In-depth proteomics analysis using a *timsTOF* with the PASEF method and Deep Learning

ASMS 2018, ThP388

Zia Rahman(1), Hieu Tran(2), Clark Chen(1), Gary Kruppa(3), and Paul Shan (1)
(1) Bioinformatics Solutions, Waterloo, Canada, (2) Department of Computer Science, University of Waterloo, Waterloo, Canada, (3) Bruker Daltonics Inc, Billerica, USA

Introduction

Data *dependent* acquisition (DDA) has been the most widely used strategy in LC-MS for over 20 years. However, the depth of the analysis is foremost limited by the sequencing speed and sensitivity of the mass spectrometer. In contrast to DDA, every detectable precursor is fragmented in data independent acquisition strategies (DIA). No gaps in the data are observed. However, analyzing DIA datasets is challenging due to the multiplexing of MS/MS spectra. An alternative approach – PASEF on a TIMS-QTOF was reported to addresses this issue. All precursor ions are accumulated in parallel and released sequentially as a function of their ion mobility. Here we report on a deep-learning approach for PASEF data analysis with more than one peptide per MS/MS spectrum.

Methods

Four-dimensional (m/z, retention time, ion-mobility, intensity) feature detection was performed. Each peptide feature was associated with a set of MS/MS spectra.
Two layers of neural networks were used for peptide identification from MS/MS. First a convolutional neural network (CNN) learns local alternative peak features of all possible next amino acid candidates given the currently predicted partial peptide sequence. The second convolutional neural network learns general features of the spectrum and passes them to the LSTM network. The LSTM learns sequence patterns of the currently predicted partial peptide sequence in association with the spectrum features from the spectrum-CNN.
A target and decoy database approach was used for the estimation of false-discovery rate. Feature intensity was used for peptide quantitative analysis.

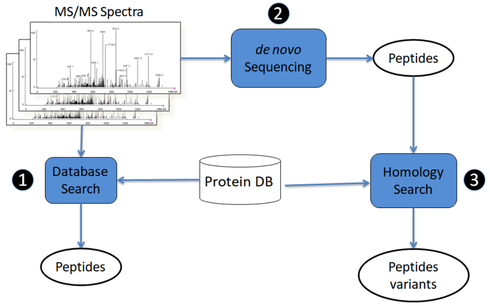

Fig1: **Workflow of Identification in PEAKS™** *de novo* sequencing-assisted database search provides in-depth peptide identification, including peptides in the specified database, peptides with un-suspected modifications, sequence variants and novel peptides

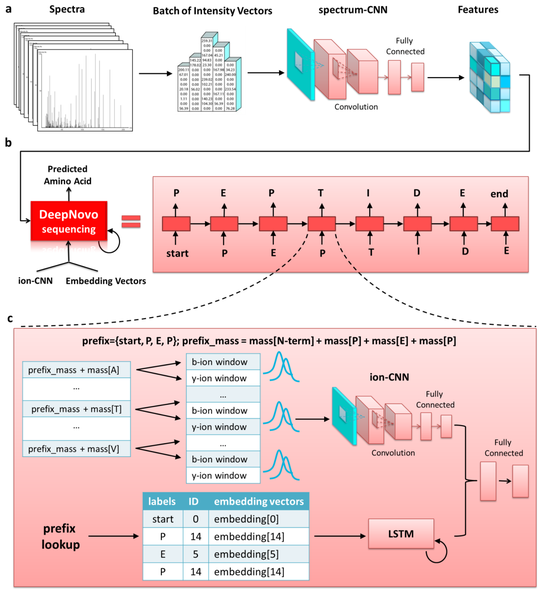

Fig2: **Deep Learning Model for *do novo* Sequencing**
Group input MS2 spectra into multi-dimensional vectors (m/z are vector index, intensity are vector value). Feed them into a convolutional neural network (CNN), output is features learned from MS2 spectra;

DeepNovo sequencing: predict one amino acid at a time. For each iteration, we predict the next amino acid based on the output of previous steps;

LSTM: long-short term memory network; RNN: recurrent neural network.

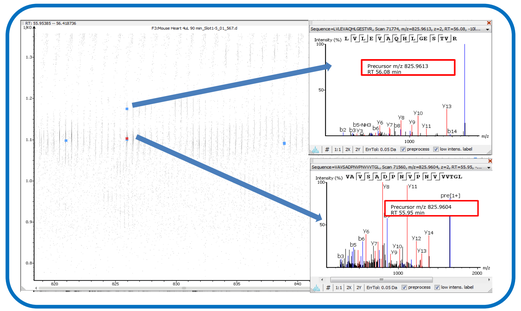

Fig3: **PEAKS™ display of PASEF ion-mobility separation** The two co-eluting parent ions (0.009 Da apart) have been separated in the ion mobility dimension prior to quadrupole isolation and fragmentation, revealing two distinct peptides which are successfully identified.

Results & Discussion

PASEF enables hundreds of MS/MS events per second at full sensitivity, and the resulting spectra are fully precursor mass resolved. The deep learning model was implemented in PEAKS product. High throughput DDA measurements acquired from 200 ng of a HeLa tryptic digest and separated using 90 - 120 min gradient length measured on a TIMS-QTOF (timsTOF Pro, Bruker Daltonics) using the PASEF method resulted in identifications of 65,000 – 70,000 unique peptides and 5000 to 5300 protein groups. 304651, 278724 and 273244 peptide features were detected in three technical replicate runs, and associated with 205483, 198404 and 195784 MS/MS spectra, respectively. At 1% FDR, 67323, 65985 and 65122 unique peptides were identified. Feature-based identification powered by deep learning enabled 1.2 identified peptides per MS/MS spectrum.
High throughput DDA measurements using the PASEF method acquired from yeast tryptic digest spiked with UPS2 proteins were used for label-free quantification experiments. Extremely high depth of 4 orders and excellent reproducibility (R^2 = 0.98) and high accuracy (15% of CV) in the label-free intensities could be achieved.

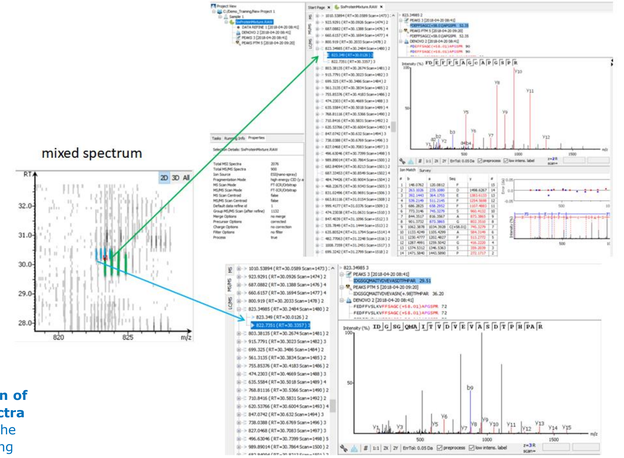

Fig4: **Identification of Mixture Spectra in PEAKS™** The two overlapping parent ions have been detected and associated one MS/MS spectrum, revealing two distinct peptides which are successfully identified.

Conclusions

- PASEF method and Deep Learning enabled the progress of in-depth proteomics analysis

timsTOF Pro