific cleavage in the C-terminal peptide of RNAase A. Further confirmation of their peptide sequence was also achieved by tandem MS analysis. A few more singlets with very low relative intensity compared to the peptides eluting with in the chromatographic peak were observed. These singlets were primarily attributed to partial in-source fragmentation of the eluting peptides. A total of 96% sequence coverage was observed for the RNAase A protein.

The C-terminal peptide was successfully determined for all the four model proteins. Table 2 summarizes the characteristic MS signature observed for the C-terminal peptides. Fig. 3a depicts the lack of incorporation of the O^{18} tag into the C-terminal peptide for RNAase A. The overlap of theoretical and observed isotope envelopes indicates absence of any O^{18} incorporation. Obviously with trypsin as a protease, this approach will not be useful if the C-terminal amino acid of the protein is either lysine or arginine. It would also be of limited utility when the C-terminal peptide orig-

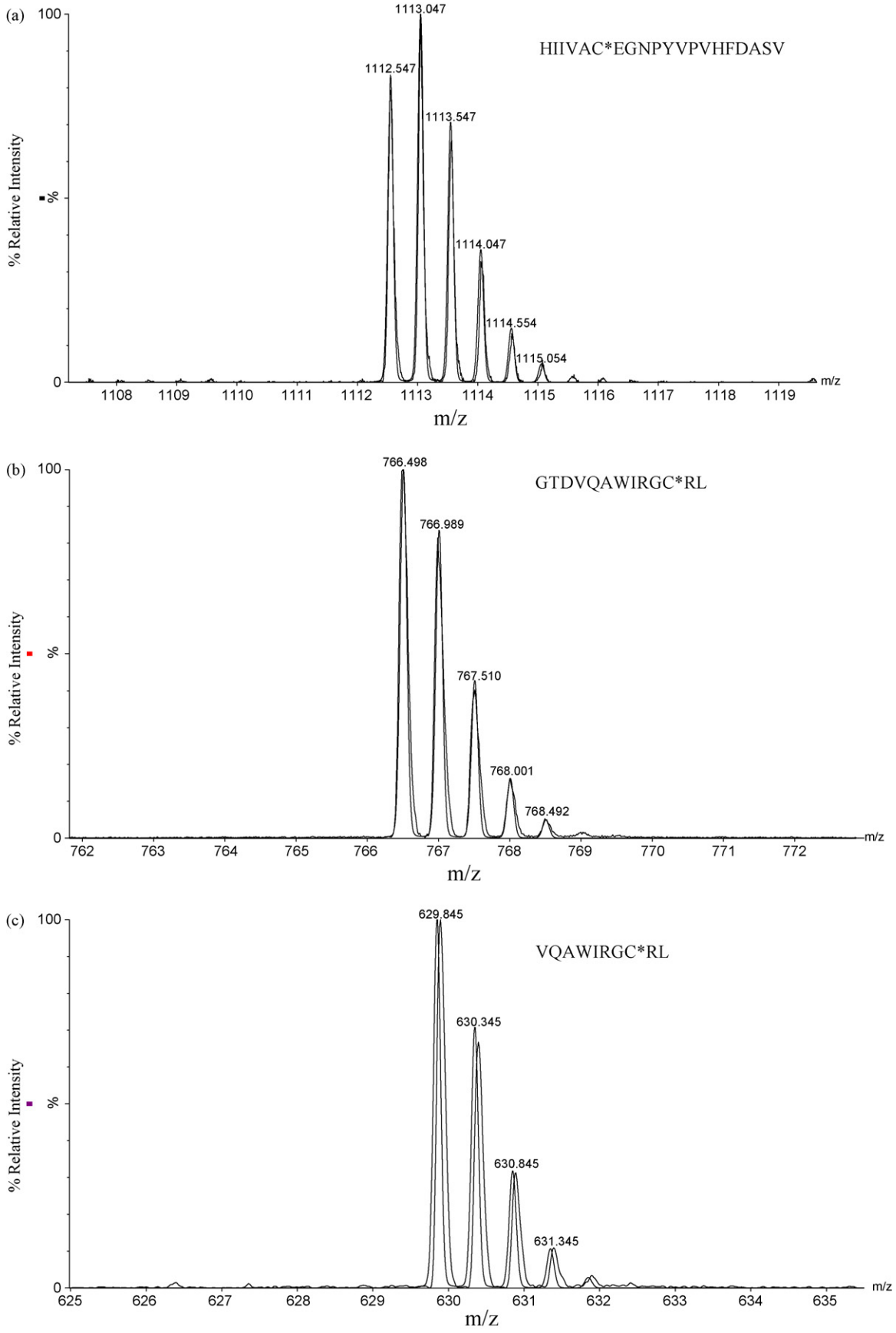

Fig. 3. Overlaid mass spectra of theoretical and experimental isotope envelopes of C-terminal peptides for (a) RNAase A cleaved with trypsin, lysozyme cleaved with (b) Lys-C and (c) Glu-C, respectively.

Table 1
Peptide mass fingerprinting of model protein; RNase A coupled with O^{18} strategy utilizing trypsin as a protease

Frag #	Residue #	Sequence	Charge state	Theoretical m/z	Observed m/z	Observed m/z (one O^{18} incorporation)	Observed m/z (two O^{18} incorporation)
T2–0	2–7	(K)ETAAAK(F)	1	590.314	590.32	592.321	594.332
T3–0	8–10	(K)FER(Q)	1	451.23	451.228	453.232	455.239
T4–0	11–31	QHMDSSTSAASSS NYCNQMMK	2	1182.965	1182.978	1183.979	1184.981
T4–0	11–31	QHMDSSTSAASSSNYC NQMMK(1-ox)	2	1190.963	1190.973	1191.977	1192.883
T4–0	11–31	QHMDSSTSAASSSN YCNQMMK	3	788.977	788.99	789.651	790.321
T4–0	11–31	QHMDSSTSAASSSNYCN QMMK(1-ox)	3	794.308	794.313	794.985	795.656
T5–1	32–37	(K)SRNLTK(D)	1	718.421	718.423	720.439	722.429
T7–0	40–61	(R)CKPVNTFVHESLADV QAVCSQK(N)	2	1259.111	1259.113	1260.108	1261.114
T7–0	40–61	(R)CKPVNTFVHESLADVQ AVCSQK(N)	2	839.74	839.745	840.397	841.064
T8–0	62–66	(K)NVACK(N)	1	591.291	591.294	593.294	595.303
T9–1	67–91	(K)NGQTNCYQSYSTMSITD CRETGSSK(Y)	3	959.063	959.072	959.736	960.4
T10–0	86–91	(R)ETGSSK(Y)	1	608.288	608.289	610.288	612.296
T11–0	92–98	(K)YPNCAYK(T)	1	915.402	915.402	917.405	919.41
T11–0	92–98	(K)YPNCAYK(T)	2	458.201	458.206	459.209	460.208
T12–0	99–104	(K)TTQANK(H)	1	662.346	662.348	664.345	666.358
T13–0	111–124	(C)EGNPYVPVHFDASV	1	1530.722	1530.716	Non-specific cleavage- C-terminal peptide	
T13–0	111–124	(C)EGNPYVPVHFDASV	2	765.861	765.861	Non-specific cleavage- C-terminal peptide	
T13–0	105–124		2	1112.542	1112.5415	C-terminal Peptide	
T13–0	105–124	(N)HIIVACE GNPYV PVHFDASV	3	742.028	742.023	C-terminal Peptide	

Fragments nomenclature: e.g., T4-0 represents fourth tryptic peptide with zero missed cleavage site. T5–1 represents fifth tryptic peptide with one missed cleavage site.

inating from tryptic digestion is very short and hydrophilic and is lost in the void volume. In both these cases, use of an alternate enzyme is recommended. During digestion, it is expected that a few non-specific cleavages may occur. However, with accurate mass data (and MS/MS analyses if required) on only these few non-specific cleavages, the C-terminal peptide can be easily detected.

The O^{18} approach was also used with Lys-C digestion to detect the C-terminal peptide of proteins. Three model proteins, RNAase A, lysozyme and lactoglobulin were evaluated for the feasibility of O^{18} strategy with Lys-C as a choice of enzyme. The C-terminal peptide was successfully detected for the three proteins as shown in Table 2. Fig. 3b depicts the overlaid mass spectra of theoretical and observed isotope envelopes for the C-terminal peptide of lysozyme. Further confirmation of the peptide sequence of the three candidates was performed using subsequent MS/MS analysis.

A minor downside of using Lys-C as a protease is that peptide sequences tend to be relatively long. This results in multiply charged peptides. For peptides having a charge greater than 4, it becomes difficult to infer the presence or absence of O^{18} tag incorporation, even after comparing the theoretical and experimental isotopic envelopes. It may not be possible to ascertain whether these internal peptides successfully incorporated the O^{18} tag which is necessary to eliminate the possibility of these internal peptides as potential false positives. One alternative approach can be to

Table 2
Summary of C-terminal peptide identification of model proteins based (i) O^{18}, (ii) acylation and (iii) esterification based strategies

Isotope coding strategy	Protein	Sequence	Theoretical (m/z = 1 amu)	Observed C-terminal peptide	Characteristic MS signature
O^{18} strategy					
With trypsin	RNAase A	HIIVAC*EGNPYVPVHFDASV	2224.086	1112.541(2+)	No incorporation of O^{18} in the C-terminal peptide
	Lactoglobulin	LSFNPTQLEEQC*HI	1715.806	858.405(2+)	
	Cytochrome *c*	KATNE	562.603	562.283(1+)	
	Myoglobin	ELGFQG	650.314	650.312(1+)	
With Lys-C	RNAase A	HIIVAC*EGNPYVPVHFDASV	2224.086	1112.542(2+)	
	Lysozyme	GTDVQAWIRGC*RL	1531.78	766.4006(2+)	
	Lactoglobulin	LSFNPTQLEEQC*HI	2534.264	845.403(3+)	
With Glu-C	Myoglobin	LGFQG	521.271	521.275(1+)	
	Lysozyme	VQAWIRGC*RL	1258.683	1258.678(2+)	
	Lactoglobulin	pyroQC*HI	540.247	540.232(1+)	
Acylation strategy					
With Lys-C	RNAase A	HIIVAC*EGNPYVPVHFDASV	2224.086	1133.561 (d_0);1135.0815 (d_3); Δ = 1.5; m/z = 2	Single acylation of C-terminal peptide; $\Delta m/z$ = 3 between peptide pairs (z = 1)
	Lysozyme	GTDVQAWIRGC*RL	1531.78	787.399 (d_0); 788.907 (d_3); Δ = 1.5; m/z = 2	
	Myoglobin	ELGFQG	650.314	692.332 (d_0); 695.344 (d_3); Δ = 3; m/z = 1	
Esterification strategy					
With Glu-C	Myoglobin	LGFQG	521.271	535.285 (d_0); 538.309 (d_3); Δ = 3; m/z = 1	Single esterification of C-terminal peptide; $\Delta m/z$ = 3 between peptide pairs (z = 1)
	RNAase A	ASV	276.154	n.d.	
	Lactoglobulin	pyroQC*HI	557.25	554.244 (d_0); 557.258 (d_3); Δ = 3; m/z = 1	

n.d.: Not detected.
* Iodoacetamide derivatized cysteine.

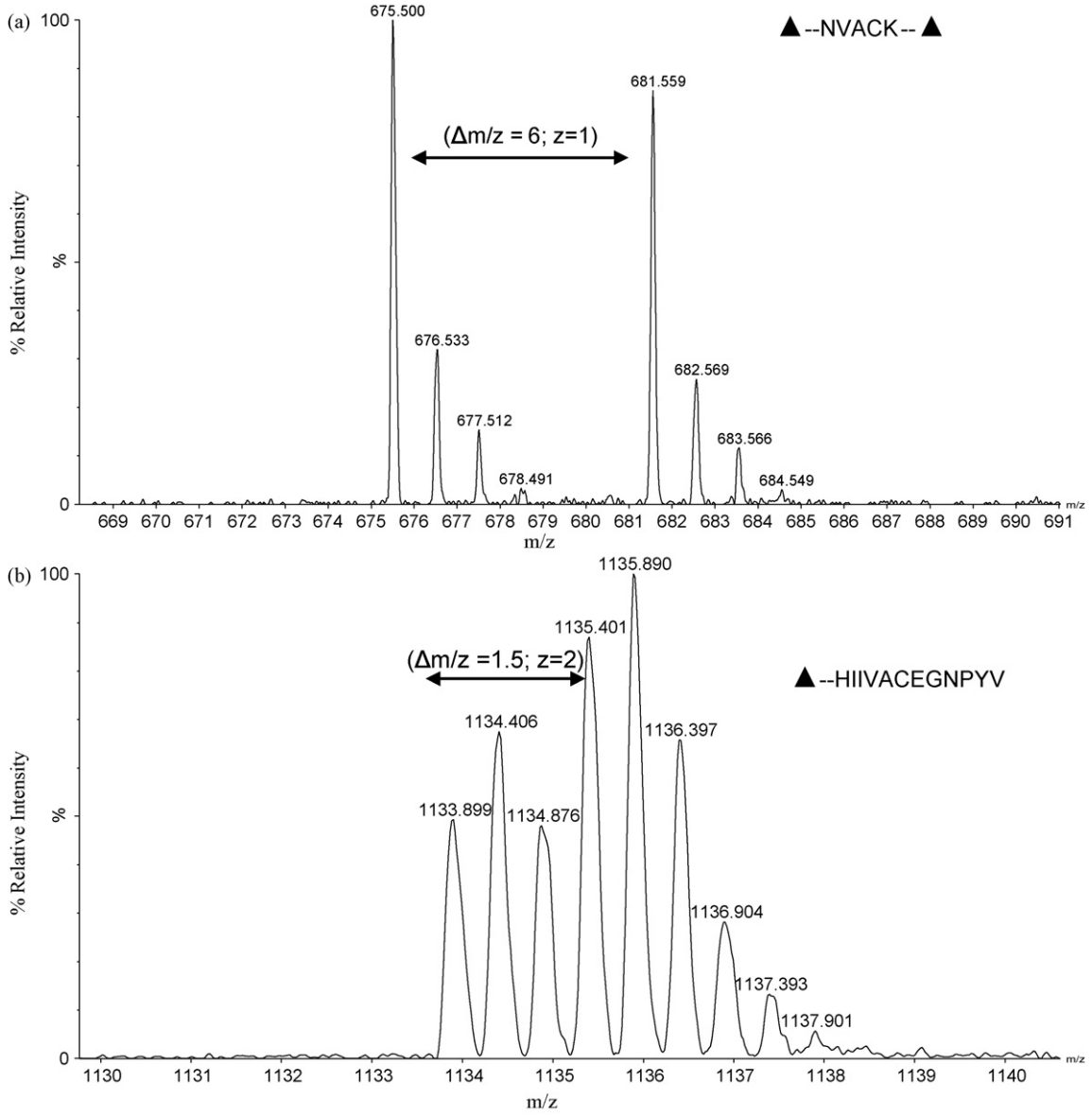

Fig. 4. Mass spectra of differentially acylated (a) internal peptide; NVACK and (b) C-terminal peptide, HIIVACEGNPYV, of RNAase A cleaved using Lys-C protease. (▲) Represents the acylating reagent.

carry out the proteolytic digestion independently in H_2O^{16} and H_2O^{18} buffer followed by mixing in order to get a $\Delta m/z$ (mass difference/charge) of 4 at $z=1$ between the coded internal peptide pairs. Another approach is to couple the O^{18} approach to sequential enzymatic digestion, for e.g., Lys-C digestion followed by a Glu-C digestion. With serial digestions with Lys-C and Glu-C proteases, all internal peptides terminating in lysine and glutamic acid and cleavages at aspartic acid can incorporate the O^{18} tag. An advantage of using Lys-C as a protease is that several Lys-C cleaved peptides have one or more arginine residues in their internal peptide sequence, thereby inherently having higher ionization efficiency for these peptides. Obviously, the O^{18} strategy coupled with Lys-C protease will not be useful if the protein is terminating in lysine.

Similarly, the feasibility of the O^{18} coding strategy with Glu-C as a protease was also evaluated [22]. Three model proteins, myoglobin, lysozyme and lactoglobulin were selected. The C-terminal peptide was successfully detected for all three proteins. Fig. 3c depicts the overlaid mass spectra of theoretical and experimental isotopic envelopes of C-terminal peptide for lysozyme. Further confirmation of the peptide sequence of the three candidates was performed using subsequent MS/MS analyses. The characteristic C-terminal peptide signature for Glu-C cleaved peptides is summarized in Table 2.

Since the protease Glu-C also has partial enzymatic activity at the C-terminus of aspartic acid residue, in some instances, peptides terminating in aspartic acid were also observed to incorporate the O^{18} tag. It was observed that Glu-C cleaved peptides are often short since both glutamic acid and aspartic acid residues are abundant amino acids in most of the proteomes. Obviously, the O^{18} strategy coupled with Glu-C protease will not be useful if the protein is terminating either in glutamic acid and/or aspartic acid residues.

Based on the above analyses, it was concluded that the O^{18} strategy is a very simple and a powerful approach to detect the C-terminal peptide of a protein. The main advantage of the O^{18} approach is the flexibility of using different enzymes, thereby expanding the utility of the approach to a broader range of protein sequences. No notable side-reactions were observed with the O^{18} strategy. Moreover, the approach is not sample intensive. The

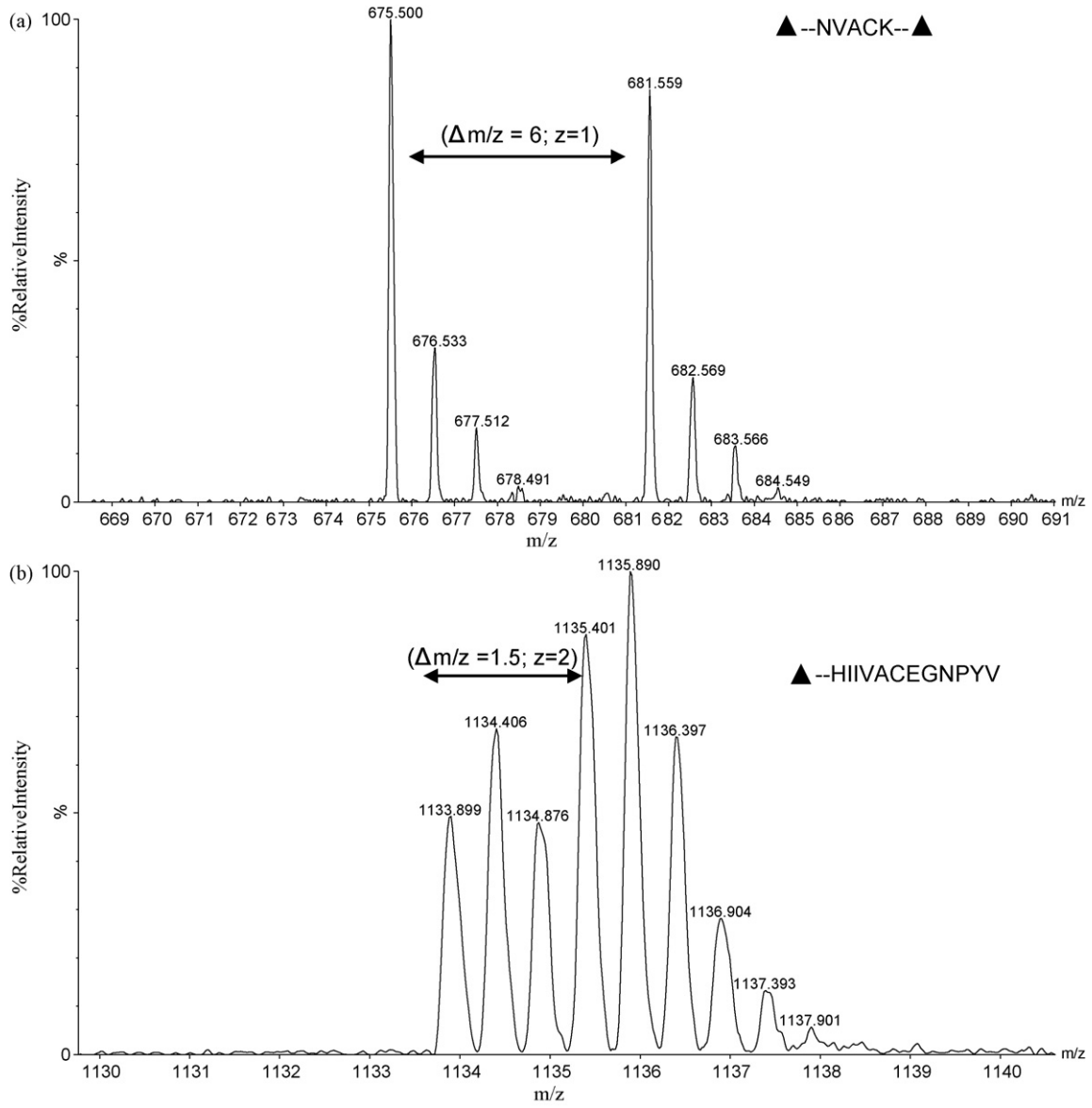

Fig. 5. Mass spectra of differentially esterified (a) internal peptide; VLIRLFTGHPE and (b) C-terminal peptide; LGFQG, of myoglobin (equine) cleaved using Glu-C protease. (■) Represents the esterification reagent.

O^{18} approach will be successful even if the C-terminal peptide is blocked as a result of post translational modifications [23]. The O^{18} isotope coding approach is also useful in detecting the number of non-specific cleavages occurring during a proteolytic digestion. This approach can be useful in quality control of proteolytic digestions. One of the limitations with the O^{18} approach is that the mass separation is smaller. The mass separation between peptide pairs can be increased by carrying out proteolyic digestion and O^{18} coding in light and heavy water separately. It has been recently reported that chymotrypsin [21], Asp-N [21], Lys-N [24] incorporate only one O^{18} atom into the C-termini of resulting internal peptides. The above proteases can also be used for detection of the C-terminal peptide, however, the mass difference between the internal peptide pairs will be smaller.

3.2. Acylation approach

The acylation approach has been reported for quantification in various global proteomic applications [13]. The approach, when used with Lys-C as a protease, has the potential to detect the C-terminal peptide of a protein. The premise of the strategy, as summarized in Fig. 1, is that all internal peptides cleaved after proteolysis with Lys-C provide two primary amine centers, one at the N-terminus of the peptide and the other at the γ-amino group of lysine residue. In contrast, the C-terminal peptide only provides with a single nucleophilic primary amine center at the N-terminus of the peptide. The strategy involves separating the Lys-C digest into two equal aliquots. Each aliquot is then differentially derivatized with light and heavy isoforms of the acylating reagent. This is followed by mixing the two aliquots and subsequent analyses with RPC–MS. For proof of concept, light (d_0) and heavy (d_3) isoforms of acetic anhydride were utilized. Based on the above strategy, differentially derivatized Lys-C cleaved internal peptide pairs will depict a $\Delta m/z$ of 6 at $z = 1$. In contrast, the C-terminal peptide pair will result in a $\Delta m/z$ of 3 at $z = 1$ due to derivatization at the N-terminal site.

Three model proteins; RNAase A, lysozyme, and lactoglobulin were selected for evaluating the acylation strategy. As shown in

Fig. 4a and b, a $\Delta m/z$ of 6 at $z=1$ was observed between differentially coded internal peptide pair; "NVACK" due the presence of two derivatizable centers. In contrast, a $\Delta m/z$ of 3 at $z=1$ ($\Delta m/z=1.5$ at $z=2$) was observed for peptide, HIIVACEGNPYV, due to only a single derivatization at the N-terminus of the C-terminal peptide. Acylation of the hydroxyl group of amino acids, threonine (Thr), serine (Ser) and tyrosine (Tyr), was observed as a side reaction. It is critical to remove any O-esterification occurring on Thr, Tyr and Ser residues. Presence of one of these residues in the C-terminal peptide may lead to multiple derivatizations. This may result in a $\Delta m/z=6$ or 9 at $z=1$ between C-terminal peptide pairs which could be misinterpreted for an internal peptide. It has been previously reported that the side reaction can be eliminated by either boiling the derivatized aliquots or by increasing the pH to 12 with hydroxylamine/NaOH followed by reversing the pH to 8 [19,25]. In this study, after completion of acylation, the aliquots were heated at 100 °C for approximately 3 h to remove possible O-esterification.

Using the acylation strategy, a characteristic MS signature was observed for the C-terminal peptide for the three evaluated proteins as depicted in Table 2. An advantage of the derivatization-based approach is that the sensitivity of the method can be improved more than 10-fold by intelligent choice of the derivatizing reagent [25]. Since the Lys-C peptides are relatively larger in size and can be multiply charged, $\Delta m/z=3$ at $z=1$ ($\Delta m/z=1.5$, 1 at $z=2$, 3, respectively) can be considered small for detection of the C-terminal peptide. This can again be overcome by utilizing a larger acylating tag. The main shortcoming of the approach is that it can only be coupled with Lys-C as a protease. Thus, the strategy will not be effective if the C-terminal amino acid of the protein is a lysine or if the C-terminal peptide generated due to Lys-C digestion is a short hydrophilic peptide. A potential missed cleavage in the (N-1) lysyl peptide (due to Pro-Lys) will also result in two primary amine centers in the C-terminal peptide.

3.3. Esterification approach

Esterification based isotope coding approach was also utilized for identification of C-terminal peptide of proteins. The approach involves Glu-C digestion followed by esterification of free carboxylate groups in order to differentiate the C-terminal peptide from internal peptides. Just like acylation strategy, the premise of esterification-based strategy is that all internal peptides provide two derivatizable centers while the C-terminal peptide provides only one derivatizable center. Based on the above strategy, the internal Glu-C cleaved peptides will be derivatized twice, at the C-terminus and at the glutamic acid or aspartic acid side chain. As a proof of concept, light (d_0) and heavy (d_3) isoforms of methanol were selected as esterification reagents. These labels will provide a $\Delta m/z=6$ at $z=1$ between differentially coded internal peptide pairs. In contrast, the C-terminal peptide will only be derivatized once producing a $\Delta m/z=3$ at $z=1$ which can serve as a characteristic signature for the C-terminal peptide. It is critical that enzymatic cleavage with Glu-C protease occurs after both glutamic acid and aspartic acid residues. The strategy involves separating Glu-C digest into two equal aliquots. Each aliquot is then differentially esterified with light (d_0) and heavy (d_3) isoforms of methanol. This is followed by mixing the two aliquots and subsequent analyses with RPC–MS.

Three model proteins, lactoglobulin, RNAase A and myoglobin were evaluated for the esterification-based strategy. The characteristic C-terminal peptide signature was observed for both lactoglobulin and myoglobin. As presented in Fig. 5a and b, different MS profiles for differentially esterified internal peptide; VLIRLFTGHPE, and C-terminal peptide, LGFQG, were observed. However, a characteristic C-terminal peptide signature was not detected for the C-terminal peptide of RNAase A due to a missed cleavage after the aspartic acid residue. The resulting C-terminal peptide; GNPYVPVHFDASV was derivatized twice and thus had the same MS signature ($\Delta m/z=6$ at $z=1$) as observed for other internal peptides. Moreover; during peptide mapping with Glu-C/esterification strategy, several other missed cleavages after aspartic acid residues were observed. This resulted in multiple derivatizable sites in peptides containing aspartic acid residues. It is known that cleavage of endoproteinase Glu-C after aspartic acid residues is considerably slower after glutamic acid residues. This is a major shortcoming of this approach, unlike the O^{18} coding and acylation based approaches. Just like the acylation approach, the sensitivity of the method and the mass separation between peptide pairs can be improved by utilizing a suitable esterification reagent.

4. Conclusions

Based on the analyses, it was observed that the isotope based coding approaches followed by reversed phase/mass spectrometric analyses is a powerful tool to detect the C-terminal peptide from a proteolytic mixture. The advantages and limitations of detecting the C-terminal peptide with each evaluated isotope coding based approach are summarized in Table 3. In each method, a characteristic MS signature was observed for the C-terminal peptide in comparison to the other internal peptides. The C-terminal peptide was detected for the analyzed proteins in 15 of the 16 isotope coding/enzyme strategies. Of the three approaches studied, the O^{18} strategy provides the flexibility to use different enzymes and therefore expands the applicability of this technique to a broader range of protein sequences. The other advantage of the O^{18} approach for detection of C-terminal peptide is the absence of any side reactions. It was observed that when analyzing puri-

Table 3
Comparison of different isotope coding approaches for detection of C-terminal peptide

	O^{18} coding	Acylation	Esterification
Applicability	Approach can be coupled with Trypsin, Glu-C, Lys-C digestions	Approach can be coupled only with Lys-C digestion	Approach can be coupled only with Glu-C digestion
Strategy	Isotope coding is carried out during proteolysis	Isotope coding is carried out after proteolysis	Isotope coding is carried out after proteolysis
Side Reactions	No notable side-reaction	Partial O-esterification on Ser, Thr, Tyr residues	Partial loss of esterification [26]
Mass difference between differentially coded peptide pairs	Small	Dependent on choice of derivatizing reagent	Dependent on choice of derivatizing reagent
Sensitivity	Similar to underivatized peptide	Dependent on choice of derivatizing reagent	Dependent on choice of derivatizing reagent
Missed cleavages	Missed cleavage will not limit the detection of the C-terminal peptide	Missed cleavage at penultimate peptide may limit detection of the C-terminal peptide	Missed cleavage after (i) aspartic acid residues and (ii) penultimate peptide may limit detection of the C-terminal peptide

fied proteins using high mass accuracy mass spectrometers, isotope coding based strategies coupled with reversed phase chromatography–mass spectrometry were sufficient to detect the C-terminal peptide. This will preclude the use of further MS/MS analyses on each peptide eluting within chromatographic peaks. This will be the case with most of the accurate mass peptide mapping analyses on biopharmaceuticals.

Acknowledgment

The authors would like to acknowledge Dr. Jeff Gilbert of Dow AgroSciences for technical assistance.

References

[1] C.M. Murphy, C. Fenselau, Anal. Chem. 67 (1995) 1644.
[2] T. Hayashi, T. Sasagawa, Anal. Biochem. 209 (1993) 163.
[3] S. Julka, F. Regnier, Anal. Chem. 76 (2004) 5799.
[4] C.-X. Zhang, B.V. Weber, J. Thammavong, T.A. Grover, D.S. Wells, Anal. Chem. 78 (2006) 1636.
[5] K.A. Johnson, P-F. Kari, B.S. Tangarone, T.J. Porter, J.C. Rouse, Anal. Biochem. 360 (2007) 75.
[6] Y. Iwafune, H. Kawasaki, H. Hirano, Arch. Biochem. Biophys. 431 (2004) 9.
[7] B. Munch-Petersen, W. Knecht, C. Lenz, L. Sondergaard, J. Piskur, J. Biol. Chem. 275 (2000) 6673.
[8] J. Li, S. Liang, Anal. Biochem. 302 (2002) 108.
[9] T. Isobe, T. Ichimura, T. Okuyama, Anal. Biochem. 155 (1986) 135.
[10] B. Thiede, B. Wittmann-Liebold, M. Bienert, E. Krause, FEBS Lett. 357 (1995) 65.
[11] K. Miyazaki, A. Tsugita, Proteomics 4 (2004) 11.
[12] M. Yamaguchi, M. Oka, K. Nishida, M. Ishida, A. Hamazaki, H. Kuyama, E. Ando, T.-A. Okamura, N. Ueyama, S. Norioka, O. Nishimura, S. Tsunasawa, T. Nakazawa, Anal. Chem. 78 (2006) 7861.
[13] S. Julka, F.E. Regnier, J. Proteome Res. 3 (2004) 350.
[14] K. Rose, L. Savoy, M.G. Simona, R.E. Offord, P. Wingfield, Biochem. J. 250 (1988) 253.
[15] K. Rose, M.G. Simona, R.E. Offord, C.P. Prior, B. Otto, D.R. Thatcher, Biochem. J. 215 (1983) 273.
[16] T. Kosaka, T. Takazawa, T. Nakamura, Anal. Chem. 72 (2000) 1179.
[17] A. Panchaud, E. Guillaume, M. Affolter, F. Robert, P. Moreillon, M. Kussmann, Rapid Commun. Mass Spectrom. 20 (2006) 1585.
[18] R. Zhang, C.S. Sioma, R.A. Thompson, L. Xiong, F.E. Regnier, Anal. Chem. 74 (2002) 3662.
[19] N. Abello, H.A.M. Kerstjens, D.S. Postma, R. Bischoff, J. Proteome Res. 6 (2007) 4770.
[20] D.R. Goodlett, A. Keller, J.D. Watts, R. Newitt, E.C. Yi, S. Purvine, J.K. Eng, P.V. Haller, R. Aebersold, E. Kolker, Rapid Commun. Mass Spectrom. 15 (2001) 1214.
[21] Y. Xudong, A. Freas, J. Ramirez, P.A. Demirev, C. Fenselau, Anal. Chem. 73 (2001) 2836.
[22] K.J. Reynolds, X. Yao, C. Fenselau, J. Proteome Res. 1 (2002) 27.
[23] A. Fujiyama, S. Tsunasawa, F. Tamanoi, F. Sakiyama, J. Biol. Chem. 266 (1991) 17928.
[24] K.C.S. Rao, R.T. Carruth, M. Miyagi, J. Proteome Res. 4 (2005) 507.
[25] H. Mirzaei, F. Regnier, Anal. Chem. 78 (2006) 4175.
[26] C. Ji, L. Li, J. Proteome. Res. 4 (2005) 734.