

“De Novo” Peptide Sequencing by MALDI-Quadrupole-Ion Trap Mass Spectrometry: A Preliminary Study

Wenzhu Zhang, Andrew N. Krutchinsky, and Brian T. Chait

The Rockefeller University, New York, New York, USA

Collision-induced dissociation of singly charged peptide ions produced by resonant excitation in a matrix-assisted laser desorption/ionization (MALDI) ion trap mass spectrometer yields relatively low complexity MS/MS spectra that exhibit highly preferential fragmentation, typically occurring adjacent to aspartyl, glutamyl, and prolyl residues. Although these spectra have proven to be of considerable utility for database-driven protein identification, they have generally been considered to contain insufficient information to be useful for extensive de novo sequencing. Here, we report a procedure for de novo sequencing of peptides that uses MS/MS data generated by an in-house assembled MALDI-quadrupole-ion trap mass spectrometer (Krutchinsky, Kalkum, and Chait *Anal. Chem.* **2001**, *73*, 5066–5077). Peptide sequences of up to 14 amino acid residues in length have been deduced from digests of proteins separated by SDS-PAGE. Key to the success of the current procedure is an ability to obtain MS/MS spectra with high signal-to-noise ratios and to efficiently detect relatively low abundance fragment ions that result from the less favorable fragmentation pathways. The high signal-to-noise ratio yields sufficiently accurate mass differences to allow unambiguous amino acid sequence assignments (with a few exceptions), and the efficient detection of low abundance fragment ions allows continuous reads through moderately long stretches of sequence. Finally, we show how the aforementioned preferential cleavage property of singly charged ions can be used to facilitate the de novo sequencing process. (*J Am Soc Mass Spectrom* 2003, *14*, 1012–1021) © 2003 American Society for Mass Spectrometry

The use of mass spectrometry combined with database searching has gained wide acceptance for rapid, sensitive protein identification [1]. This method requires the availability of extensive protein, cDNA, or genomic sequence from the organism of interest (or at least from a closely related species). If such sequence information is not available, as remains the case for the vast majority of extant organisms, it is desirable to utilize de novo peptide sequencing to identify and obtain information about their protein(s). Although Edman sequencing has for a long time been a method of choice for de novo sequencing of peptides and proteins, mass spectrometry (MS) has also been shown to have considerable utility for this purpose.

Indeed, mass spectrometric de novo sequencing based on the fragmentation of derivitized peptide ions dates back to the late 1960s and on underivatized peptide ions to the early 1980s (reviewed in [2] and [3]). More recently, MS sequencing has been given consid-

erable impetus by the development of ESI and MALDI, ionization techniques whose efficiencies for producing intact peptide ions are far higher than those that were previously available. Again, both derivitized and underivatized peptides have been used in this newer work. Thus, chemical derivitization has been applied to simplify/enhance the fragmentation spectra and/or to differentiate N- from C-terminal fragment ions [4–13]. Notable in this regard is the use of ^{18}O incorporation at the carboxyl-termini of peptides during protein hydrolysis [14–17]. For underivatized peptides, ESI has found use for de novo sequencing under a variety of conditions, which include collision-induced dissociation (CID) in the triple quadrupole mass analyzer [18–20] and the QqTOF mass analyzer [18, 19, 21–26], resonant excitation in the quadrupole ion trap mass analyzer [14, 27–29], and electron capture dissociation in the Fourier transform ion cyclotron resonance mass analyzer [30–34]. MALDI de novo sequencing has been carried out using CID and metastable decay in single- [35] and double-stage [36–39] TOF instruments and with CID in QqTOF instruments [40–44]. A number of computer algorithms have also been devised that seek to infer sequence de novo from MS/MS data [45–50].

Published online July 24, 2003

Address reprint requests to Dr. B. T. Chait, The Rockefeller University, 1230 York Avenue, Box 170, New York, NY 10021, USA. E-mail: chait@rockefeller.edu

In mass spectrometric de novo sequencing, the candidate peptide sequence is derived directly from its MS/MS fragmentation spectrum. The accuracy of the method depends on the accuracy with which the mass differences between the relevant ion peaks are determined, the completeness of the observed fragment ion series, and the extent to which the fragmentation spectrum can be correctly interpreted. The completeness of the observed fragmentation series depends on the particular sequence of the peptide under study, the charge state of the ion produced, and the mode by which the fragmentation is induced within the mass spectrometer. By and large, doubly charged peptide ions tend to fragment more evenly across a given sequence than do singly charged ion species ([40, 44, 51] and extensive personal observations). Thus, a large proportion of de novo sequencing has been performed on doubly charged ions, which are readily produced from tryptic peptides by electrospray ionization (ESI). By contrast, MALDI produces predominantly singly charged ions from peptides. These singly charged ion species often yield relatively low complexity CID mass spectra that tend to be dominated by a few preferred fragmentation pathways [40, 52–56]. When this fragmentation is induced on the slow time scale of resonant excitation within an ion trap mass spectrometer, the dissociation of singly-charged ions becomes especially selective, occurring largely adjacent to any aspartyl, glutamyl, or prolyl residues that may be present in the peptide [54, 56]. Although the resulting low complexity spectra have proven to be of considerable utility for database-driven protein identification [57–60], we have generally considered these spectra to contain insufficient information to be useful for extensive de novo sequence determination.

MALDI ionization has several attractive properties for the analysis of peptides. It has a relatively high tolerance to impurities and common biochemical additives and salts, reducing the demands on sample preparation compared to that required for ESI. In addition, the MALDI sample is reusable until it is depleted. Thus, provided that the instrument used for mass analysis has sufficient sensitivity, data can be collected from many different ions species produced from a single peptide mixture and this data can be collected until the desired quality of spectra have been obtained. Towards these goals, we recently assembled a MALDI-ion trap mass spectrometer by adding a MALDI source and a quadrupole interface to a commercial ion trap mass analyzer (Finnigan Thermoquest LCQ DECA XP, San Jose, CA) [59, 61]. This configuration has proved sufficiently efficient to allow us to detect ions resulting from relatively low probability cleavage pathways and to observe more complete sets of fragment ion series than we previously thought possible. Using this instrument, we demonstrate that de novo sequencing is feasible using resonant excitation of singly charged ions in an ion trap and that the rules governing preferential cleav-

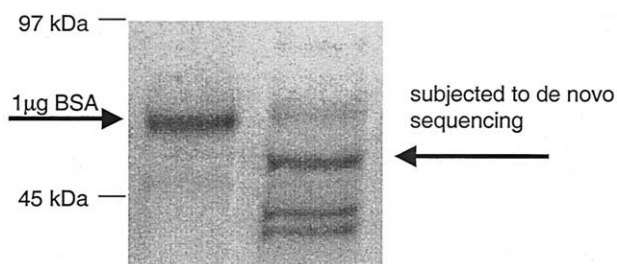


Figure 1. Portion of SDS-PAGE gel of proteins isolated from pine tree (*Xylem pinus*) visualized with Coomassie blue stain (right lane). The 55 kDa protein band indicated by the arrow was subjected to in-gel digestion with trypsin, and several of the resulting peptides (see Figure 2) were subjected to de novo sequencing. One microgram of BSA was loaded in the left lane as a standard.

age can actually be helpful in the de novo sequencing process.

Experimental

Sample preparation

The resolved protein band at ~55 kDa on the Coomassie stained SDS-PAGE of a sample derived from pine tree (*Xylem pinus*) is estimated to contain ~0.5 µg (~10 pmol) of protein based on visual comparison to the 1 µg BSA band (Figure 1), while the 60 kDa band from cultured human HeLa cells is estimated to contain 50–100 ng (1–2 pmol) (Figure 5). These two bands were excised and subjected to an in-gel digestion procedure described elsewhere using between 0.1–0.2 µg of trypsin on each band [62]. The resulting peptide mixture was extracted from the gel band as described [62] using Poros 50 reversed phase beads (Perseptive Biosystems, MA) and eluted with matrix solution (×2 dilution of a saturated solution of 2,5-dihydroxybenzoic acid in 60% Methanol/5%acetic acid) onto a compact disc MALDI sample probe [59, 61]. This compact disc MALDI sample probe was first loaded into the MALDI-QqTOF mass spectrometer for peptide mapping and subsequently into the MALDI-quadrupole ion trap mass spectrometer for MS/MS analysis. This ability to examine the same sample in both mass spectrometers allows us to take advantage of the particular strengths of each instrument—i.e., the high mass determination accuracy of the MALDI-QqTOF mass spectrometer and the high efficiency for obtaining product ion mass spectra of the MALDI-quadrupole ion trap mass spectrometer. Thus, we use the first to obtain peptide maps with high mass accuracy, and subsequently obtain product ion spectra on any or all of the individual peptide ions observed in the peptide map.

MALDI QqTOF Mass Spectrometer

Peptide map spectra were collected on an in-house modified MALDI-QqTOF (Centaur, Sciex, Concord, ON) instrument [62]. The nitrogen laser (VSL-337, Laser

Science Inc., Spectra-Physics, Franklin, MA) was operated at 30 Hz. This instrument was modified to accept a MALDI compact disc (CD) target [59, 61] so that the CD target is readily exchangeable between the MALDI-QqTOF and MALDI-quadrupole ion trap instruments.

MALDI-Quadrupole Ion Trap Mass Spectrometer

Collision-induced fragmentation spectra were collected on a MALDI-quadrupole ion trap mass spectrometer, which we recently assembled by adding an in-house built MALDI source and a quadrupole interface to a LCQ DECA XP (Finnigan Thermoquest) ion trap mass spectrometer [59, 61]. The nitrogen laser (VSL-337, Laser Science Inc, MA) was operated at 20 Hz.

Product Ion Experiments

All product ion spectra were obtained using the following settings of the LCQ DECA XP: automatic gain control (AGC) off, ion injection time 500 ms, m/z window 4 ("isolation width"), excitation energy 30% ("normalized collision energy"), q of excitation 0.25 ("activation q ") and excitation time 300 ms ("activation time"). The ion count displayed on the "Tune Plus" window was in the range 8×10^5 – 3×10^6 . The number of microscans for the collection of product ion spectra ranged from 60–150 (collection times between 1–2.5 min). The averaged spectra were displayed in the "Tune Plus" window during data acquisition to allow us to assess the quality of the spectra during collection and to judge when to stop acquisition.

Data Analysis

De novo sequencing was performed manually. After obtaining the product ion spectrum, we first inspected the low mass region to visually identify peaks with good signal-to-noise ratio. Assuming that each peak corresponds to a y -ion (or b -ion) fragment, we look for its complementary b -ion (or y -ion) by subtracting the m/z value of the fragment ion from the m/z value of the singly protonated parent ion and adding a hydrogen mass. If the complementary ion exists in the spectrum, we mark the pair on the spectrum. Then, we inspect the whole spectrum to identify other major peaks and look for their complementary fragment ions and mark the pairs. Thus the spectrum in the region from the low mass cut-off to the parent mass minus the low-cut mass cutoff is greatly simplified. De novo sequencing is performed starting with these marked complementary pairs.

To assess the accuracy of our de novo sequencing results, the product ion spectra were converted to the "dta format" and submitted to the Mascot search engine (Matrix Science Ltd., London, UK) to search the NCBI EST database. For sequences that were not identified using the search engine (because of unanticipated cleav-

ages), we searched the databases directly with the de novo-determined sequences.

Results and Discussion

We began the present de novo sequencing investigation during a study of proteins from the pine tree (*Xylem pinus*) at a time when relatively few full-length pine protein sequences were available and when we were unable to gain access to the pine EST database. Subsequent access to this database allowed us to objectively assess the accuracy of our de novo sequencing efforts using the same tandem mass spectrometric data and conventional database searches.

The example, which we will discuss in detail, involves sequence from the 55 kDa protein band shown in Figure 1, which we estimate to contain $\sim 0.5 \mu\text{g}$ (10 pmol) of protein. Figure 2 shows the MALDI-QqTOF peptide map obtained following in-gel tryptic digestion of the 55 kDa band. Because our in-house assembled MALDI-quadrupole-ion trap mass spectrometer is sensitive and efficient, we speculated that it might be feasible to detect the less abundant fragment ions resulting from the less favorable peptide bond cleavages. We therefore attempted de novo peptide sequencing by collecting product ion spectra with high signal-to-noise for the seven strongest peaks in the spectrum. We used spectral accumulation times that were 10–25 times longer than those that we normally use for protein identification by regular database searching. Thus we used collection times of 1–2.5 min rather than the 5–7 s used for database searching.

Figure 3 shows a product ion spectrum of the parent ion with m/z 1705.882 (accumulation time 60 s). The large amount of sample (~ 10 pmol) and the longer than normal spectral accumulation time yielded a spectrum with high signal-to-noise ratio, enabling detection of fragment ions originating from cleavage at every peptide bond. To simplify spectral interpretation, we identified complementary b - and y -ions in the low mass region of the spectrum using the procedure described under the heading Data Analysis. We then started to build sequence using these complementary ions and extended the sequence into the high mass region where complementary ion information is unavailable. In this way, two long b - and y -ion series were identified. The sequences derived from the b - and y -ion series overlap over a stretch of four amino acid residues. In this way, the peptide was manually sequenced to be VVDEE[I/L]FD[I/L][I/L]E[Q/K]EK, where [I/L] denotes either I or L and [Q/K] either Q or K. Note that the high signal-to-noise has made relatively weak peaks in Figure 3 clearly visible. These weak peaks correspond to y or b fragment ions without which de novo sequencing would be impossible. The high signal-to-noise also allowed us to determine the mass difference between adjacent pairs of fragment ion peaks with accuracy generally better than 0.1 Da (Figure 3), further facilitating the de novo sequencing process. Finally, we found

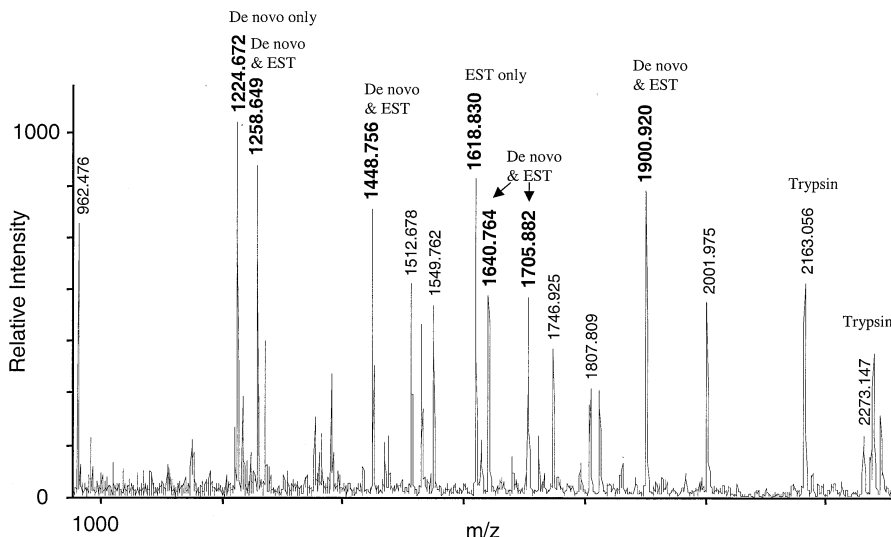


Figure 2. MALDI-QqTOF tryptic peptide mass map of the 55 kDa pine protein band shown in Figure 1. The seven peaks labeled with bold face were selected for MALDI-quadrupole ion trap tandem mass spectrometry. The label “De novo only” indicates that the labeled peptide was identified by de novo sequencing, but not by the EST database search. The label “EST only” indicates that the labeled peptide was identified by the EST database search, but not by de novo sequencing. The label “De novo and EST” indicates that the labeled peptide was identified by both de novo sequencing and EST database searching.

that previously described rules governing preferential fragmentation [54, 56] can be useful in assessing the plausibility of the amino acid assignment and the assignment of ion series to which an intense ion peak belongs. For example, the strong peak at 759.59 Da is consistent with the assignment that it is a y-ion frag-

ment resulted from a preferential cleavage between the peptide bond D[I/L]. The strong peak at 1559.85 Da is a b-ion fragment from a preferential cleavage between the peptide bond EK. Thus, the preferential cleavage property of MALDI ion in the ion trap can be useful in distinguishing y-ion and b-ion series. After obtaining

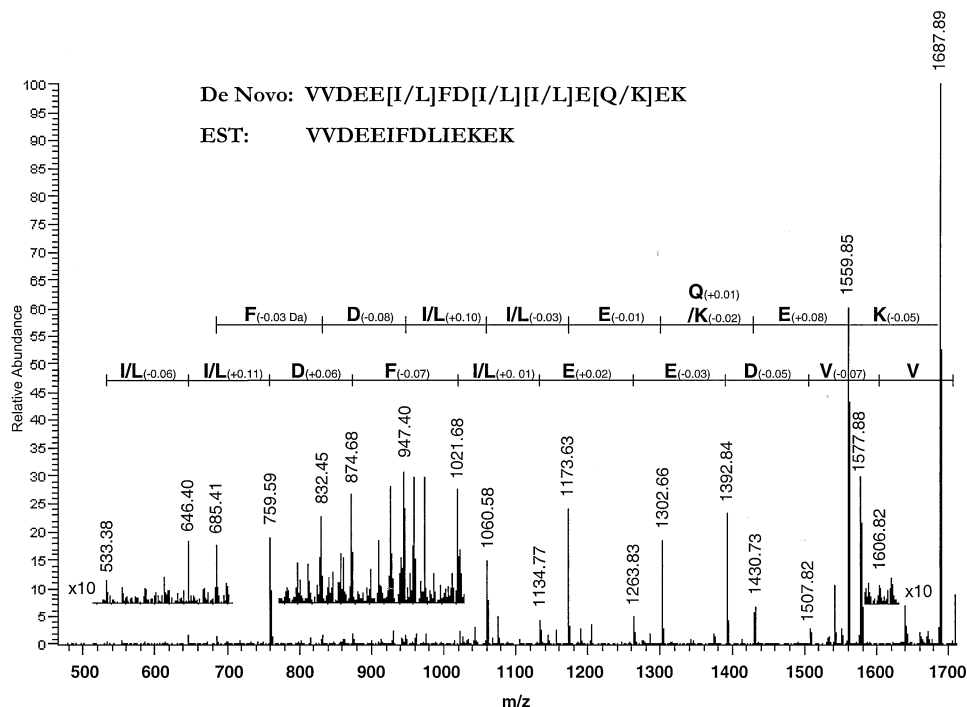


Figure 3. MALDI-quadrupole ion trap product ion mass spectrum of the m/z 1705.882 ion from the 55 kDa pine protein band digest (Figure 2). The numbers within the brackets provide differences between the experimental and theoretical masses of the assigned amino acids.

Table 1. Peptides identified by de novo sequencing and EST database searching

Peptide	<i>m/z</i> (Da)	De novo sequencing	EST database
1	1224.672	PVNVWGNTPLK	PVNVWGNTPLK ^a
2	1258.649	[(AL)/(SP)][Q/K]AFH[I/L]DSEK	ALQAFHLDSEK
3	1448.756	[(VV)/(PT)]DEE[I/L]FD[I/L][I/L]EK	VVDEEIFDLIEK
4	1618.830	No available	ISATSIYFESLPYK
5	1640.777	[Q/K]G[Q/K]PE[Q/K]LYDYEDR	KGQPEGALYDYEDR
6	1705.882	VVDEE[I/L]FD[I/L][I/L]E[Q/K]EK	VVDEEIFDLIEKEK
7	1900.920	[Q/K]WETGF[I/L]DYD[Q/K][I/L]EEK	VSQETGFIDYDKLEEK

^aThe sequence of this non-tryptic peptide can be mapped to an existing EST, but was not identified by EST database searching (see text).

the de novo sequence, the EST database became available. We thus searched this database with the same tandem mass spectrometric data, using the Mascot search engine. The EST database search yielded the sequence VVDEEIFDLIEKEK, which with the exception of certain ambiguities (discussed below), matches and confirms our de novo sequencing result.

Table 1 summarizes our de novo sequencing results on the peptides indicated in Figure 2 together with the subsequently obtained EST database search results. Out of the seven product ion spectra investigated, de novo sequencing yielded entire or nearly entire sequences for the six peptides (1, 2, 3, 5, 6, 7 in Table 1) and did not yield sequence for one of the peptides (4 in Table 1). EST database searching also identified six peptides, but missed peptide 1 (Table 1). Note that a different peptide was missed by each technique and that the de novo sequencing and EST database search results overlap for five of the seven peptides. The de novo-deduced sequences are consistent with those identified in the EST database, with the exception of certain ambiguities and two incorrect assignments. In addition to the well known isobaric I/L ambiguity, the mass accuracy of the ion trap mass spectrometer is not sufficiently high to distinguish between K and Q and between the N-terminal amino acid pairs (AL)/(SP) for the peptide at *m/z* 1258.649 and (VV)/(PT) for the peptide at *m/z* 1448.756. Ambiguities among these pairs of amino acids are caused by the absence of an observable y_{n-1} ion fragment peak corresponding to loss of the N-terminal amino acid residue. Certain of these non-isobaric ambiguities can be resolved by determining the peptide masses with high accuracy. Thus, we were able to resolve the ambiguity between Q and K at residue 12 of peptide 6 through examination of the mass of the peptide determined in the MALDI-QqTOF measurement (Figure 2). For example, the *m/z* of the 1705.882 peptide was determined with an accuracy of ± 5 ppm. If one assumes that residue 12 of this peptide is Q, the theoretical mass differs from the experimental mass by 17 ppm, whereas the difference is only 4 ppm if we assume that residue 12 is K. Resolution of this ambiguity allows us to deduce the sequence VVDEE[I/L]FD[I/L][I/L]EKEK, again in agreement with that deduced from the EST search. In a similar manner, we were able to resolve the ambiguity between the pair VV and PT in peptide 3 because the sequence with VV agrees with the

experimental mass to within 3 ppm whereas the sequence with PT differs by 24 ppm. However, in those cases where we find more than one non-isobaric ambiguity (i.e., peptides 5 and 7), the accurate mass may not help resolve the ambiguities.

One apparent error in our de novo sequence for the peptide with *m/z* 1640.777 (peptide 5) was the assignment of the single amino acid residue Q (or K) in place of the two amino acid residues, GA, deduced from the EST database. We can trace the origin of this error to the fact that we missed the small peak that corresponds to cleavage between the G and A peptide bond, a reaction pathway that we have previously reported to occur with relatively low probability [63].

In some cases, knowledge of the fragmentation systematics of singly charged ions can be used to assist in the resolution of ambiguities. For example, in extensive unreported work, we have observed that a proline residue in the second position from the amino terminus invariably results in an intense y_{n-2} fragment. Note that this preferred fragmentation pathway on the C-terminal side of a P residue is specific to this second position (P residues elsewhere result in preferred fragmentation on the N-terminal side of P [53, 56]). In this case, failure to observe an intense y_{n-2} fragment indicates that the terminal residues are likely AL or LA rather than SP (we can rule out PS because trypsin only cleaves [K/R]P sequences with very low efficiency.) Once this ambiguity was partially resolved, it proved possible to resolve the remaining Q/K ambiguity at the 3rd position since Q yields a peptide mass that agrees to within 5 ppm and K one that differs by 24 ppm.

Because of the frequent absence of the y_{n-1} ion in the spectrum (Table 1) and our inability to detect the low mass b_1 residue, determination of the N-terminal residues can present a special challenge. This is illustrated by our mis-assignment of the N-terminal residues of peptide 7, where we assigned [Q/K]W in place of VSQ (as in the EST-derived sequence). We can trace this error to the fact that while no y_{n-1} fragment ion was in fact observed (corresponding to the loss of the N-terminal V residue), an intense peak was seen that was incorrectly interpreted to arise from the loss of Q/K at the N-terminal. In retrospect, we can see that the ion in question arose from elimination of a K residue at the C-terminus. Indeed, we frequently observe an intense

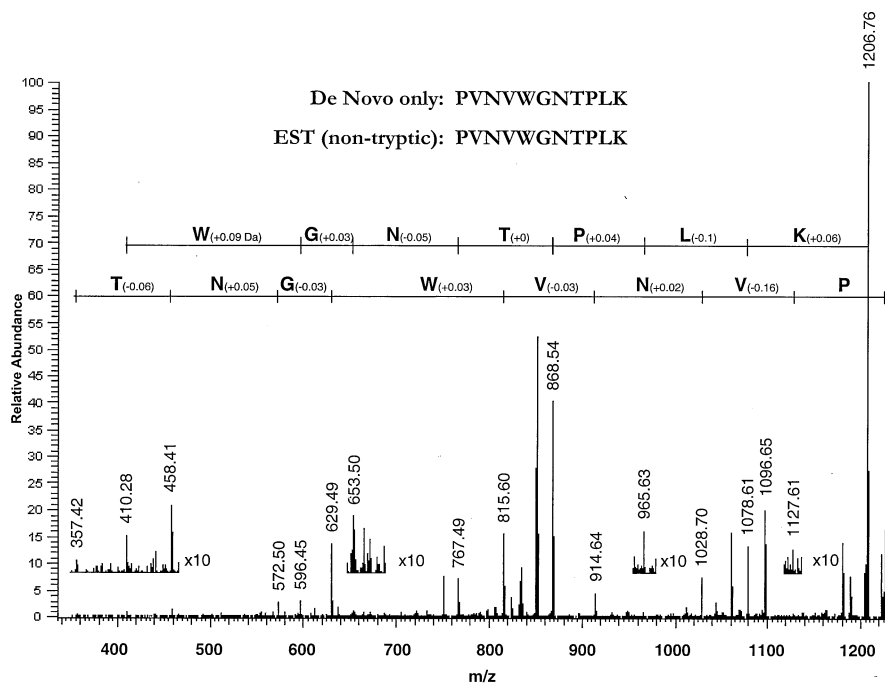


Figure 4. MALDI-quadrupole ion trap product ion mass spectrum of the m/z 1224.672 ion from the 55 kDa pine protein band digest (Figure 2).

loss of 128 Da (or 156 Da), which is formally equivalent to hydrolysis of the terminal K residue (or R residue).

The case of peptide 4 (Table 1) illustrates the kind of complication that can hinder de novo sequencing. Although we were able to correctly deduce the partial sequence YFESL (or LSEFY), we were unable to extend this sequence any further because the next critical y -ion fragment (y_9) between I and S at m/z 1159.60 was hidden in the middle of the isotopic pattern of a strong ion fragment at m/z 1158.61, which appears to arise by the loss of three water molecules from the b_{11} ion fragment (m/z 1212.67) occurring between L and P (data not shown).

Full-length de novo sequencing requires the observation of fragment ions from every peptide bond along the peptide backbone. Despite the presence of factors that cause ambiguities, we have found that product ion spectra from the MALDI-quadrupole ion trap can often be used for de novo sequencing. In addition to obtained five sequences that partially match those obtained from EST database searches, this utility was convincingly demonstrated for peptide 1 (m/z 1224.672). This is a case in which de novo sequencing succeeded where the EST database search failed. As shown in Figure 4, the peptide was sequenced to be PVNVWGNTPLK. (The unassigned major peaks are due to losses of ammonia/water and the C-terminal elimination of the lysine residue.) After we deduced this sequence, we used it to search the contiguous EST sequence that we assembled from our six MS/MS hits in the EST database. We found that the peptide does indeed exist within the assembled sequence, but that it is a non-tryptic peptide, which explains why the search engine could not find it. Since

the EST sequence is . . .DPNVWGNTPLK, we think it likely that this non-tryptic peptide arose by cleavage of the acid labile DP bond during extraction by acidic solution (4% formic acid/0.06% TFA) of tryptic fragments from the gel slice.

A practical issue in data acquisition is to collect as many useful spectra for de novo sequencing as possible from a single protein digest sample. Not every peak in the spectrum of the resulting peptide mixture will result in a useful product ion spectra because certain peptides do not fragment in a manner that yields the needed information. For these, it is preferable to discontinue data collection so as not to waste sample. Conversely for peptides that do produce informative product ion spectra, we wish to spend an appropriate length of time to collect sufficiently high signal-to-noise ratio data for the small peaks. To determine when to continue collecting data and when to stop, we display the averaged spectrum during data acquisition and amplify the spectrum to inspect the small peaks. If the peaks are sparsely distributed, we stop the acquisition. If the peaks are distributed across the whole mass region, we examine the weak peaks (normally in the low mass region) and stop acquisition when a useable signal-to-noise ratio has been achieved.

We have stated that more material is needed for successful de novo sequencing in the MALDI-quadrupole ion trap compared to that required for protein identification using product ion spectra database searching. To get a better feeling for the minimum quantity of material required for de novo sequencing, we show a second example involving a 60 kDa human protein where we estimate the amount of protein

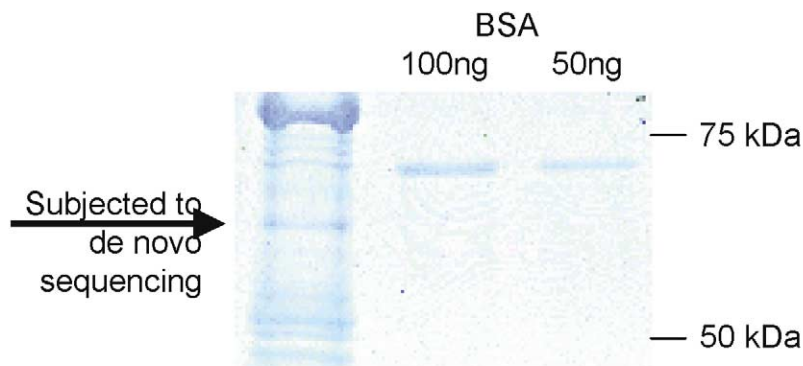


Figure 5. Portion of SDS-PAGE gel of proteins isolated from cultured human cells visualized with Coomassie blue stain (left lane). The 60 kDa protein band indicated by the arrow was subjected to in-gel digestion with trypsin, and one of the resulting peptides was subjected to de novo sequencing (Figure 6). BSA (100 and 50 ng) was loaded onto the two right-hand lanes as standards.

present in the band of interest to be 50–100 ng (i.e., 1–2 pmol) (Figure 5). Conventional database searches using tandem mass spectrometric data from several tryptic peptides identify the protein MDM2 to be present in the band. However, one of the peptides from the band digest (having m/z 1859.848 as measured by QqTOF-MS) did not map to any known tryptic peptide of MDM2. Thus, we performed de novo sequencing on this peptide. The procedure yielded the partial sequence $^{352}\text{QAEEGFDPD}^{361}$ (Figure 6), corresponding to the C-terminal portion of a peptide from MDM2 ($^{345}\text{AKLENSTQAEEGFDPD}^{361}$). This peptide appears to originate from tryptic cleavage between K^{344} and A^{345} and an unexpected cleavage between D^{361} and C^{362} . However, the theoretical m/z of this peptide is 1849.846, which is 10 Da lower than the measured m/z

(1859.848 as measured by QqTOF-MS). Assuming our de novo C-terminal sequence to be correct, this discrepancy should occur between residues 345–351. The simplest explanation for this discrepancy is a single base mutation that yields a P rather than an S residue at position 350. This mutation yields the peptide $^{345}\text{AKLENPTQAEEGFDPD}^{361}$ with theoretical m/z 1859.858, which is within 10 ppm of the measured value. Inspection of the tandem mass spectrum shown in Figure 6 appears to confirm this hypothesis since the expected preferential cleavage on the N-terminal side of P^{350} is observed, as is the partial sequence LEN. In this case, we were able to use de novo sequencing to obtain partial sequence of a mutant form of a peptide from MDM2 that was cleaved in an unexpected position, and further to use the MS/MS data to elucidate the nature of the mutation.

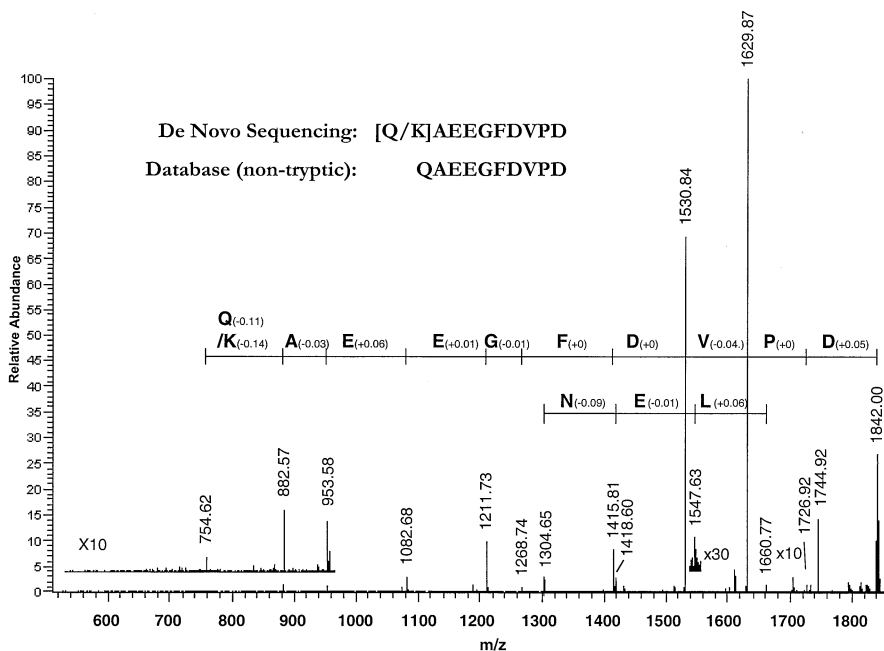


Figure 6. MALDI-quadrupole ion trap product ion mass spectrum of the m/z 1859.848 ion from the 60 kDa human protein band digest.

From the data provided above, we also conclude that we currently need at least one order of magnitude more material in the band of interest for de novo sequencing than we normally require for conventional database-driven protein identification.

Conclusion

We have demonstrated that tandem mass spectrometric data generated in our in-house assembled MALDI-quadrupole ion trap mass spectrometer can be used for de novo sequencing of peptides from in-gel digested proteins (although ambiguities and mis-assignments are to be expected, especially for residues at the N-terminal.) We have found that despite a proclivity to fragment preferentially at certain amino acid residues, singly charged MALDI ions within the ion trap can fragment at nearly every peptide bond along the peptide backbone. Given sufficient protein (≥ 1 pmol in the gel), the highly sensitive MALDI-quadrupole ion trap mass spectrometer is capable of detecting ions that result from the less favorable fragmentation channels with usable signal-to-noise ratios. The resulting tandem mass spectra tend to be sufficiently simple to allow straightforward de novo sequencing. Moreover, we have found that the preferential cleavage of singly charged ions can actually facilitate the de novo sequencing process.

Acknowledgments

This work was supported by a grant from the National Institutes of Health (RR00862). The authors thank Drs. David Calhoun and Ming Jin for the SDS-PAGE separated *Xylem pinus* proteins and Drs. Wei Gu and Chris Brooks for the SDS-PAGE separated human proteins, which were identified during our collaboration.

References

- Mann, M.; Hendrickson, R. C.; Pandey, A. Analysis of Proteins and Proteomes by Mass Spectrometry. *Annu. Rev. Biochem.* **2001**, *70*, 437–473.
- Biemann, K.; Martin, S. A. Mass Spectrometric Determination of the Amino Acid Sequence of Peptides and Proteins. *Mass Spectrom. Rev.* **1987**, *6*, 1–76.
- Papayannopoulos, I. A. The Interpretation of Collision-Induced Dissociation Tandem Mass Spectra of Peptides. *Mass Spectrom. Rev.* **1995**, *14*, 49–73.
- Keough, T.; Lacey, M. P.; Youngquist, R. S. Solid-Phase Derivatization of Tryptic Peptides for Rapid Protein Identification by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1003–1015.
- Keough, T.; Lacey, M. P.; Strife, R. J. Atmospheric Pressure Matrix-Assisted Laser Desorption/Ionization Ion Trap Mass Spectrometry of Sulfonic Acid Derivatized Tryptic Peptides. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2227–2239.
- Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate de Novo Sequencing of Lysine-Terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2348–2356.
- Keough, T.; Lacey, M. P.; Fieno, A. M.; Grant, R. A.; Sun, Y.; Bauer, M. D.; Begley, K. B. Tandem mass spectrometry methods for definitive protein identification in proteomics research. *Electrophoresis* **2000**, *21*, 2252–2265.
- Bauer, M. D.; Sun, Y.; Keough, T.; Lacey, M. P. Sequencing of Sulfonic Acid Derivatized Peptides by Electrospray Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 924–929.
- Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 7131–7136.
- Shen, T. L.; Huang, Z. H.; Laivenieks, M.; Zeikus, J. G.; Gage, D. A.; Allison, J. Evaluation of Charge Derivatization of a Proteolytic Protein Digest for Improved Mass Spectrometric Analysis: De Novo Sequencing by Matrix-Assisted Laser Desorption/Ionization Post-Source Decay Mass Spectrometry. *J. Mass Spectrom.* **1999**, *34*, 1154–1165.
- Lindh, I.; Hjelmqvist, L.; Bergman, T.; Sjovall, J.; Griffiths, W. J. De novo Sequencing of Proteolytic Peptides by a Combination of C-terminal Derivatization and Nano-Electrospray/Collision-Induced Dissociation Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 673–686.
- Muenchbach, M.; Quadroni, M.; Miotto, G.; James, P. Quantitation and Facilitated de Novo Sequencing of Proteins by Isotopic N-Terminal Labeling of Peptides with a Fragmentation-Directing Moiety. *Anal. Chem.* **2000**, *72*, 4047–4057.
- Cagney, G.; Emili, A. De Novo Peptide Sequencing and Quantitative Profiling of Complex Protein Mixtures Using Mass-Coded Abundance Tagging. *Nat. Biotechnol.* **2002**, *20*, 163–170.
- Qin, J.; Herring, C. J.; Zhang, X. De Novo Peptide Sequencing in an Ion Trap Mass Spectrometer with 18O Labeling. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 209–216.
- Wilm, M.; Neubauer, G.; Taylor, L.; Shevchenko, A.; Bachi, A. De Novo Sequencing of Proteins with Mass Spectrometry Using the Differential Scanning Technique. *Proteome Prot. Anal.* **2000**, 65–79.
- Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. Rapid "De Novo" Peptide Sequencing by a Combination of Nano-electrospray, Isotopic Labeling, and a Quadrupole/Time-of-Flight Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1015–1024.
- Uttenweiler-Joseph, S.; Neubauer, G.; Christoforidis, S.; Zerial, M.; Wilm, M. Automated De Novo Sequencing of Proteins Using the Differential Scanning Technique. *Proteomics* **2001**, *1*, 668–682.
- Shevchenko, A.; Chernushevich, I.; Shevchenko, A.; Wilm, M.; Mann, M. "De Novo" Sequencing of Peptides Recovered from In-Gel Digested Proteins by Nano-electrospray Tandem Mass Spectrometry. *Mol. Biotechnol.* **2002**, *20*, 107–118.
- Shevchenko, A.; Chernushevich, I.; Wilm, M.; Mann, M. De Novo Peptide Sequencing by Nano-electrospray Tandem Mass Spectrometry Using Triple Quadrupole and Quadrupole/Time-of-Flight Instruments. *Methods Mol. Biol.* **2000**, *146*, 1–16.
- Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **1996**, *379*, 466–469.
- Morris, H. R.; Paxton, T.; Panico, M.; McDowell, R.; Dell, A. A Novel Geometry Mass Spectrometer, the Q-TOF, for Low-Femtomole/Attomole-Range Biopolymer Sequencing. *J. Protein Chem.* **1997**, *16*, 469–479.
- Bateman, R. H.; Blackstock, W.; Bordoli, R. S.; Gilbert, A. J.; Hoyes, J. B.; Langridge, J.; Ward, M. Protein Characterization Using 2-D Gel Electrophoresis with Nanospray MS and MS/MS on a Q-TOF. *Adv. Mass Spectrom.* **1998**, *14*.

23. van Der Wel, H.; Morris, H. R.; Panico, M.; Paxton, T.; North, S. J.; Dell, A.; Thomson, J. M.; West, C. M. A Non-Golgi α 1,2-Fucosyltransferase that Modifies Skp1 in the Cytoplasm of *Dictyostelium*. *J. Biol. Chem.* **2001**, *276*, 33952–33963.
24. Romaris, F.; North, S. J.; Gagliardo, L. F.; Butcher, B. A.; Ghosh, K.; Beitin, D. P.; Panico, M.; Arasu, P.; Dell, A.; Morris, H. R.; Appleton, J. A. A Putative Serine Protease Among the Excretory-Secretory Glycoproteins of *L1 Trichinella spiralis*. *Mol. Biochem. Parasitol.* **2002**, *122*, 149–160.
25. Vandenberghe, I.; Kim, J.-K.; Devreese, B.; Hacısalihoglu, A.; Iwabuki, H.; Okajima, T.; Kuroda, S.; Adachi, O.; Jongejan, J. A.; Duine, J. A.; Tanizawa, K.; Van Beeumen, J. The Covalent Structure of the Small Subunit from *Pseudomonas putida* Amine Dehydrogenase Reveals the Presence of Three Novel Types of Internal Cross-Linkages, All Involving Cysteine in a Thioether Bond. *J. Biol. Chem.* **2001**, *276*, 42923–42931.
26. Ou, K.; Seow, T. K.; Liang, R. C.; Ong, S. E.; Chung, M. C. Proteome Analysis of a Human Hepatocellular Carcinoma Cell Line, HCC-M: An Update. *Electrophoresis* **2001**, *22*, 2804–2811.
27. Arnott, D.; Henzel, W. J.; Stults, J. T. Rapid Identification of Comigrating Gel-Isolated Proteins by Ion Trap-Mass Spectrometry. *Electrophoresis* **1998**, *19*, 968–980.
28. Zhang, Z.; McElvain, J. S. De Novo Peptide Sequencing by Two-Dimensional Fragment Correlation Mass Spectrometry. *Anal. Chem.* **2000**, *72*, 2337–2350.
29. Sonsmann, G.; Romer, A.; Schomburg, D. Investigation of the Influence of Charge Derivatization on the Fragmentation of Multiply Protonated Peptides. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 47–58.
30. McLafferty, F. W.; Horn, D. M.; Breuker, K.; Ge, Y.; Lewis, M. A.; Cerda, B.; Zubarev, R. A.; Carpenter, B. K. Electron Capture Dissociation of Gaseous Multiply Charged Ions by Fourier-Transform Ion Cyclotron Resonance. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 245–249.
31. Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated de Novo Sequencing of Proteins by Tandem High-Resolution Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10313–10317.
32. Horn, D. M.; Ge, Y.; McLafferty, F. W. Activated Ion Electron Capture Dissociation for Mass Spectral Sequencing of Larger (42 kDa) Proteins. *Anal. Chem.* **2000**, *72*, 4778–4784.
33. Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. *Anal. Chem.* **2000**, *72*, 563–573.
34. Zubarev, R. A.; Fridriksson, E. K.; Horn, D. M.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. Electron Capture Dissociation Produces Many More Protein Backbone Cleavages than Collisional and IR Excitation. *Mass Spectrom. Biol. Med.* **2000**, *111*–120.
35. Schilling, B.; Wang, W.; McMurray, J. S.; Medzihradzsky, K. F. Fragmentation and Sequencing of Cyclic Peptides by Matrix-Assisted Laser Desorption/Ionization Post-Source Decay Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 2174–2179.
36. Medzihradzsky, K. F.; Campbell, J. M.; Baldwin, M. A.; Falick, A. M.; Juhasz, P.; Vestal, M. L.; Burlingame, A. L. The Characteristics of Peptide Collision-Induced Dissociation Using a High-Performance MALDI-TOF/TOF Tandem Mass Spectrometer. *Anal. Chem.* **2000**, *72*, 552–558.
37. Bienvenut, W. V.; Deon, C.; Pasquarello, C.; Campbell, J. M.; Sanchez, J. C.; Vestal, M. L.; Hochstrasser, D. F. Matrix-Assisted Laser Desorption/Ionization-Tandem Mass Spectrometry with High Resolution and Sensitivity for Identification and Characterization of Proteins. *Proteomics* **2002**, *2*, 868–876.
38. Yergey, A. L.; Coorsen, J. R.; Backlund, P. S. Jr.; Blank, P. S.; Humphrey, G. A.; Zimmerberg, J.; Campbell, J. M.; Vestal, M. L. De Novo Sequencing of Peptides Using MALDI/TOF-TOF. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 784–791.
39. Juhasz, P.; Campbell, J. M.; Vestal, M. L. MALDI-TOF/TOF Technology for Peptide Sequencing and Protein Identification. *Mass Spectrom. Hyphen. Tech. Neuropeptide Res.* **2002**, 375–413.
40. Wattenberg, A.; Organ, A. J.; Andrew, J.; Schneider, K.; Tyldesley, R.; Bordoli, R.; Bateman, R. H. Sequence Dependent Fragmentation of Peptides Generated by MALDI Quadrupole Time-of-Flight (MALDI Q-TOF) Mass Spectrometry and Its Implications for Protein Identification. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 772–783.
41. She, Y.-M.; Wang, G.-Q.; Loboda, A.; Ens, W.; Standing, K. G.; Burczynski, F. J. Sequencing of Rat Liver Cytosolic Proteins by Matrix-Assisted Laser Desorption Ionization-Quadrupole Time-of-Flight Mass Spectrometry Following Electrophoretic Separation and Extraction. *Anal. Biochem.* **2002**, *310*, 137–147.
42. She, Y.-M.; Haber, S.; Seifers, D. L.; Loboda, A.; Chernushevich, I.; Perreault, H.; Ens, W.; Standing, K. G. Determination of the Complete Amino Acid Sequence for the Coat Protein of Brome Mosaic Virus by Time-of-Flight Mass Spectrometry: Evidence for Mutations Associated with Change of Propagation Host. *J. Biol. Chem.* **2001**, *276*, 20039–20047.
43. Shevchenko, A.; Sunyaev, S.; Loboda, A.; Shevchenko, A.; Bork, P.; Ens, W.; Standing, K. G. Charting the Proteomes of Organisms with Unsequenced Genomes by MALDI-Quadrupole Time-of-Flight Mass Spectrometry and BLAST Homology Searching. *Anal. Chem.* **2001**, *73*, 1917–1926.
44. Cramer, R.; Corless, S. The Nature of Collision-Induced Dissociation Processes of Doubly Protonated Peptides: Comparative Study for the Future Use of Matrix-Assisted Laser Desorption/Ionization on a Hybrid Quadrupole Time-of-Flight Mass Spectrometer in Proteomics. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2058–2066.
45. Taylor, J. A.; Johnson, R. S. Implementation and Uses of Automated de Novo Peptide Sequencing by Tandem Mass Spectrometry. *Anal. Chem.* **2001**, *73*, 2594–2604.
46. Johnson, R. S.; Taylor, J. A. Searching Sequence Databases via de Novo Peptide Sequencing by Tandem Mass Spectrometry. *Methods Mol. Biol.* **2000**, *146*, 41–61.
47. Hines, W. M.; Falick, A. M.; Burlingame, A. L.; Gibson, B. W. Pattern-Based Algorithm for Peptide Sequencing from Tandem High Energy Collision-Induced Dissociation Mass Spectra. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 326–336.
48. Fernandez-De-Cossio, J.; Gonzalez, J.; Satomi, Y.; Shima, T.; Okumura, N.; Besada, V.; Betancourt, L.; Padron, G.; Shimomishi, Y.; Takao, T. Automated Interpretation of Low-Energy Collision-Induced Dissociation Spectra by SeqMS, a Software Aid for de Novo Sequencing by Tandem Mass Spectrometry. *Electrophoresis* **2000**, *21*, 1694–1699.
49. Dancik, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. De Novo Peptide Sequencing via Tandem Mass Spectrometry. *J. Comput. Biol.* **1999**, *6*, 327–342.
50. Chen, T.; Kao, M.-Y.; Tepel, M.; Rush, J.; Church, G. M. A Dynamic Programming Approach to de Novo Peptide Sequencing via Tandem Mass Spectrometry. *J. Comput. Biol.* **2001**, *8*, 325–337.
51. Tabb, D. L.; Smith, L. L.; Breci, L. A.; Wysocki, V. H.; Lin, D.; Yates, J. R. 3rd. Statistical Characterization of Ion Trap Tandem Mass Spectra from Doubly Charged Tryptic Peptides. *Anal. Chem.* **2003**, *75*, 1155–1163.
52. 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.

53. Yu, W.; Vath, J. E.; Huberty, M. C.; Martin, S. A. Identification of the Facile Gas-Phase Cleavage of the Asp-Pro and Asp-Xxx Peptide Bonds in Matrix-Assisted Laser Desorption Time-of-Flight Mass Spectrometry. *Anal. Chem.* **1993**, *65*, 3015–3023.
54. Qin, J.; Chait, B. T. Preferential Fragmentation of Protonated Gas-Phase Peptide Ions Adjacent to Acidic Amino Acid Residues. *J. Am. Chem. Soc.* **1995**, *117*, 5411–5412.
55. Tsapralis, G.; Nair, H.; Somogyi, A.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. Influence of Secondary Structure on the Fragmentation of Protonated Peptides. *J. Am. Chem. Soc.* **1999**, *121*, 5142–5154.
56. Qin, J.; Chait, B. T. Collision-Induced Dissociation of Singly Charged Peptide Ions in a Matrix-Assisted Laser Desorption Ionization Ion Trap Mass Spectrometer. *Int. J. Mass Spectrom.* **1999**, *190*, 313–320.
57. Cronshaw, J. M.; Krutchinsky, A. N.; Zhang, W.; Chait, B. T.; Matunis, M. J. Proteomic Analysis of the Mammalian Nuclear Pore Complex. *J. Cell Biol.* **2002**, *158*, 915–927.
58. Zhang, J.; Kalkum, M.; Chait, B. T.; Roeder, R. G. The N-CoR-HDAC3 Nuclear Receptor Corepressor Complex Inhibits the JNK Pathway Through the Integral Subunit GPS2. *Mol. Cell* **2002**, *9*, 11–623.
59. Krutchinsky, A. N.; Kalkum, M.; Chait, B. T. Automatic Identification of Proteins with a MALDI-Quadrupole Ion Trap Mass Spectrometer. *Anal. Chem.* **2001**, *73*, 5066–5077.
60. Qin, J.; Fenyó, D.; Zhao, Y.; Hall, W. W.; Chao, D. M.; Wilson, C. J.; Young, R. A.; Chait, B. T. A Strategy for Rapid, High-Confidence Protein Identification. *Anal. Chem.* **1997**, *69*, 3995–4001.
61. Krutchinsky, A. N.; Chait, B. T. On the Nature of the Chemical Noise in MALDI Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 129–134.
62. Krutchinsky, A. N.; Zhang, W.; Chait, B. T. Rapidly Switchable Matrix-Assisted Laser Desorption/Ionization and Electrospray Quadrupole-Time-of-Flight Mass Spectrometry for Protein Identification. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 493–504.
63. Chait, B. T.; Gisin, B. F.; Field, F. H. Fission Fragment Ionization Mass Spectrometry of Alamethicin I. *J. Am. Chem. Soc.* **1982**, *104*, 5157–5162.