Anal. Chem. **2008,** *80,* 4825–4835

Comprehensive Comparison of Collision Induced Dissociation and Electron Transfer Dissociation

Henrik Molina,*,† Rune Matthiesen,‡ Kumaran Kandasamy,†,§ and Akhilesh Pandey*,†

McKusick-Nathans Institute of Genetic Medicine and Departments of Biological Chemistry, Pathology and Oncology, The Johns Hopkins University, Baltimore, Maryland 21205, CIC bioGUNE, Parque Tecnológico de Bizkaia Edificio 801 A, 48160 Derio, Spain, and Institute of Bioinformatics, International Tech Park, Bangalore 560 066, India

Electron transfer dissociation (ETD) is a recently introduced mass spectrometric technique which has proven to be an excellent tool for the elucidation of labile post-translational modifications such as phosphorylation and O-GlcNAcylation of serine and threonine residues. However, unlike collision induced dissociation (CID), which has been studied for decades, the intricacies of ETD-based fragmentation have not yet been firmly established or systematically addressed. In this analysis, we have systematically compared the CID and ETD fragmentation patterns for the large majority of the peptides that do not contain such labile modifications. Using a standard 48 protein mix, we were able to measure false-positive rates for the experiments and also assess a large number of peptides for a detailed comparison of CID and ETD fragmentation pattern. Analysis of ~19 000 peptides derived from both standard proteins and complex protein samples revealed that (i) CID identified 50% more peptides than ETD; (ii) ETD resulted in ~20% increase in amino acid sequence coverage over CID; and (iii) combining CID and ETD fragmentation increased the sequence coverage for an average tryptic peptide to 92%. Interestingly, our analysis revealed that nearly 60% of all ETD-identified peptides carried two positive charges, which is in sharp contrast to what has been generally accepted. We also present a novel strategy for automatic validation of peptide assignments based on identification of a peptide by consecutive CID and ETD fragmentation in an alternating mode.

The most common approach to tackle the identification of proteins is the bottom-up approach in which proteins are digested with specific enzymes followed by collision induced dissociation (CID) of the resulting peptides. The collision of peptides with gas molecules results in a vibrational excitement which gives rise to dissociation of the peptide backbone amide bonds, between the carbonyl and the amine groups. The fragment ions that are generated are termed b-ions if they originate from the N-terminal part of the peptide and y-ions if they contain the C-terminal part of the peptide.[1] In electron transfer dissociation (ETD), the fragmentation is induced by converting the positively charged peptides into radicals in an electron transfer reaction using radical anions often generated from fluoranthene.[2] The resulting peptide cation radicals are unstable and typically undergo fragmentation with a dissociation of bonds between amines and α-carbons in the peptide backbone. These fragment ions are termed c- and z-ions when containing the N-terminus and the C-terminus of the peptide, respectively.[1] Another fragmentation technique, electron capture dissociation (ECD),[3] also gives rise to c- and z-ions. ETD has primarily been used in the elucidation of serine/threonine phosphorylations and O-GlcNAcylation since ETD has proven to be an excellent tool for the identification of these labile post-translational modifications.[4–12]

From the fragmentation pattern of both CID and ETD spectra, the experimentalist can either read a partial or complete amino acid sequence or use automated search algorithms that match the fragment ions to a database of in silico fragmented peptides. The latter has become the preferred approach because today's instrumentation coupled with global proteomic approaches are capable of generating thousands of MS/MS spectra in a single analysis making manual interpretation an impractical task and automation/streamlining a necessity. To evaluate and validate results generated by automated analysis of fragmented peptides,

* To whom correspondence should be addressed.
† The Johns Hopkins University.
‡ CIC bioGUNE.
§ Institute of Bioinformatics.

(1) Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*, 601.
(2) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528–9533.
(3) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266.
(4) Zhang, Q.; Tang, N.; Brock, J. W.; Mottaz, H. M.; Ames, J. M.; Baynes, J. W.; Smith, R. D.; Metz, T. O. *J. Proteome Res.* **2007**, *6*, 2323–2330.
(5) O'Connor, P. B.; Cournoyer, J. J.; Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 15–19.
(6) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2199–2204.
(7) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E.; Shabanowitz, J.; Hunt, D. F. *Biochim. Biophys. Acta* **2006**, *1764*, 1811–1822.
(8) Hogan, J. M.; Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *J. Proteome Res.* **2005**, *4*, 628–632.
(9) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. *Mol. Cell. Proteomics* **2007**, *6*, 1942–1951.
(10) Coon, J. J.; Ueberheide, B.; Syka, J. E.; Dryhurst, D. D.; Ausio, J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9463–9468.
(11) Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2193–2198.
(12) Khidekel, N.; Ficarro, S. B.; Clark, P. M.; Bryan, M. C.; Swaney, D. L.; Rexach, J. E.; Sun, Y. E.; Coon, J. J.; Peters, E. C.; Hsieh-Wilson, L. C. *Nat. Chem. Biol.* **2007**, *3*, 339–348.

10.1021/ac8007785 CCC: $40.75 © 2008 American Chemical Society
Published on Web 06/10/2008

empirically as well as statistically derived criteria plays a vital role (see review by Nesvizhskii *et al.*[13]). CID has been used for the identification of peptides for more than two decades, and search engine algorithms, interpretation, as well as validation methodologies have been refined for this type of data. On the other hand, ETD is a relatively new technique where data analysis methodologies have not yet been established. To date, only a handful of studies have been of a global nature, i.e., generating more than a thousand peptides identified by ETD.

The aim of our study was to investigate if subjecting peptide precursors to both CID and ETD experiments (alternating CID/ETD) is a viable approach for the analysis of nonmodified peptides. To test the idea of an alternating CID/ETD approach, we first needed to learn the consequences of subjecting each precursor to two fragmentation events. Alternating CID/ETD, compared to a CID-only approach, can halve the number of different precursors fragmented in LC−MS/MS studies of complex mixtures. The evaluation of such an approach also requires a detailed analysis of CID and ETD identified peptides which include studying the properties of the peptides identified by the two methods, their fragmentation characteristics, and the complementary nature of CID and ETD. In a comparison like this, it is important that the characteristics of the test sample are well defined. The best defined sample that we had available is the Universal Proteomics Standard protein mixture containing 48 known human proteins (UPS1, Sigma). To increase the number of peptides in our analysis, we also subjected a number of complex protein samples to alternating CID/ETD and analyzed those data using parameters derived from the analysis of the 48 known human proteins.

The use of two consecutive fragmentation experiments has previously been suggested by Olsen and Mann[14] and Nielsen *et al.*[15] Olsen and Mann used the second stage in a two stage CID experiment to increase the confidence of identified peptides, whereas Nielsen *et al.* identified "golden complementary pairs" from alternating CID/ECD experiments to create high-quality fragment ion peak lists for faster database searching. Our alternating fragmentation approach differs by using ETD which is today commercially available on ion trap mass spectrometers of significantly lower cost than the FT equipped ECD instruments. An additional important difference is that in our approach CID and ETD data are searched independent of each other which allows us to use the search results for internal validation.

MATERIALS AND METHODS

Universal Proteomics Standard mix (UPS1, Sigma) containing 48 known proteins, was separated on an SDS-PAGE gel or digested in-solution. For digestion of SDS-PAGE separated proteins, we followed a standard published protocol.[16] In short, protein bands were excised, reduced and alkylated, and subjected to trypsin (Promega, Madison, WI) digestion. The resulting peptides were extracted, dried down (SpeedVac, Eppendorf), and redissolved in 20 μL of 3% acetonitrile/0.1% formic acid. For in-solution digestions, we followed a published trifluoroethanol based protocol.[17] In short: proteins were denaturized in 50% trifluoroethanol (Sigma-Aldrich), followed by reduction and alkylation. Prior to digestion, the trifluoroethanol concentration was diluted to 3% with 100 mM ammoniumbicarbonate (Sigma). Trypsination was performed using a trypsin to protein ratio of 1:20. Peptides different from the Universal Proteomics Standard mix were either generated by the SDS-PAGE approach (pancreas sample and immunoprecipitated proteins) or by in-solution digestion (*Escherichia coli* lysate from BioRad, Hercules, CA, and pancreas sample) using the above-described and referenced protocols. The additional in-solution generated peptides were further separated by strong cation exchange chromatography (PolyLC, Columbia, MD) into 40 fractions.

LC−MS/MS Analysis. All LC−MS/MS analyses were performed using an ETD equipped Agilent 1100 series HPLC-Chip/MS system (Agilent 6340, Agilent Technologies, CA), a Paul type ion trap connected to pumps/autosampler using a chip integrated microreversed-phase column/emitter. An 8 min (short) or 15 min (longer) gradient (10% solvent B to 45% solvent B) was run at 300 nL/min (solvent A, 0.1% formic acid in water; solvent B, 90% acetonitrile with 0.1% formic acid). Samples were loaded onto a precolumn using 3% acetonitrile/0.1% formic acid at 4 μL/min. For each of the analyses, 0.5 μL were injected, corresponding to a maximum amount of 125−500 fmol per protein. Peptides were analyzed by either CID-only or alternating CID/ETD. The numbers of microscans used were 2. When operating in ETD mode, Smart Decomposition was activated (Auto) and accumulation time set at 40 ms. Smart Decomposition (Auto) introduces a low vibrational energy into doubly charged peptide after the electron transfer reaction has taken place. This low vibrational energy activation has been shown to improve the quality of the ETD spectra of doubly charged peptides.[18] It is believed that for this group of peptides, the electrostatic forces are too low to result in dissociation but that the low vibrational energy helps this process. For fragmentation experiments, no charge states were favored except that singly charged peptides were excluded.

MS/MS Database Searching. MS/MS data was analyzed using Spectrum Mill Proteomics Workbench version A.03.03.080 (Agilent Technologies, Santa Clara, CA). We used Spectrum Mill because this software is integrated with our MS platform and has proven to perform well when searching ETD data.[6] CID and ETD data were extracted using default extraction parameters and searched against human or *E. coli* subsets of the RefSeq database (NCBI, March 2007). Searches were performed allowing for three missed cleavages with complete proteolytic specificity (trypsin), ±1.5 Da for the precursor mass, ±0.5 Da for the fragment masses, 40% minimum scored peak intensity, 4+ as the maximum ambiguous charge state for the spectra with precursors of unassigned charge state, variable N-terminal modification of glutamine (pyro-Glu), and variable oxidized methionine. A maximum precursor mass shift range (due to modifications) of −18 Da and +130 Da were used to accommodate pyro-Glu and oxidation of methionine. All data were simultaneously searched against a reversed database.

(13) Nesvizhskii, A. I.; Vitek, O.; Aebersold, R. *Nat. Methods* **2007**, *4*, 787–797.
(14) Olsen, J. V.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13417–13422.
(15) Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. *Mol. Cell. Proteomics* **2005**, *4*, 835–845.
(16) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850–858.
(17) Horth, P.; Miller, C. A.; Preckel, T.; Wenz, C. *Mol. Cell. Proteomics* **2006**, *5*, 1968–1974.
(18) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E.; Coon, J. J. *Anal. Chem.* **2007**, *79*, 477–485.

For CID data, proton mobility scoring[19] was used. The retrieved peptides were validated using either default Spectrum Mill autovalidation criteria or an empirically derived set of criteria. The autovalidation of a peptide match is a validation system based on four parameters in which all thresholds must be fulfilled. The parameters are (1) the assigned score of a peptide match, (2) the score difference between a forward and a reverse database search, (3) the score difference between the best and the second best peptide matched to a spectrum, and (4) a scored peak intensity percent, which is the fraction of a spectrum that can be assigned to the matched peptide. These thresholds serve the same function as when filtering, e.g., Mascot results using scores and *E*-values or Sequest results using delta correlation and cross-correlation values. As for Sequest, the Spectrum Mill thresholds are charge state dependent. The above-described thresholds are listed in Supporting Information Table 2.

Scripts. A PERL script was created to compare the identified peptides to the sequences of the 48 standard proteins and commonly observed keratins. Another PERL script was created to extract consecutive pairs of CID and ETD spectra matching the exact same peptide from Spectrum Mill result files. The script reads the data in the tagSummary.tsv file generated by Spectrum Mill and makes a list of all consecutive pairs of CID and ETD spectra. The Spectrum Mill output files (.spo files in the "result-_mstag" folder) for all the pairs are then analyzed for the peptide information. During this step, the fragmentation information from the CID and ETD spectra are merged in such a way that it accommodates maximum information from the two spectra. This is presented as a combined peptide sequence coverage map for consecutive CID/ETD pairs. All the scripts are available for use upon request.

RESULTS AND DISCUSSION

The aim of our study was to compare CID and ETD as alternative fragmentation methods and to determine the feasibility of an alternating CID/ETD strategy in a proteomics shotgun approach. However, before we could initiate our analysis, we needed to test whether an alternating CID/ETD approach would result in too many missed identifications due to a lower number of different peptides selected for fragmentation. If too many peptides are missed compared to a standard CID-only approach, alternating CID/ETD might be less meaningful. Further, in a comparison of CID and ETD, we rely on the number of peptides that we validate and must therefore ensure that the criteria we use to validate the CID part are comparable to the criteria used for the ETD part of the data. In the following sections these two issues are addressed.

Comparing Alternating CID/ETD with CID-Only in an LC−MS/MS Experiment. To establish the effect of alternating CID/ETD on the number of peptide identifications, we first compared this approach to typical CID-only based LC−MS/MS analysis. The test samples were a standard protein mix (UPS1, Sigma) either digested by trypsin in-solution or separated by SDS-PAGE and then digested by trypsin. The in-solution digest resembles the complexity of fractions resulting from experiments such as SCX separation of peptides or reversed phase separation of proteins. The amounts used for the tests were 125 (in-gel samples) or 200 fmol (in-solution digest sample) of each of the standard proteins. The MS methods used for the comparison was our standard one survey MS scan followed by up to six CID events and one survey MS scan followed by up to three consecutive CID/ETD events. Both approaches resulted in approximately 600 validated peptide identifications. However, taking peptide redundancy into account, the CID-only approach out-performed the alternating CID/ETD approach by 7−17%. Though the number of unique peptide ions identified was higher for the CID-only approach, we reasoned that the difference is small enough that alternating CID/ETD can be pursued as a valid alternative to a CID-only approach. The data from this comparison are summarized in Supporting Information Table 1.

Optimizing Validation Criteria for ETD Identification. To ensure that the criteria used to validate CID and ETD data are comparable, we measured false-positive and false-negative rates for the two fragmentation techniques. Measuring these rates is easy when the sample is known and knowledge of these rates is important to ensure that the validation tool does not introduce a bias. Combining of all alternating CID/ETD analyses of the standard protein mixture (35 000 spectra) and using default Spectrum Mill validation criteria, a total of 5 001 peptides were identified (3 235 from CID and 1 766 from ETD). For the CID-identified peptides, we measured a false-positive rate of 0.8% versus 0% for ETD. The differences in false-positive rates prompted us to adjust the validation criteria for the ETD part of the experiment to reach a similar false-positive rate as for the CID part. Since four parameters are used to validate a peptide match, it was necessary to establish which of these four required adjustment: (1) score, (2) difference between forward and reverse database score, (3) score difference between best match and second best match, and (4) the scored peak intensity percent (the fraction of a MS/MS signal that can be explained by the matched peptide). We extracted all spectra that were matched to the standard proteins regardless of scores and examined the above parameters. A total of 4 236 CID spectra and 3 152 ETD spectra could be matched to the standard proteins. For both data sets, approximately 60% had been matched as doubly charge peptides (Figure 1A,B).

When we compared the score distributions for CID and ETD matched peptides (Figure 1C), it became clear that the doubly charged peptides matched by ETD, in particular, scored much lower on average (∼7) than doubly charged peptide ions in CID experiments (∼14). Differences were also observed for 3+ peptides but were less pronounced (ETD, ∼11 versus CID, ∼13). The score difference between a forward and reverse search for the CID and ETD matched peptides showed fewer differences. Here, 3+ ETD matched peptides together with 2+ and 3+ CID peptides showed very similar distributions where the 2+ ETD matched peptides differed with a tighter distribution (Figure 1D). Also the distributions of score differences between best and second best match was very similar, although again, the ETD matched doubly charged peptides deviated by a tighter distribution and smaller differences (Figure 1E). For scored peak intensity percent, ETD matched peptides (2+ and 3+) and the 3+ CID matched peptides showed similar distributions while many of the doubly charged CID matched peptides had a lower value (Figure

(19) Kapp, E. A.; Schutz, F.; Reid, G. E.; Eddes, J. S.; Moritz, R. L.; O'Hair, R. A.; Speed, T. P.; Simpson, R. J. *Anal. Chem.* **2003**, *75*, 6251–6264.

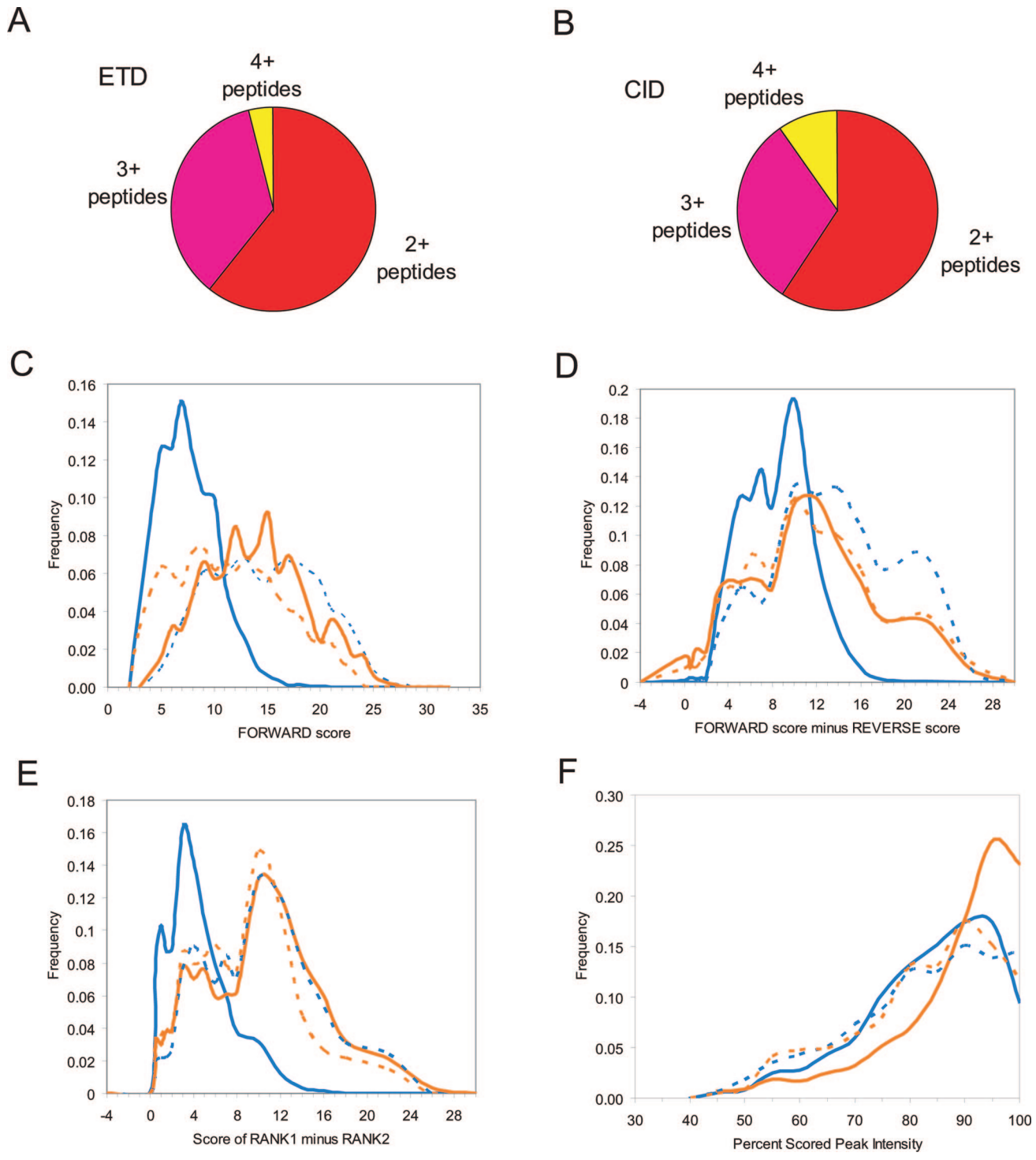

Figure 1. Analysis of extracted standard peptides fragmented by CID and ETD. Extraction was performed without any validation. Spectra that could be matched to the standard proteins (Universal Proteomics Standard, UPS1) were analyzed with respect to charge state distribution (A and B). These data were also analyzed with respect to (C) score; (D) score difference between a normal search and a reverse search; (E) score difference between best match and second best match, and (F) the percent of assigned fragmentation signal − scored peak intensity. In the plots, ETD distributions are shown in blue and CID distributions in orange. Distributions for doubly charge peptides are marked in a bold full line where distributions for triply charged peptides are truncated.

1F). It was evident from this analysis that the score threshold for doubly charged ETD matched peptides needed to be adjusted (toward a lower value) for us to perform a less biased comparison of CID and ETD. As is clear from the charge state distributions in Figure 1A and B, this is especially important because a large portion of doubly charged peptides had been subjected to fragmentation. The reason for the dramatically lower score of doubly charged ETD peptides could be explained by the nature of ETD and the scoring algorithm of Spectrum Mill. Because a doubly charged peptide is charge reduced in the electron transfer process, it cannot dissociate into both a C- and an N-terminal fragment ion since only one charge is available. For 2+ peptide

Table 1. Using Known Peptides for the Optimization of Validation Criteria[a]

	validated peptide ions			
fragmentation mode	no. of peptides validated	false positives	false positive rate (%)	false negative rate[d] (%)
alternating CID/ETD[b]	5001	26	0.520	10.303
CID[b]	3235	26	0.804	7.606
ETD[b]	1766	0	0.000	13.849
alternating CID/ETD[c]	5237	47	0.879	8.888
ETD[c]	2161	21	0.972	10.641

[a] The table contains the number of peptides identified and the respective measured false positive and false negative rates. [b] Default Spectrum Mill validation criteria. [c] Empirical derived ETD validation criteria (see Supporting Information Table 2) using known peptides. [d] Percentages of not-validated peptides that could be matched to the standard proteins.

ions identified by CID, this is possible in a charge separation fragmentation and the scores will therefore have the potential of being twice the score of the corresponding ETD peptide. Since the number of matched fragments ions is the main contributor to the score assigned by Spectrum Mill, doubly charged ETD-generated peptides will be assigned lower scores. We also lowered the scored peak intensity percent thresholds for the ETD validation. Our empirical derived thresholds (listed in Supporting Information Table 2) were used in a new validation of the ETD data, which resulted in 22% additional identifications with a false positive rate of <1%, very similar to the CID data. These results, before and after the adjustments of the ETD validation criteria, are summarized in Table 1.

Comparison of ETD and CID with Alternating CID/ETD. After optimization of ETD validation parameters as described above, it was now possible to perform a less biased comparison of ETD and CID experiments. To gain a better statistical basis for this comparison, we subjected additional samples to the alternating CID/ETD approach. The samples were tryptic digests of immunoprecipitated proteins separated by SDS-PAGE 1D-gel, *E. coli* lysates and proteins from pancreatic juice. The analysis of these samples resulted in a total of 18 999 validated peptide identifications (11 591 by CID and 7 408 by ETD). Defining a unique peptide ion based on its sequence, charge state, and modified residue(s), CID identified 3 518 unique peptide ions while ETD identified 2 235. A total of 1 890 unique peptides were identified by both fragmentation techniques.

Physical Properties of the Peptide (Size and Charge State). An analysis of physical properties of the identified peptides revealed that the CID and ETD peptides were very similar with regard to size. On average, a CID identified peptide was composed of 14.4 amino acids where the number was 14.5 for an average ETD peptide. However, as seen from the amino acid length distribution of the CID and ETD peptides (Figure 2A), there is a small trend toward ETD being favorable for peptides containing >13 amino acids. The data, that were searched allowing for two, three, and four positive charges, showed that 57% of the ETD identified peptides carried two positive charges compared to 65% for the CID data set. For triply charged peptides, the numbers for the CID and ETD identifications were 28% and 39%, respectively (parts C and D of Figure 2). The ETD charge state distribution was consistent with our results from the analysis of the standard protein mix (see Figure 1A).

Fragmentation. In any tandem mass spectrometry experiment, the nature of fragmentation itself is the vital component. One way to assess the quality of peptide identification is to calculate how much of the MS/MS signal that can be explained by the matched peptide (the scored peak intensity). Another method is to calculate the percentage of peptide sequence coverage. We argue that the latter is a valid measure since the purpose of a fragmentation experiment is to ultimately provide a peptide's amino acid sequence coverage. With calculation of the sequence coverage for an average peptide identified by either CID or ETD, the average ETD-identified peptide had 22% higher sequence coverage (67% for CID versus 82% for ETD). The sequence coverage distributions for ETD and CID are shown in Figure 2B. We also observed that generally the ETD identified peptides had a much lower frequency of possessing complementary C- and N-terminal fragment ions at the same location in the sequence. Only 21% of cleavage sites produced complementary fragments for ETD as compared to 56% for CID. However, this is explained by the large number of doubly charged peptides in this analysis that become singly charged during the electron transfer process. With only one charge available, only one fragment ion is possible in contrast to the doubly charged peptides subjected to CID (see previous section). Repeating this calculation for ETD identified peptides with charge states greater than 2+ showed that half of all dissociations were matched by both z- and c-type ions, very similar to CID. In a recent study,[9] a correlation between percent fragmentation (number of observed fragment ions/number of theoretical fragment ions) and mass-to-charge ratio was reported. We were curious to see if our data set would show the same correlation. However, because of the large number of doubly charged peptides, we decided to use percent sequence coverage since this measurement is less biased toward the "one-fragment-ion-per-ETD-event" for doubly charged peptides. The overall picture of these plots (Figure 2E,F) showed (1) centers of gravity are observed at a mass-to-charge ratio of 600−700 for both CID and ETD; (2) ETD handles higher mass-to-charge ratios better than CID; and (3) higher sequence coverage is associated with ETD identified peptides. Dividing the data into the charge states 2+ (red), 3+ (violet), and 4+ (yellow) peptides, we observed for both CID and ETD a decrease in sequence coverage as a function of increasing m/z values, but only for the peptides with charge state 3+ or 4+. On the contrary, 2+ peptides, especially for ETD, appear to be independent of mass-to-charge (in the m/z range measured). Scatter plots (data not shown) with percent fragmentation in place of sequence coverage were very similar to the plots in parts E and F of Figure 2. The only exception was that the percent fragmentation for 2+ ETD identified peptides were shifted toward lower values, compared to peptides with charge states greater than 2+.

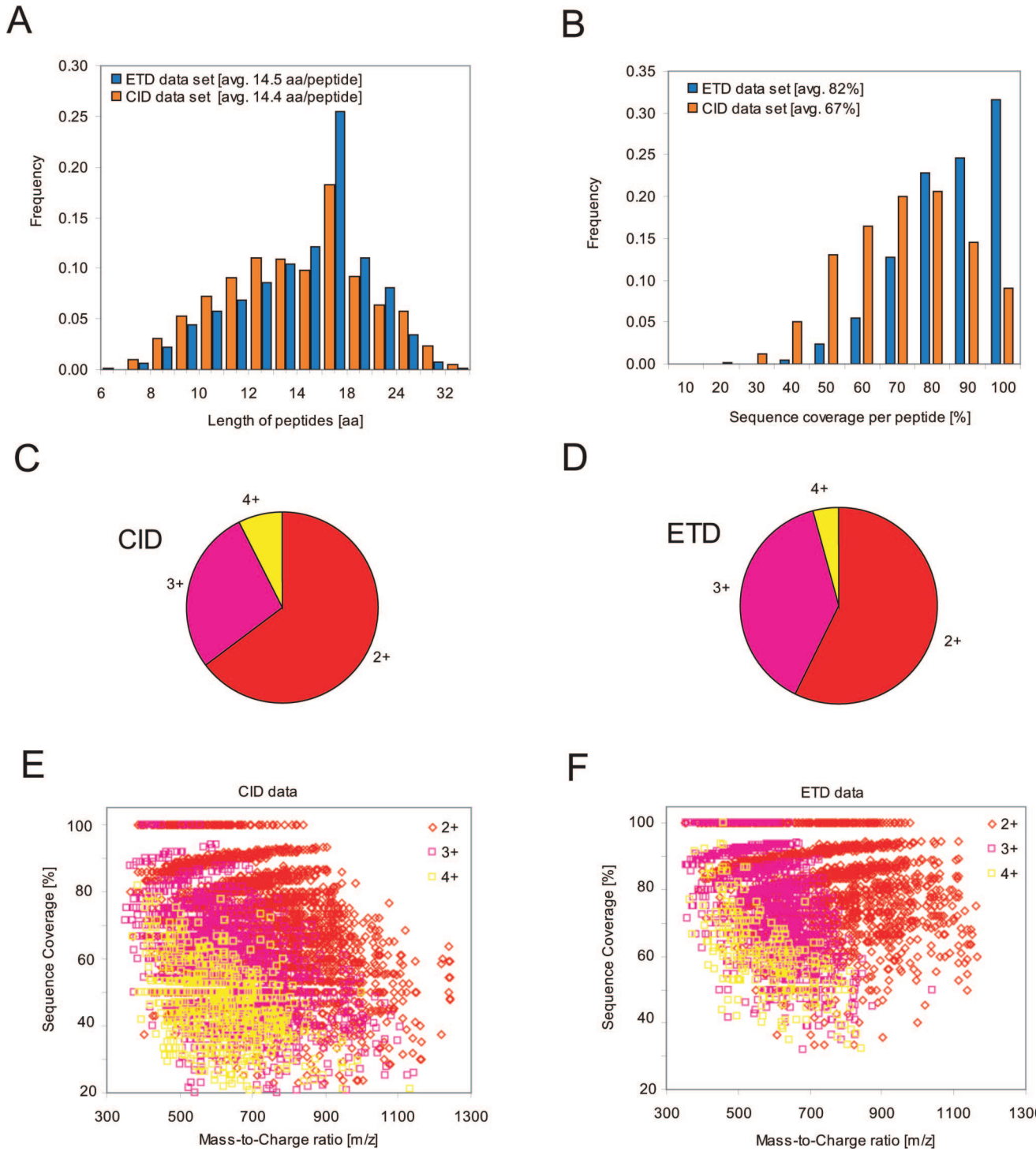

Figure 2. Analysis of a large alternating CID/ETD data set. The CID data were validated using default Spectrum Mill validation criteria whereas the ETD data were validated using a threshold specific for ETD. (A) Peptide length distribution of CID and ETD identified peptides, (B) the distribution of percent sequence coverage for CID and ETD identified peptides. ETD distribution is marked in blue, CID in orange. (C and D) Charge state distributions for CID (C) and ETD (D) identified peptides. (E and F) Scatter plots of sequence coverage for CID (E) and ETD (F) identified peptides as a function for mass-to-charge ratio. Color code: red, 2+ peptides; violet, 3+ peptides; and yellow, 4+ peptides.

From our analysis of this large CID/ETD data set, the most pronounced differences between CID and ETD were that ETD identified fewer peptides but with better sequence coverage. This might seem counterintuitive but can be explained by the sum of (1) a more uniform fragmentation in ETD (observed here and previously reported) and (2) the fact that the generated ETD fragment ions generally have intensities that are 5−10 times lower than the corresponding CID generated fragments. Another important observation is that the CID and ETD identified and validated peptides do not differ greatly with respect to neither charge state nor size. In Table 2, key statistics from this large data set are summarized.

Table 2. Key Values Obtained from a Large Alternating CID/ETD Dataset

fragmentation technique	CID	ETD
number of identifications	11591	7408
average peptide length (number of amino acids)	14.5	14.4
charge state distribution (2+, 3+, 4+)	65%, 28%, 8%	57%, 38%, 4%
N-terminal fragment ions	60%	58%
C-terminal fragment ions	40%	42%
average number of fragments per amino acid	0.8	1.0
fragmentation percentages (number of observed fragments in respect to possible fragments)	49%	51%
amino acid pairs identified by both a N- and a C-terminal fragment ion	55%	21% (50%[a])
average sequence coverage	67%	82%

[a] For ETD identified peptide ions with charge states higher than 2+.

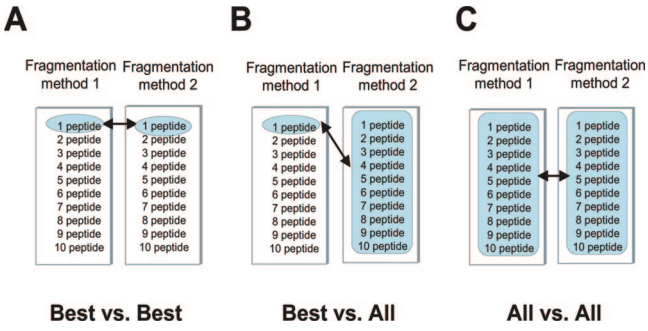

Figure 3. A schematic of the proposed and tested conditional consecutive CID/ETD validation. (A) The most stringent type of validation is to require that both CID and ETD matches the same peptide as the best match, (B) validation where at least one of the fragmentation methods has identified the peptide as being the best match. (C) The requirement for the least stringent validation is that a peptide is matched by both CID and ETD spectra irrespectively of ranking.

Validation Based on Peptides Identified by Both CID and ETD. Of the above 1 890 unique peptide ions identified by both CID and ETD, 93% were identified in consecutive CID/ETD experiments. We expected that the confidence of these identified peptide ions were high since they passed the Spectrum Mill validation criteria thresholds and were also identified by two different MS/MS techniques. This was indeed confirmed by a false positive rate of 0% calculated for the data that was associated with the standard protein mix (~1 500 peptides). If this was to be the criterion employed for reporting final results, the identifications would be highly reliable although many peptides would be missed (50% compared to the score based autovalidation alone). Since the above analysis showed that the physical properties of the CID and ETD identified tryptic peptides were relatively similar, we speculated if an alternating CID/ETD data set could be validated by a conditional consecutive CID/ETD validation alone. By conditional consecutive CID/ETD validation, we mean that a peptide is valid only if the particular peptide was identified from the exact same precursor by both CID and ETD but regardless of the score. By the term "exact same precursor", mean to that only the CID and ETD spectra obtained in the same MS-survey/fragmentation cycle are compared. In this scenario, the most stringent criterion for such validation would require that the identified peptide was assigned as the best match among all peptides matched to a given MS/MS spectral pair (Figure 3A). Less stringent criteria would include the possibility that one (Figure 3B) or both (Figure 3C) of the fragmentation methods resulted in a given peptide assignment but did not necessarily score that peptide as the best match. We considered these latter, less stringent criteria because most search engines provide several peptide matches to every spectrum along with their ranking, which is essentially based on how well they fit the spectrum. To test the conditional consecutive CID/ETD validation method, we extracted 10 841 CID/ETD MS/MS spectral pairs fulfilling one of the three criteria. Those 10 841 CID/ETD spectra were extracted from ~100 000 spectra that had led to ~70 000 peptide matches. A total of 9 446 of those (Supporting Information Table 3) were of the most stringent type meaning that a matched peptide was top ranked by both CID and ETD. More than $^1/_4$ (2 784) of these most stringent spectral pairs originated from the standard protein mix analysis which allowed us to calculate a false positive rate of 0.86% for these conditional consecutive CID/ETD validated peptide assignments. The less stringent validation (parts B and C of Figure 3) resulted in an 18% and 26% increase in assigned standard peptides, but the false-positive rate increased dramatically, by 10−100- fold to 6% and 10%, respectively. Thus, with respect to false positive rates, clearly only the most stringent criterion was deemed valid.

Where the Spectrum Mill score based autovalidation resulted in 3 855 unique peptide ions identified by CID, ETD, or both, conditional consecutive CID/ETD validations resulted in 2 867 unique peptide ions. Of the 2 867, only 197 unique peptides had not been validated by the Spectrum Mill approach. Although the conditional consecutive CID/ETD validation only identified a few additional peptides and missed many, it is important to note that this type of validation increased the number of autovalidated peptides identified by two orthogonal fragmentation techniques from 50% to 70%. The Venn diagram in Figure 4 depicts unique peptide ions with respect to numbers, overlaps, validation method, and calculated false-positive rates. As we had observed previously, peptides having passed both the Spectrum Mill validation and conditional consecutive CID/ETD validation are very confident identifications. In Figure 4, this group of peptides is encompassed with a bold line. Again, using the standard peptides, we were able to calculate a false positive rate for these peptides and the result was 0.11%, 8−10 times lower than both Spectrum Mill validation and conditional consecutive CID/ETD validation alone.

Clearly, identifying a peptide with two different fragmentation methods increases the accuracy of the identification. We have, in the previous section, used sequence coverage as a measure of the accuracy of a peptide identification. It is, therefore, natural to further test if merging the CID/ETD pairs will increase the sequence coverage of a peptide and thereby increase the quality of an identification. Although a peptide can be confidently

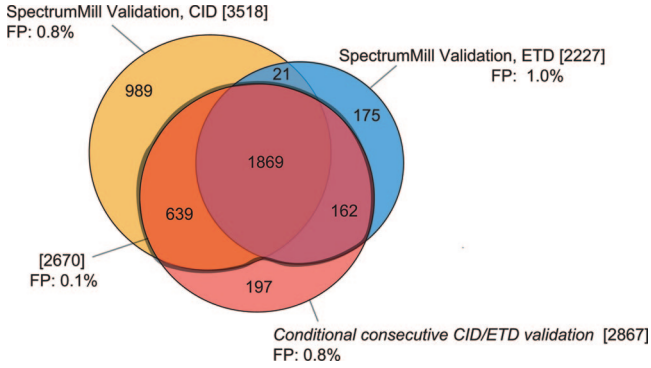

Figure 4. Venn diagram illustrating the overlap of unique peptide ions identified and validated in an alternating CID/ETD experiment. Orange area corresponds to Spectrum Mill validated peptides identified by CID where blue is the corresponding ETD peptides. Red area is peptides ions validated by *conditional consecutive CID/ETD validation*. In the diagram, the area encompassed by a bold Lare peptides that have been validated by the Spectrum Mill validation and *conditional consecutive CID/ETD validation*. False positive rates were calculated using the standard (UPS1) peptides contained in the data sets.

identified without high sequence coverage, this is still a desirable goal and is very valuable especially when the task is to pinpoint post-translational modified residues. In the conditional consecutive CID/ETD validation, we are comparing search results from CID and ETD spectra of the exact same precursor. A different way of using an alternating CID/ETD experiment is to combine the available fragmentation information offered by the two techniques for a peptide. We created a script to merge the fragment ions matched to the CID and ETD spectra for a given peptide and were able to increased sequence coverage for $^3/_4$ of all CID/ETD pairs. For the whole data set, this resulted in sequence coverage of 92% for an average peptide identified by consecutive CID and ETD. This number is to be compared to sequence coverages of 68% and 77%, respectively, for CID and ETD alone. The increase in sequence coverage is exemplified by the ETD and CID spectra pairs shown in Figure 5. Combining CID and ECD data has previously been described[20] and has proven valuable in *de novo* sequencing.[21,22]

Stringent Comparison of CID and ETD Identified Peptides. With a CID spectrum and an ETD spectrum of the exact same precursor, identified as the exact same peptide (conditional consecutive CID/ETD validation), it is possible to compare CID and ETD in the most direct manner. First, we wanted to verify our previous observation: ETD identified peptides scores lower than their corresponding CID peptides. For 2+ peptides, the average difference between the CID and ETD scores was 8.5, compared to 3.0 and 3.2 for 3+ and 4+ peptides, respectively. All average differences were in favor of the CID score and in accordance with what we measured for the standard peptides. Of the 9 446 CID/ETD pairs, 3 781 exhibited sequence coverage differences of ≥20% between the two techniques. As expected, for the majority of those CID/ETD pairs (80%), the ETD spectrum provided a better coverage than CID. Comparing the length of the peptides revealed that ETD resulted in a better sequence coverage when peptides were longer (an average length of 14.6 amino acids versus 12.3 for the CID favored peptides) as shown in Figure 6A. The charge state distribution for the peptides where ETD provided a better sequence coverage than CID were very similar to the overall distribution for all ETD identified peptides. However, for the peptides where CID resulted in higher sequence coverage than ETD, we counted that 90% of those peptides were doubly charged. The two distributions are shown in parts B (CID) and C (ETD) of Figure 6.

To assess the complementary nature of CID and ETD fragmentation, we compared their fragmentation patterns in relation to neighboring amino acids using the 9 446 pairs of spectra obtained in the conditional consecutive CID/ETD validation. This was done by extracting the signals of a-, b-, y-, (y − 17)-, and (y − 18)-ions (CID) and c-, c′-, z-, z′-, and z″-ions (ETD) for each peptide. A 21 × 21 matrix of average and normalized ion intensities for fragmentation between two amino acids A and B was created. Methionine oxidation and all natural occurring amino acids were considered. The matrices for CID and ETD data are calculated by the following formulas:

$$F_{\mathrm{AB,CID}} = \frac{I_{\mathrm{a}}^{x} + I_{\mathrm{b}}^{x} + I_{\mathrm{y}}^{n-x} + I_{\mathrm{y-17}}^{n-x} + I_{\mathrm{y-18}}^{n-x}}{I_{\mathrm{total}}} \quad (1)$$

$$F_{\mathrm{AB,ETD}} = \frac{I_{\mathrm{c}}^{x} + I_{\mathrm{c'}}^{x} + I_{\mathrm{z}}^{n-x} + I_{\mathrm{z'}}^{n-x} + I_{\mathrm{z''}}^{n-x}}{I_{\mathrm{total}}} \quad (2)$$

where x is the position in the sequence of the amino acids A and B. n is the length of the sequence, I is the intensity of the different ion types, and I_{total} is the sum of all observed intensity from every ion type generated from all possible positions of fragmentation. Because our analysis is conducted using the exact same peptides subjected to both CID and ETD, we chose to focus on the relative differences between the CID and ETD data sets by using the ratios of $F_{\mathrm{AB,ETD}}/F_{\mathrm{AB,CID}}$ (eqs 1 and 2). In a heatmap (Figure 7), where all combinations of amino acid pairs are shown, we plotted three states being either (i) no change defined as less than a 2-fold difference between CID and ETD (black), (ii) a 2-fold or greater difference in favor of ETD (red), and (iii) a 2-fold or greater difference in favor of CID (green). From the heatmap, it is clear that ETD fragmentation is more pronounced then CID fragmentation when amino acids are flanked by the basic residues: lysine (K), arginine (R), and histidine (H). In the heatmap, this is shown by red boxes both vertically and horizontally for the amino acids K, R, and H. Similar observation has been made earlier for ECD fragmentation of ubiquitin.[23] A possible explanation is that the positive charged residues facilitate the capture of the transferred electrons. Another striking pattern is the lack of fragmentation in ETD, compared to CID, when proline is C-terminal to any preceding residue (vertical green area for C-terminal proline). This observation has previously been reported for ECD.[24] Though less

(20) Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A. *Angew. Chem., Int. Ed.* **2006**, *45*, 5301–5303.
(21) Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Zubarev, R. A. *J. Proteome Res.* **2005**, *4*, 2348–2354.
(22) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10313–10317.
(23) Breuker, K.; Oh, H.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002**, *124*, 6407–6420.
(24) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 563–573.

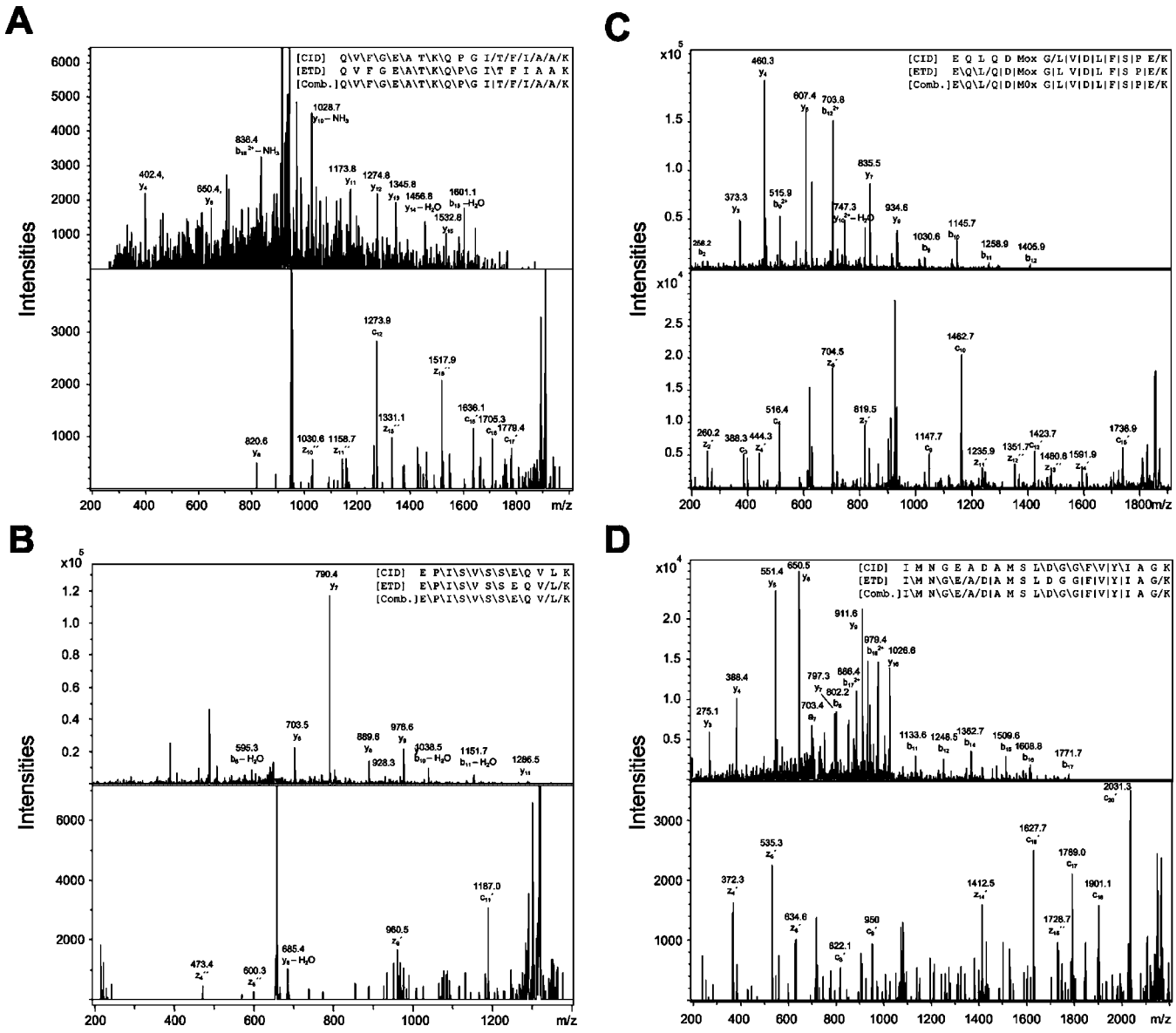

Figure 5. Four representative peptides identified by both CID (top panels) and ETD (lower panels) in an alternating CID/ETD experiment. Each set of spectra are recorded of the same precursor ion. The inserts in each figure (A−D) show the amino acid sequence coverage obtained from the CID experiments and the ETD experiments. The sequence maps marked [Comb.] are the combined CID/ETD sequence map. In all the four examples, the amino acid sequence coverage increases when combining the CID and ETD experiments. Only selected fragment ions are marked. Legends for the peptide sequence maps: (\) C-terminal fragment ion, (/) N-terminal fragment ion, and (|) both C- and N-terminal fragment ions.

prominent, the heat map also suggests that ETD fragmentation is less favorable than CID for amino acids flanked by the sulfur containing amino acids methionine and cysteine. It has been shown that the presence of alkylated cysteines and methionines in a peptide sequence lead to dominant side chain fragmentation.[25] Thus, it is likely that the sulfur containing side chain is favored over the adjacent backbone carbonyl site, inhibiting the fragmentation at these sites and leading to these dominant neutral losses.

CONCLUSIONS

In this report, we compared LC coupled alternating CID/ETD to a CID-only approach using standard peptides and the search engine Spectrum Mill. Similar numbers of peptides are identified by both approaches but the CID-only experiments resulted in approximately 12% more unique peptide identifications. Using a standard universal protein mix, we observed that Spectrum Mill scored ETD data lower than the corresponding CID data. Thus we established new validation criteria for ETD data. To gain a statistically sound basis for a comparison of CID and ETD, we included a number of medium complexity samples. In our study, more than half of all ETD identified peptide ions carried two positive charges. This is in sharp contrast to a recent study by Good et al.[9] where less than 1% of ETD identified peptides carried two positive charges. This difference might be because Good et al. measured peptides generated by both trypsin and Lys-C but can also be attributed to the design of the two types of instruments exemplified by different efficiencies in post-ETD low vibrational

(25) Chalkley, R. J.; Brinkworth, C. S.; Burlingame, A. L. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1271–1274.
(26) Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *Anal. Chem.* **2005**, *77*, 5662–5669.

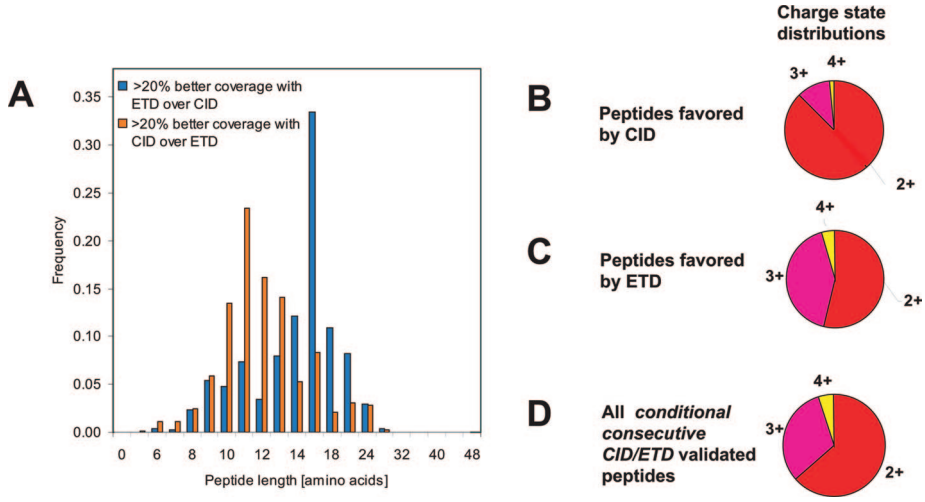

Figure 6. Histogram showing the distribution of peptide length of CID (orange) and ETD (blue) matched peptides. Only peptides for which the sequence coverage differed by 20% or more in-between the two fragmentation techniques are counted. The pie chart inserts show the charge state distribution of these two sets of peptides.

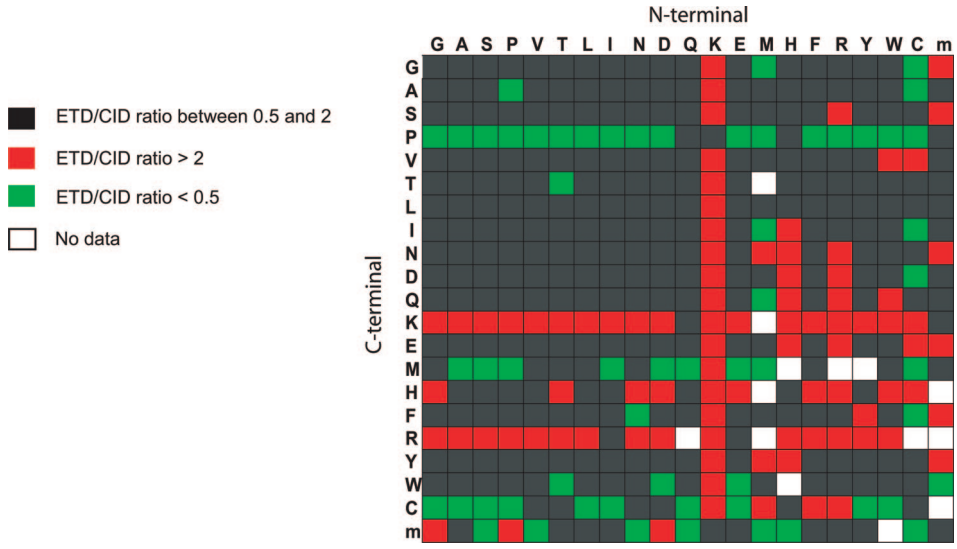

Figure 7. Heatmap showing averaged ETD over CID intensity ratios. Green indicates favorable ETD fragmentation over CID, while red indicates a favorable CID over ETD fragmentation. Black indicates differences of less than 2-fold. The heatmap is generated from the spectra of all the >9 000 peptides identified by consecutive CID/ETD. "*m*" indicates oxidized methionine. All cysteines, C, are alkylated. Similar diagrams, but divided into charge states (2+ and 3+) are shown in Supporting Information Figure 1.

energy collisions, also named ETcaD.[18] Post-ETD low-vibrational energy collisions have been shown to enhance the ETD fragmentation pattern of especially doubly charged peptides.[26] Yet, different explanations could be a bias in the ETD data extraction or in the used search engines (OMSSA versus Spectrum Mill). The charge state distribution for the CID identified peptides in our study showed a far better agreement with Good et al. results. Our comparison of CID and ETD showed that the main difference between the two techniques was higher sequence coverage for ETD identified peptides. Comparing the backbone fragmentation of CID and ETD with respect neighboring amino acids, we observed that ETD showed a favorable fragmentation for residues neighboring basic amino acids. Another clear pattern was a lower fragmentation frequency for ETD over CID with proline as a C-terminal neighbor to any residue. We also tested if peptides could be validated based on identification by both CID and ETD in an alternating CID/ETD experiment. This is indeed possible but the number of validated peptides is lower than when performing validation based on score, differences between forward and a reverse database searches, and best and second best peptide matches, together with the scored peak intensity percent. Validation based on complementary identification by both CID and ETD (conditional consecutive CID/ETD validation) only identified 3% additional peptides not already validated by the Spectrum Mill based validation. However, combining conditional consecutive CID/ETD validations and Spectrum Mill's validation tool allowed us to pinpoint 70% of the peptides that had been identified by two complementary fragmentation techniques. These peptides were remarkable in that the identification accuracy was 10-fold better than using scored based validation alone. Additionally, combining the fragment information for these peptides resulted in 92% sequence coverage for an average tryptic peptide.

Taking the (slightly) lower number of peptides identified by the alternating CID/ETD approach (compared to CID-only) and the

higher confidence and peptide sequence coverage into account the difference between these approaches can roughly be described as "quality versus quantity." A higher confidence in peptides independently identified by two orthogonal MS/MS techniques makes alternating CID/ETD a valid approach for samples of low to medium complexity or samples where a high number of so-called one-hit-wonders[27] are expected. In addition, the higher peptide sequence coverage obtained by combining CID and ETD search results is an important benefit when the aim of an analysis is to pinpoint post-translational modifications.

(27) Veenstra, T. D.; Conrads, T. P.; Issaq, H. J. *Electrophoresis* **2004**, *25*, 1278–1279.

ACKNOWLEDGMENT

This work was supported by Contract N01-HV-28180 from the National Heart Lung and Blood Institute and Grants U54RR020839 and R01CA106424 from the National Institutes of Health. We thank David M. Horn and Harsha Gowda for their valuable input.

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review April 19, 2008. Accepted May 23, 2008.

AC8007785