EXPERT
REVIEWS

ETD fragmentation features improve algorithm

Wenzhou Li and
Vicki H Wysocki*

*Department of Chemistry and
Biochemistry, University of Arizona,
Tucson, AZ, USA*
*Author for correspondence:
vwysocki@email.arizona.edu*

Evaluation of: Sridhara V, Bai DL, Chi A *et al.* Increasing peptide identifications and decreasing search times for ETD spectra by pre-processing and calculation of parent precursor charge. *Proteome Sci.* 10(1), 8 (2012).

Electron transfer dissociation (ETD) is an alternative technique used in mass spectrometry-based proteomics experiments. Because it is newer, most of the protein identification algorithms for ETD are still a simple derivation of well-established collision-activated dissociation algorithms without the consideration of many unique ETD spectral features. Sridhara and coworkers recently reported removing the charge-reduced precursors and corresponding neutral loss peaks to improve ETD peptide identification with the Open Mass Spectrometry Search Algorithm (OMSSA). These peaks were also used to deduce the charge of the precursors for low resolution data. The scheme is a concrete example of implementing known ETD fragmentation features to improve a protein identification algorithm.

KEYWORDS: algorithm • electron transfer dissociation • fragmentation features • precursor charge • spectrum preprocessing

Collision-activated dissociation (CAD) and electron transfer dissociation (ETD) are two frequently used peptide fragmentation methods in modern mass spectrometry-based proteomics experiments. CAD fragments peptides by colliding them with inert gas atoms or molecules, resulting in energy randomization and subsequent dissociation of weaker bonds such as amide bonds. In ETD, however, multiply charged peptide cations receive electrons from radical anions to form an aminoketyl radical and dissociate further, primarily at N–C bonds [1]. This distinct mechanism gives the ETD spectra many unique features: N–C cleavage generates c, z ions rather than the b, y ions of CAD; peptides with higher charge states generally fragment better with ETD than with CAD; ETD spectra normally have intense charge-reduced precursor peaks ([ET-no-D] products) as well as corresponding neutral loss peaks; charge transfer products, especially c-1 and z+1 ions, are frequency observed in ETD spectra; cleavage with ETD is less selective, generating more extensive ion series; and labile post-translational modifications that are often lost in CAD can be retained in ETD [1–4]. Understanding these unique features is extremely helpful for the correct interpretation of ETD data.

However, because ETD is a newer technique, most of the protein identification algorithms for ETD are still a simple derivation of well-established CAD algorithms, only searching with c, z ions instead of b, y ions [5]. Although this model works, it is over-simplified because it only considers the first ETD feature mentioned above while ignoring all the others. This could be problematic in many cases, for example, the strong charge-reduced precursor peaks can be accidently matched to fragment ions, resulting in a higher false discovery rate. Clearly, specialized algorithms should be developed for ETD, or existing algorithms should be adapted for ETD, by considering unique ETD features, and efforts have already been made by many groups [6–8]. The paper by Sridhara *et al.* describes an effort to improve the Open Mass Spectrometry Search Algorithm (OMSSA) [9].

Summary of methods & results

Sridhara and coworkers used several ETD features to modify the current OMSSA algorithm. This involved the preprocessing of ETD spectra to remove charge-reduced precursor peaks as well as associated neutral loss peaks, using a new noise filter for ETD and using linear discriminant analysis to determine precursor charge states for low-resolution spectra. More specifically, different options for removing precursor and neutral loss peaks, such as using a fixed or variable

mass window, were compared, and the optimized parameters were applied to determine the sensitivity of the modified approach. They also reported that an extra ion type (y ion, which is consistent with the ETD statistical study [2]) should be considered in ETD, so the noise filter in OMSSA was adjusted to allow more peaks in a specific mass window (+/-27 for 1+ and +/-14 for 2+). In addition, the author discussed the relationship between precursor charge states and the distribution of charge-reduced precursor series as well as the frequently observed neutral loss series (e.g., loss of water and ammonia). Using charge states as a grouping variable, discriminant analysis was performed to predict precursor charge states from charge-reduced precursors and neutral loss series. The authors set thresholds for charge state identification and showed that it was necessary to search only one charge state for those spectra above the strictest threshold, two for those between the top and second threshold, and a range of charge states for those spectra that are below both thresholds. It was shown that this method can reduce search times by 3.5-fold for low-resolution data because of the reduced number of candidates to be searched.

At a false discovery rate of 1%, removing precursor and neutral loss peaks allowed the authors to identify 9.8% more peptides using OMSSA. With the new noise filter applied, another 4.2% were identified. The precursor charge determination contributed an additional 3.8% peptides. The overall improvement (18.8%) is significant and pointedly illustrates the necessity of incorporating more ETD features into algorithms.

Discussion

OMSSA is one of the most frequently used algorithms for ETD, and its performance has been characterized in many studies. For instance, Good and coworkers compared OMSSA, Zcore, Mascot and Sequest and found that OMSSA's performance was most similar to that of Zcore, but better than that of Mascot and Sequest [6]. By contrast, Kandasamy and colleagues searched ETD datasets with OMSSA, Mascot, Spectrum Mill and X!Tandem, and noticed more peptide identifications from Spectrum Mill and Mascot [5]. Although the conclusions vary, the method described in this paper is a great improvement to the OMSSA algorithm. ETD researchers can directly benefit from the increased sensitivity and specificity of this modified OMSSA algorithm.

The database search method used in peptide identification can be simply described as comparing an experimental spectrum with a set of theoretical spectra or peak lists derived from candidate sequences pulled from a sequence library by setting a particular mass tolerance for the precursor ion. To do this, nearly all the current algorithms involve the preprocessing of an experimental spectrum, generation of a set of theoretical spectra, and using certain scoring functions to evaluate the similarities of the theoretical spectra to the experimental spectrum to determine the best match. The transition of CAD algorithms to ETD mainly focuses on the first two aspects, which will be discussed here.

The major purpose of preprocessing an experimental spectrum is to remove peaks that are less indicative of the peptide sequences or can lead to false peak matches, including nonproduct ions, isotope ions and noise peaks. In CAD, the fragmentation is so

efficient that the remaining precursor is not a concern, while in ETD, charge-reduced precursor ions and corresponding neutral loss ions are too abundant to be ignored. These ions are double-edged swords for ETD peptide identification: on one hand, these derivatives of precursors are confounding because they account for a large portion of total ion current but contain little information about backbone fragmentation. Good and coworkers first reported improved peptide identification by removing these interfering ions before submitting the data for search [6,10], and Sridhara's work here further exploited the filtering conditions and implemented them into OMSSA. On the other hand, these ions may be used to deduce the property of the precursor. Xia reported using the amino acid side chain neutral loss in ETD as a fingerprint for amino acid composition [11]. For instance, a neutral loss of 43 Da from the precursor suggests the presence of arginine, which can greatly reduce the search space. Sridhara here also showed that the distribution of charge-reduced precursors and neutral losses can be used to predict precursor charge. An ideal method for ETD spectrum preprocessing is probably a combination of the two: using charge-reduced precursors and neutral loss to get hints of the sequence and charge, then removing them during spectra comparison.

In terms of theoretical ETD spectra generation, an in-depth understanding of the ion types and fragment intensities in ETD is required. Chalkley and coworkers did a statistical study on ETD spectra and showed that besides c, z ions, y, z+1 and c-1 ions are also abundant and their occurrence varies by charge states [2]. They later implemented a charge- and sequence-dependent scoring method and reported an 80% increase in peptide identification [7]. A similar concept has also been implemented into algorithms such as pFind [8]. Sridhara's work here utilized an extra y ion series but it could also be worthwhile to take into account proton transfer products and charge states. Fragment intensity in ETD is less understood and could be the next catalyst to boost ETD peptide identification, and it is already well accepted that intensity patterns can improve peptide identification for CAD [12]. Recent intensity pattern studies show that selective cleavage also exists for ETD, and is dependent upon both the amino acid composition and the position of cleavage sites [4]. This information can be incorporated into ETD identification algorithms to further improve peptide identification.

In summary, this work by Sridhara is a concrete example of how to implement known fragmentation features, or chemical knowledge, into algorithms. The process involves both deep understanding of the chemical knowledge as well as numerous trials and errors for optimization. It is important to note the findings described in this paper can essentially benefit most, if not all, ETD algorithms, by simply implementing the spectrum preprocessing scheme. The same 'from fragmentation features to algorithms' concept can also be applied to all the other established or future fragmentation methods.

Five-year view

We anticipate that as more researchers obtain instruments with ETD and as researchers better define ETD acquisition

methods and search algorithms, ETD will continue to increase in popularity owing to its complementary fragmentation patterns to CAD and the ability to retain post-translational modifications. The exponential growth of ETD datasets as well as higher mass accuracy require faster algorithms that are optimized for high-resolution data. It is important to make large and high-resolution ETD datasets available so that researchers can extract more ETD features using statistical methods (clustering, linear discriminant analysis and so on) to add into algorithms and evaluate their performance. Other peptide identification methods, such as ETD spectral library searches and ETD *de novo* sequencing, will also benefit from a deeper understanding of ETD features. Finally, more postprocessing tools such as Scaffold should be developed to integrate data from multiple fragmentation methods and multiple algorithms [13].

Financial & competing interests disclosure

Key issues

- Spectrum preprocessing to remove precursor and neutral loss peaks reduced the probability of false fragment matches.
- y ions are considered in electron-transfer dissociation fragmentation in addition to c, z ions.
- The precursor charge states from low-resolution instruments can be estimated using the linear discriminant analysis results from training of the distribution of charge-reduced precursors and neutral losses.
- The overall performance of the Open Mass Spectrometry Search Algorithm is improved by 18.8% and is 3.5-times faster.
- The exploited spectrum preprocessing conditions can be transplanted to essentially all other algorithms.
- This work demonstrated how known fragmentation features can lead to better algorithms.

References

Papers of special note have been highlighted as:
•• of considerable interest

1   Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl Acad. Sci. USA* 101(26), 9528–9533 (2004).

2   Chalkley RJ, Medzihradszky KF, Lynn AJ, Baker PR, Burlingame AL. Statistical analysis of peptide electron transfer dissociation fragmentation mass spectrometry. *Anal. Chem.* 82(2), 579–84 (2010).

••   **Analyzed the ion types and charge states of electron transfer dissociation.**

3   Savitski MM, Kjeldsen F, Nielsen ML, Zubarev RA. Complementary sequence preferences of electron-capture dissociation and vibrational excitation in fragmentation of polypeptide polycations. *Angew. Chem. Int. Ed. Engl.* 45(32), 5301–5303 (2006).

4   Li W, Song C, Bailey DJ, Tseng GC, Coon JJ, Wysocki VH. Statistical analysis of electron transfer dissociation pairwise fragmentation patterns. *Anal. Chem.* 83(24), 9540–9545 (2011).

5   Kandasamy K, Pandey A, Molina H. Evaluation of several MS/MS search algorithms for analysis of spectra derived from electron transfer dissociation experiments. *Anal. Chem.* 81(17), 7170–7180 (2009).

6   Good DM, Wenger CD, Coon JJ. The effect of interfering ions on search algorithm performance for electron-transfer dissociation data. *Proteomics* 10(1), 164–167 (2010).

••   **Demonstrated that removing interference ions in electron transfer dissociation can lead to improved peptide identification for multiple algorithms.**

7   Baker PR, Medzihradszky KF, Chalkley RJ. Improving software performance for peptide electron transfer dissociation data analysis by implementation of charge state- and sequence-dependent scoring. *Mol. Cell Proteomics* 9(9), 1795–1803 (2010).

8   Sun RX, Dong MQ, Song CQ *et al.* Improved peptide identification for proteomic analysis based on comprehensive characterization of electron transfer dissociation spectra. *J. Proteome Res.* 9(12), 6354–6367 (2010).

9   Sridhara V, Bai DL, Chi A *et al.* Increasing peptide identifications and decreasing search times for ETD spectra by pre-processing and calculation of parent precursor charge. *Proteome Sci.* 10(1), 8 (2012).

10   Good DM, Wenger CD, McAlister GC, Bai DL, Hunt DF, Coon JJ. Post-acquisition ETD spectral processing for increased peptide identifications. *J. Am. Soc. Mass Spectrom.* 20(8), 1435–1440 (2009).

11   Xia Q, Lee MV, Rose CM *et al.* Characterization and diagnostic value of amino acid side chain neutral losses following electron-transfer dissociation. *J. Am. Soc. Mass Spectrom.* 22(2), 255–264 (2011).

12   Li W, Ji L, Goya J, Tan G, Wysocki VH. SQID: an intensity-incorporated protein identification algorithm for tandem mass spectrometry. *J. Proteome Res.* 10(4), 1593–1602 (2011).

13   Searle BC. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* 10(6), 1265–1269 (2010).