Analysis of EThcD Spectra with PEAKS® Platform for *de novo* Sequencing and Peptide Identification

Lei Xin, Lin He, Wen Zhang, Baozhen Shan
Bioinformatics Solutions Inc., Waterloo, ON, Canada

Summary
Recently a combined fragmentation scheme by combining electron transfer and higher-energy collision dissociation, called EThcD, was introduced and implemented on high resolution mass spectrometers such as LTQ Orbitrap Elite and Orbitrap Fusion.

However, the data analysis tools lack efficiency for this new type of fragmentation method. The PEAKS® platform integrates a new model for EThcD data analysis, which outperforms current tools for peptide identification and *de novo* sequencing [1].

Introduction
The EThcD method first applies ETD in the linear ion trap and then transfers both precursors and product ions to the collision cell for HCD fragmentation. An EThcD spectrum is approximately equivalent to the union of an ETD and an HCD spectrum, containing both b/y and c/z ions. The richer MS/MS spectra derived from EThcD appears advantageous for peptide *de novo* sequencing and identification of post-translational modifications (PTMs).

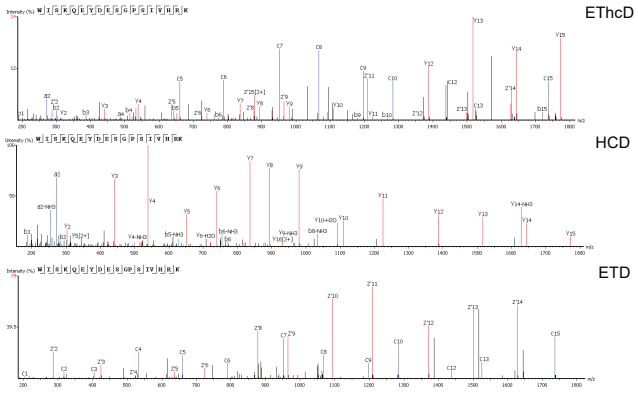
Method
Linear discriminant function (LDF) was used to score peptide spectrum matches (PSMs). Six different fragment ions (a, b, c, y, z, z+1) were selected as main features of the LDF. For each ion, not only its intensity but also its rank in the spectrum were used when scoring a PSM. Besides the six main ions, some minor ions such as y-H2O and y-NH3were also considered in the scoring function. Besides ion information, some other features such as peptide length and precursor mass error were also added to the LDF to achieve higher sensitivity and accuracy. The final coefficients of the LDF were trained by using support vector machine. The training was performed separately for each charge state.

The model was tested by using a published dataset [2]. Briefly, a human leukocyte antigen (HLA) Class II ligandomesample was isolated, fractionated by strong cation exchange chromatograph and analyzed on an LTQ Orbitrap Elite mass spectrometer by using either EThcD, HCD, or ETD fragmentation.

The MS data were analyzed using the searching parameters specified by the original publication [2]. Briefly, the data was searched against a human database with 3 ppm of mass error tolerance for precursor ions and 0.02 Da for product ions with PEAKS® EThcD, ETD or HCD model. No enzyme was specified. Oxidation (M) and deamidation(N, Q) were set as variable modifications. Peptide-spectrum matches (PSMs) identified with 0.5% of false discovery rate (FDR) were applied.

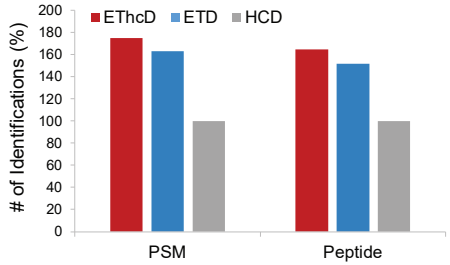
Results
The EThcD fragmentation model was implemented in PEAKS® 8.5 and tested with the published dataset [2] for *de novo* sequencing and database searching.

Spectra of an Unmodified Peptide Analyzed by EThcD, HCD, or ETD

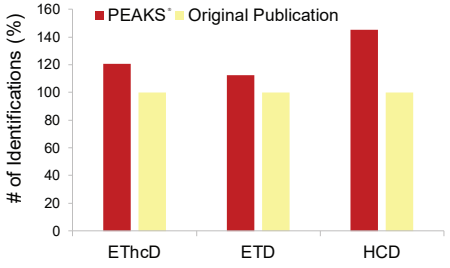


Dataset: PMID 26764012, PXD002951 [2]

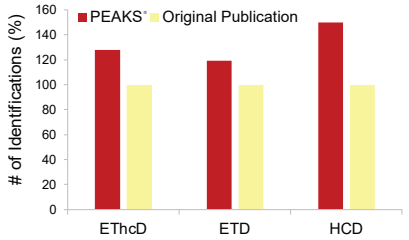
1. EThcD data searched by using EThcD model resulted in 60% more IDs than using ETD or HCD model in PEAKS®



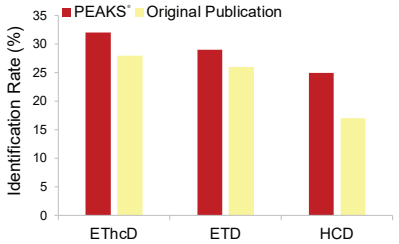
2. EThcD, ETD and HCD models in PEAKS® identified 20% more PSMs than the original publication



3. EThcD, ETD and HCD models in PEAKS® identified 20% more peptides than the original publication



4. EThcD, ETD and HCD models in PEAKS® gave higher identification rates



De novo Sequencing Accuracy Was Improved for EThcD Data

Data Set	*de novo* Amino Acids		Peptide Sensitivity
	Sensitivity	Precision	
EThcD	74.25	77.09	49.89
ETD	56.83	63.90	26.59
HCD	55.09	56.98	28.86

De novo accuracy was evaluated by identified peptides at 0.5% FDR from database searching.

Conclusion
- PEAKS® 8.5 EThcD fragmentation method resulted in more identifications and higher *de novo* sequencing accuracy.
- EThcD, ETD, and HCD fragmentation models in PEAKS® 8.5 offered more identified PSMs and peptides and higher identification rates.

References
[1] Zhang J, et al. Mol Cell Proteomics. 2012 Apr; 11(4):M111.010587.
[2] Mommen GP, et al. Mol Cell Proteomics. 2016 Apr; 15(4):1412-23.