Research Article

Rapid Communications in Mass Spectrometry

Received: 15 September 2014 Revised: 12 December 2014 Accepted: 22 December 2014 Published online in Wiley Online Library

Rapid Commun. Mass Spectrom. **2015**, *29*, 659–666
(wileyonlinelibrary.com) DOI: 10.1002/rcm.7137

Paired single residue-transposed Lys-N and Lys-C digestions for label-free identification of N-terminal and C-terminal MS/MS peptide product ions: ultrahigh resolution Fourier transform ion cyclotron resonance mass spectrometry and tandem mass spectrometry for peptide *de novo* sequencing

Naomi C. Brownstein[1,2†], Xiaoyan Guan[1†], Yuan Mao[3,4], Qian Zhang[3], Peter A. DiMaggio[5], Qiangwei Xia[4,6], Lichao Zhang[7], Alan G. Marshall[1,3] and Nicolas L. Young[1*]

[1]National High Magnetic Field Laboratory, Florida State University, 1800 East Paul Dirac Drive, Tallahassee, FL 32310, USA
[2]Department of Statistics, Florida State University, 117 N. Woodward Ave., Tallahassee, FL 32306, USA
[3]Department of Chemistry and Biochemistry, Florida State University, 95 Chieftain Way, Tallahassee, FL 32303, USA
[4]Regeneron Pharmaceuticals Inc., 777 Old Saw Mill River Road, Tarrytown, NY 10591, USA
[5]Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK
[6]CMP Scientific, Corp, 760 Parkside Ave, Brooklyn, NY 11226, USA
[7]Department of Chemistry, University of Virginia, McCormick Rd., Charlottesville, VA 22904, USA

RATIONALE: Paired Lys-N and Lys-C proteases produce peptides of identical mass and similar retention time, but different tandem mass spectra. Data from these parallel experiments provide constraints that are applied before data analysis. With this approach, we can find matched spectra before analysis, distinguish ion type, and determine residue level confidence.

METHODS: Aliquots are digested separately by Lys-N and Lys-C peptidases, and analyzed by reversed-phase nano-flow liquid chromatography, collision-induced dissociation, and 14.5 T Fourier transform ion cyclotron resonance mass spectrometry. Matched pairs of fragmentation spectra with equal precursor mass and similar retention times from each digestion are compared, leveraging single-residue transposed information with independent interferences to confidently identify fragment ion type, residues, and peptides. The paired spectra are solved together as a single *de novo* sequencing problem.

RESULTS: Two pairs of spectra of a *de novo* sequenced 18-mer are presented. In one example, the 18-mer has coverage of all residues except the N- and C- terminal lysines and their adjacent residues. The confidence level is high due to six pairs of transposed ions. In the other example, the coverage is incomplete. Nonetheless, nine pairs of transposed ions facilitate identification of two trimer sequence tags with high confidence, one with medium confidence, and additional sequence information with residue-by-residue confidence, thus demonstrating the value of residue-by-residue confidence.

CONCLUSIONS: Sequence identity and variability, such as post-translational modifications (PTMs), are essential to understanding biological function and disease. The present method facilitates discovery of new peptides with multiple levels of confidence, promises potential characterization of PTMs, and validates peptides from databases. Independent validation may be of interest for a number of applications.

Tandem mass spectrometry is a powerful tool for protein and peptide characterization.[1–5] Confident sequencing of peptides and proteins is the crux of proteomics; however, complete characterization of peptides and proteins, including all sources of variation, such as post-translational modifications (PTMs), splice variants, and novel sequence mutations, is not entirely addressed by most current methods.

Database search techniques[6–15] compare acquired mass spectra to a database of known protein sequences to identify the sequence tag, which serves as a tool for protein identification. Although database-dependent identification of proteins by mass spectrometry is well established, the methods do not apply if the analyte sequence does not exist in the current databases. For example, unsequenced genomes, polymorphisms, and splice variants[16] are not generally amenable to such methods. Protein homology, incorrect

*Correspondence to: N. L. Young, National High Magnetic Field Laboratory, Florida State University, 1800 East Paul Dirac Drive, Tallahassee, FL 32310, USA.
E-mail: nyoung@magnet.fsu.edu

†These authors contributed equally to this work.

assignment of charge state for precursor ions, incorrect identification of isotopic peaks, and enzymatic digestion at non-traditional sites can result in incorrect database matches.[17] Database-centered protein identification methods are not well designed to detect variations beyond the amino acid sequence, such as the presence and location of PTMs.[18–20] Many database search methods currently allow for the presence and *de novo* detection of variable PTMs in addition to the database search; however, a *de novo* component often makes it difficult to determine the statistical significance of the finding.

De novo sequencing[16,21–25] circumvents these limitations; *de novo* sequencing is a process in which amino acid sequences are derived directly from tandem mass spectra without the assistance of a sequence database.[26–32] Independence from a sequence database makes *de novo* sequencing the only method for identifying novel peptides, unsequenced organisms, and other sources of variance, which database-search methods were not designed to detect.[22,33] However, peptide *de novo* sequencing has been a more challenging problem than the traditional database search approach, due to ambiguous assignments and incomplete fragmentation, leading to low sequence coverage, and difficulty in distinguishing ion series, notably N-terminal from C-terminal MS/MS product ions (*b* ions from *y* ions) – the main topic of this paper.

One way to improve upon *de novo* sequencing is to chemically derivatize a sample, enabling distinction between ion series. For example, Keough and colleagues[34] employed guanidination of the ε-amino group of lysine-terminated tryptic peptides to identify proteins for matrix-assisted laser desorption/ionization (MALDI) data. Conrotto and Hellman[35] similarly targeted the ε-amino group of lysine-terminated tryptic peptides but used an imidazole derivative to minimize chemical side effects. Kim and colleagues[36] introduced bromide signatures on the C-terminus, facilitated by oxazolone chemistry, which enables identification of *y* ions. Warwood and colleagues[37] implemented guanidination, which increases selectivity and allows for relative quantitation as well as identification.

Other methods interrogate the same sample multiple times to increase confidence in the identification. Liu and colleagues[33] capitalized on both bottom-up and top-down mass spectrometry. Two types of fragmentation, collision-induced dissociation (CID) and electron capture dissociation (ECD), have been leveraged in multiple settings.[38,39] Guthals and colleagues[40] combined information from three types of fragmentation: CID, ETD, and higher energy collisional dissociation (HCD). In each case, the complementarity of multiple data types provided the most confident assignment. Finally, it is possible to distinguish C-terminal (z•) from N-terminal (c′) ECD product ions based on the ratio of prime to radical ion abundance in ECD vs activated-ion ECD (AI-ECD) MS/MS product ion mass spectra.[41]

Traditional *de novo* methods are prone to false positives in part due to limited mass resolution and accuracy. Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS)[42] offers the highest mass resolution and accuracy,[43] and has been used previously[44] for *de novo* sequencing with ECD to study bromide loss and resulting fragmentation patterns. As the constraints in mass tolerance are tightened, the resulting assignments become increasing confident.

Here, we present a method for *de novo* sequencing based on a combination of paired complementary digestion, high-resolution FTICR-MS, and assignment of confidence in the identification at the individual ion level. We demonstrate the method for bovine serum albumin (BSA) digested by Lys-N and Lys-C, a unique pair of enzymes that produce single residue-transposed peptides of the same molecular weight and similar retention time in liquid chromatography (LC). Paired digestions enable unequivocal distinction between N-terminal and C-terminal ions for increased confidence in the overall assignment as well as finely partitioned confidence along a particular peptide segment. The resulting "granular" detail allows for anchor points that reduce the length of sequences that must be assigned. Lacking such advantages, even long peptides that are not tractable with traditional *de novo* methods may be sequenced accurately as a series of shorter segments.

Finally, in an alternative label-free technique, if mass measurement is sufficiently accurate to determine peptide elemental composition ($C_cH_hN_nO_oS_s$), then N-terminal and C-terminal ions may be distinguished according to whether the number of nitrogen atoms plus hydrogen atoms is even or odd ("valence parity" method[45]). However, experimental mass measurement accuracy is generally insufficient to determine elemental compositions for peptides containing more than 6–8 amino acids.[46]

CONCEPTUAL BASIS

Identification of *b* and *y* ions

To illustrate the basis of the method, consider the segment: …KATEEQLK… Assuming there are no missed cleavages, Lys-N digestion will yield the precursor peptide KATEEQL whereas (separate) Lys-C digestion will yield ATEEQLK (See Fig. 1). The two precursor ions have exactly the same mass and differ only in the position of the lysine residue. Subsequent CID will yield pairs of ions between the two spectra that differ by (+/−) the residue mass of lysine (128.0950 Da), for cleavages that take place at the same peptide linkage (one fragment ion index away). For example, the *y*5 ion from ATEEQLK will be EEQLK and the *y*4 ion from KATEEQL will be EEQL, which is *lower* in mass by one lysine residue. Similarly, the *b*5 fragment ion from ATEEQLK will be ATEEQ and the *b*6 fragment ion from KATEEQL will be KATEEQ, which is *higher* in mass by one K residue. Let M(Lys-C) and M(Lys-N) denote the monoisotopic neutral masses corresponding to fragment ions from MS/MS from Lys-C and Lys-N digestions. The CID product ion spectra are then searched for ions that satisfy Eqn. (1) or (2):

$$\mathrm{M(Lys\text{-}N)}-\mathrm{M(Lys\text{-}C)} = +128.0950 \quad (1)$$

$$\mathrm{M(Lys\text{-}N)}-\mathrm{M(Lys\text{-}C)} = -128.0950 \quad (2)$$

As shown in Fig. 1, an M(Lys-N) – M(Lys-C) mass difference of +128.0950 identifies both members of the pair as *b* ions, and an M(Lys-N) – M(Lys-C) mass difference of –128.0950 identifies both members of the pair as *y* ions.

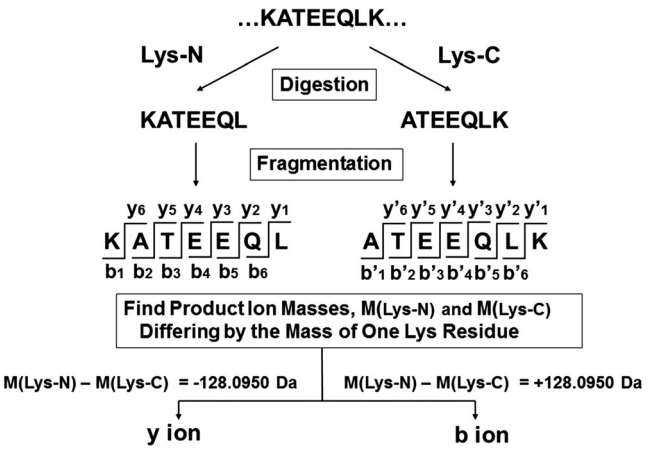

Figure 1. Schematic procedure for the *de novo* method. Parallel Lys-N and Lys-C digestions produce single residue-transposed peptides, which are then fragmented by collision-induced dissociation in separate LC/MS/MS analyses. Comparison of paired spectra that meet multiple constraints serves to distinguish N-terminal from C-terminal fragment ion types for confident identification of the amino acid sequence. The two spectra are considered simultaneously with the assumption, based on mass and retention time, that they are single residue-transposed peptides. Fragment ions that differ by 128.0950 Da are related by the equations listed: e.g., Lys-N b_3 – Lys-C b_2 = +128.0959, identifying both as *b* ions. Ions that are now of known ion type are combined into a single Lys-C fragment spectrum via the listed relationships and used as anchor points in a traditional *de novo* algorithm that includes the undistinguished ions of the original Lys-C spectrum.

Knowledge of fragment ion type *a priori* can significantly improve the confidence of *de novo* sequencing. These transposed fragment ions contain the same information about the sequence, with the exception of the placement of the lysine residue. Thus, the presence of both ions in a pair increases confidence in the identification. Distinguishing between *b* and *y* ions without prior sequence knowledge reduces the dimensionality of the search space for peptide/protein identification and sequencing. Ion type is determined without labeling, based simply upon the mass difference between two MS/MS fragment ions cleaved from complementary precursor ions from the paired digestions.

Specific matched pairs from CID product ion mass spectra from Lys-N and Lys-C digestion are selected for comparison of their fragment ions. In particular, a spectrum pair is considered only if the precursor masses are identical and HPLC retention times are similar. We used a relatively wide retention time window of 5 min, but other windows are possible. Empirical observation shows that rarely if ever do the Lys-N and Lys-C peptides have the same retention time, nor is there a consistent retention order, but they consistently elute at similar times, typically within a couple of minutes. Such restriction greatly reduces the computational time required to analyze the data compared to a complete analysis of all pairs. Moreover, interferences for either Lys-C or Lys-N digests alone will be reduced by the above restriction, because most interferants will have either different retention time or different precursor mass. Consequently, the matched spectrum pairs are more robust with respect to chemical noise compared to the spectra for individual Lys-C or Lys-N digestion.

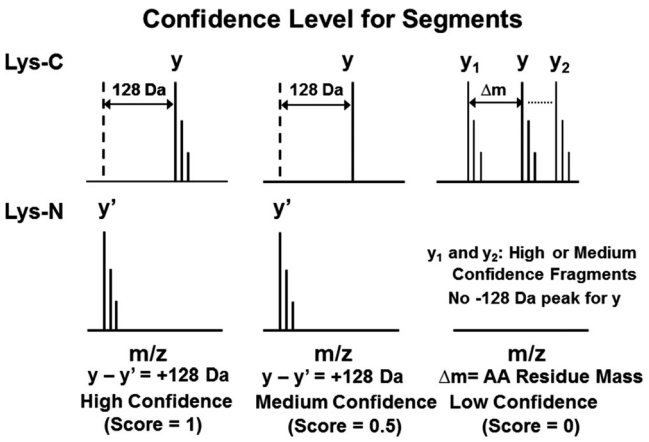

Figure 2. Confidence levels for each set of fragment ions, individually assigned based on ion presence or absence and the quality of their isotopic distributions in both Lys-N and Lys-C spectra. In this manner, we can distinguish high versus low confidence segments of the sequence independent of the complete sequence assignment, and reduce the low confidence sections to a series of smaller *de novo* problems.

Assignment confidence scoring

A confidence level may be assigned to each inter-amino acid fragmentation site according to two factors: the ability or inability to classify the ion by type based on Eqns. (1) and (2) and the presence or absence of an isotopic distribution (see Fig. 2). A high confidence (score = 1) fragment ion is one for which (a) ion type can be determined by finding a complementary fragment ion related by Eqn. (1) or (2); and (b) both fragment ions exhibit isotopic distributions. A medium confidence (score = 0.5) fragment ions is one for which (a) ion type can be determined by finding a complementary fragment ion related by Eqn. (1) or (2) and (b) one of the fragment ions from two spectra lacks an isotopic distribution (e.g., electronic noise or heavy-atom peak magnitude below a defined threshold). A low confidence (score = 0) fragment ion is one that appears in the Lys-C spectrum with an isotopic distribution but lacks any corresponding fragment ion differing by 128.0950 Da in the Lys-N spectrum and is thus of unknown ion type.

We have applied the previously described criteria to generate confidence scores at both the amino acid and peptide levels. Amino acid level confidence is the average of the confidence for the fragment ions on each side of the residue. The overall peptide confidence is given by the average confidence level for all fragment ions used to determine the putative peptide identity, multiplied by 100. Multiple levels of confidence establish the depth and quality of the data at each level.

Proper interpretation of the peptide score is as follows. The lowest possible peptide score is zero, which indicates that the use of paired spectra did not add any additional information above standard *de novo* sequencing methods based on only one spectrum. To be clear, a score of zero does not imply that there is no sequence information. For example, the best single spectrum *de novo* possible would receive a score of zero. Zero simply means there was no added information from the experimental scheme we describe here. On the other hand, the best possible score is 100, and indicates that all of the ion

information in one spectrum is supported by complementary ions in the other spectrum, and all ions forming each pair are complete with full isotopic distributions. Scores between zero and 100 indicate some pairs of complementary ions and allow the possibility that exactly one of the ions in a pair lacks an isotopic distribution.

EXPERIMENTAL

In-solution Lys-C and Lys-N digestion of BSA and protein mixture

Endopeptidase Lys-C (Lys-C) and bovine serum albumin (BSA) were purchased from Sigma Aldrich Chemical Co. (St. Louis, MO, USA) and recombinant endopeptidase Lys-N (Lys-N) was purchased from U-ProteinExpress (Utrecht, The Netherlands). BSA (100 µg) was diluted in 50 µL of 100 mM ammonium bicarbonate (pH 8.5) and reduced with 5 mM dithiothreitol (incubated at 95 °C for 10 min), followed by alkylation with 12 mM iodoacetamide (incubated in the dark at room temperature for 30 min). For Lys-C digestion, the resulting solution was digested with Lys-C re-suspended in 100 µL HPLC grade water at a concentration of 0.05 µg/µL at a 1:66 (w/w enzyme/substrate) ratio and incubated overnight at 37 °C. For Lys-N digestion, the resulting solution was digested with Lys-N re-suspended in 73 µL HPLC grade water at a concentration of 0.37 µg/µL at a 1:85 (w/w enzyme/substrate) ratio and incubated overnight at 37 °C.

Data acquisition: on-line LC/CID MS/MS

Each Lys-C and Lys-N digest (~4 µg) was subjected to reversed-phase nano-flow high-performance liquid chromatographic (HPLC) separation using an Ultra Nano HPLC system (Eksigent, Livermore, CA, USA). Digests were separated on an in-house packed Xterra C18 column (i.d. 200 µm, o.d. 357 µm, 5 µm, 125 Å; Waters, MA, USA) at a flow rate of 200 nL/min with mobile phase A containing 95% H_2O and 4.9% acetonitrile with 0.1% formic acid (FA) and mobile phase B containing 95% acetonitrile and 4.9% H_2O with 0.1% FA. Beginning with 0% mobile phase B, phase B linearly increased to 60% after 120 min and to 90% at 121 min. After washing at 90% for 21 min, the column was equilibrated with 0% B for 38 min. A blank solution containing 95% H_2O and 4.9% acetonitrile with 0.1 % formic acid was run between each sample injection for quality control. Beginning with 0% mobile phase B, phase B linearly increased to 90% after 16 min, wash at 90% for 8 min and the column was equilibrated with 0% B for 16 min.

Separate Lys-C and Lys-N digests were electrosprayed at 3 kV and injected into a modified linear ion trap (LTQ, Thermo Fisher Scientific, San Jose, CA, USA) 14.5 T FTICR mass spectrometer[47] equipped with a Predator data station.[48] Precursor ions were isolated in an external octopole ion trap[49] with an isolation window width, $[(m/z)2-(m/z)1)]=2$, followed by CID fragmentation at a normalized collision energy of 35% for the five most abundant precursor ions. Ions resulting from five cycles of CID were accumulated in an adjacent storage octopole (2.2 MHz, 250 $V_{p\text{-}p}$) before being transferred to the ICR cell (external ion accumulation increases the number of ions that are transferred to the ICR cell). The ion accumulation period was typically less than 100 ms during peptide elution and the FTICR time-domain signal acquisition period was 1400 ms (i.e., an overall duty cycle of ~0.6 Hz per acquisition). The FTICR time-domain signal was acquired from *m/z* 400 to 2000 at a mass resolving power ($m/\Delta m_{50\%}$, in which $\Delta m_{50\%}$ is the full width of a mass spectral peak at half-maximum peak height) of 200 000 at *m/z* 400. Following Hanning apodization and zero-filling, fast Fourier transformation, and phase correction,[50] the mass spectra were externally calibrated by Pierce LTQ Velos ESI positive ion calibration solution (Thermo Fisher Scientific, San Jose, CA, USA) based on the quadrupolar trapping approximation.[51,52] The mass spectrometer incorporates various improvements in sensitivity, speed, mass resolution, and mass accuracy,[53,54] and is ideally suited for *de novo* sequencing with high confidence.

Data analysis

Ions were manually identified and assigned as *b* or *y* ions from mass spectral segments selected as described above.

RESULTS AND DISCUSSION

Figure 3 shows CID product ion mass spectra and cleavage maps for isobaric precursor ions containing 18 identical residues (VPQVSTPTLVEVSRSLGK and KVPQVSTPTLVEVSRSLG) from Lys-C and Lys-N digested samples of BSA. Sequence coverage is nearly complete for MS/MS of the Lys-C digest precursor ion. On the other hand, coverage is sparse for the Lys-N digest precursor ion, lacking any fragmentation within the first seven N-terminal residues. Nevertheless, the method allows us to confidently identify the relatively long peptide sequence by leveraging the complementarity of pairs of *b* and *y* ions from each spectrum. For example, the y_4 ion from the Lys-N spectrum and the y_5 ion from the Lys-C spectrum differ by the mass of a lysine residue, and correspond to the fragment RSLG(K). We thus assign these paired ions a high confidence score of one. The y_5 ion from the Lys-N digest and the y_6 ion from the Lys-C spectrum are also both present, leading to a confidence score of one for the N-terminal fragment of SRSLG(K). Because both sides of the S residue are identified with confidence score of one, the amino acid is also identified with high confidence (score = 1). For other amino acids, such as the R, the fragment ions are identified with high confidence on one side, with lower confidence on the other side. The concept of medium confidence amino acid identification may be extended to segments comprised of adjacent medium-confidence amino acids, such as TP, VEV, and RS. Still other amino acids are fragmented on both sides but lack complementary ions. These ions have the lowest confidence of those that match known residue masses. The six pairs of confidently identified corresponding ions serve as anchors to aid in confident identification of the entire 18-mer based on the paired mass spectra. The peptide score for the first replicate is 12, showing that the use of paired spectra resulted in a small amount of complementary information beyond simply sequencing the Lys-C spectrum alone.

Figure 4 shows another useful application of our method for the same peptide from a different biological replicate. In this case, the sequence coverage is not optimal from either spectrum. Both the Lys-C and Lys-N digest CID product ion

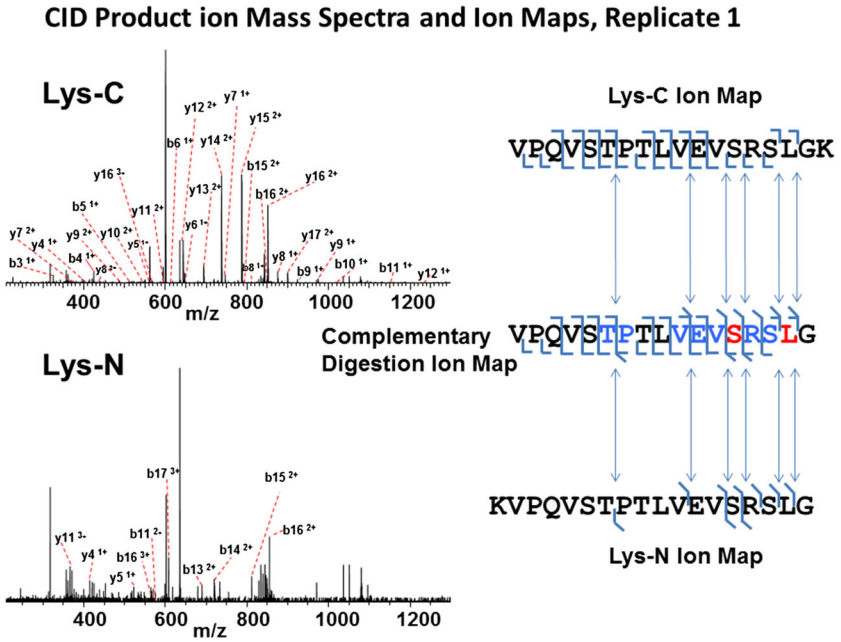

Figure 3. Paired annotated CID positive product ion mass spectra for the same peptide digested separately by Lys-C (top) and Lys-N (bottom), showing complementary transposed ion pairs and illustrating how different levels of confidence (red, blue, and black denote high, medium, and low) for various amino acids and segments of the peptide give a more complete picture of the quality of the identification.

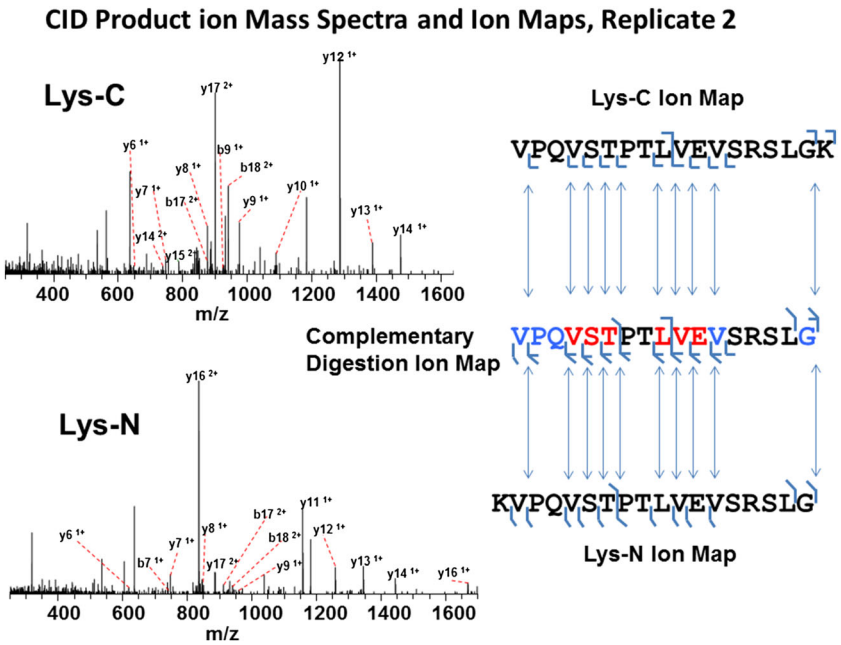

Figure 4. Paired annotated CID positive product ion mass spectra for another replicate of the same peptide, demonstrating how large numbers of paired transposed ions serve as anchors to aid in confident identification of peptide segments even if individual and total sequence coverage are suboptimal.

mass spectra lack at least four fragment ions. Nevertheless, the presence of nine pairs of complementary ions increases the confidence in the peptide identification. The confidence score for this second replicate is 67. The higher score reflects the additional pairs of complementary ions, which help to mitigate the low sequence coverage and aid the *de novo* sequencing confidence. In fact, the presence of two pairs of complementary ions on both sides of an amino acid results not only in amino acids identified with high confidence, but also in high confidence sequence tags (VST and LVE). Such

trimers could be used to check the final identification with known databases. The paired spectra also have a trimer (VPQ) and two single amino acids identified with medium confidence. Note that although the sequence coverage for the same peptide shown for two replicate data sets in Figs. 3 and 4 is different, the ability and flexibility of this method for *de novo* sequencing data of different quality yields several high-confidence residue identifications. More important, the fact that the spectra are not the same is one of the major points of the method. The fragment ion peaks provide common information whereas spectral noise in the two spectra is unrelated.

Finally, although the entire peptide may not be *de novo* sequenced in this case, the sequence coverage for the complementary digestion is higher due to the additional information from the transposed product ion pairs. Moreover, the length of the sequential residues determined is the critical outcome of any *de novo* sequencing experiment. The objective of most previous *de novo* experiments is to determine the complete sequence of tryptic peptides. Thus, they strive to identify the entire sequence of each peptide even if there is insufficient information to do so. With our method we are able to derive relatively long sequence tags with enhanced and known confidence, and correctly identify the portions of the sequence that are inherently ambiguous in lower quality data. The relatively longer peptides derived from Lys-C/Lys-N compared to a tryptic digest also potentially provide longer sequence tags than tryptic peptides. Although the sequence tags shown in Fig. 4 do not cover the entire sequence of the peptide, our method can provide long sequence tags with high confidence. Thus, sequence tags of useful length may be identified with high local confidence despite incomplete fragmentation. Such high confidence local sequences are of at least equal value to confidently identified proteolytic peptides of the same length and have the added advantage of being linked to additional lower confidence or even incomplete sequence information that may be confirmed or completed by subsequent data.

CONCLUSIONS

The present methodology leverages paired single residue-transposed digestion with Lys-C and Lys-N peptidases and high-resolution FTICR-MS/MS for *de novo* peptide sequencing. The special relationships inherent in the transposed fragment ions of the two peptidases enable identification of ion type (N-terminal vs C-terminal), knowledge of which simplifies the *de novo* sequencing problem. The chosen pair of peptidases is unique in producing specific mass relationships, such as those in Eqns. (1) and (2), which are known before the experiment and data analysis begin. Other peptidases, such as trypsin and the endoproteinase, Glu-C, do not have this property.

Paired fragments are required to have isobaric precursor ions, and similar HPLC retention times, thereby providing confidence scores for sequence assignment. Unlike other approaches, our method provides measures of confidence not only for peptide identification, but also at the ion, amino acid, and sequence tag levels. We demonstrate the method for manually annotated experimental BSA CID MS/MS positive-ion mass spectra to confidently identify a peptide that is 18 amino acids long.

Sequences that are confidently identified by the paired single residue-transposed digestion method may be fed into a subsequent database search. The value in such an approach is that the initial identity is independent of the database, as in Figs. 3 and 4. The database serves as an *ex post facto* confirmation rather than a constraint. An unconstrained *de novo* search may include other sources of variation, such as PTMs, not necessarily included in the database. The underlying sequence, however, must be confirmed by the database. This approach differs from current database search methods, in which additional variations simply relax the constraints of the database search, increasing the likelihood of a hit. Here, the constraints of the *de novo* search are relaxed, not the database search. The *de novo* step results in more false positives, but the database search can filter out the false positives.

The method should be a useful tool for analysis of proteomes of both known and unknown genomes. Algorithmic development for automatic processing and validation is in progress. The method may be expanded to allow for *de novo* protein identification for complex samples and post-translational modification localization for peptides. Additional information as noted in the text is available free of charge via the internet.[55]

Acknowledgements

This work was supported by NSF Division of Materials Research through DMR-11-57490 and the State of Florida. The authors thank Christopher L. Hendrickson for his early assistance with the project.

REFERENCES

[1] R. Aebersold, M. Mann. Mass spectrometry-based proteomics. *Nature.* **2003**, *422*, 198.
[2] A. Doerr. Mass spectrometry-based targeted proteomics. *Nat. Methods* **2013**, *10*, 23.
[3] J. R. Yates, C. I. Ruse, A. Nakorchevsky. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **2009**, *11*, 49.
[4] Z. Zhang, S. Wu, D. L. Stenoien, L. Paša-Tolić. High-throughput proteomics. *Annu. Rev. Anal. Chem.* **2014**, *7*, 427.
[5] M. Mann, R. C. Hendrickson, A. Pandey. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **2001**, *70*, 437.
[6] H. Steen, M. Mann. The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell. Biol.* **2004**, *5*, 699.
[7] J. K. Eng, A. L. McCormack, J. R. Yates III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976.
[8] R. Craig, R. C. Beavis. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics.* **2004**, *20*, 1466.
[9] Y. Fu, Q. Yang, R. Sun, D. Li, R. Zeng, C. X. Ling, W. Gao. Exploiting the kernel trick to correlate fragment ions for peptide identification via tandem mass spectrometry. *Bioinformatics* **2004**, *20*, 1948.
[10] L. Y. Geer, S. P. Markey, J. A. Kowalak, L. Wagner, M. Xu, D. M. Maynard, X. Yang, W. Shi, S. H. Bryant. Open mass spectrometry search algorithm. *J. Proteome Res.* **2004**, *3*, 958.

[11] D. Li, Y. Fu, R. Sun, C. X. Ling, Y. Wei, H. Zhou, R. Zeng, Q. Yang, S. He, W. Gao. pFind: a novel database-searching software system for automated peptide and protein identification via tandem mass spectrometry. *Bioinformatics* **2005**, *21*, 3049.
[12] D. N. Perkins, D. J. Pappin, D. M. Creasy, J. S. Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551.
[13] L. H. Wang, D. Q. Li, Y. Fu, H. P. Wang, J. F. Zhang, Z. F. Yuan, R. X. Sun, R. Zeng, S. M. He, W. Gao. pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 2985.
[14] R. Craig, R. C. Beavis. A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2310.
[15] A. I. Nesvizhskii. Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol. Biol.* **2007**, *367*, 87.
[16] J. Allmer. Algorithms for the de novo sequencing of peptides from tandem mass spectra. *Expert Rev. Proteomics* **2011**, *8*, 645.
[17] Y. Chen, J. Zhang, G. Xing, Y. Zhao. Mascot-derived false positive peptide identifications revealed by manual analysis of tandem mass spectra. *J. Proteome Res.* **2009**, *8*, 3141.
[18] M. Mann, O. N. Jensen. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **2003**, *21*, 255.
[19] G. T. Cantin, J. R. Yates III. Strategies for shotgun identification of post-translational modifications by mass spectrometry. *J. Chromatogr. A* **2004**, *1053*, 7.
[20] J. Seo, J. Jeong, Y. M. Kim, N. Hwang, E. Paek, K. J. Lee. Strategy for comprehensive identification of post-translational modifications in cellular proteins, including low abundant modifications: application to glyceraldehyde-3-phosphate dehydrogenase. *J. Proteome Res.* **2008**, *7*, 587.
[21] V. Dancik, T. A. Addona, K. R. Clauser, J. E. Vath, P. A. Pevzner. De novo peptide sequencing via tandem mass spectrometry. *J. Comput. Biol.* **1999**, *6*, 327.
[22] K. F. Chong, H. W. Leong. Tutorial on de novo peptide sequencing using MS/MS mass spectrometry. *J. Bioinform. Comput. Biol.* **2012**, *10*, 1231002.
[23] C. Hughes, B. Ma, G. A. Lajoie. De novo sequencing methods in proteomics. *Methods Mol. Biol.* **2010**, *604*, 105.
[24] K. G. Standing. Peptide and protein de novo sequencing by mass spectrometry. *Curr. Opin. Struct. Biol.* **2003**, *13*, 595.
[25] B. Ma, R. Johnson. De novo sequencing and homology searching. *Mol. Cell. Proteomics* **2012**, *11*, O111.014902.
[26] A. Frank, P. Pevzner. PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal. Chem.* **2005**, *77*, 964.
[27] B. Ma, K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby, G. Lajoie. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337.
[28] P. A. DiMaggio Jr, C. A. Floudas. De novo peptide identification via tandem mass spectrometry and integer linear optimization. *Anal. Chem.* **2007**, *79*, 1433.
[29] B. Fischer, V. Roth, F. Roos, J. Grossmann, S. Baginsky, P. Widmayer, W. Gruissem, J. M. Buhmann. NovoHMM: a hidden Markov model for de novo peptide sequencing. *Anal. Chem.* **2005**, *77*, 7265.
[30] P. A. Dimaggio Jr, C. A. Floudas. A Mixed-integer optimization framework for de novo peptide identification. *AIChE J.* **2007**, *53*, 160.
[31] J. Zhang, L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, W. Zhang, Z. Zhang, G. A. Lajoie, B. Ma. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteomics* **2012**, *11*, M111.010587.
[32] K. Ning, N. Ye, H. W. Leong. On preprocessing and antisymmetry in de novo peptide sequencing: improving efficiency and accuracy. *J. Bioinform. Comput. Biol.* **2008**, *6*, 467.
[33] X. Liu, L. J. Dekker, S. Wu, M. M. Vanduijn, T. M. Luider, N. Tolic, Q. Kou, M. Dvorkin, S. Alexandrova, K. Vyatkina, L. Paša-Tolić, P. A. Pevzner. De novo protein sequencing by combining top-down and bottom-up tandem mass spectra. *J. Proteome Res.* **2014**, *13*, 3241.
[34] T. Keough, M. P. Lacey, R. S. Youngquist. 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.
[35] P. Conrotto, U. Hellman. Lys Tag: an easy and robust chemical modification for improved de novo sequencing with a matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1823.
[36] J. S. Kim, M. Shin, J. S. Song, S. An, H. J. Kim. C-terminal de novo sequencing of peptides using oxazolone-based derivatization with bromine signature. *Anal. Biochem.* **2011**, *419*, 211.
[37] S. Warwood, S. Mohammed, I. M. Cristea, C. Evans, A. D. Whetton, S. J. Gaskell. Guanidination chemistry for qualitative and quantitative proteomics. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3245.
[38] D. M. Horn, R. A. Zubarev, F. W. McLafferty. Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10313.
[39] M. M. Savitski, M. L. Nielsen, F. Kjeldsen, R. A. Zubarev. Proteomics-grade de novo sequencing approach. *J. Proteome Res.* **2005**, *4*, 2348.
[40] A. Guthals, K. R. Clauser, A. M. Frank, N. Bandeira. Sequencing-grade de novo analysis of MS/MS triplets (CID/HCD/ETD) from overlapping peptides. *J. Proteome Res.* **2013**, *12*, 2846.
[41] Y. O. Tsybin, H. He, M. R. Emmett, C. L. Hendrickson, A. G. Marshall. Ion activation in electron capture dissociation to distinguish between N-terminal and C-terminal product ions. *Anal. Chem.* **2007**, *79*, 7596.
[42] A. G. Marshall, C. L. Hendrickson, G. S. Jackson. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **1998**, *17*, 1.
[43] F. Xian, C. L. Hendrickson, A. G. Marshall. High resolution mass spectrometry. *Anal. Chem.* **2012**, *84*, 708.
[44] S. S. Nair, C. L. Nilsson, M. R. Emmett, T. M. Schaub, K. H. Gowd, S. S. Thakur, K. S. Krishnan, P. Balaram, A. G. Marshall. De novo sequencing and disulfide mapping of a bromotryptophan-containing conotoxin by Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **2006**, *78*, 8082.
[45] S. L. Hubler, A. Jue, J. Keith, G. C. McAlister, G. Craciun, J. J. Coon. Valence parity renders z(*)-type ions chemically distinct. *J. Am. Chem. Soc.* **2008**, *130*, 6388.
[46] Y. Mao, J. D. Tipton, G. T. Blakney, C. L. Hendrickson, A. G. Marshall. Valence parity to distinguish c′ and z* ions from electron capture dissociation/electron transfer dissociation of peptides: effects of isomers, isobars, and proteolysis specificity. *Anal. Chem.* **2011**, *83*, 8024.
[47] T. M. Schaub, C. L. Hendrickson, S. Horning, J. P. Quinn, M. W. Senko, A. G. Marshall. High-performance mass spectrometry: Fourier transform ion cyclotron resonance at 14.5 Tesla. *Anal. Chem.* **2008**, *80*, 3985.

[48] G. T. Blakney, C. L. Hendrickson, A. G. Marshall. Predator data station: A fast data acquisition system for advanced FT-ICR MS experiments. *Int. J. Mass Spectrom.* **2011**, *306*, 246.
[49] M. W. Senko, C. L. Hendrickson, M. R. Emmett, S.-D. Shi, A. G. Marshall. External accumulation of ions for enhanced electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 970.
[50] F. Xian, C. L. Hendrickson, G. T. Blakney, S. C. Beu, A. G. Marshall. Automated broadband phase correction of Fourier transform ion cyclotron resonance mass spectra. *Anal. Chem.* **2010**, *82*, 8807.
[51] E. B. Ledford Jr, D. L. Rempel, M. L. Gross. Space charge effects in Fourier transform mass spectrometry. Mass calibration. *Anal. Chem.* **1984**, *56*, 2744.
[52] S. D. Shi, J. J. Drader, M. A. Freitas, C. L. Hendrickson, A. G. Marshall. Comparison and interconversion of the two most common frequency-to-mass calibration functions for Fourier transform ion cyclotron resonance mass spectrometry. *Int. J. Mass Spectrom.* **2000**, *195/196*, 591.
[53] B. E. Wilcox, C. L. Hendrickson, A. G. Marshall. Improved ion extraction from a linear octopole ion trap: SIMION analysis and experimental demonstration. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 1304.
[54] N. K. Kaiser, J. J. Savory, A. M. McKenna, J. P. Quinn, C. L. Hendrickson, A. G. Marshall. Electrically compensated Fourier transform ion cyclotron resonance cell for complex mixture mass analysis. *Anal. Chem.* **2011**, *83*, 6907.
[55] Available: http://pubs.acs.org.

