Locating and identifying posttranslational modifications by in-source decay during MALDI-TOF mass spectrometry

JOHN J. LENNON¹ AND KENNETH A. WALSH

Department of Biochemistry, University of Washington, Seattle, Washington 98195-7350

(RECEIVED May 7, 1999; ACCEPTED August 9, 1999)

Abstract

A technique is described for identifying and locating posttranslational modifications (PTMs) in peptides and proteins of known sequence by interpretation of c_n ion signals generated by in-source decay during delayed ion extraction in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Sites of phosphorylation in seven synthetic peptides were determined, as was the location of both the heme group and *N,N,N*-trimethyllysine in yeast cytochrome *c*. A semi-automated data analysis process facilitates the identification of segments of the sequence on each side of the PTM, permitting its placement at the junction of the segments and definition of the added mass. A graphical display facilitates illustration of both the location and mass of the PTM.

Keywords: delayed ion extraction; in-source decay; MALDI; posttranslational modifications; protein sequencing by MALDI

Posttranslational modifications (PTMs) are known to be important as mechanisms for controlling intracellular protein function. Nearly every mature eukaryotic protein is covalently modified during its synthesis, cellular targeting, or regulation. The modifications may regulate enzymatic activity, stabilize protein structure, or modulate the chemical functionality of the protein (Graves et al., 1994). With the growing access to genomic sequence information, it is important to establish the covalent structure of each mature protein by determining both the nature and specific site of their covalent modifications. Method developments have emphasized localization of phosphorylation sites because of their widespread regulatory roles.

Mass spectrometry currently provides the best general method for investigating, identifying, and locating post-translationally modified amino acid residues. Phosphorylation sites are often identified after enzymatic digestion of a phosphoprotein by either matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI TOF-MS) or electrospray mass spectrometry. Comparison of observed peptide masses with those calculated from the known sequences draw attention to peptides with masses 80 Da greater than predictions. Alternatively, Covey et al. (1991) detected phosphopeptides in complex mixtures by their neutral loss of 98

Da (mass of H_3PO_4) during collision-induced dissociation. Phosphopeptides have also been identified by their release of diagnostic negative fragment ions at m/z 79 after collisions induced in the orifice during HPLC/MS (Huddleston et al., 1993; Carr et al., 1996).

An alternative technique involves post-source decay (PSD) in the MALDI TOF instrument (Annan & Carr, 1996), which examines metastable fragment ions. By comparing mass spectra of an HPLC fraction in linear and reflecting modes, peptides that lose 98 Da appear to be phosphorylated. The same PSD process that gives rise to the loss of phosphate from phosphopeptides can be employed (Kaufmann et al., 1993, 1994) to obtain primary sequence information from predictable cleavages along the backbone of the peptide.

All of these methods require prior enzymatic fragmentation of a protein to identify the specific site of phosphorylation in a small peptide. Recent studies (Brown & Lennon, 1995; Lennon & Walsh, 1997; Katta et al., 1998; Reiber et al., 1998) have shown that partial amino acid sequence information can be deduced from c_n fragment patterns without prior digestion from in-source decay (ISD) of metastable, MALDI-generated ions in an instrument with delayed ion extraction capability. Zubarev et al. (1998, 1999) have observed similar extended patterns of c ions by electron capture dissociation of multiply charged protein ions (from an electrospray source) during Fourier transform mass spectrometry. They propose that energetic short-lived hydrogen atoms are intermediates in the process. During MALDI-ISD, segments of 40 or more continuous residues are displayed in the mass spectrum as small but coherent ion signals (Lennon & Walsh, 1997; Katta et al., 1998). These

Reprint requests to: Kenneth A. Walsh, Department of Biochemistry, Box 357350, University of Washington, Seattle, Washington 98195-7350; e-mail: walsh@u.washington.edu.

¹Current address: Large Scale Biology Corp., 9620 Medical Center Drive, Rockville, Maryland 20850-3338.

appear to involve energy transfer from matrix rather than prompt photodissociation (Brown et al., 1996). An approach to the identification of phosphorylation sites and other PTMs is made possible by the analysis of ISD data from small proteins using delayed ion extraction in a MALDI-TOF instrument. We describe here this general technique for locating and identifying PTMs from ISD mass spectral data.

Results and discussion

We examined seven synthetic phosphopeptides of known sequence, but without advance knowledge of their sites of phosphorylation. In each case, 2 pmol was applied to the target and subjected to MALDI TOF with delayed ion extraction as described. A representative ISD mass spectrum from a 24-residue peptide is shown in Figure 1. A nearly continuous series of c_n , y_n , and $z_n + 2$ ions (nomenclature of Biemann, 1988) was observed. The largest signals corresponded to c_n ions from c_7 to c_{19} (Table 1), allowing recognition of a large segment of the primary sequence. Although the spacing of signals labeled c_{20} to c_{23} corresponded to a sequence recognized as EDE, the gap between c_{19} and c_{20} was larger (by 80 Da) than expected for the serine residue known to occupy that position. Clearly, this placed the phosphoryl group on Ser20.

A semiautomated data analysis approach was used to identify the site of phosphorylation. Initially, the peak table generated from the labeling of the ion signals in the ISD mass spectrum was saved in a spreadsheet (.txm file) format. These data were imported into

Table 1. Theoretical and observed c_n ion fragments from the ISD mass spectrum of the phosphopeptide shown in Figure 1

Residue retained	n	Theoretical mass (unmodified)	Observed mass	Delta mass ^a (observed – theoretical)	Mean (SD) ^b
R	1	174.23			
R	2	330.41			
A	3	401.49			
A	4	472.57			
V	5				
P	6				
P	7	765.94	765.77	–0.17	
S	8				
P	9	950.13	950.18	0.05	
S	10	1,037.21	1,037.11	–0.10	
L	11	1,150.37	1,150.76	0.39	
S	12	1,237.45	1,237.52	0.07	
R	13	1,393.64	1,393.63	–0.01	
H	14	1,530.78	1,531.18	0.40	
S	15	1,617.86	1,618.18	0.32	
S	16				
P	17	1,802.05	1,802.37	0.32	
H	18	1,939.19	1,939.76	0.57	0.57
Q	19	2,067.32	2,067.71	0.39	0.20 (0.24)
S	20	2,154.40	2,234.51	80.11	
E	21	2,283.52	2,363.48	79.96	
D	22	2,398.60	2,479.07	80.47	
E	23	2,527.72	2,608.07	80.35	80.22 (0.23)
E					

^aThe increase in Delta Mass at Ser20 indicates the location and mass of the posttranslational modification.

^bMean delta mass, residues 7–19, 20–23.

Fig. 1. ISD mass spectrum of a synthetic phosphopeptide in DHB at 25 kV ion source bias potential, 150 ns pulsed extraction delay time, and 1.20 kV pulsed extraction voltage. One hundred laser pulses are averaged. The site of the phosphoserine (pS) is Ser20.

a spreadsheet program, where differences between successive c_n peaks (the major peaks) were calculated and compared with residue masses (± 0.4 Da) to give a tentative “observed” sequence. Since the actual primary sequence was known, segments of that sequence could be readily recognized among the “observed” sequence data. Moreover, since the primary sequence was known, all theoretical fragment ions could be predicted using the Sherpa software package (Taylor et al., 1996). If there were no PTM, the observed c_n series should correspond to the calculated c_n series. If there is a PTM, one would expect correspondence of observed and calculated values until the PTM was reached, but observed values after that would carry the extra mass of the PTM. Such a discontinuity in the c_n series is recognized by examining the tentative “observed” sequence (from the spreadsheet calculations) for a second segment of the known sequence. By calculating the difference (“Delta Mass”) between the theoretical and observed fragment ion masses (Table 1), one sees that only c_n fragments that included Ser20 have the extra 80 Da diagnostic of phosphorylation. This is graphically illustrated in Figure 2F, which reveals both the site and mass of the modification. The method relies on observing two

segments of identifiable sequence and identifying the site of “delta mass” offset at their junction.

Gaps in the c_n series, and its spreadsheet-derived tentative sequence, do occur when an N-C α bond does not cleave, as in X-Pro sequences. When mass differences between ion signals did not match a single residue mass, it was usually possible to explain in terms of an uncleaved bond (e.g., at Ser16–Pro17 in Table 1).

Figure 2 shows graphically the sites of phosphorylation determined for six other synthetic phosphopeptides as well. In each case only the delta mass data from the c_n fragment ion series is plotted. Analogous plots for the y_n ion series (not shown) yielded confirming data, except for the peptides shown in Figures 2A and 2F where the modification was too close to the C-terminus.

An inherent problem with any MALDI process for peptides is its inability to accurately identify ISD fragment ions below m/z of about 400–500 (Lennon & Walsh, 1997; Reiber et al., 1998). It has been shown (Brown & Lennon, 1995; Reiber et al., 1998) that both c_n and y_n fragment ions dominate ISD spectra for peptides. If the site of modification is within four to five residues of the N-terminus of the peptide, identification and location of the modified residue

Fig. 2. Graphical representation of the data from Table 1 for the c_n ion series produced from ISD experiments of seven synthetic phosphopeptides. The vertical offset (~ 80 Da) in each graph indicates the site (Residue Number) and mass of the modification (Delta Mass). Phosphoserine and phosphothreonine are represented as pS and pT, respectively. The ISD mass spectrum for peptide F is shown in Figure 1.

would not be possible using the c_n ion series alone. Although the y_n ion series is typically less pronounced, it may provide placement near the N-terminus.

An ISD analysis of 10 pmol of undigested cytochrome *c* from yeast (Fig. 3) illustrates direct application to a protein with three modified residues (heme attachment on Cys19 and Cys22, *N,N,N*-trimethylation of Lys77). Although y_n and $z_n + 2$ series were also observed, for simplicity, analysis is focused on the c_n ion series, which usually comprises the larger signals. Treatment of the ISD data as above indicated the absence of c_{19} through c_{21} , as would be consistent with heme covalently bound to both Cys19 and Cys22 (Table 2). The c_n series was then recognized again by the c_n spacing among c_{22} to c_{76} (Fig. 3; Table 2) but each had an extra mass of $616.7 \text{ Da} \pm 0.9$ (theoretical 616.5). Similarly, there was a break in the c_n pattern at c_{77} but a new segment of sequence could be recognized in the pattern of c_{77} to c_{90} , and a further increase in mass ($42.5 \text{ Da} \pm 1.2$) was identified at c_{77} and beyond. This increase is consistent with the trimethylation of Lys77 (Table 2). The residue attachment sites of heme and the trimethylation were readily recognized. The extra mass of 616.7 Da associated with

residues 22 to 76 identified the heme. However, the additional 42.5 Da beginning at residue 77 was not accurate enough to discriminate between trimethylation (43 Da) and acetylation (42 Da).

Identifying the c_n ion series C-terminal to the heme could have been very labor intensive because of the difficulty of distinguishing the c_n series from the y_n and $z_n + 2$ series. The data analysis time was greatly reduced by applying a spreadsheet subroutine that automatically identified segments of primary sequence based on the mass difference between the largest successive ISD ion signals. These segments of sequence are easiest to discern when the most intense ISD ion signals in this m/z range are of the c_n type.

The delta mass data are plotted against the residue number for the ISD analysis of yeast cytochrome *c* in Figure 4, illustrating the location of the heme group and the *N,N,N*-trimethyllysine. Although the trimethyllysine modification was confirmed in the y_n ion data (not shown), the y_n ion series diminishes beyond a m/z of 6,000–7,000 and did not provide evidence confirming the heme site.

In conclusion, in-source decay of metastable ions during delayed ion extraction generates extended series of c_n ions (and a

Fig. 3. ISD mass spectrum of yeast cytochrome *c* in DHB at 25 kV ion source bias potential, 150 ns pulsed extraction delay time, and 2.45 kV pulsed extraction voltage. One hundred laser pulses are averaged. Note the lack of c_{19} – c_{21} ions (cf. Table 2), which is consistent with the covalent attachment of heme through Cys19 and Cys22. The *N,N,N*-trimethylation of Lys77 is observed in both the c_n and y_n ion series.

Table 2. A portion of the theoretical and observed c_n ion fragments from the ISD mass spectrum of yeast cytochrome *c* shown in Figure 3

Residue retained	n	Theoretical mass (unmodified)	Observed mass	Delta mass ^a (observed – theoretical)	Mean (SD) ^b
A	12	1,194.38	1,194.25	–0.13	
T	13	1,295.48	1,296.02	0.54	
L	14	1,408.64	1,409.44	0.80	
F	15	1,555.82	1,554.56	–1.26	
K	16	1,683.99	1,683.61	–0.38	
T	17	1,785.10	1,786.07	0.97	
R	18	1,941.29	1,940.35	–0.94	–0.06 (0.86)
C	19	2,044.43			
L	20	2,157.59			
Q	21	2,285.72			
C	22	2,388.86	3,003.96	615.10	
H	23	2,526.01	3,141.91	615.90	
T	24	2,627.11			
V	25	2,726.24	3,343.27	617.03	
E	26	2,855.36	3,472.98	617.62	
K	27	2,983.53			
G	28	3,040.59	3,658.17	617.58	
G	29				
P	30	3,194.75	3,811.67	616.92	
H	31	3,331.90	3,947.29	615.39	
S	70	7,662.66	8,280.46	617.80	
E	71	7,791.78	8,408.30	616.52	
Y	72	7,954.96	8,571.59	616.63	
L	73	8,068.12	8,685.76	617.64	
T	74	8,169.22	8,785.54	616.32	
N	75				
P	76	8,380.44	8,997.03	616.59	616.70 (0.87)
K	77	8,508.62	9,167.45	658.83	
K	78	8,636.79	9,295.90	659.11	
Y	79	8,799.96	9,459.10	659.14	
I	80				
P	81	9,010.24	9,669.44	659.20	
G	82	9,067.29	9,726.45	659.16	
T	83	9,168.40	9,828.24	659.84	659.21 (0.33)

^aEach increase in Delta mass, at Cys22 and at Lys77, indicates the location and the mass of a posttranslational modification.^bMean delta mass, residues 12–18, 22–76, 77–83.

lesser series of y_n ions) that enables the discernment of segments of primary sequence of peptides and proteins. PTMs can be located and identified by examining these data for residue-specific mass increases that exceed theoretical values. This technique is effective for locating PTMs in pure peptides and small proteins of known sequence. Identification of the chemical nature of the PTM depends upon the accuracy of the instrument since it is based on differences between large m/z values.

Materials and methods

Matrices purchased from Aldrich Chemical Co. (Milwaukee, Wisconsin) were used without further purification. Lyophilized synthetic phosphopeptides were generously donated by the Department of Pharmacology, University of Washington. The sequences were provided but the sites of phosphorylation were not divulged until after our analyses. Cytochrome *c* (from *Saccharomyces cerevisiae*) was purchased from Sigma Chemical Co. (St. Louis, Missouri) and used without further purification.

In-source decay experiments were performed on a PerSeptive Biosystems Voyager-Elite MALDI-TOF mass spectrometer with a 2.0 m linear flight path and delayed ion extraction, as described previously (Lennon & Walsh, 1997). All samples were prepared in duplicate using either 2,5-dihydroxybenzoic acid (DHB) or sinapinic acid, and applied to multiple target areas on a smooth sample plate by the dried droplet method (Karas & Hillenkamp, 1988). Matrix solutions were saturated in 40% CH₃CN, 0.1% trifluoroacetic acid (TFA) in water. Phosphopeptides (2 μM) and yeast cytochrome *c* (10 μM) were dissolved in water containing 5% acetic acid. Analyte solution (1.0 μL) was applied first to the target, then 1.5 μL of matrix solution was stirred in with a pipette tip and the sample allowed to air evaporate.

Delayed ion extraction (150 ns) was employed, using a 25 kV bias potential and linear TOF. The delayed pulse extraction voltage was optimized for the singly protonated pseudomolecular ion (typically 1–3 kV). Each analysis was first externally calibrated at two points using a laser fluence near the ionization threshold. One of the calibrant ions was the singly protonated pseudomolecular ion

Fig. 4. Graphical representation of the data from Table 2 for two different portions of the c_n ion series produced from an ISD experiment (Fig. 3) of yeast cytochrome *c*. The two vertical offsets indicate the sites (Residue Number) and masses (Delta Mass) of the modifications. The large offset on the left corresponds to the heme (616.5 Da) bound to residues 19 and 22 and the smaller offset on the right is due to *N,N,N*-trimethylation (43.1 Da) of Lys77.

of the analyte since an accurate mass was known. The other calibrant ion was the singly protonated pseudomolecular ion of either the matrix for the phosphopeptide or Angiotensin II for yeast cytochrome *c*. The ISD spectrum was then optimized by increasing the laser fluence approximately 10% (by reducing the laser attenuation by about 250 U of a 2,500 U range).

A spreadsheet program was used to deduce partial sequence data from patterns of c_n fragments. A description of the program is available at http://www.lsbcl.com/papers/auto_fragISD.htm

Acknowledgments

J.J.L. was supported by the E.W. Davie/Zymogenetics endowment fund. The authors are grateful for helpful discussions with Lowell Ericsson, the assistance of Sam Purvine and the generous donation by Anita Colvin (Department of Pharmacology, University of Washington) of the synthetic phosphopeptides.

References

- Annan RS, Carr SA. 1996. Phosphopeptide analysis by matrix-assisted laser desorption time-of-flight mass spectrometry. *Anal Chem* 68:3413–3421.
- Biemann K. 1988. Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom* 16:99–111.
- Brown RS, Carr BL, Lennon JJ. 1996. Factors that influence the observed fast fragmentation of peptides in matrix-assisted laser desorption. *J Am Soc Mass Spectrom* 7:225–232.
- Brown RS, Lennon JJ. 1995. Sequence-specific fragmentation of matrix-assisted laser-desorbed protein/peptide ions. *Anal Chem* 67:3990–3999.
- Carr SA, Huddleston MJ, Annan RS. 1996. Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry. *Anal Biochem* 239:180–192.
- Covey T, Sushan R, Bonner R, Schröder W, Hucho F. 1991. LC/MS and LC/MS/MS screening for the sites of posttranslational modification in proteins. In: Jörnvall H, Höög JO, Gustavsson, eds. *Methods in protein sequence analysis*. Basel: Birkhäuser Verlag. pp 249–256.
- Graves DJ, Martin BL, Wang JH. 1994. *Co- and post-translational modification of proteins. Chemical principles and biological effects*. New York: Oxford University Press.
- Huddleston MJ, Annan RS, Bean MF, Carr SA. 1993. Selective detection of phosphopeptides in complex mixtures by electrospray liquid chromatography/mass spectrometry. *J Am Soc Mass Spectrom* 4:710–717.
- Karas M, Hillenkamp F. 1988. Laser desorption/ionization of proteins with molecular masses exceeding 10000 daltons. *Anal Chem* 60:2299–2301.
- Katta V, Chow DT, Rohde MF. 1998. Applications of in-source fragmentation of protein ions for direct sequence analysis by delayed extraction MALDI-TOF mass spectrometry. *Anal Chem* 70:4410–4416.
- Kaufmann R, Kirsch D, Spengler B. 1994. Sequencing of peptides in a time-of-flight mass spectrometer: Evaluation of postsource decay following matrix-assisted laser desorption/ionization (MALDI). *Int J Mass Spectrom Ion Processes* 131:355–385.
- Kaufmann R, Spengler B, Lützenkirchen F. 1993. Mass spectrometric sequencing of linear peptides by product-ion analysis in a reflectron time-of-flight mass spectrometer using matrix-assisted laser desorption/ionization. *Rapid Comm Mass Spectrom* 7:902–910.

- Lennon JJ, Walsh KA. 1997. Direct sequence analysis of proteins by in-source fragmentation during delayed ion extraction. *Protein Sci* 6:2446–2453.
- Reiber DC, Brown RS, Weinberger S, Kenny J, Bailey J. 1998. Unknown peptide sequencing using matrix-assisted laser desorption/ionization and in-source decay. *Anal Chem* 70:1214–1222.
- Taylor JA, Walsh KA, Johnson RS. 1996. Sherpa: A Macintosh-based expert system for the interpretation of ESI LC/MS and MS/MS of protein digests. *Rapid Comm Mass Spectrom* 10:679–687.
- Zubarev RA, Kelleher NL, McLafferty FW. 1998. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc* 120:3265–3266.
- Zubarev RA, Kruger NA, Fridriksson EK, Lewis MA, Horn DM, Carpenter BK, McLafferty FW. 1999. Electron capture dissociation of gaseous multiply-charged proteins is favored at disulfide bonds and other sites of high hydrogen atom affinity. *J Am Chem Soc* 121:2857–2862.