

# A Method To Define the Carboxyl Terminal of Proteins

Salvatore Sechi and Brian T. Chait\*

Laboratory for Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, 1230 York Avenue, New York, New York 10021

**Accurate definition of the carboxyl terminal of proteins is necessary for elucidating posttranslational processing at the C-terminal and more generally for characterizing protein primary structures. Here, we describe a strategy for isolating and characterizing the C-terminal peptide of a protein after proteolysis with endoprotease Lys-C. Isolation is achieved using anhydrotrypsin, a catalytically inert derivative of trypsin that binds peptides containing lysine or arginine residues at their C-termini without cleaving them. Rapid, accurate characterization of the isolated C-terminal peptide is achieved by mass spectrometry. Initial identification of the C-terminal peptide is obtained by comparing matrix-assisted laser desorption/ionization time-of-flight mass spectra of the digest prior to and after incubation with anhydrotrypsin. Characterization of the C-terminal sequence is achieved by capillary-HPLC electrospray ionization tandem mass spectrometry of the isolated peptide using a quadrupole ion trap mass spectrometer in the selective reaction monitoring mode. This strategy was successfully applied to the characterization of the C-terminal of proteins with molecular masses ranging up to 56 kDa.**

Definition of the carboxyl terminal (C-terminal) of proteins is necessary for accurately describing the results of posttranslational processing and more generally for characterizing protein primary structures. Posttranslational processing at the C-terminus plays a critical role in a variety of biological processes. For example, prenylation, which occurs on cysteine residues near the C-terminus, functions to anchor proteins to lipid membranes,<sup>1</sup> while aberrant C-terminal processing of the amyloid  $\beta$ -protein in humans is believed to be involved in the pathogenesis of Alzheimer's disease.<sup>2</sup> Characterization of the C-terminus can also be important for identifying sequence motifs that function as organelle targeting signals (e.g., the consensus C-terminal tripeptide peroxisomal targeting signal<sup>3</sup>). Furthermore, oligonucleotides derived from the C-terminal sequence are useful for the cloning of cDNAs as well as for defining the termini of proteins where the only information available is genomic DNA sequence.

Methods that have been developed to define the C-terminal of proteins include chemical degradation, enzymatic degradation, and chemical/isotopic labeling. In one set of chemical degradation methods, the C-terminal amino acid is chemically modified, excised from the polypeptide chain, and characterized. This cycle is repeated a number of times to provide the sequence of the C-terminus.<sup>4–6</sup> The relatively modest sensitivity attainable to date has limited the utility of this approach, although improvements are steadily being made.<sup>7,8</sup> An alternative set of chemical procedures has been developed that utilize acid-catalyzed chemical degradation to produce peptide ladders that yield information on the C-terminal of peptides and proteins.<sup>9–12</sup> The application of these acid-catalyzed degradation methods is limited by the occurrence of nonspecific hydrolysis and the difficulty of analyzing ladders with high molecular mass. The use of carboxypeptidases for protein sequencing was introduced more than 25 years ago.<sup>13,14</sup> With the advent of improved methods of mass spectrometric (MS) readout, the utility of the exopeptidase approach has increased considerably.<sup>15–25</sup> Limitations of this approach include the possible inaccessibility of the C-terminal of the protein to the enzyme, the

\* To whom correspondence should be addressed: (e-mail) chait@rockvax.rockefeller.edu.

- (1) Glomset, J. A.; Gelb, M. H.; Farnsworth, C. C. *Trends Biochem. Sci.* **1990**, *15*, 139–142.
- (2) Selkoe D. J. *Trends Cell Biol.* **1998**, *8*, 447–53
- (3) Gould, S. J.; Keller, G. A.; Hosken, N.; Wilkinson, J.; Subramani, S. *J. Cell. Biol.* **1989**, *108*, 1657–1664.

- (4) Inglis, A. S. *Anal. Biochem.* **1991**, *195*, 183–196.
- (5) Boyd, V. L.; Bozzini, M.; Zon, G.; Noble, R. L.; Mattaliano, R. J. *Anal. Biochem.* **1992**, *206*, 344–352.
- (6) Bailey, J. M.; Tu, O.; Issai, G.; Ha, A.; Shively, J. E. *Anal. Biochem.* **1995**, *224*, 588–596.
- (7) Hardeman, K.; Samyn, B.; Van Der Eycken, J.; Van Beeumen, J. *Protein Sci.* **1998**, *7*, 1593–1602.
- (8) Samyn, B.; Hardeman, K.; Van der Eycken, J.; Van Beeumen, J. *Anal. Chem.* **2000**, *72*, 1389–1399.
- (9) Tsugita, A.; Takamoto, K.; Iwadata, H.; Kamo, M.; Yano, H.; Miyatake, N.; Satake, K. In *Methods in Protein Sequence Analysis*; Imahori, K., Sakiyama, F., Eds.; Plenum Press: New York, 1993; pp 55–62.
- (10) Takamoto, K.; Kamo, M.; Kubota, K.; Satake, K.; Tsugita, A. *Eur. J. Biochem.* **1995**, *228*, 362–372.
- (11) Knierman, M. D.; Coligan, J. E.; Parker, K. C. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 1007–1010.
- (12) Gobom, J.; Mirgorodskaya, E.; Nordhoff, E.; Hojrup, P.; Roepstorff, P. *Anal. Chem.* **1999**, *71*, 919–927.
- (13) Liao, T.-H.; Sahnikow, J.; Moore, S.; Stein, W. H. *J. Biol. Chem.* **1973**, *248*, 1489–1495.
- (14) Hermodson, M. A.; Kuhn, R. W.; Walsh, K. A.; Neurath, H.; Eriksen, N.; Benditt, R. *Biochemistry* **1972**, *11*, 2934–2938.
- (15) Tsugita, A.; Van den Broek, R.; Przybylski, M. *FEBS Lett.* **1982**, *137*, 19–24.
- (16) Bradley, C. V.; Williams, D. H.; Hanley, M. R. *Biochem. Biophys. Res. Commun.* **1982**, *104*, 1223–1230.
- (17) Caprioli, R. M.; Fan, T. *Anal. Biochem.* **1986**, *154*, 596–603.
- (18) Chait, B. T.; Chaudhary, T.; Field, F. In *Methods in Protein Sequence Analysis*; Walsh, K. A., Ed.; The Human Press Inc.: Clifton, NJ, 1986; pp 483–492.
- (19) Craig, A. G.; Engstrom, A.; Bennich, H.; Kamensky, I. *Biomed. Environ. Mass Spectrom.* **1987**, *14*, 669–673.
- (20) Caprioli, R. M.; Klarskov, K.; Breddam, K.; Roepstorff, P. *Anal. Biochem.* **1989**, *180*, 28–37.

strong dependence of the enzyme cleavage kinetics on the amino acid sequence, and the inadequate mass accuracy and resolution of most commonly used mass spectrometers for analyzing the C-terminus of proteins with molecular mass of >15 kDa. A combination of chemical and enzymatic degradation has also been used with MS readout.<sup>26</sup> Procedures for specifically identifying the C-terminal peptide include tritium labeling,<sup>27</sup> labeling of carboxyl groups with methyl ester,<sup>28</sup> and enzymatic cleavage of proteins in the presence of H<sub>2</sub><sup>18</sup>O.<sup>18,29–36</sup> An esterification procedure has also been reported that facilitates the specific recognition of the C-terminal peptide generated by cyanogen bromide digests of proteins.<sup>37</sup> More recent approaches for C-terminal peptide sequence analysis include multistage MS of alkali-cationized peptides in an ion trap<sup>38</sup> and fragmentation of intact protein ions in Fourier transform mass spectrometers.<sup>39</sup>

Notwithstanding the utility of the above-described methods, the C-terminus remains a problematic region of proteins that often goes unanalyzed because of the lack of a sufficiently facile, robust procedure. In an endeavor to provide such a procedure, we developed a method in which the C-terminal region of proteins is specifically isolated, identified by matrix-assisted laser desorption/ionization time-of-flight MS (MALDI-TOF-MS), and ultimately sequenced by tandem MS. Anhydrotrypsin, a catalytically inert derivative of trypsin in which Ser-195 is converted to dehydroalanine, has the capability of binding peptides containing a lysine (K) or arginine (R) residue at their C-termini.<sup>40</sup> The modified enzyme has been used for the isolation of the C-terminal portion of proteins following trypsin digestion.<sup>40–43</sup> In the method described here, we digest the protein with lysyl endopeptidase (Lys-

C) and use anhydrotrypsin to capture all the resulting proteolytic peptides with the exception of the C-terminal peptide (provided that it does not terminate in K or R). The novelty of the present approach is in coupling the isolation of the C-terminal peptide (by anhydrotrypsin) to a rapid, robust MS readout procedure (i.e., MALDI-TOF-MS) in order to facilitate its identification. Tandem MS is then used to obtain information concerning the sequence of the C-terminal peptide. Here, we demonstrate the characterization of the C-termini of several proteins, including a protein with molecular mass of 56 kDa.

## EXPERIMENTAL SECTION

**Materials and Reagents.** Anhydrotrypsin coupled to agarose beads was purchased from Panvera (Madison, WI). Trifluoroacetic acid and acetonitrile were from Pierce (Rockford, IL). The MALDI matrix, 2,5-dihydroxybenzoic acid (DHB), was obtained from Sigma (St. Louis, MO). Poros RII, for the removal of sodium ion, was from Perseptive BioSystems (Framingham, MA). Sequencing grade Lys-C was from Boehringer Mannheim (Indianapolis, IN). Water was purified using a Millipore (Bedford, MA) Milli-Q UV plus purification system. All the organic solvents were HPLC grade, and all the other chemicals were reagent grade.

The recombinant construct of the DNA-binding protein max, residues 22–113, (provided by A. R. Ferre-D'Amare and S. K. Burley) was expressed and purified as described.<sup>44</sup> The helix-loop-helix region of the *Drosophila melanogaster* protein hairy, residues 30–96, was provided by S. K. Burley. Human karyopherin  $\alpha$ -2, residues 40–510, (provided by E. Conti and J. Kuriyan) was expressed and purified as described.<sup>45</sup> Protein concentrations were determined by absorbance at 280 nm, using an extinction coefficient calculated as described.<sup>46</sup>

**Isolation of the C-Terminal Peptide (Figure 1).** Anhydrotrypsin coupled to agarose beads was washed 3 times with 50 mM sodium acetate buffer. This wash was necessary to get rid of the NaN<sub>3</sub> present in the original storage buffer. A volume of the sodium acetate buffer equal to the bead volume was then added to the beads. The proteins of interest were dissolved in 25 mM Tris-HCl buffer (pH 8.5) and incubated for 4 h at 37 °C with Lys-C (1/20 enzyme/substrate (w/w)). Between 1 and 2  $\mu$ L of proteolytic digest solution, obtained from ~20 pmol of protein, was added to the suspension containing anhydrotrypsin and the resultant mixture was vortexed for 5 min. The amount of anhydrotrypsin was adjusted for each of the proteins investigated. In every case, the first trial was with 20  $\mu$ L of bead suspension. When this volume of beads proved insufficient to capture the bulk of the non-C-terminal Lys-C peptides and thus isolate the C-terminal peptide in the supernatant, additional anhydrotrypsin suspension was added. (The level of capture was rapidly assayed by MALDI-MS analysis.) In the examples shown, the amount added for the proteins hairy, max, and karyopherin  $\alpha$ -2 were respectively 20, 35, and 100  $\mu$ L. To ensure success of the current procedure, it is important that anhydrotrypsin be in sufficient excess to efficiently

- (21) Patterson, D. H.; Tarr, G. E.; Regnier, F. E.; Martin, S. A. *Anal. Chem.* **1995**, *67*, 3971–3978.
- (22) Woods, A. S.; Huang, A. Y.; Cotter, R. J.; Pasternack, G. R.; Pardoll, D. M.; Jaffee, E. M. *Anal. Biochem.* **1995**, *226*, 15–25.
- (23) Thiede, B.; Wittman-Liebold, B.; Michael, B.; Krause, E. *FEBS Lett.* **1995**, *357*, 65–69.
- (24) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1–8.
- (25) Bonetto, V.; Bergman, A. C.; Jornvall, H.; Sillard, R. *Anal. Chem.* **1997**, *69*, 1315–1319.
- (26) Thiede, B.; Salnikow, J.; Wittmann-Liebold, B. *Eur. J. Biochem.* **1997**, *244*, 750–754.
- (27) Matsuo, H.; Narita, K. In *Protein Sequence Determination*; Needleman, S. B., Ed.; Springer-Verlag: Berlin, 1975; pp 104–112.
- (28) Yates, J. R., III; Shabanowitz, J.; Griffin, P. R.; Zhu, N. Z.; Hunt, D. H. In *Techniques in Protein Chemistry*; Hugly, T. E., Ed.; Academic Press: San Diego, CA, 1989; pp 168–175.
- (29) Desiderio, D.; Kai, M. *Biomed. Mass. Spectrom.* **1983**, *10*, 471–479.
- (30) Rose, K.; Simona, M.; Offord, R.; Prior, C.; Thatcher, D. *Biochem. J.* **1983**, *215*, 273–277.
- (31) Gaskell, S. J.; Haroldsen, P. E.; Reilly, M. H. *Biomed. Environm. Mass. Spectrom.* **1988**, *16*, 31–33.
- (32) Rose, K.; Savoy, L.; Simona, M.; Offord, R.; Wingfield, P. *Biochem. J.* **1991**, *250*, 253–259.
- (33) Takao, T.; Hori, H.; Okamoto, K.; Harada, A.; Kamachi, M.; Shimonishi, Y. *Rapid Commun. Mass. Spectrom.* **1991**, *5*, 312–315.
- (34) Whaley, B.; Caprioli, R. M. *Biol. Mass. Spectrom.* **1991**, *20*, 210–214.
- (35) Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D. *Electrophoresis* **1996**, *17*, 945–953.
- (36) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Commun. Mass. Spectrom.* **1997**, *11*, 1015–1024.
- (37) Murphy, C. M.; Fenselau, C. *Anal. Chem.* **1995**, *67*, 1644–1645.
- (38) Lin, T.; Glish, G. L. *Anal. Chem.* **1998**, *70*, 5162–5165.
- (39) Kelleher, N. L.; Taylor, S. V.; Grannis, D.; Kinsland, C.; Chiu, H. J.; Begley, T. P.; McLafferty, F. W. *Protein Sci.* **1998**, *7*, 1796–1801.
- (40) Ishii, S.; Yokosowa, H.; Kumasaki, T.; Nakamura, I. *Methods n Enzymol.* **1983**, *91*, 378–383.
- (41) Yokosowa, H.; Ishii, S. *Biochem. Biophys. Res. Commun.* **1976**, *72*, 1443–1449.

- (42) Ishii, S.; Yokosowa, H.; Shiba, S.; Kasai, K. *Adv. Exp. Med. Biol.* **1979**, *120A*, 15–27.
- (43) Kumasaki, T.; Nakako, T.; Arisaka, F.; Ishii, S. *Proteins* **1986**, *1*, 100–107.
- (44) Ferre-D'Amare, A. R.; Prendergast, G. C.; Ziff, E. B.; Burley, S. K. *Nature* **1993**, *363*, 38–45.
- (45) Conti, E.; Uy, M.; Leighton, L.; Blobel, G.; Kuriyan, J. *Cell* **1998**, *94*, 193–204.
- (46) Perkins, S. J. *Eur. J. Biochem.* **1986**, *157*, 169–180.

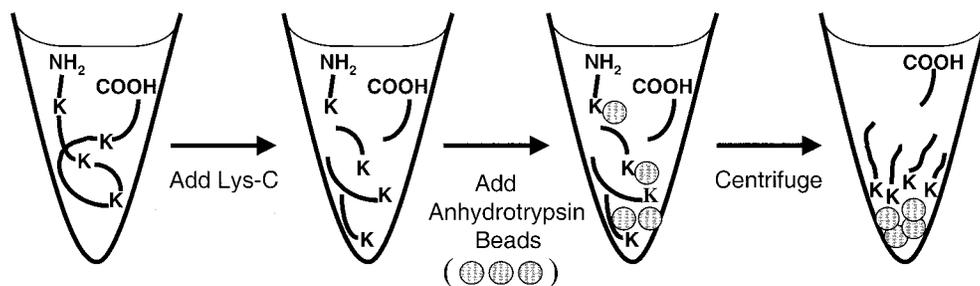


Figure 1. Scheme for defining the C-terminus of a protein. The protein is dissolved, digested with the endoprotease Lys-C, and incubated with anhydrotrypsin coupled to agarose beads. After rapid centrifugation, the C-terminal peptide is recovered from the supernatant while all the other peptides are bound to the anhydrotrypsin-agarose beads. Comparison of the MALDI-TOF mass spectra obtained before and after incubation with anhydrotrypsin allows the identification of the C-terminal peptide. Characterization of the sequence of the C-terminal peptide is obtained by analyzing a fraction of the supernatant using LC-MS/MS, selecting a single precursor species corresponding to the expected  $m/z$  of the C-terminal peptide.

capture the bulk of the internal Lys-C fragments. We note, however, that too large an excess of anhydrotrypsin might ultimately remove the C-terminal peptide from the solution as a result of nonspecific binding to the beads. To ensure maximal binding of the Lys-C peptides, it is important that the pH of the solution containing the anhydrotrypsin be maintained close to pH 5.0.<sup>40</sup> Prior to taking aliquots for MALDI-MS and LC-MS analysis, the samples were centrifuged at 14000g for 30 s in order to separate the beads from the supernatant containing the C-terminal peptide.

**MALDI-TOF Mass Spectrometry.** The MALDI mass spectra were obtained with a Perseptive BioSystems Voyager-DE STR mass spectrometer (Framingham, MA). The matrix solution was 25 mg/mL of DHB in 0.1% trifluoroacetic acid (TFA)/acetonitrile 2/1 (v/v). One microliter of the Lys-C digest (or of the Lys-C digest supernatant after incubation with anhydrotrypsin) was mixed with 4  $\mu$ L of matrix solution, and 0.5  $\mu$ L of the resulting solution was added to the MS sample probe. Because the karyopherin preparation contained a high concentration of salt, 2  $\mu$ L of Poros beads was added to the supernatant containing the C-terminal peptide prior to MALDI-MS analysis. The beads were washed with water, and the C-terminal peptide eluted from the beads with acetonitrile prior to mixing with matrix solution. The MALDI mass spectra were externally calibrated using a mixture of known synthetic peptides. After data collection, the spectra were processed using the program "m/z".<sup>47</sup>

#### LC-MS/MS Characterization of the C-Terminal Peptides.

To further characterize the C-terminal peptide, a fraction of the supernatant from the anhydrotrypsin incubation was analyzed by LC-MS/MS using a Michrom UMA HPLC (Michrom BioResources Inc. Auburn, CA) coupled to an LCQ ion trap mass spectrometer (ThermoQuest, San Jose, CA). The HPLC was equipped with a precolumn splitter to yield a flow of  $\sim 2$   $\mu$ L/min through a 0.2  $\times$  50 mm MAGIC C18 column (Michrom). The column was equilibrated with a solvent containing 0.01% TFA, 0.1% acetic acid, and 5% acetonitrile. The peptides were eluted using a 15-min linear gradient from 5 to 100% with a solvent containing 0.01% TFA, 0.1% acetic acid, and 95% acetonitrile. To maximize the sensitivity of the analysis, the ion trap mass spectrometer was operated exclusively in the MS/MS mode, selecting on a single

precursor  $m/z$  value for the entire LC-MS/MS run. The  $m/z$  values used in these MS/MS analyses were determined by MALDI-MS characterization, initially assuming that the charge state of the peptide ions of interest observed in the electrospray ionization measurement was 2+. For the MS/MS analyses shown here, only 2  $\mu$ L of each sample was injected, leaving a large proportion of the sample available for further analysis (if needed).

The present C-terminal characterization procedure is shown schematically in Figure 1. For the isolation of the C-terminal peptide, we digest the protein with Lys-C and then add anhydrotrypsin coupled to agarose beads. This enzyme binds peptides containing K or R at the C-terminus. If the C-terminal amino acid is not K or R, then all proteolytic peptides except for the C-terminal peptide will bind to the anhydrotrypsin beads. Approximately 84% of all protein sequences present in the SwissProt database do not contain K or R at the C-terminus (personal communication, David Fenyo) and should therefore be accessible to the present procedure (although an additional subset of the C-terminal peptides will be too short for ready analysis by MALDI-MS.) Comparison of the MALDI mass spectra obtained prior to and after incubation with anhydrotrypsin allows the identification of the C-terminal peptide. Once the C-terminal peptide is identified, its sequence can be characterized by tandem MS.

We first used this strategy to characterize the C-terminus of a low molecular mass ( $\sim 8$  kDa) transcription factor hairy construct. Figure 2A shows the MALDI-MS spectrum of the Lys-C digest of the hairy construct. The observed peptide ions cover 64 out of the 67 residues of hairy. After incubation with anhydrotrypsin, the spectrum is dominated by a peak arising from the peptide encompassing residues 55–67 (Figure 2B). Therefore, this peptide is the prime candidate for the C-terminal peptide. From a comparison of the spectra obtained prior to and after incubation, it may appear that 13–26 should also be considered as a possible candidate. In such cases, we found that we could objectively assess the best candidate by evaluation of the peak height ratios,  $R_{x-y}$  = peak height of peptide  $x-y$  after incubation/peak height of peptide  $x-y$  prior to incubation. Because  $R_{13-26} = 0.5$  and  $R_{55-67} = 1.1$ , peptide 55–67 is the more probable C-terminal peptide candidate. The appearance in Figure 2B of small amounts of the non-C-terminal peptide 13–26 can be attributed to its relatively low affinity for anhydrotrypsin.<sup>40</sup> It can be seen that the signal-to-noise ratio of the peptide peaks after incubation is lower than that before

(47) Fenyo, D.; Ens, W. E.; Carroll, J.; Beavis, R. C. *Proceedings of the 45th ASMS Conference on Mass Spectrometry and Allied Topics*, Palm Springs, CA, 1997.

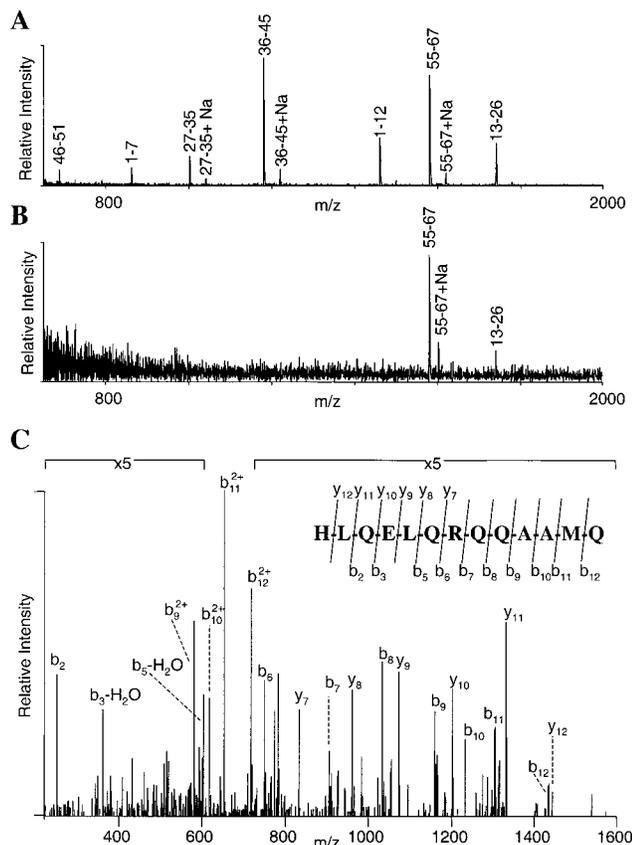


Figure 2. Characterization of the C-terminal-peptide of a  $\sim 8$  kDa hairy construct. A preparation containing residues 30–96 of hairy was digested with endoprotease Lys-C and subjected to MALDI-TOF-MS analysis prior to (A) and after (B) incubation with anhydrotrypsin-agarose beads. The peaks are labeled with their corresponding amino acid sequence (here renumbered 1–67 for simplicity). Comparison of (A) and (B) indicates that 55–67 is the C-terminal peptide (see text). A fraction of the supernatant after incubation with anhydrotrypsin was also analyzed by LC-MS/MS using an LCQ mass spectrometer (C). The  $m/z$  value of the precursor ion used in the MS/MS analysis corresponded to that calculated for the doubly charged C-terminal peptide (55–67) $^{2+}$ , as determined from the MALDI analysis shown in (A) and (B). The amino acid sequence of the expected C-terminal peptide is shown together with b and y fragment ions identified from the MS/MS spectrum.

incubation (Figure 2), presumably as a consequence of losses due to nonspecific binding to the anhydrotrypsin-agarose beads as well as dilution effects. To further characterize the C-terminal peptide, a fraction of the supernatant was analyzed by tandem MS (Figure 2C). The MS/MS spectrum was obtained using an ion trap selecting a single precursor  $m/z$  value for the entire run—i.e., applying an approach similar to that described previously for the characterization of phosphorylation sites.<sup>48</sup> The advantage of this mode of operation is that useful MS/MS can be obtained when the single-stage MS peak of interest is very small (or even in the noise). The results of the MS/MS analysis shown in Figure 2C are consistent with the expected sequence of the hairy construct and allowed complete sequence characterization of the C-terminal peptide.

The same approach was used for the characterization of the C-terminal peptide of a DNA-binding protein max construct

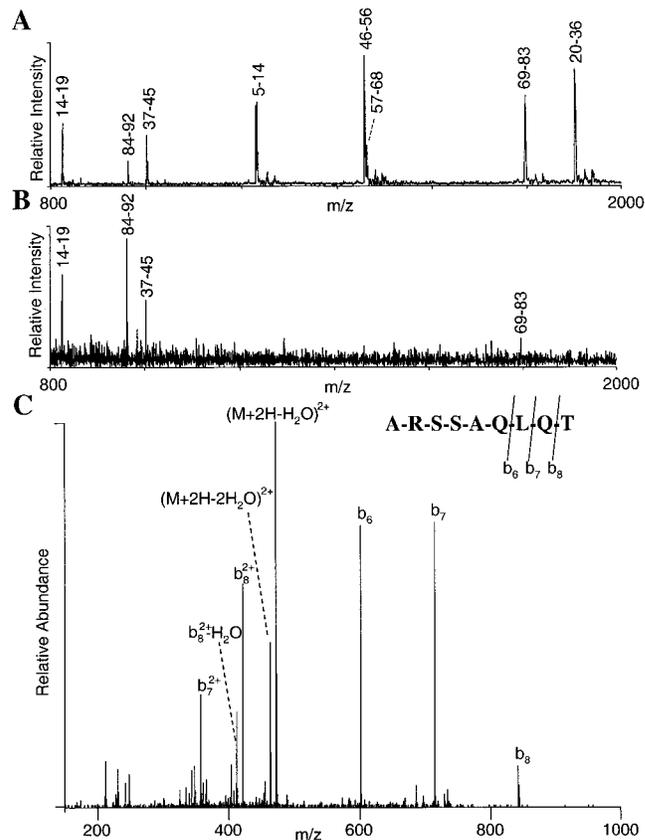


Figure 3. Characterization of the C-terminal peptide of an  $\sim 11$  kDa max construct. A preparation containing max 22–113 was digested with endoprotease Lys-C and subjected to MALDI-TOF-MS analysis prior to (A) and after (B) incubation with anhydrotrypsin-agarose beads. The peaks are labeled with their corresponding amino acid sequence (here renumbered 1–92 for simplicity). Comparison of (A) and (B) indicates that 84–92 is the C-terminal peptide. A fraction of the supernatant after incubation with anhydrotrypsin was also analyzed by LC-MS/MS using an LCQ mass spectrometer (C).

(molecular mass  $\sim 11$  kDa). Here, again the C-terminus could be readily identified by comparison of the MALDI spectra obtained prior to and after incubation with anhydrotrypsin (Figure 3A and B). The peptide labeled 84–92 corresponds to the only peptide in the spectrum whose relative peak height increases after incubation with anhydrotrypsin and is therefore identified as the C-terminal peptide. The tandem mass spectral data were consistent with the expected terminal sequence of the max construct (Figure 3C) although the MS/MS spectrum provided only partial sequence information. Nevertheless, the three C-terminal amino acids were confirmed and all the major ions in the spectrum were consistent with the expected fragmentation for the C-terminal peptide.

We also applied the present procedure to characterize the C-terminus of a higher molecular mass protein, karyopherin  $\alpha$ -2 (56 kDa). Two peptides were observed after incubation with anhydrotrypsin (Figure 4). Comparison of the spectra obtained prior to and after incubation with anhydrotrypsin revealed that the relative intensity of the peptide 495–510 increased  $\sim 3$ -fold after incubation, while the relative intensity of the peptide 84–102 did not change significantly. Thus, it would appear that the C-terminus of the protein is 495–510. The tandem MS data of peptide 495–510 are shown in Figure 4C. Preferential fragmentation is observed to occur at the V–P peptide bond. Among the

(48) Zhang, X.; Herring, C. J.; Romano, P. R.; Szczepanowska, J.; Brzeska, H.; Hinnebusch, A. G.; Qin J. *Anal. Chem.* **1998**, *70*, 2050–2059

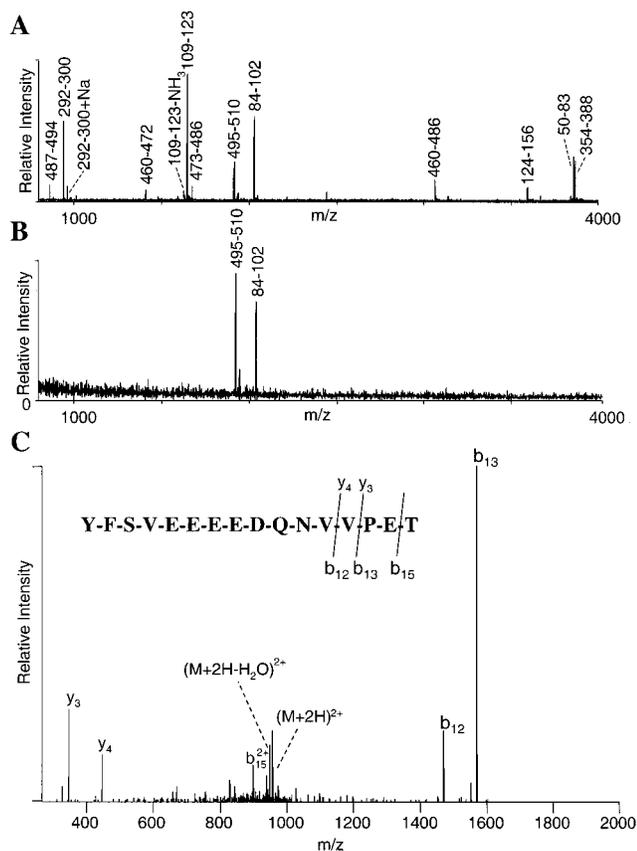


Figure 4. Characterization of the C-terminal peptide of the 56 kDa protein karyopherin  $\alpha$ -2. A preparation containing karyopherin  $\alpha$ -2 was digested with endoprotease Lys-C and subjected to MALDI-TOF-MS analysis prior to (A) and after (B) incubation with anhydrotrypsin-agarose beads. Comparison of (A) and (B) indicates that 495–510 is the C-terminal peptide. A fraction of the supernatant after incubation with anhydrotrypsin was also analyzed by LC–MS/MS using an LCQ mass spectrometer (C).

few fragments that are observed,  $b_{15}$  identifies the C-terminal amino acid to be threonine. This experiment demonstrates that the present procedure can be used to characterize the C-terminus of proteins with molecular masses in excess of 50 kDa.

## CONCLUSION

We have demonstrated a facile procedure for defining the C-terminal fragment of proteins. A notable strength of this procedure is the isolation of the C-terminal peptide, enabling the characterization of the terminal of high molecular mass proteins (e.g., karyopherin  $\alpha$ -2, molecular mass 56 kDa). Indeed, provided that the proteins of interest do not contain lysine or arginine residues at the C-terminus (true for  $\sim 84\%$  of proteins) or that the C-terminal peptides are not too short for analysis by MALDI-MS, there does not appear to be an obvious molecular mass limit to its application. The present methodology was reliably applied at the 20-pmol level, which should be sufficient for solving many current problems relating to C-terminal characterization. However, we believe that the sensitivity can be further improved either by engineering a catalytically inert form of trypsin to have a higher binding affinity or by increasing the density of the stationary affinity phase. The present method should be of considerable interest for proteomic studies that require the identification of the C-termini of protein components of complex mixtures—an application that would benefit from the use of  $^{18}\text{O}/^{16}\text{O}$  labeling to provide an added level of discrimination between the C-terminal peptides and internal proteolytic fragments that are not completely removed by the anhydrotrypsin beads.<sup>49</sup>

## ACKNOWLEDGMENT

This work was supported by the National Institute of Health (Grant RR00862) and the National Science Foundation (Grant 9630936). We thank A. R. Ferre-D'Amare, S. K. Burley, E. Conti, and J. Kuriyan (The Rockefeller University) for providing the proteins used in this study. We are grateful to R. Wang, D. Fenyo, Y. Zhao, Y. Oda, and S. Cohen for helpful suggestions.

Received for review January 10, 2000. Accepted April 21, 2000.

AC000045I

(49) Toshiyuki, K.; Takazawa, T.; Nakamura, T. *Anal. Chem.* **2000**, *72*, 1179–1185.