

Peptide identification by tandem mass spectrometry with alternate fragmentation modes

Adrian Guthals¹ and Nuno Bandeira^{1,2,*}

¹Department of Computer Science and Engineering, ²Skaggs School of Pharmacy and Pharmaceutical Sciences

University of California, San Diego

* Corresponding author: Nuno Bandeira, bandeira@ucsd.edu

Abstract

The high-throughput nature of proteomics mass spectrometry is enabled by a productive combination of data acquisition protocols and the computational tools used to interpret the resulting spectra. One of the key components in mainstream protocols is the generation of tandem mass (MS/MS) spectra by peptide fragmentation using collision induced dissociation (CID), the approach currently used in the large majority of proteomics experiments to routinely identify hundreds to thousands of proteins from single mass spectrometry runs. Complementary to these, alternative peptide fragmentation methods such as electron capture/transfer dissociation (ECD/ETD, respectively) and higher-energy collision dissociation (HCD) have consistently achieved significant improvements in the identification of certain classes of peptides, proteins and post-translational modifications. Recognizing these advantages, mass spectrometry instruments now conveniently support fine-tuned methods that automatically alternate between peptide fragmentation modes for either different types of peptides or for acquisition of multiple MS/MS spectra from each peptide. But while these developments have the potential to substantially improve peptide identification, their routine application requires corresponding adjustments to the software tools and procedures used for automated downstream processing. This review discusses the computational implications of alternative and alternate modes of MS/MS peptide fragmentation and addresses some practical aspects of using such protocols for identification of peptides and post-translational modifications.

Introduction

Technological and computational developments continue to expand the fundamental role of tandem mass spectrometry (MS^2) in high throughput proteomics [1]. Current protocols regularly identify thousands of proteins and post-translational modifications (PTMs [2]) per experiment and can deliver very high levels of reproducibility [3] (e.g. using targeted approaches). In the large majority of cases, these advances have been enabled by a combination of standard protocols based on trypsin digestion followed by protein identification from collision-induced dissociation (CID) MS^2 spectra using database search tools [4]. The efficiency and reliability of trypsin digestion remains one of the major reasons for the success of high-throughput proteomics. Protein digestion results in multiple peptides per protein and, in the limit, only one peptide needs to be significantly identified to be able to identify the corresponding protein [5, 6]. Trypsin further contributes to spectrum identifiability in CID MS^2 by cleaving C-term of K/R (Lysine/Arginine) and thus yielding rich and characteristic MS^2 peptide fragmentation patterns [7, 8] for a substantial subset of all peptides. Incorporating knowledge of these peptide fragmentation patterns into algorithms for MS^2 identification has led to a variety of database search software tools [9, 10] for peptide identification by scoring matches between acquired MS^2 spectra and predicted spectra for peptide sequences extracted from a protein sequence database.

Despite the success of the popular trypsin/CID tandem mass spectrometry protocols, it is well known that this approach generally biases experiments towards the identification of certain types of peptides, such as doubly- and triply-charged peptide precursors from medium-length peptides [11, 12, 13, 14, 15]. Historically this has been advantageous because CID generates identifiable spectra from these types of peptides (low charge, high m/z) with the greatest efficiency. But a disadvantage of relying upon a single specific protease like trypsin is that many digested peptides are too short and can thus lead to incomplete coverage of the proteome. In the case of trypsin, the enzyme cleaves K/R-rich regions into peptides that are too short for reliable identification (about half of tryptic yeast peptides are ~ 6 residues long [15]). Alternative

protease(s) can be utilized to obtain a wider distribution of peptides and increase coverage. For example, some have demonstrated the value of using of non-specific proteases [16, 17, 18, 19], but these can also decrease experimental reproducibility. Another limiting factor is that non-specific proteases can greatly increase the computational search space when matching MS² spectra to all possible peptides in the database. This not only increases the time required to search the database, but it can also decrease the sensitivity of spectrum identification at a given false discovery rate (FDR [6]). Others have had success using multiple specific proteases (e.g. LysC, AspN, GluC, and ArgC) in a more targeted approach [15, 20, 21]. Compared to the use of a single specific protease, utilizing multiple specific proteases allows for a greater percentage of the proteome (94% for yeast) to be covered by at least one peptide suitable for mass spectrometry sequencing technology [15]. If a separate MS² run is executed for each enzymatic digestion, there is little increase in computational complexity because MS² spectra from each run can be separately matched to a set of peptides cleaved by a single specific protease.

A disadvantage of using multiple proteases is that they often yield peptides that are longer and contain one or more internal basic residues, which are poorly fragmented by CID [22]. But alternative fragmentation strategies such as higher-energy collision dissociation (HCD [23], also formerly known as higher-energy C-trap dissociation) and electron transfer dissociation (ETD [24]) are known to improve identification of long, highly-charged peptides, peptides containing basic residues, and peptides containing many or highly-labile PTMs [25, 26, 27, 28]. A popular strategy has been to use Lys-C in combination with ETD because compared to trypsin, Lys-C generates a larger portion of peptides amenable to efficient ETD fragmentation [15]. The particular combination of ETD with Lys-N digestion yields peptide coverage complementary to trypsin and very simple fragmentation patterns that can aid in manual de novo peptide sequencing [29]. Although the reduction of C-terminal ions in Lys-N ETD peptides reduces the difficulty of manual sequencing, it can also hinder automated approaches that utilize symmetry between N- and C-terminal ions. The complementarity of CID, HCD, and ETD dissociation

strategies has been assessed in a variety of contexts [30, 31, 32, 33] and the recognition that each improves identifications of different types of peptides underlies the decision tree [34, 35, 15, 33] approach for real-time selection of fragmentation mode(s) based on each precursor's m/z and charge. Alternatively, CID/ETD [36] or CID/ETD/HCD [31] alternating MS^2 acquisition for every precursor have also been shown to substantially improve peptide identification and enable otherwise difficult analysis of PTMs [37]. Underlying the utility of all tandem mass spectrometry approaches to high-throughput proteomics is the need to calculate false discovery rates, usually estimated using the Target/Decoy approach (TDA [38, 39]). This approach is a key control of statistical significance in high-throughput proteomics but, as discussed below, requires careful adjustments before it can be meaningfully applied to the analysis of MS^2 data acquired with alternate peptide fragmentation modes.

Peptide fragmentation modes

While there are several comparisons of peptide fragmentation modes in terms of the resulting numbers of spectrum and peptide identifications, only some [40, 14, 41, 36, 42] attempt to characterize the observed differences in terms of their underlying MS^2 fragmentation statistics. These matter because the ability of a database search tool to identify MS^2 spectra is proportional to how well it models MS^2 fragmentation statistics and how it uses them to score Peptide Spectrum Matches (PSMs). The basic types of information captured in these peptide fragmentation models are illustrated here using spectra from a recent comparison [33] of how CID, HCD and ETD MS^2 acquisition modes affect Mascot [10] identifications. In brief, a HeLa tryptic digest was analyzed in three separate Thermo LTQ Orbitrap runs, one for each of CID, HCD and ETD MS^2 acquisition; survey scans were acquired in the Orbitrap with resolution 30000 and MS^2 spectra were acquired in the Orbitrap at resolution 7500. Peptide identifications were obtained here using MS-GFDB as previously described [36] and resulted in 17,378 CID PSMs (out of 33,586 spectra), 21,246 HCD PSMs (out of 37,810 spectra) and 12,834 ETD PSMs (out of

25,734 spectra). As shown in Table 1, all fragmentation modes achieved comparable spectrum identification rates of ~50% but also exhibit the expected correlations with precursor charge where ETD tends to perform better than CID or HCD on higher charge states. These show that ETD tends to achieve the highest identification rates for precursors of charge 3 or higher but yet also yields the lowest number of total identifications due to its slower scan rate (only $\sim \frac{2}{3}$ as many MS² spectra as in HCD acquisition). Longer acquisition time remains a disadvantage of ETD, however recent work suggests that optimizing ETD acquisition parameters can significantly reduce the scan rate without loss in coverage [43]. The complementarity of the fragmentation modes is further illustrated in Figure 1 and supported by the observation that most peptides are not identified by at least one acquisition mode (see supplementary table SM-2). This is partly due to the different scan rates and the stochastic nature of data-dependent MS² precursor selection but, as clearly illustrated for precursor charge 2 (for CID/HCD) and precursor charge 5 (for ETD), this also shows that different peptide fragmentation modes work best for different types of peptides. Of peptides that can be identified by all three acquisition modes, the *breaks* (observed cleavages along the peptide backbone, supported by either N- or C-terminal fragments) captured by ETD tend to complement those captured by CID and HCD, especially for precursors of charge 3 or higher. This can be seen in statistics from Figure 2, which show how the union of observed peptide breaks increases by 24-72% from CID/HCD to CID/HCD/ETD for precursors of charge 3 or higher. Supplementary table SM-1 details how many breaks are unique to every possible combination of fragmentation modes: ETD alone accounts for 19% of all possible peptide breaks in CID/HCD/ETD triplets, the intersection of breaks seen in CID and HCD accounts for 17%, and the intersection of breaks seen in CID, HCD, and ETD accounts for 30%.

The connection between the identifications in Table 1 and the underlying MS² peptide fragmentation statistics in Figure 2 is easily established for MS-GFDB PSM scoring models. As previously described [44, 36], MS-GFDB uses the exact same computational model to learn MS²

fragmentation statistics for CID, HCD and ETD and weigh different MS² fragment ion types based on their observed propensity in spectra from each fragmentation mode. The predominant CID, HCD and ETD ion types and their relative propensities are shown in Figure 2. As expected, the most prominent differences in peptide fragmentation pertain to the contrast between the dominant *b/y*-ions in CID/HCD spectra and *c/z*-ions in ETD spectra. These ions are the most important in peptide identification because of their direct indication of peptide breaks; Figure 2 compares the fractions of observed breaks on peptides identified on all three fragmentation modes. Not surprisingly, the per-precursor-charge fractions of observed breaks on CID, HCD and ETD show a very high correlation with the relative identification rates in Table 1 as higher fractions of observed breaks almost always result in higher rates of identified spectra.

The complementarity of different peptide fragmentation modes in yielding identifications for different classes of peptides was the underlying principle behind the Decision Tree (DT) acquisition mode [34, 33] where MS² acquisition modes are selected in real time based on precursor *m/z* and charge. Decision tree parameters are usually set to maximize the resulting number of peptide identifications and thus implicitly encode instrument-specific tradeoffs between scan rates and the expected rate of success in post-acquisition spectrum identification. Since the latter is highly dependent on the type of mass spectrometry experiment (e.g., shotgun proteomics [34] vs. phosphoproteomics [30, 45]) and on the success rates of the software tools chosen for spectrum identification, it is important to note that optimal decision tree parameters may vary between instrument models/methods and depend on the choice of software tools used for peptide identification.

Peptide identification

The complementarity of multiple peptide fragmentation modes can also be used to improve peptide identification by combining multiple spectra for the same peptides. One of the earliest such approaches was the utilization of MS/MS/MS (MS³) acquisition for de novo peptide

sequencing [46], later automated with heuristic [47] and optimal [48] algorithms. Similar applications to de novo peptide sequencing were also introduced at about the same time [49] for paired CID and Electron Capture Dissociation (ECD) spectrum acquisition from the same precursors, also later automated for CID/ECD pairs [50] and CID/ETD pairs [51, 52]. In all cases, automated de novo sequencing was improved by combining ions in the MS^2/MS^3 or in the paired CID/ExD spectra using either combined peak intensities [47, 49, 50, 52] or statistical scoring models [48, 51]. Current de novo applications of CID/HCD/ETD and MS^3 remain heavily dependent on manual interpretation with some assistance from automated methods [53].

Although database search tools developed for CID can be used for HCD and easily adapted to process ETD (by simply examining c/z° ion offsets instead of b/y), they would likely perform much worse than tools that are sensitive to unique features of HCD and ETD. Features of HCD that are not captured by CID scoring models include peaks in the low m/z range (including immonium ions), high fragment mass resolution (most CID spectra have low fragment mass resolution) and the presence of internal ions. Aside from c/z° ions, ETD spectra typically contain charge-reduced precursor peaks with high intensity, characteristic losses from charge-reduced precursors and additional related ions at offsets $\pm H$ from c and z° ions. Thus, CID scoring models should be re-designed for HCD and ETD in order to be most effective. The difficulty of the adaptation depends on the algorithm being considered. For example, MS-GFDB [36] can be automatically re-trained for new types of spectra with only 1,000 PSMs from unique peptides (per precursor charge state). Most CID database search tools [4, 9, 54, 55, 56, 57, 58, 59] have been extended to support identification of ETD MS^2 spectra: OMSSA [60], ProteinProspector [61], MS-GFDB [36], PeaksDB [62], Trans-Proteomic Pipeline [63]. However, database search using alternate fragmentation modes remains a less explored approach due to two key issues: *i*) how to best combine multiple spectra from the same precursor and *ii*) how to calculate experiment-wide false discovery rates when CID, HCD and ETD PSM scores may not be directly comparable.

Estimation of false discovery rates

As suggested by the MS² fragmentation statistics in Figure 2, database search tools trained to identify CID spectra will not perform well if given ETD or CID/ETD merged peak lists [45]. A more careful combination of CID/ECD MS² spectra for de novo sequencing was first described over a decade ago [49, 50] and later approaches [51, 36] have extended their statistical scoring models to incorporate CID- and ETD-specific ions into merged spectra that are then searched appropriately. In these cases PSM scores are directly comparable and standard FDR calculations [38] suffice to reveal substantial gains in peptide identification. An alternative approach to the creation of merged spectra is to separately search the multiple spectra from the same precursors and later merge the search results. This approach simplifies the reutilization of existing database search tools for additional alternate fragmentation modes but complicates the FDR calculations because, for example, a score threshold of 40 may yield a 1% FDR for CID matches but a 4% FDR for ETD matches. One possible way to address such discrepancies is to derive statistical e-value models for each type of search [64, 65] and use the resulting normalized values to combine search results, similarly to approaches devised to combine search results from multiple search tools [66]. To avoid the need for score normalization, intersection-based approaches [40] address this issue by requiring matching identifications from *both* CID and ETD spectra to accept an identification for each precursor but unfortunately this is known [36] to lower the number of resulting identifications by requiring significant matches from both CID and ETD spectra. Union-based approaches are often also mentioned where one imposes a 1% FDR on separate CID and ETD searches and reports the union of results; unfortunately this approach can result in a combined FDR higher than 1% because correct identifications will match the same peptide but incorrect identifications mostly accumulate. In practice, a simple strategy for 1% FDR estimation with k alternate acquisition modes would be to impose an FDR threshold of $1/k$ on each separate search, thus leading to merged results with accumulated false positives at $\leq 1\%$ FDR. However, this is known to be a conservative strategy that is likely to be less sensitive than existing

methods for combining search results from different search tools [66, 67, 59, 68], which can also be adapted for the identification of spectra from experiments using alternate fragmentation modes, including Decision Tree-based MS² acquisition protocols [34]. Another possible approach to this issue is to combine the search scores prior to FDR calculations. This approach was used in the first extension to multi-spectrum database search by deriving combined Mascot/probability scores for MS²/MS³ spectrum pairs [69]. These approaches [70, 36] avoid the intersection/union difficulties by using combined scores to facilitate FDR calculations and allow identifications where one fragmentation mode results in a good spectrum even if the other mode results in a poor spectrum. It is expected that combining search engines for CID, HCD and/or ETD also improves results, but it is important to combine these approaches with appropriate estimation of FDR (such as iProphet [68]).

Post-translational modifications

Identification of post-translational modifications stands to benefit substantially from MS² analysis with alternate fragmentation modes, especially those involving electron-based dissociation modes [71, 72]. One of the earlier such approaches [73] is the still popular phosphoproteomics protocol [74] where MS³ acquisition is triggered by the dominant loss of phosphate from precursor ions observed in MS² spectra. Alternating MS² modes further improve identification of phosphorylated peptides [35, 45, 74] and, in addition, enable otherwise challenging experiments such as co-identification of glycans and peptides from glycosylated peptides using ETD [75] or alternating CID/ETD acquisition [37, 76]. The complementarity of CID and ETD fragmentation modes is especially useful for PTMs such as glycosylation since CID leads to preferential fragmentation of the more labile glycosidic bonds and generally poor peptide fragmentation, thus facilitating glycan identification but complicating peptide identification. Conversely, ETD fragmentation of glycosylated peptides tends to result in series of c- and z-ions much like with

unmodified peptides and thus facilitates peptide identification and localization of the site of glycosylation.

Accurate localization of PTM sites is also an important area that stands to gain from alternate [30, 74, 32] and alternative [77] peptide fragmentation modes. As recently reviewed by Chalkley and Clauser [78], the problem of PTM site localization was first addressed with the AScore approach [79] which assigns a probabilistic score to a site assignment based on the number of observed ions distinguishing the top-scoring site assignment from the runner-up site assignment. More recent approaches have slightly adapted this concept to assign site assignment scores based on the difference of database search scores between the top and runner-up peptide-spectrum matches to each spectrum from a modified peptide. In all cases, the key factor determining the significance of site assignments is the presence or absence of MS^2 ions in between the possible sites of post-translational modification. Since it is clear from Figure 2 that alternate fragmentation modes tend to increase observation of peptide breaks, it is expected that the utilization of such modes will result in different numbers and quality of site assignments [30, 80], even though strategies for estimation of false-positive site assignments (False Localization Rates [78]) are still in their infancy. In addition to improving PTM site localization by increasing the numbers of observed *b/y*-ions, one especially interesting feature of HCD MS^2 acquisition is its generation of *x*-ions that are very specific indicators of phosphorylation sites [81]. These are hypothesized to derive from the phosphoric acid being a much better leaving group than water on serines and threonines and hence was observed to be a precise indicator of phosphorylation sites, though it was found in only 33% of all phosphorylated peptides.

Conclusions and outlook

The substantial advantages of peptide identification with alternate fragmentation modes have been clearly demonstrated in a variety of contexts [49, 37, 51, 82, 36, 31, 83, 84] but their widespread adoption remains limited by two significant hurdles: i) the scan rate tradeoff between

increasing the chances of identifying each peptide versus just acquiring spectra for more distinct peptides and *ii*) the evolving but limited support of peptide identification tools for taking advantage of alternate fragmentation modes.

The scan rate tradeoff is a challenge that will continue to be addressed with technological developments such as those that brought us the current generation of mass spectrometry instruments, some of which are already able to generate tens of thousands of MS² spectra per hour [85]. Nevertheless, alternate fragmentation has already been shown to be useful in key areas or potential therapeutic relevance [86, 87, 88, 89] such as peptidomics [31] and post-translational modifications analysis such as glycosylation [37], phosphorylation identification and site localization [84, 30] and histone modifications [83].

The challenge of limited software support is well on its way to resolving itself as more and more peptide identification tools [51, 41, 61, 36, 63, 62] add support for alternative MS² fragmentation modes. Still, too few tools support integrated analysis of multiple fragmentation modes [36]. Although the concept of combining fragmentation modes is over a decade old [49], most existing scoring functions for peptide identification process CID, HCD, and/or ETD spectra individually. But possibly the biggest hurdle in the development of new software tools is the recurring limited public availability of mass spectrometry data. For example, only 5 out of over 120 published papers on CID/ETD have deposited their raw data [34, 90, 36, 45, 91] on Tranche/ProteomeCommons [92] or PeptideAtlas [93]. This is a limitation that disproportionably affects the most novel and least common types of mass spectrometry experiments. We hope that the growing trend of making raw data publicly available will catch up with the dominant public availability of software tools and continue the self-reinforcing cycle towards better and more robust peptide identification strategies.

Acknowledgements

This work was partly supported by the National Institutes of Health grant 1-P41-RR024851 from the National Center for Research Resources.

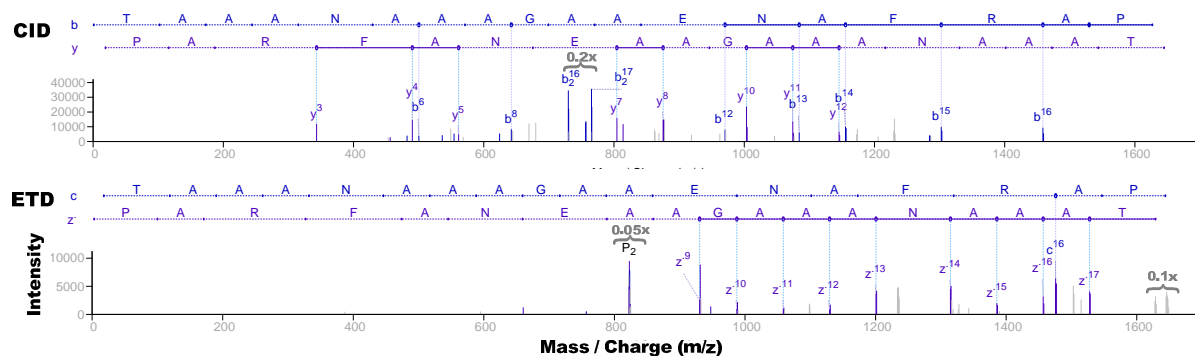
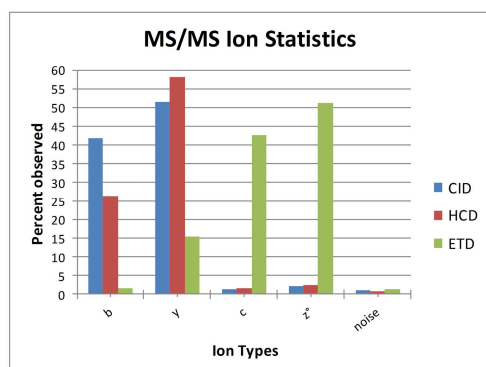


Figure 1: Complementary fragmentation in CID and ETD for peptide TAAANAAAGAAENAFRAP. CID (top) and ETD (bottom) spectra were separately identified against this C-terminal tryptic peptide at 1% FDR. Enough ions were separately detected in each spectrum to identify the peptide (65% of breaks in CID, 53% in ETD). But combining the two yields full coverage of all possible breaks, thus giving higher confidence to breaks observed in both spectra and possibly enabling full-length de novo sequencing. See Figure 2 and supplementary table SM-1 for evidence of CID/ETD complementarity over all identified spectra.

	CID		HCD		ETD	
All Charges	17378	51.7%	21246	56.2%	12834	49.9%
Charge 2	10031	56.0%	1880	59.4%	6058	44.1%
Charge 3	5719	49.2%	7485	56.7%	5076	57.0%
Charge 4	1408	43.6%	1644	44.9%	1391	56.8%
Charge 5	206	29.3%	213	26.6%	262	47.5%
Charge 6	13	12.9%	23	20.2%	42	51.9%

Table 1: Identified Peptide-Spectrum Matches (PSMs) for alternative fragmentation modes. Each column shows the number of identified PSMs and the corresponding fraction of identified spectra per precursor charge state. As expected, the identification rate for ETD spectra is higher for precursors with charge states 3 or higher, in contrast with CID/HCD which tend to perform best for doubly-charged precursors. Nevertheless, the currently slower ETD scan rate (an issue that may be resolved soon [43]) still leads to lower numbers of identified spectra with precursor charges 3 and 4 even though its identification rate is consistently higher than those of CID and HCD spectra; this tradeoff seems to disappear at precursor charge states 5 and higher, where ETD consistently identified more spectra and a higher percentage of precursors than either CID or ETD.



Prec. Charge	Peptide Breaks				
	2	3	4	5	All
# Unique Pep.	4361	2873	712	87	8037
CID	72.3	45.7	32.2	25.7	55.8
HCD	74.1	54.0	34.9	27.1	60.3
ETD	64.1	63.6	55.9	40.4	62.4
CID/HCD	83.9	61.9	42.5	33.4	69.0
CID/ETD	88.9	80.7	70.7	56.5	82.8
HCD/ETD	89.1	84.3	72.1	57.4	84.4
CID/HCD/ETD	93.2	86.8	75.4	61.1	87.8

Figure 2: Ion statistics for alternative peptide fragmentation modes. **(Left)** Peptide MS² ion statistics for alternative fragmentation modes - This shows the percentage of breaks observed by each ion type over all identified MS² spectra with precursor charge 2 or 3 for each fragmentation method. z^0 corresponds to peaks at offset +H from z ions [94]. Ions were counted if observed peak masses were within 20 ppm of expected ion masses. The "noise" ion corresponded to offset $b+0.5$, which was counted to show the level of noise in each type of MS² spectra. **(Right)** Peptide break statistics for combinations of alternative fragmentation modes - peptide breaks were counted for all unique peptides identified by all three fragmentation modes. The 6 columns show the percentage of breaks detected by each fragmentation mode and combination of fragmentation modes per precursor charge state. In CID and HCD spectra, the presence of breaks was indicated by the presence of b or y ions. For ETD, c , z^0 , or z^0+H ions indicated the presence of a break. Multiply charged ions (up to the spectrum's precursor charge) were also considered in each spectrum. Prior to this analysis, peak filtering was applied all CID, HCD, and ETD spectra such that each peak was retained only if its intensity was ranked 5th or higher over all neighboring peaks in a ± 56 Da radius. If a peptide was identified by more than one CID, HCD, or ETD spectrum, a single representative spectrum was randomly chosen for each fragmentation mode.

References

- [1] T Nilsson, M Mann, R Aebersold, J R Yates, A Bairoch, and J J Bergeron. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat Methods*, 7:681–685, 2010.
- [2] M R Larsen, M B Trelle, T E Thingholm, and O N Jensen. Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques*, 40:790–798, 2006.
- [3] B Domon and R Aebersold. Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotechnol*, 28:710–721, 2010.
- [4] J K Eng, B C Searle, K R Clauser, and D L Tabb. A face in the crowd: recognizing peptides through database search. *Mol Cell Proteomics*, 10, 2011.
- [5] N Gupta and P A Pevzner. False discovery rates of protein identifications: a strike against the two-peptide rule. *J Proteome Res*, 8:4173–4181, 2009.
- [6] A I Nesvizhskii. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics*, 73:2092–2123, 2010.
- [7] B Paizs and S Suhai. Fragmentation pathways of protonated peptides. *Mass Spectrom Rev*, 24:508–548, 2005.
- [8] A C Gucinski, E D Dodds, W Li, and V H Wysocki. Understanding and exploiting peptide fragment ion intensities using experimental and informatic approaches. *Methods Mol Biol*, 604:73–94, 2010.
- [9] JK. Eng, McCormack AL., and JR. Yates. An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *Journal Of The American Society For Mass Spectrometry*, 5:976–989, 1994.
- [10] DN. Perkins, DJ. Pappin, DM. Creasy, and JS. Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20:3551–3567, 1999.
- [11] D L Tabb, Y Huang, V H Wysocki, and J R Yates. Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. *Anal Chem*, 76:1243–1248, 2004.
- [12] Z. Zhang. Prediction of low-energy collision-induced dissociation spectra of peptides. *Anal Chem*, 76:3908–3922, 2004.
- [13] Z Zhang. Prediction of low-energy collision-induced dissociation spectra of peptides with three or more charges. *Anal Chem*, 77:6364–6373, 2005.
- [14] R A Zubarev, A R Zubarev, and M M Savitski. Electron capture/transfer versus collisionally activated/induced dissociations: solo or duet? *J Am Soc Mass Spectrom*, 19:753–761, 2008.

- [15] D L Swaney, C D Wenger, and J J Coon. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res*, 9:1323–1329, 2010.
- [16] R. Aebersold, J. Leavitt, R. A. Saavedra, L. E. Hood, and S. B. H. Kent. Internal amino-acid sequence-analysis of proteins separated by one-dimensional or two-dimensional gel-electrophoresis after insitu protease digestion on nitrocellulose. *Proc. Natl. Acad. Sci.*, 84:6970–6974, 1987.
- [17] MJ. MacCoss, WH. McDonald, A. Saraf, R. Sadygov, JM. Clark, JJ. Tasto, KL. Gould, D. Wolters, M. Washburn, A. Weiss, JI. Clark, and JR. Yates. Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci U S A*, 99:7900–7905, 2002.
- [18] A. Schlosser, J. T. Vanselow, and A. Kramer. Mapping of Phosphorylation Sites by a Multi-Protease Approach with Specific Phosphopeptide Enrichment and NanoLC-MS/MS Analysis. *Anal. Chem.*, 77:5243–5250, 2005.
- [19] B. Wang, R. Malik, E. A. Nigg, and R. Kolli,ner. Evaluation of the low-specificity protease elastase for large-scale phosphoproteome analysis. *Anal. Chem.*, 80:9526–9533, 2008.
- [20] R. G. Biringer, H. Amato, M. G. Harrington, A. N. Fonteh, J. N. Riggins, and A. F. Huhmer. Enhanced sequence coverage of proteins in human cerebrospinal fluid using multiple enzymatic digestion and linear ion trap LC-MS/MS. *Brief Funct. Genomic Proteomic*, 5:144–153, 2006.
- [21] G. Choudhary, S.-L. Wu, P. Shieh, and W. S. Hancock. Multiple Enzymatic Digestion for Enhanced Sequence Coverage of Proteins in Complex Proteomic Mixtures Using Capillary LC with Ion Trap MS/MS. *J. Proteome Res.*, 2:59–67, 2003.
- [22] Y Huang, J M Triscari, G C Tseng, L Pasa-Tolic, M S Lipton, R D Smith, and V H Wysocki. Statistical characterization of the charge state and residue dependence of low-energy cid peptide dissociation patterns. *Anal Chem*, 77:5800–5813, 2005.
- [23] J V Olsen, B Macek, O Lange, A Makarov, S Horning, and M Mann. Higher-energy c-trap dissociation for peptide modification analysis. *Nat Methods*, 4:709–712, 2007.
- [24] J E Syka, J J Coon, M J Schroeder, J Shabanowitz, and D F Hunt. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A*, 101:9528–9533, 2004.
- [25] J J Coon, B Ueberheide, J E Syka, D D Dryhurst, J Ausio, J Shabanowitz, and D F Hunt. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci U S A*, 102:9463–9468, 2005.
- [26] R J Chalkley, A Thalhammer, R Schoepfer, and A L Burlingame. Identification of protein o-glcnacylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc Natl Acad Sci U S A*, 106:8894–8899, 2009.
- [27] N L Young, P A Dimaggio, and B A Garcia. The significance, development and progress of high-throughput combinatorial histone code analysis. *Cell Mol Life Sci*, 67:3983–4000, 2010.

- [28] M S Kim and Pandey A. Electron transfer dissociation mass spectrometry in proteomics. *Proteomics*, page Epub ahead of print, 2012.
- [29] N Taouatas, A J R Heck, and S Mohammed. Evaluation of metalloendopeptidase lys-n protease performance under different sample handling conditions. *J Proteome Res*, 9:4282–4288, 2010.
- [30] M Aguiar, W Haas, S A Beausoleil, J Rush, and S P Gygi. Gas-phase rearrangements do not affect site localization reliability in phosphoproteomics data sets. *Journal of Proteome Research*, 9:3103–3107, 2010.
- [31] Y Shen, N Tolic', F Xie, R Zhao, S O Purvine, A A Schepmoes, R J Moore, G A Anderson, and R D Smith. Effectiveness of CID, HCD, and ETD with FT MS/MS for degradomic-peptidomic analysis: comparison of peptide identification methods. *J Proteome Res*, 10:3929–3943, 2011.
- [32] M P Jedrychowski, E L Huttlin, W Haas, M E Sowa, R Rad, and S P Gygi. Evaluation of hcd- and cid-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Mol Cell Proteomics*, (in press), 2011.
- [33] C K Frese, A F Altelaar, M L Hennrich, D Nolting, M Zeller, J Griep-Raming, A J Heck, and S Mohammed. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J Proteome Res*, 10:2377–2388, 2011.
- [34] D L Swaney, G C McAlister, and J J Coon. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods*, 5:959–964, 2008.
- [35] D L Swaney, C D Wenger, J A Thomson, and J J Coon. Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A*, 106(4):995–1000, 2009.
- [36] S Kim, N Mischerikow, N Bandeira, J D Navarro, L Wich, S Mohammed, A J Heck, and P A Pevzner. The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Mol Cell Proteomics*, 9:2840–2852, 2010.
- [37] J M Hogan, S J Pitteri, P A Chrisman, and S A McLuckey. Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. *J Proteome Res*, 4:628–632, 2005.
- [38] J E Elias and S P Gygi. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods*, 4:207–214, 2007.
- [39] N Gupta, N Bandeira, U Keich, and Pevzner PA. Target-decoy approach and false discovery rate: When things may go wrong. *J Am Soc Mass Spectrom*, 22:1111–1120, 2011.
- [40] H Molina, R Matthiesen, K Kandasamy, and A Pandey. Comprehensive comparison of collision induced dissociation and electron transfer dissociation. *Anal Chem*, 80:4825–4835, 2008.
- [41] H Chi, R X Sun, B Yang, C Q Song, L H Wang, C Liu, Y Fu, Z F Yuan, H P Wang, S M He, and M Q Dong. pNovo: de novo peptide sequencing and identification using HCD spectra. *J Proteome Res*, 9:2713–2724, 2010.

- [42] S R Hart, K W Lau, S J Gaskell, and S J Hubbard. Distributions of ion series in ETD and CID spectra: making a comparison. *Methods Mol Biol*, 696:327–337, 2011.
- [43] P D Compton, J V Strukl, D L Bai, J Shabanowitz, and D F Hunt. Optimization of Electron Transfer Dissociation via Informed Selection of Reagents and Operating Parameters. *Anal Chem*, page Epub ahead of print, 2012.
- [44] S Kim, N Gupta, N Bandeira, and P A Pevzner. Spectral dictionaries: Integrating de novo peptide sequencing with database search of tandem mass spectra. *Mol Cell Proteomics*, 8:53–69, 2009.
- [45] M S Kim, J Zhong, K Kandasamy, B Delanghe, and A Pandey. Systematic evaluation of alternating CID and ETD fragmentation for phosphorylated peptides. *Proteomics*, 11:2568–2572, 2011.
- [46] T Lin and G L Glish. C-terminal peptide sequencing via multistage mass spectrometry. *Anal Chem*, 70:5162–5165, 1998.
- [47] Z. Zhang and JS. McElvain. De novo peptide sequencing by two-dimensional fragment correlation mass spectrometry. *Anal Chem*, 72:2337–2350, 2000.
- [48] N. Bandeira, J. Olsen, M. Mann, and P.A. Pevzner. Multi-Spectra Peptide Sequencing and its Applications to Multistage Mass Spectrometry. *Bioinformatics (ISMB 2008 special issue)*, 24:i416–23, 2008.
- [49] D M Horn, R A Zubarev, and F W McLafferty. Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc Natl Acad Sci U S A*, 97:10313–10317, 2000.
- [50] M M Savitski, M L Nielsen, F Kjeldsen, and R A Zubarev. Proteomics-grade de novo sequencing approach. *J Proteome Res*, 4:2348–2354, 2005.
- [51] R Datta and M Bern. Spectrum fusion: using multiple mass spectra for de novo peptide sequencing. *J Comput Biol*, 16:1169–1182, 2009.
- [52] A. Bertsch, A. Leinenbach, A. Pervukhin, M. Lubeck, R. Hartmer, C. Baessmann, Y. A. Elnakady, R. Muller, S. Bocker, C. G. Huber, and O. Kohlbacher. De novo peptide sequencing by tandem MS using complementary CID and electron transfer dissociation. *Electrophoresis*, 30:3736–3747, 2009.
- [53] K F Medzihradzsky and C J Bohlen. Partial De Novo Sequencing and Unusual CID Fragmentation of a 7 kDa, Disulfide-Bridged Toxin. *J Am Soc Mass Spectrom*, page Epub ahead of print, 2012.
- [54] D N Perkins, D J Pappin, D M Creasy, and J S Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20:3551–3567, 1999.
- [55] R Craig and R C Beavis. Tandem: matching proteins with tandem mass spectra. *Bioinformatics*, 20:1466–1467, 2004.

- [56] M Bern, Y Cai, and D Goldberg. Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Anal Chem*, 79:1393–1400, 2007.
- [57] D L Tabb, C G Fernando, and M C Chambers. Myrimatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J Proteome Res*, 6:654–661, 2007.
- [58] L Wang, DQ Li, Y Fu, HP Wang, JF Zhang, ZF Yuan, RX Sun, R Zeng, SM He, and W Gao. pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 21:2985–2991, 2007.
- [59] M Spivak, J Weston, L Bottou, L Kal`Il, and WS Noble. Improvements to the percolator algorithm for peptide identification from shotgun proteomics data sets. *Journal of Proteome Research*, 8:3737–3745, 2009.
- [60] L Y Geer, S P Markey, J A Kowalak, L Wagner, M Xu, D M Maynard, X Yang, W Shi, and S H Bryant. Open mass spectrometry search algorithm. *Journal of Proteome Research*, 3:958–964, 2004.
- [61] P R Baker, K F Medzihradzsky, and R J Chalkley. Improving software performance for peptide electron transfer dissociation data analysis by implementation of charge state- and sequence-dependent scoring. *Mol Cell Proteomics*, 9:1795–1803, 2010.
- [62] X Liu, B Shan, L Xin, and B Ma. Better score function for peptide identification with ETD MS/MS spectra. *BMC Bioinformatics*, 11:S4, 2010.
- [63] E W Deutsch, D Shteynberg, H Lam, Z Sun, J K Eng, C Carapito, P D von Haller, N Tasman, L Mendoza, T Farrah, and R Aebersold. Trans-proteomic pipeline supports and improves analysis of electron transfer dissociation data sets. *Proteomics*, 10:1190–1195, 2010.
- [64] G Alves, A Y Ogurtsov, W W Wu, G Wang, R F Shen, and Y K Yu. Calibrating e-values for ms2 database search methods. *Biol Direct*, 2:26–26, 2007.
- [65] G Alves, A Y Ogurtsov, and Y K Yu. Raid_aps: Ms/ms analysis with multiple scoring functions and spectrum-specific statistics. *PLoS One*, 5, 2010.
- [66] G Alves, W W Wu, G Wang, R F Shen, and Y K Yu. Enhancing peptide identification confidence by combining search methods. *J Proteome Res*, 7:3102–3113, 2008.
- [67] B C Searle, M Turner, and A I Nesvizhskii. Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies. *J. Proteome Res.*, 7:245–253, 2008.
- [68] D Shteynberg, EW Deutsch, H Lam, JK Eng, Z Sun, N Tasman, L Mendoza, RL Moritz, R Aebersold, and AI Nesvizhskii. iProphet: Multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol Cell Proteomics*, page Epub ahead of print, 2011.
- [69] J V Olsen and M Mann. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci U S A*, 101:13417–13422, 2004.

- [70] P J Ulintz, B Bodenmiller, P C Andrews, R Aebersold, and A I Nesvizhskii. Investigating MS2-MS3 matching statistics: A model for coupling consecutive stage mass spectrometry data for increased peptide identification confidence. *Mol Cell Proteomics*, 7:71–87, 2008.
- [71] R A Zubarev. Reactions of polypeptide ions with electrons in the gas phase. *Mass Spectrom Rev*, 22:57–77, 2003.
- [72] L M Mikesch, B Ueberheide, A Chi, J J Coon, J E Syka, J Shabanowitz, and D F Hunt. The utility of etd mass spectrometry in proteomic analysis. *Biochim Biophys Acta*, 1764:1811–1822, 2006.
- [73] S A Beausoleil, M Jedrychowski, D Schwartz, J E Elias, J Villén, J Li, M A Cohn, L C Cantley, and S P Gygi. Large-scale characterization of hela cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A*, 101:12130–12135, 2004.
- [74] A M Palumbo, S A Smith, C L Kalcic, M Dantus, P M Stemmer, and G E Reid. Tandem mass spectrometry strategies for phosphoproteome analysis. *Mass Spectrom Rev*, 30:600–625, 2011.
- [75] R Chen, F Wang, Y Tan, Z Sun, C Song, M Ye, H Wang, and H Zou. Development of a combined chemical and enzymatic approach for the mass spectrometric identification and quantification of aberrant N-glycosylation. *J Proteomics*, page Epub ahead of print, 2012.
- [76] F G Hanisch. O-glycoproteomics: site-specific O-glycoprotein analysis by CID/ETD electrospray ionization tandem mass spectrometry and top-down glycoprotein sequencing by in-source decay MALDI mass spectrometry. *Methods Mol Biol*, page Epub ahead of print, 2012.
- [77] N Nagaraj, R C D’Souza, J Cox, J V Olsen, and M Mann. Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J Proteome Res*, 9:6786–6794, 2010.
- [78] R J Chalkley and K R Clauser. Modification site localization scoring: Strategies and performance. *Mol Cell Proteomics*, page (in press), 2012.
- [79] S A Beausoleil, J Villén, S A Gerber, J Rush, and S P Gygi. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol*, 24:1285–1292, 2006.
- [80] T Taus, T Köcher, P Pichler, C Paschke, A Schmidt, C Henrich, and K Mechtler. Universal and confident phosphorylation site localization using phosphors. *J Proteome Res*, 10:5354–5362, 2011.
- [81] C D Kelstrup, O Hekmat, C Francavilla, and J V Olsen. Pinpointing phosphorylation sites: Quantitative filtering and a novel site-specific x-ion fragment. *J Proteome Res*, 10:2937–2948, 2011.
- [82] S M Chowdhury, X Du, N Tolic’, S Wu, R J Moore, M U Mayer, R D Smith, and J N Adkins. Identification of cross-linked peptides after click-based enrichment using sequential collision-induced dissociation and electron transfer dissociation tandem mass spectrometry. *Anal Chem*, 81:5524–5532, 2009.

- [83] A Jufvas, P Strålfors, and A V Vener. Histone variants and their post-translational modifications in primary human fat cells. *PLoS One*, 6, 2011.
- [84] Y Lu, X Zhou, P Stemmer, and G Reid. Sulfonium ion derivatization, isobaric stable isotope labeling and data dependent CID- and ETD-MS/MS for enhanced phosphopeptide quantitation, identification and phosphorylation site characterization. *Journal of The American Society for Mass Spectrometry*, pages 1–17, 2011.
- [85] M W Duncan, R Aebersold, and R M Caprioli. The pros and cons of peptide-centric proteomics. *Nat Biotechnol*, 28:659–664, 2010.
- [86] F. R. DePontieu, J. Qian, A. L. Zarling, T. L. McMiller, T. M. Salay, A. Norris, A. M. English, J. Shabanowitz, V. H. Engelhard, D. F. Hunt, and S. L. Topalian. Identification of tumor-associated, MHC class II-restricted phosphopeptides as targets for immunotherapy. *Proc. Natl. Acad. Sci. U.S.A.*, 106:12073–12078, 2009.
- [87] G Walsh. Post-translational modifications of protein biopharmaceuticals. *Drug Discov Today*, 15:773–780, 2010.
- [88] C. R. Robertson, S. P. Flynn, H. S. White, and G. Bulaj. Anticonvulsant neuropeptides as drug leads for neurological diseases. *Nat Prod Rep*, 28:741–762, 2011.
- [89] G. F. King. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert Opin Biol Ther*, 11:1469–1484, 2011.
- [90] M E Burkard, J Maciejowski, V Rodriguez-Bravo, M Repka, D M Lowery, K R Clauser, C Zhang, K M Shokat, S A Carr, M B Yaffe, and P V Jallepalli. Plk1 self-organization and priming phosphorylation of hscyk-4 at the spindle midzone regulate the onset of division in human cells. *PLoS Biol*, 7, 2009.
- [91] M M Savitski, S Lemeer, M Boesche, M Lang, T Mathieson, M Bantscheff, and B Kuster. Confident phosphorylation site localization using the mascot delta score. *Mol Cell Proteomics*, 10:(in press), 2011.
- [92] J A Falkner and P C Andrews. P6-T Tranche: secure decentralized data storage for the proteomics community. *J Biomol Tech.*, 18:3, 2007.
- [93] F Desiere, E W Deutsch, N L King, A I Nesvizhskii, P Mallick, J Eng, S Chen, J Edes, S N Loevenich, and R Aebersold. The PeptideAtlas project. *Nucleic Acids Res*, 34:655–658, 2006.
- [94] M Savitski, F Kjeldsen, M Nielsen, and R Zubarev. Hydrogen rearrangement to and from radical z fragments in electron capture dissociation of peptides. *Journal of The American Society for Mass Spectrometry*, 18:113–120, 2007.