Generating Protein Sequence Tags by Combining Cone and Conventional Collision Induced Dissociation in a Quadrupole Time-of-Flight Mass Spectrometer

Joy M. Ginter, Feng Zhou, and Murray V. Johnston
Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware, USA

The goal of proteomics research is to be able to identify and quantify the vast numbers of proteins within an organism or tissue. "Top-down" methods address this goal without the need for proteolytic digestion prior to mass analysis. We report here an approach for top-down protein identification that has been implemented on a commercially available, unmodified Qq-TOF mass spectrometer. Intact protein molecular ions first undergo cone fragmentation in the electrospray inlet. Conventional MS/MS is then performed on a mass selected cone fragment using CID in the Qq interface of the Qq-TOF mass spectrometer to generate a sequence tag through a pseudo-MS^3 experiment. Seven proteins varying in molecular weight between 11 and 66 kDa were chosen to demonstrate applicability of method. After the molecular weight of the intact protein was determined, the cone voltage was varied to induce fragmentation. Cone fragment ions were then further dissociated using conventional CID, and the resulting MS/MS spectra were processed and analyzed for sequence tags. Sequence tags were easily identified from a MS/MS spectrum of a cone induced fragment ion both manually and through a de novo sequencing program included in the software associated with the mass spectrometer. Sequence tags were subjected to database searching using the PeptideSearch program of EMBL, and all protein sequence tags gave unambiguous search results. In all cases, sequence tags were found to originate from the n- and/or c-termini of the proteins.   (J Am Soc Mass Spectrom 2004, 15, 1478–1486) © 2004 American Society for Mass Spectrometry

The field of proteomics is as dynamic as the proteome itself, constantly changing and evolving to meet the needs of researchers. The demands placed on the field increase with the yearly advancements, but the ultimate goal remains the same—global identification and quantitation of proteins in an organism or tissue. The most widely used methods for identifying whole proteomes are two-dimensional gel electrophoresis (2D-GE) and the more recently developed, "shot gun" approach [1, 2]. The 2D-GE approach involves separating proteins first by their isoelectric point and subsequently by their molecular weight in a SDS gel. Individual protein spots are enzymatically digested within the gel to yield peptides that are easily extracted from the gel matrix. The peptide solution is then mass analyzed using either MALDI-MS or ESI-MS to yield a peptide mass fingerprint, sometimes after further chromatographic separation. Protein identifications are then determined from database searching on the resulting peptide mass fingerprints and/or using sequence tags obtained from MS/MS experiments on individual peptides [3, 4]. In the shot gun approach, hundreds of proteins are simultaneously identified after global digestion of a protein mixture. The resultant peptide solution is then separated through multidimensional chromatography (2D-LC); strong cation exchange coupled with reverse phase is most common. The time limitations in these and most methods for proteomic analysis are found in sample preparation and separation procedures. The mass spectrometry step that is used ultimately to generate the sequence tags or peptide mass fingerprints for identification takes relatively little time in comparison.

Identifying proteins by introducing proteolytically cleaved fragments into a mass spectrometer has become known as 'bottom-up' proteomics, and remains the method of choice for many researchers in the proteomics field due to its proven record. Alternatively, introducing intact proteins into a mass spectrometer and fragmenting through the use of the instrument has been termed 'top-down' proteomics [5]. McLafferty and co-workers were the first to introduce the concept of top-down proteomics by utilizing the high resolving power of FT-ICR MS to explore the fragmentation of intact proteins [6, 7]. Various dissociation techniques

have been used to facilitate fragmentation. An early example was the work of Kelleher and McLafferty using nozzle-skimmer fragmentation of a 42kDa protein, thiaminase I to elucidate enzymatic activity information as well as active site location [8]. More recently, electron capture dissociation (ECD) has proven suitable for sequencing entire proteins as well as determining structural characteristics and modifications [9–11]. Tertiary structure information has also been determined through the use of a combination of chemical cross linkers and top-down strategies employing FT-ICR MS [12].

While FT-ICR MS has demonstrated its superiority in the realm of top-down proteomics, practically speaking many researchers do not have the instrumentation to support such experiments. McLuckey and coworkers extended the practice of top-down proteomics to ion traps. They have shown that although ion traps do not typically have the resolving power needed to identify highly charged protein fragments, ion traps utilizing ion–ion chemistry, specifically proton transfer reactions, can reduce multiply charged fragments to singly charged species enabling easier spectral interpretation with less ambiguity [13–16]. Early reports on the use of this technique focused generally on smaller molecular weight proteins but a recent report demonstrated the technique to be suitable for a somewhat larger protein at 25.9 kDa and also compared fragmentation behavior of the reduced and native form [17]. Clemmer et al. also addressed the problem of complicated spectra resulting from protein dissociation in an ion trap by coupling an ion trap to an ion mobility time-of-flight (TOF) mass spectrometer to separate the multiply charged fragments [18].

Other instruments that have been shown to have the capability of dissociating protein ions include triple quadrupole [19–22], quadrupole/time-of-flight (Qq-TOF) [23], and most recently matrix assisted laser desorption/ionization—time-of-flight/time-of-flight (MALDI-TOF/TOF) mass spectrometers [24, 25]. None of these instruments have equivalent resolving power to FT-ICR MS nor do they use significant ion–ion chemistry to simplify product ion spectra. The dissociation of protein molecular ions in Qq-TOF and MALDI-TOF/TOF instruments has been shown to lead to the identification of limited sequence tags although far less efficiently than by FT-ICR MS. Nemeth-Cawley et al. demonstrated the utility of the Qq-TOF for accurately measuring intact protein molecular masses [23]. Additionally, in some instances sequence information for a protein could be generated from CID studies of multiply charged protein molecular ions. In most cases, however, several precursor ions needed to be dissociated to give a sufficient number of resolved product ions to identify a sequence tag. In principle, the capabilities of some of these instruments can be extended through the use of cone fragmentation (also referred to as in-source CID, up-front CID, or nozzle-skimmer dissociation). Cone fragmentation can lead to pseudo-

MS³ capabilities for instruments not typically equipped for such experiments. Examples include the use of cone fragmentation for sequencing oligonucleotides with a triple quadrupole mass spectrometer and peptide sequencing in a MALDI-QqTOF [26, 27]. The application of cone fragmentation for larger biopolymers was initially explored by Loo et al. demonstrating the favored dissociation of higher charge states over lower charge states as cone voltage is increased [28]. Loo and coworkers continued exploring the use of cone fragmentation in a triple quadrupole and later demonstrating feasibility in a Qq-TOF reporting its use for determining primary sequence information of proteins via a pseudo MS³ experiment [21, 22, 29]. Recently, Thevis, Loo, R., and Loo, J. [30] demonstrated the utility of a Qq-TOF for top-down characterization of transferrins which had previously been explored through the use of nozzle-skimmer dissociation [31]. In this most recent report, sequence information from conventional MS/MS of intact transferrins was compared to sequence information generated through cone fragmentation in the Qq-TOF, and both approaches were found to provide comparable results.

We report here the use of cone fragmentation in a commercially available Qq-TOF for pseudo-MS³ experiments on standard proteins ranging in molecular weight between 11 and 66 kDa. We show that this dual cone fragmentation and conventional CID approach provides large sequence tags (greater than 10 amino acids in most cases) near the termini of proteins which enable unique protein identification when subjected to database searching. Results reported herein also agree quite well with those of Suckau and Resemann who used in-source decay in a MALDI-TOF/TOF mass spectrometer for protein identification ("T³-sequencing") [25]. Pseudo-MS³ represents a potentially new approach for high throughput protein characterization that combines several attributes of both bottom-up and top-down methods. As with top-down methods, intact molecular mass measurements are obtained prior to inducing fragmentation. As with bottom-up methods, CID of the cone fragment ions leads to the identification of sequence tags for database searching, but without the need for proteolytic digestion.

Experimental

Instrumentation

Data were collected on a Micromass Q-TOF Ultima API-US (Manchester, UK) equipped with a z-spray configuration for electrospray ionization. Samples were loaded into 4 μm glass emitters (New Objective, Woburn, MA). In most cases, the cone voltage was increased from 45 to 65 V or until significant fragmentation occurred. The capillary voltage was between 1.5–2.1 kV. Scan time was 2.4 s with an interscan time of 0.1 s. Total acquisition time depended on quality of MS/MS spectra, but on average total time of acquisition

Table 1. Proteins chosen for analysis with average molecular weights,[a] cone fragment masses,[b] and the corresponding sequence[c] with PeptideSearch results

Protein	M_r (Da)	Cone fragment M^b	Corresponding sequence[c]	PeptideSearch results
Cytochrome *c* (equine)	12230.00	4658.20	***MEYLENPKKYIPGTKMIFAGIKKKTE*** REDLIAYLKKATNE(c-terminus) (4658.50)	9/9 matches to cytochrome *c*
Thioredoxin (e. coli recombinant)	11675.00	2798.50	(n-terminusM)S***DKIIHLTDDSFDTDVLKA DGAILVD*** (2816.42)	12/12 matches to thioredoxin
Alpha-Lactalbumin (bovine)	14172.00	2651.41	(n-terminus)***EQLTKCEVFRELKDLKGYGGV*** SL (2669.40)	9/9 matches to α-lactalbumin
		1821.89	***LCSEKLDQWLCE***KL (c-terminus) (1821.89)	17/17 matches to α-lactalbumin
Trypsin Inhibitor (soybean)	19978.00	1393.81 (unresolved)	(n-terminusDF)***VLDNEGNPLENG*** (seq. tag[d])	9/11 matches to trypsin inhibitor
Actin (bovine)	41951.00	3205.49	***STFQQMWITK WEYDEAGPSIVHRKC***F(c-terminus) (3187.51)	23/26 matches to actin
Carbonic Anhydrase II (bovine)	29027.00	2959.76	***PELLMLAN***WRPAQPLKNR QVRGFPK(c-terminus) (2959.66)	2/2 matches to carbonic anhydrase
Bovine Serum Albumin	66433.00	3497.76	(n-terminus)DTHKSEIAHRFKDLG ***EEHFKGLVLIAFSQY*** (3515.81)	2/2 matches to serum albumin
		3334.80	(n-terminus)DTHKSEIAHRFKDLGEEHFK GL***VLIAFSQ*** (3352.74)	12/16 matches to serum albumin

[a]For each protein, the average molecular weight determined through deconvoluting charge states with the MaxEnt1 function is listed.
[b]Additionally, the monoisotopic masses of cone fragments chosen for further CID studies is listed with the cone fragment mass of trypsin inhibitor being an exception as this ion was unresolved.
[c]In instances where there is no agreement a difference of 18 Da is observed. Bold italicized portions of the sequence indicate the identified sequence tag and corresponds to the amino acid sequence searched against the database with PeptideSearch.
[d]The corresponding sequence listed within the table is the sequence of the protein relating to the cone fragment produced followed by the calculated monoisotopic mass of that sequence.

was not more than 5 min or less than 200 scans. The resolving power in these experiments was ~10,000. For CID experiments argon gas was used, and the collision energy varied from 25–45 eV. MS/MS spectra were analyzed using Micromass's MassLynx (version 3.5) software's de novo sequencing function (MassSeq) and through manual de novo sequencing. Protein molecular weights were calculated from the deconvolution function of the MassLynx software (MaxEnt1).

Materials

All proteins and reagents were purchased from Sigma (St. Louis, MO). A table of the seven proteins examined as well as their molecular weights is included. (Table 1) Solvents were of HPLC grade and purchased from Fisher (Pittsburgh, PA). Proteins were dissolved in a 50:50:2 mixture of $MeOH:H_2O:HOAc$ at a concentration of 50 pmol/μL with actin being an exception at 15 pmol/μl due to protein quantity limits. Proteins containing more than one disulfide bond were also reduced (DTT) and alkylated (IAA) by standard procedures [32] if it was determined cone fragmentation would not occur.

Results and Discussion

Methods

The seven proteins chosen for this study are given in Table 1. Proteins were first electrosprayed into the mass spectrometer at a low cone voltage (40–45 V) to determine molecular mass and subsequently electrosprayed at sequentially higher cone voltages up to 75–100 V. The cone voltage was stepped to higher voltages to induce fragmentation of the multiply charged molecular ions within the protein envelope. Once established that the protein would fragment at a high cone voltage, conventional CID experiments were carried out on selected cone fragment ions. Cone fragment ions were identified by first examining the charge state envelope of a protein at a low cone voltage and subsequently at a higher cone voltage. Ions that appeared as the cone voltage was increased were termed cone fragment ions.

The MS/MS spectra produced by CID were processed using MassLynx software and identified sequence tags were subjected to database searching using the PeptideSearch program (http://www.narrador. embl-heidelberg.de/GroupPages/PageLink/peptide searchpage.html). MassSeq, a de novo sequencing program within the MassLynx software, was used to sequence CID spectra. MassSeq requires that all peaks within the spectrum are first deconvoluted from multiply charged to singly charged ions (using MaxEnt3 within the MassLynx software) before the sequencing is executed. A list of potential sequences is returned and the quality of hits returned was examined manually by the user. In most instances, MassSeq data was combined with results from manual de novo sequencing by the user. The PeptideSearch program was used to search
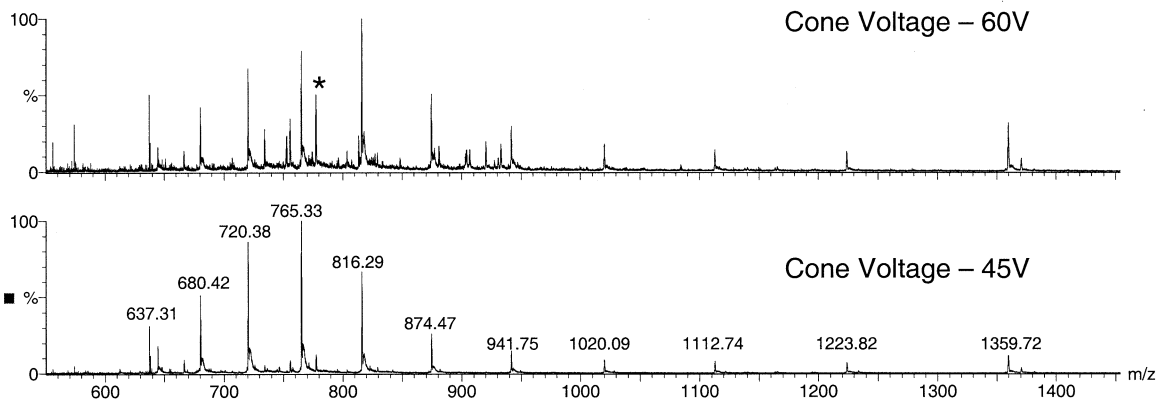
Figure 1. Charge state envelope of cytochrome *c*. The peak denoted with a star in the top spectrum is a cone induced fragment that was further subjected to conventional CID.

sequence tags against a nonredundant database. The amino acid sequence search function was used which allows a user to enter a protein's molecular weight range, partial sequence, and organism. The Peptide Search program allowed the following substitutions: leucine = isoleucine and glutamine = lysine. The latter could not be resolved in the present work because of the lack of internal lock-spray calibration on the Qq-TOF.

Cytochrome C

Figure 1 shows the charge state envelope of cytochrome *c* taken at two cone voltages (45 and 60 V). While peaks from the multiply charged molecular ions are still visible in the higher cone voltage spectrum, an increase in the number of peaks between known charge states is noticeable. When the spectrum is expanded it can be seen that most of these new peaks are isotopically resolved ions. MS/MS studies carried out on the 17+ charge state molecular ion led to the identification of a four amino acid sequence tag (MEYL) while MS/MS studies on some cone fragment ions lead to longer sequence tags of nine or more amino acids (MEYLEN PKKYIP and MIFAGIKKK). Figure 2a, shows the unprocessed MS/MS spectrum generated as a result of dissociating a 6+ cone fragment ion at 777 *m/z*. Through a manual interpretation of the spectrum the y40–y28 series consisting of 6+ to 3+ ions is easily deduced leading to a sequence tag of (MEYLENPKKYIP). Upon further processing with the Micromass MassLynx software the sequence tag was extended (MEYLENPKKY IPGTKMIFAGIKKKTE), see Table 1. Figure 2b is an example of the output of the MassSeq tool (a denovo sequencing program) in the MassLynx software and shows a y-ion series. The spectrum is dominated by y-ions although in the lower *m/z* range some b-ions are also visible.

Thioredoxin

The charge state envelope of thioredoxin (11.7 kDa) taken at a low and high cone voltage exhibits similar properties to that of cytochrome *c* (spectrum not shown). At the higher cone voltage (65 V) the baseline increases at the lower *m/z* values, and the cone induced fragments are clearly present. Performing conventional "top-down" MS/MS on both the 13+ and 12+ charge state molecular ions with the QqTOF does not generate an identifiable sequence tag, even though MALDI TOF/ TOF gives a three amino acid tag from dissociation of the [M + H]+ ion [24]. However, pseudo-MS³ with the QqTOF does generate a rather long sequence tag from the n-terminus of the protein, see Table 1. Figure 3a shows the unprocessed MS/MS spectrum of the cone fragment ion at 933 *m/z* with a 3+ charge. Processing with the MassSeq tool of MassLynx shows a y-ion series as well as a significant portion of the corresponding b-ion series (Figure 3b). The MassSeq tool led to a sequence tag of 25 amino acids (DKIIHLTDDSFDTDV LKADGAILVD). The mass of the cone induced fragment corresponds most likely to the sequence tag identified by MassLynx extended one amino acid towards the n-terminus with a water loss from the serine in amino acid position 2 [33]. As discussed below, cone fragments resulting from an 18 u difference were not unique to this protein. Additionally, sequencing results were comparable to pseudo-MS³ experiments using a triple-quadrupole by Smith and coworkers, who also reported identifying ions from the first 40 residues at the n-terminus [19].

α-Lactalbumin

α-lactalbumin (14.2 kDa) resisted cone induced fragmentation. Increasing the cone voltage to 90 V gave no enhancement in fragmentation over that taken at 45 V. (Spectra not shown.) α-Lactalbumin while not a very large protein contains four disulfide bonds which appear to inhibit fragmentation. Once the disulfide bonds
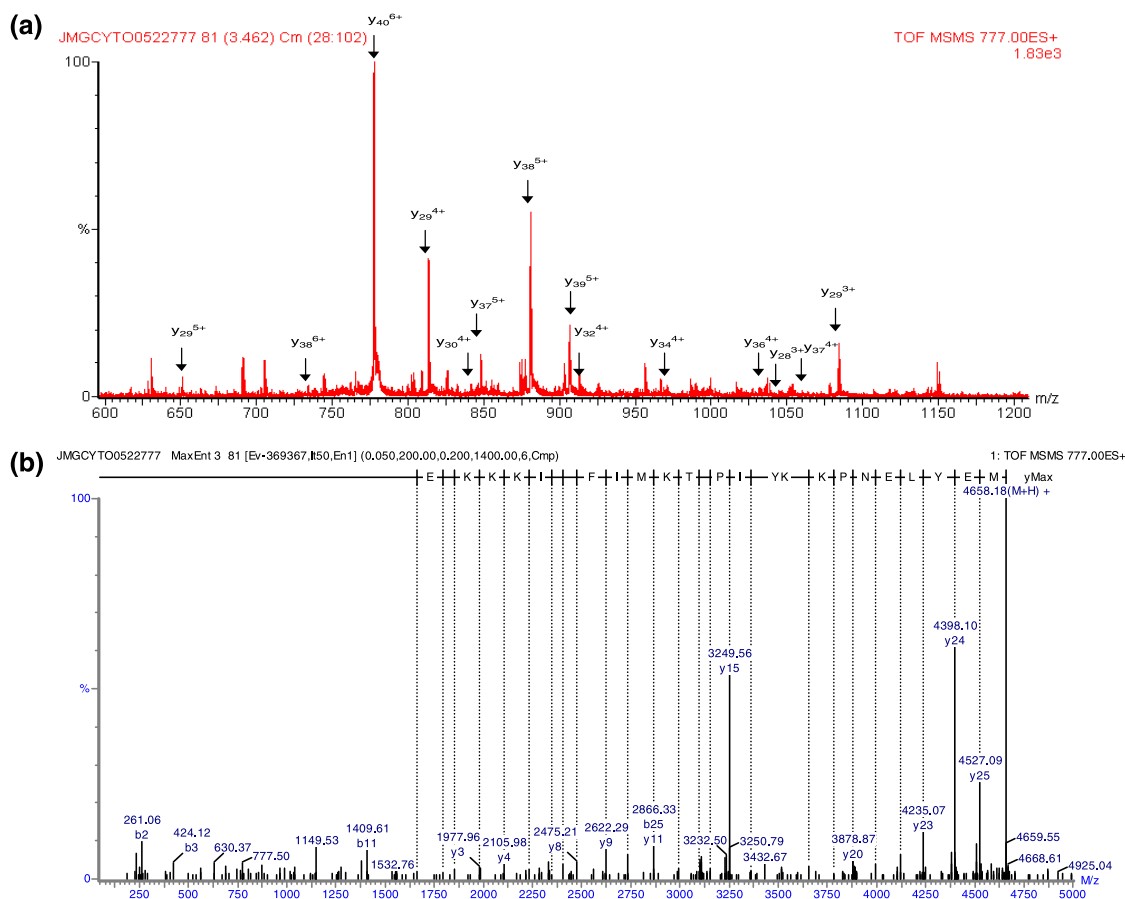
Figure 2. (**a**) Unprocessed MS/MS spectrum of a cone fragment of cytochrome *c* at *m/z* 777. Cone voltage was set to 60 V. A y-ion series is labeled where y_x refers to the y-ion series where x is indicative of amino acid position *within the entire protein sequence*. (**b**) Output from MassLynx's MassSeq (de novo sequencing tool) for MS/MS spectrum of a cone fragment at *m/z* 777. A y-ion series is labeled where y_x refers to the y-ion series *within the sequence tag that was deduced*.

are reduced and alkylated, the protein fragments readily as seen in Figure 4. Though not easily apparent in this figure, the cone fragments of interest (884 *m/z*, 3+; 911 *m/z*, 2+) are isotopically resolved. Subjecting these ions to CID gives sequence tags from two regions of the protein, a 21 amino acid tag near the n-terminus and a 12 amino acid tag at the c-terminus, see Table 1. The n-terminus cone fragment was found to have an 18 u loss which is most likely explained by cyclization at the n-terminal glutamic acid (pyroglutamic acid) [34].

Trypsin Inhibitor

Trypsin inhibitor (20.1 kDa) also was resistant to fragmentation in its unreduced form. After reduction and alkylation the protein fragmented readily with a cone voltage of 60 V. In this case, the cone fragment chosen for CID is unresolved at 1392.81 *m/z*. Examples of cone fragments up to this point included resolved ions only, ranging from a charge state of 2+ to 6+, the maximum resolved charge state for these studies. Figure 5 shows the unprocessed MS/MS spectrum that resulted from the CID experiment on the unresolved cone fragment at 1392.81

m/z. Most of the peaks seen in this unprocessed spectrum from 1200–2000 *m/z* are also unresolved. The sequence tag found by the MassSeq tool is localized to the lower *m/z* range of the spectrum. The b-series leading to this tag is labeled in Figure 5. The tag (VLDNEGNPLENG) originates from the n-terminus of the protein, see Table 1. This result shows that it is not essential to choose a resolved cone fragment ion to generate a sequence tag. The pseudo-MS³ approach described here requires isotopic resolution only among the (low mass) second-generation product ions used to generate the sequence tag. Therefore, the mass resolving power requirement of pseudo-MS³ is less stringent than that of conventional top-down analysis where the product masses are much higher.

Carbonic Anhydrase, Actin, Bovine Serum Albumin

Three higher molecular weight proteins were also chosen for this study to determine what, if any, molecular mass limit there would be. Although higher in molecular weight, each protein fragmented to give sufficient
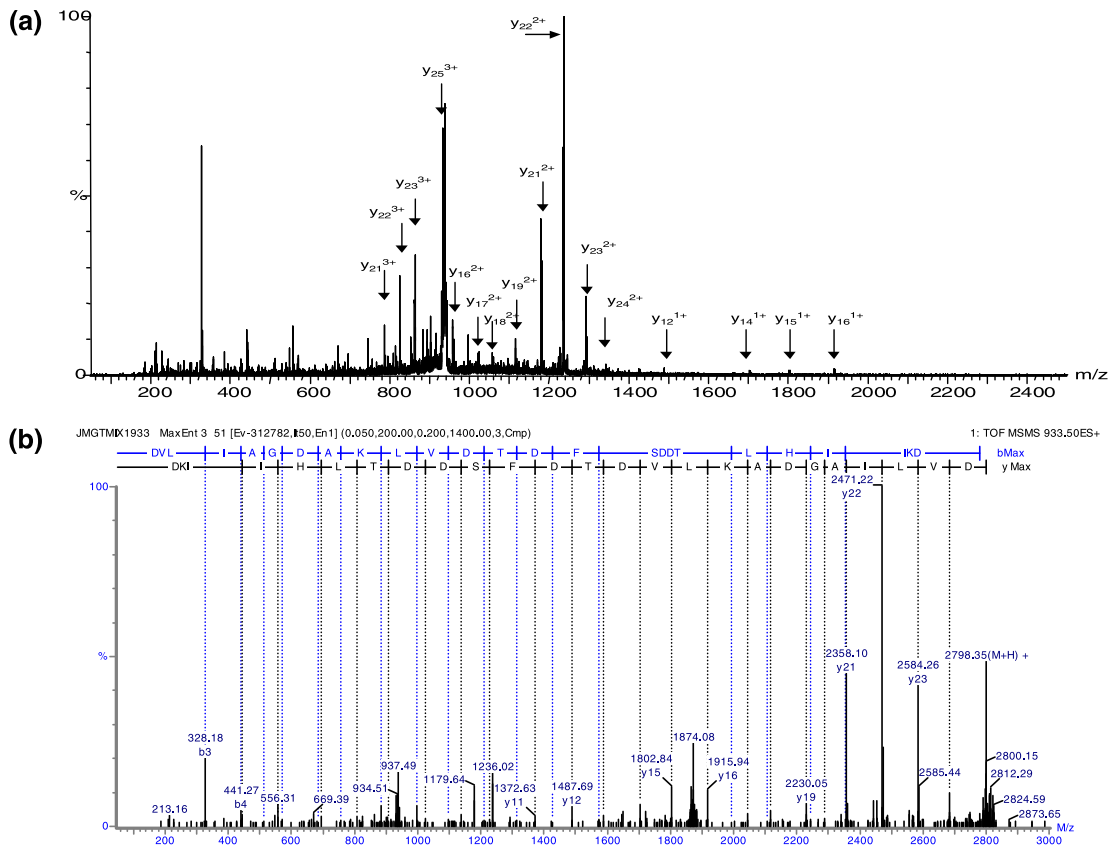
Figure 3. (**a**) MS/MS spectrum of a cone fragment of thioredoxin (*m/z* 933). Cone voltage was set to 65 V. (**b**) MassLynx's MassSeq output showing a b- and y-ion series for the dissociation of a cone fragment of thioredoxin at *m/z* 933.

sequence information for identification. Carbonic anhydrase (29 kDa) was shown to fragment well under top-down conditions by previous researchers using ECD and ion–ion interactions within an ion-trap [5]. Cone fragmentation studies using carbonic anhydrase led to an eight amino acid tag from the c-terminus unique for identification (spectrum not shown). The

recently reported T[3]-sequencing approach also identified ions from the c-terminus of carbonic anhydrase [25]. Actin (44 kDa) also fragmented successfully using pseudo-MS[3] to give a 26 amino acid tag at the c-terminus (spectrum not shown). The cone fragment for actin was found to be 18 u greater in molecular weight than the predicted sequence. This fragment was the
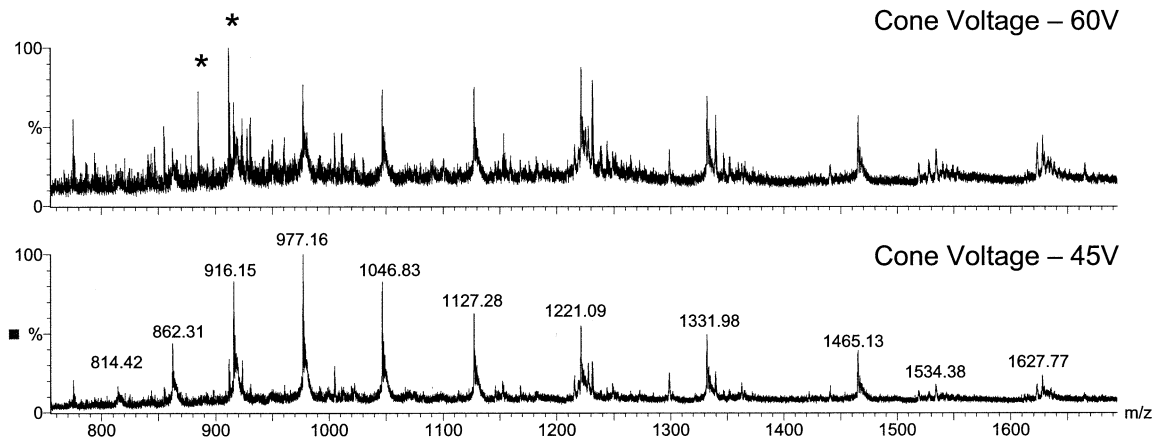

Figure 4. Charge state envelope of reduced α-lactalbumin at a cone voltage of 45V (bottom) and 60V (top). Unlike the nonreduced form, the reduced sample easily fragmented under increasing cone voltage conditions. Two fragments were chosen for CID studies in this example (indicated with a star).
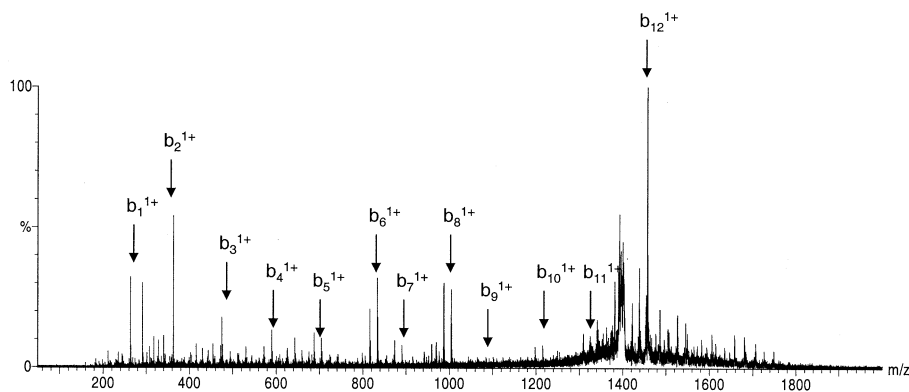
Figure 5. MS/MS spectrum resulting from CID experiments on an unresolved cone fragment of trypsin inhibitor at *m/z* 1393.

only one in this work to show a gain of 18 u instead of a loss.

After increasing the cone voltage (90 V), two sequence tags for BSA were found of 15 and 7 amino acid length from MS/MS experiments on two cone fragment ions. BSA, being a larger protein, did not fragment as readily at cone voltages used for the smaller proteins (60–65 V). Figure 6 shows the charge state envelope for BSA at three cone voltages 45, 60, and 90 V. The MS/MS spectrum for the cone fragment at 875.46 *m/z* having a 4+ charge state shows a dominant b-ion series when processed, and the sequence was found at the n-terminus (Figure 7). Additionally, another cone fragment (834.71 *m/z*, 4+) was also found to originate from the n-terminus of the protein, and both were found to have an 18 u difference in molecular weight from that predicted. This is most likely explained by the presence of the aspartic acid residue at the n-terminus of the protein; aspartic acid commonly gives a loss of 18 u [33]. Also, a likely explanation for one cone fragment (VLIAFSQ) is cyclization at the c-terminus of the fragment which is found to have glutamine [34]. Again, these results were found to be comparable to that

reported by Suckau and coworker work using T[3]-sequencing where they also reported a sequence tag from the n-terminus (residues 20–33 versus residues 16–30 for our work) [25]. Previous work by Stevenson et al. reported the existence of a 2+, 3+, and 4+ ion series originating from the n-terminus of ovine albumin when subjected to in-source CID [29]. Nemeth-Cawley et al. were able to find a six amino acid sequence tag for BSA by dissociating four charge states of BSA and combining data from MS/MS spectra [23].

Database Search Results

All identified sequence tags were subjected to database searching using the PeptideSearch program (http://www.narrador.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html), and the results are included in Table 1. In the case of cytochrome *c*, the identified four amino acid sequence tag from dissociating the 17+ charge state molecular ion was searched using PeptideSearch and found to give over 500 hits for various proteins. In contrast, searching with the 26 amino acid tag produced by pseudo-MS[3] returned only
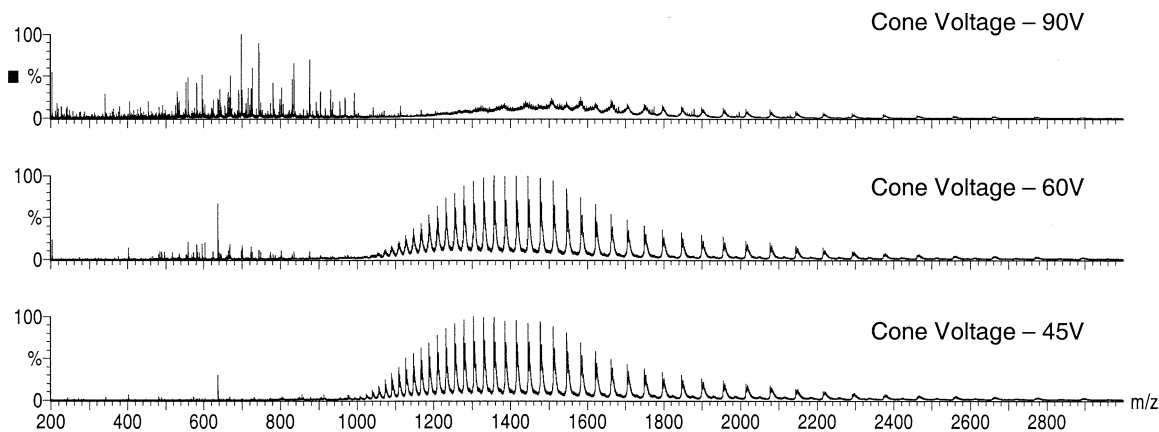

Figure 6. MS spectra of BSA taken at varying cone voltages. The increasing numbers of peaks seen in the lower *m/z* range are attributed to the fragmentation of BSA as the cone voltage is significantly increased (45 to 90 V in this instance) as well as the disappearance of the charge state envelope.
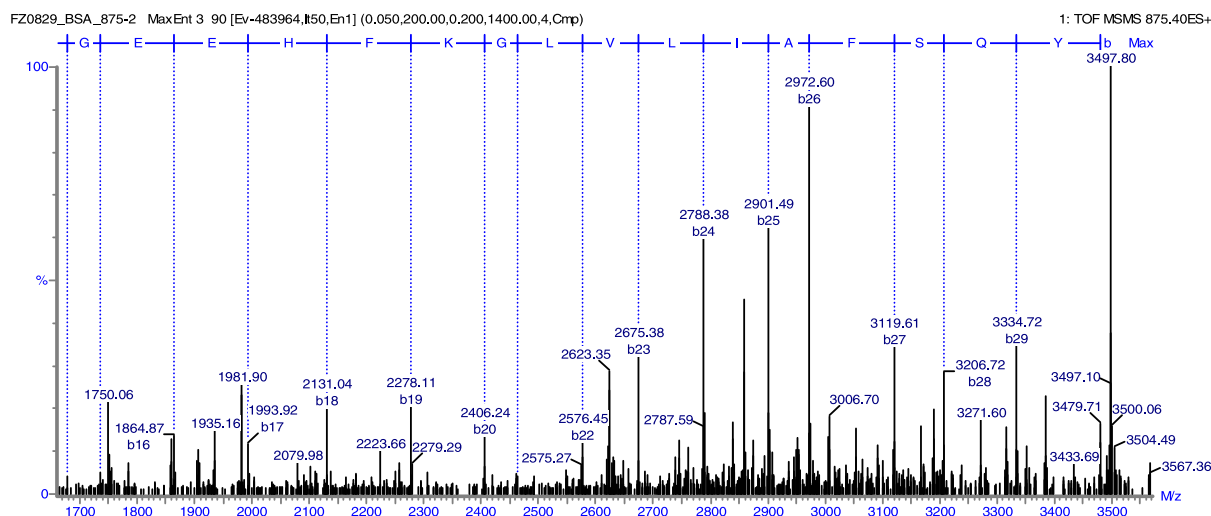
Figure 7. MassLynx's MassSeq output showing a b-ion series for the dissociation of a cone fragment of BSA at *m/z* 875.

9 hits, all matching to cytochrome *c*. While the 26 amino acid tag is clearly superior to the 4 amino acid tag, the results in Table 1 show no clear correlation between length of sequence tag and quality of PeptideSearch results, at least for the range of tag lengths studied. The search results for carbonic anhydrase were just as conclusive with an 8 amino acid tag as other proteins with tags greater than 20 amino acids.

Conclusions

Pseudo-MS³ is capable of generating sequence tags from the n- and c-termini of proteins ranging in molecular weight from 11 to 66 kDa. Individual tags are typically greater than 10 amino acids in length, which is generally sufficient for protein identification. In instances where the identification may be ambiguous, the intact molecular protein mass can aid in assignment.

Combining the measurement of an intact protein molecular mass with an extended sequence tag from pseudo-MS³ represents an approach to protein identification that may prove to be rapid, convenient, and accessible to many commercially available platforms. In principle, cone fragmentation provides a fast, simple, instrument-based alternative to conventional enzymatic or chemical digestion. The technique may be particularly well suited for probing modifications at or near termini, and much like the newly reported T³ sequencing technology by Suckau and Resemann, can also be used as a quick QC check for recombinant proteins [25]. Lastly, the pseudo-MS³ approach should be easily interfaced with many of the protein separations methods currently utilized in proteomics labs such as 2D-LC, affinity capture of intact proteins, and capillary isoelectric focusing-liquid chromatography (CIEF-HPLC).

Acknowledgments

References

1. Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. Identifying Proteins from Two-Dimensional Gels by Molecular Mass Searching of Peptide Fragments in Protein Sequence Databases. *Proc. Natl. Acad. Sci. U.S.A.* **1993,** *90,* 5011–5015.

2. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, III, J.R., John, R. Direct Analysis of Protein Complexes Using Mass Spectrometry. *Nat. Biotechnol.* **1999,** *17*(7), 676–682.

3. Yates, III, J.R., John, R.; Speicher, S.; Griffin, P. R.; Hunkapiller, T. Peptide Mass Maps: A Highly Informative Approach to Protein Identification. *Anal. Biochem.* **1993,** *214,* 397–408.

4. Mann, M.; Wilm, M. Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags. *Anal. Chem.* **1994,** *66*(24), 4390–4399.

5. Kelleher, N. L.; Lin, H. Y.; Valaskovic, G .A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.* **1999,** *121,* 806–812.

6. Moritz, E.; O'Connor, P. B.; Roepstorff, P.; Kelleher, N. L.; Wood, T. D.; McLafferty, F. W.; Mann, M. Sequence Tag Identification of Intact Proteins by Matching Tandem Mass Spectral Data Against Sequence Databases. *Proc. Natl. Acad. Sci. U.S.A.* **1996,** *93,* 8264–8267.

7. Loo, J. A.; Quinn, J. P.; Ryu, S. I.; Henry, K. D.; Senko, M. W.; McLafferty, F. W. High-Resolution Tandem Mass Spectrometry of Large Biomolecules. *Proc. Natl. Acad. Sci. U.S.A.* **1992,** *89*(1), 286–289.

8. Kelleher, N. L.; Costello, C. A.; Begley, T. P.; McLafferty, F. W. Thiaminase I (42 kDa) Heterogeneity, Sequence Refinement, and Active Site Location from High-Resolution Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1995,** *6,* 981–984.

9. Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. Plasma Electron Capture Dissociation for the Characterization of Large Proteins by Top Down Mass Spectrometry. *Anal. Chem.* **2003,** *75,* 1599–1603.

10. Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J.-H.; Begley, T. P.; McLafferty, F. W. Top Down Characterization of Larger Proteins (45 kDa) by Electron Capture Dissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2002,** *124*(4), 672–678.

11. Ge, Y.; ElNaggar, M.; Sze, S. K.; Oh, H. B.; Begley, T. P.; McLafferty, F. W.; Boshoff, H.; Barry, III. C. E. Top Down Characterization of Secreted Proteins from Mycobacterium tuberculosis by Electron Capture Dissociation Mass Spectrometry. J. Am. Soc. Mass Spectrom. **2003,** *14,* 253–261.

12. Kruppa, G. H.; Schoeniger, J.; Young, M. M. A Top Down Approach to Protein Structural Studies Using Chemical Cross-Linking and Fourier Transform Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003,** *17,* 155–162.

13. Engel, B. J.; Pan, P.; Reid, G. E.; Wells, J. M.; McLuckey, S. A. Charge State Dependent Fragmentation of Gaseous ProteinI in a Quadrupole Ion Trap: Bovine Ferri-, Ferro-, and Apo-Cytochrome c. *Int. J. Mass Spectrom.* **2002,** *219,* 171–187.

14. Stephenson, J.; James, L.; McLuckey, S. A. Simplification of Produce Ion Spectra Derived from Multiply Charged Parent Ions via Ion/Ion Chemistry. *Anal. Chem.* **1998,** *70,* 3533–3544.

15. Stephenson, J.; James, L.; McLuckey, S. A.; Reid, G. E.; Wells, J. M.; Bundy, J. L. Ion/Ion Chemistry as a Top-Down Approach for Protein Analysis. *Curr. Opin. Biotechnol.* **2002,** *13,* 57–64.

16. Schaaff, T. G.; Cargile, B. J.; Stephenson, J.; James, L.; McLuckey, S. A. Ion Trap Collisional Activation of the $(M + 2H)^{2+} - (M + 17H)^{17+}$ Ions of Human Hemoglobin β-Chain. *Anal. Chem.* **2000,** *72,* 899–907.

17. Hogan, J. M.; McLuckey, S. A. Charge State Dependent Collision-Induced Dissociation of Native and Reduced Porcine Elastase. *J. Mass Spectrom.* **2003,** *38,* 245–256.

18. Badman, E. R.; Myung, S.; Clemmer, D. E. Gas-Phase Separations of Protein and Peptide Ion Fragments Generated by Collision-Induced Dissociation in an Ion Trap. *Anal. Chem.* **2002,** *74,* 4889–4894.

19. Loo, J. A.; Edmonds, C. G.; Smith, R. D. Tandem Mass Spectrometry of Very Large Molecules. 2. Dissociation of Multiply Charged Proline-Containing Proteins from Electrospray Ionization. *Anal. Chem.* **1993,** *65,* 425–438.

20. Loo, J. A.; Edmonds, C. G.; Udseth, H. R.; Smith, R. D. Effect of Reducing Disulfide-Containing Proteins on Electrospray Ionization Mass Spectra. *Anal. Chem.* **1990,** *62,* 693–698.

21. Loo, J. A.; Edmonds, C. G.; Smith, R. D. Primary Sequence Information from Intact Proteins by Electrospray Ionization Tandem Mass Spectrometry. *Science* **1990,** *248*(4952), 201–204.

22. Smith, R. D.; Loo, J. A.; Barinaga, C. J.; Edmonds, C. G.; Udseth, H. R. Collisional Activation and Collision-Activated Dissociation of Large Multiply Charged Polypeptides and Proteins Produced by Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* **1990,** *1*(1), 53–65.

23. Nemeth-Cawley, J. F.; Rouse, J. C. Identification and Sequencing Analysis of Intact Proteins via Collision-Induced Dissociation and Quadrupole Time-of-Flight Mass Spectrometry. *J. Mass Spectrom.* **2002,** *37,* 270–282.

24. Lin, M.; Campbell, J. M.; Mueller, D. R.; Wirth, U. Intact Protein Analysis by Matrix-Assisted Laser Desorption/Ionization Tandem Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003,** *17,* 1809–1814.

25. Suckau, D.; Resemann, A. T³-Sequencing: Targeted Characterization of the N- and C-Termini of Undigested Proteins by Mass Spectrometry. *Anal. Chem.* **2003,** *75*(21), 5817–5824.

26. Chen, H.; Tabei, K.; Siegel, M. M. Biopolymer Sequencing Using a Triple Quadrupole Mass Spectrometer in the ESI Nozzle-Skimmer/Precursor Ion MS/MS Mode. *J. Am. Soc. Mass Spectrom.* **2001,** *12,* 846–852.

27. Raska, C. S.; Parker, C. E.; Huang, C.; Han, J.; Glish, G. L.; Pope, M.; Borchers, C. H. Pseudo-MS³ in a MALDI Orthogonal Quadrupole-Time of Flight Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2002,** *13,* 1034–1041.

28. Loo, J. A.; Udseth, H. R.; Smith, R. D. Charge Distribution of Ions from Large Molecules, Formed by Electrospray-Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1988,** *2*(10), 207–210.

29. Stevenson, T.I., G.W. Kilby, and J.A. Loo. Nanoelectrospray-QTOF Mass Spectrometry for Sensitive Sequencing and Determination of Phosphorylation Sites of Large Polypeptides. In *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*; Dallas, TX, June 1999.

30. Thevis, M.; Loo, R. R. O.; Loo, J. A. Mass Spectrometric Characterization of Transferrins and Their Fragments Derived by Reduction of Disulfide Bonds. *J. Am. Soc. Mass Spectrom.* **2003,** *14*(6), 635–647.

31. Loo, J.A. and R.D. Smith. Tandem Mass Spectrometry of Multiply Charged Proteins with Electrospray Ionization. In *Proceedings of the 40th ASMS Conference on Mass Spectrometry and Allied Topics*; Washington, DC, June 1992.

32. Stone, K. L.; Williams, K. R. Enzymatic Digestion of Proteins in Solution and in SDS Polyacrylamide Gels. In *The Protein Protocols Handbook*; Walker, J. M., Ed.; Humana Press Inc.: Totowa, 2000; pp 511–521.

33. Kinter, M.; Sherman, N. E. Protein Sequencing and Identification Using Tandem Mass Spectrometry. In *Wiley-Interscience Series on Mass Spectrometry*; Dominic, N.M.M.N.; Desiderio, M., Eds.; John Wiley and Sons, Inc.: New York, 2000; p 301.

34. Snyder, A. P. *Interpreting Protein Mass Spectra*; Oxford University Press, Inc.: New York, 2000; p 522.