Improvement in Analytical Software Makes a Difference on the Decision Tree Driven ETD and CID Fragmentation

Bioinformatics Solutions Inc.

Baozhen Shan[1], Lei Xin[1], Mingjie Xie[1], Bin Ma[2]
[1] Bioinformatics Solutions Inc, Waterloo, ON
[2] University of Waterloo, Waterloo, ON

Overview

Purpose: To develop an alternative fragmentation technique to improve peptide identification
Method: Use data-dependent decision tree logic to determine the fragmentation method most likely to result in a successful identification
Result: New developments in analytical software improved performance of the decision tree-driven CID/ETD fragmentation and demand adjustment of decision tree parameters

Introduction

ETD has proven to be complementary to CID for peptide identification. Coon et al. have shown in 2008 that depending on the charge state (z) and mass-to-charge ratio (m/z), fragmentation with CID or ETD may result in a higher peptide identification rate than the other fragmentation method. Thus, in so-called data dependent decision tree logic (DDDT), the choice of CID or ETD is determined on-the-fly by the m/z and z of the precursor ion (as shown in Figure 1)[1]. Since Coon's study, the ETD peptide identification software has been improved dramatically. Thus, the parameters obtained in their study may not be optimal anymore. The performance of three common search engines on ETD data are shown in Figure 2. The new PEAKS DB software improved the ETD identification significantly, making it necessary to reconsider the optimal DDDT logic. In this study we examine the effect of the advanced ETD peptide identification software on the DDDT model.

Figure 1. Data Dependent Decision Tree

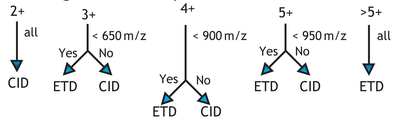
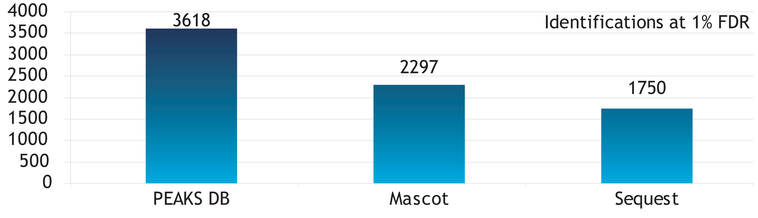


Figure 2. Numbers of Peptide-Spectrum Matches Identified by Three Common Search Engines on ETD Data from ABRF iPRG 2011



Methods

The same dataset (only one replicate) from Coon's study was used to compare the peptide identification performance using CID, ETD, and DDDT, respectively. PEAKS 5.3 and Mascot 2.3 were used together on the CID data to ensure the maximum software performance on CID spectra. The ETD and DDDT datasets were analyzed using PEAKS 5.3 alone. We wish to show that even under such an "unfair" setting, DDDT outperforms CID.

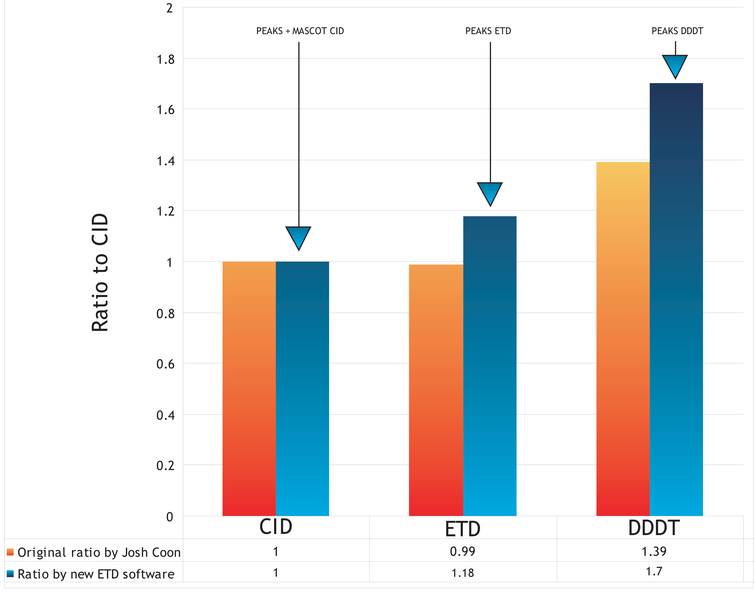
For each m/z and z range, the peptide identification success rates of using CID and ETD are determined, respectively. New parameters for the decision tree are determined by choosing the more successful method at each m/z and z range.

Preliminary Results

A portion of the same datasets used in Coon's study was downloaded from PeptideAtlas. The data was obtained from Thermo LTQ-Orbitrap on the LysC digest of a yeast lysate. It consists three LC-MS/MS runs with CID, ETD, and DDDT respectively.

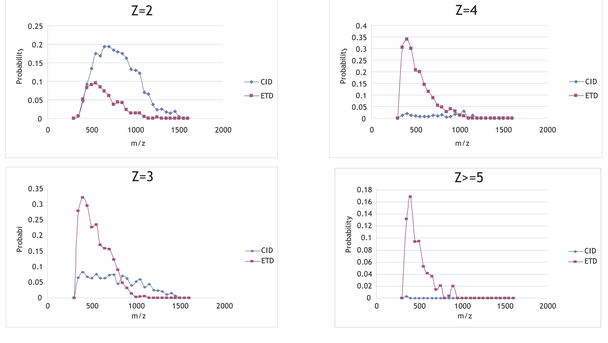
For the CID dataset, 5232 peptides were identified with either PEAKS 5.3 or Mascot 2.3 with 1% FDR. For the ETD and DDDT datasets, 6025 and 8677 peptides were identified with PEAKS 5.3 alone. The performance comparison of the three fragmentation strategies is CID: ETD : DDDT = 1: 1.18 : 1.70. This improves upon the DDDT performance reported by Josh Coon (CID: ETD : DDDT = 1: 0.99 : 1.39), as shown in Figure 3.

Figure 3. Performance Comparison of Three Fragmentation Strategies



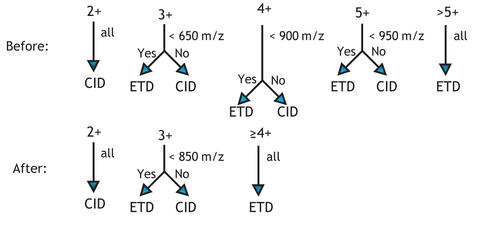
To find the new set of DDDT parameters, the 64167 CID spectra and 54852 ETD spectra from the dataset were binned by precursor m/z, and precursor charge, z. Identification rate in each bin was calculated for both CID and ETD spectra (shown in Figure 4) respectively. The DDDT logic should choose the fragmentation method with a higher identification success rate.

Figure 4. Identification Success Rates at Different Charge States and m/z Regions, using CID and ETD Spectra



Based on the identification rate in each bin we derive the new optimal parameters for the decision tree (shown in Figure 5). The improvement of ETD software significantly simplifies the logic reported by Josh Coon. The most noticeable difference caused by the ETD software improvement is that ETD is now superior than CID for z≥4 regardless of the m/z value.

Figure 5. The Data Dependent Decision Tree Logic is Simplified Due to the Improvement of ETD Identification Software



Reference

1. Swaney, D.L., McAlister, G.C., Coon, J.J., Decision tree-driven tandem mass spectrometry for shotgun proteomics. Nat. Methods (2008) 5, 959-964.