A New Algorithm for Peptide *de novo* Sequencing with Multiple Complementary Spectra

Lian Yang[1], Mingjie Xie[1], Bin Ma[2]
[1] Bioinformatics Solutions Inc, Waterloo, ON, Canada, [2] University of Waterloo, Waterloo, ON, Canada

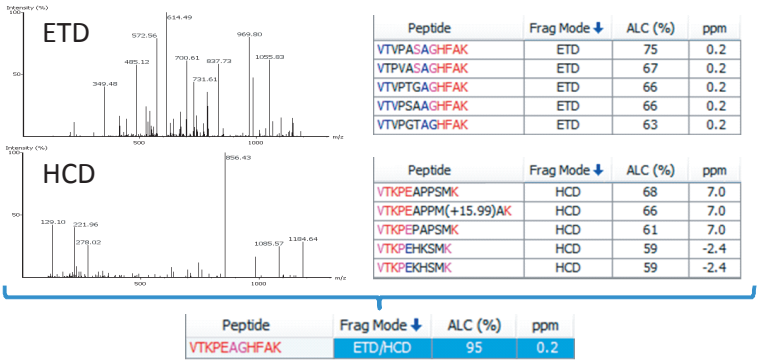
Overview

Different fragmentation methods produce complementary MS2 spectra of the same peptide. *De novo* sequencing software utilizing complementary tandem spectra can often significantly improve sequencing accuracy.

In this research, we proposed an algorithm to reconstruct a more confident sequence for spectral pair/triplet based on PEAKS *de novo* sequencing result of individual spectrum.

Introduction

De novo sequencing on complementary spectra often provides correct residues at different parts of the peptide. The algorithm uses PEAKS software to compute the top five candidate sequences for each spectra. Utilizing the PEAKS *local confidence score* assigned for residues, the algorithm selects confident *de novo* tags and assembles them into a new sequence. *De novo* tags shared by spectra of different dissociation methods are prioritized in the reconstruction process.



Peptide	Frag Mode	ALC (%)	ppm
VTVPASAGHFAK	ETD	75	0.2
VTPVASAGHFAK	ETD	67	0.2
VTVPTGAGHFAK	ETD	66	0.2
VTVPSAAGHFAK	ETD	66	0.2
VTVPGTAGHFAK	ETD	63	0.2

Peptide	Frag Mode	ALC (%)	ppm
VTKPEAPPSMK	HCD	68	7.0
VTKPEAPPM(+15.99)AK	HCD	66	7.0
VTKPEPAPSMK	HCD	61	7.0
VTKPEHKSMK	HCD	59	-2.4
VTKPEKHSMK	HCD	59	-2.4

Peptide	Frag Mode	ALC (%)	ppm
VTKPEAGHFAK	ETD/HCD	95	0.2

Method

The algorithm optimally selects a path of *de novo* tags to maximize the total tag score. Two types of tags are considered, ***single-residue tag*** and ***shared tag***.

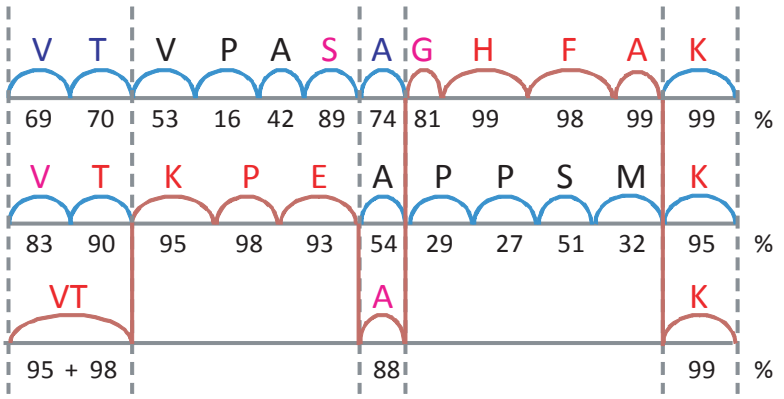
Single-residue tag - Single-residue tags are individual residues in a *de novo* sequence. We use the PEAKS *local confidence score* assigned to the residue as the tag score.

Shared tag - Shared tags are the common residues shared by a pair of individually generated *de novo* sequences. The tag score is elevated as $\Sigma(1-c_1)(1-c_2)$ for each shared residue.

Single-residue tags from ETD seq *i*

Single-residue tags from HCD seq *j*

Shared tags between seq *i* and *j*

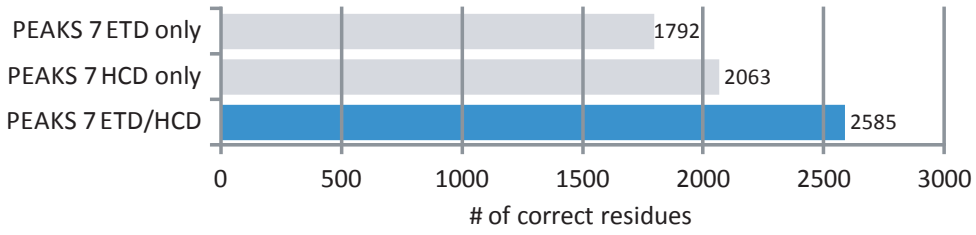


A dynamic programming algorithm is designed to efficiently calculate the optimal path, which translates to a reconstructed peptide sequence.

Result

Evaluation was performed on a dataset of 4641 pairs of ETD/HCD MS2 spectra, generated using Thermo Orbitrap Velos Pro. The ETD and HCD spectra in the dataset were respectively *de novo* sequenced using the PEAKS 7 software. Afterwards, the proposed algorithm reconstructs a new sequence for each pair using the single spectrum sequencing result.

387 out of the 4641 spectral pairs had both the ETD and HCD spectrum identified as the same sequence in database search. The 387 sequences, consisting of 5531 residues, were used to reference to examine the correctness of *de novo* sequencing result on the 387 spectral pairs. The number of correct residues was recorded.



Conclusion

- The research further confirmed that the use of complementary dissociation methods can improve the accuracy of *de novo* sequencing.
- The proposed algorithm correctly sequenced more residues, essentially provided a more reliable *de novo* sequencing result.

The described algorithm has been integrated into the PEAKS 7 software, allowing *de novo* sequencing with spectra pair/triplet.