Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

DeconMSn – A Software Tool for Determination of Accurate Monoisotopic Masses of Parent Ions of Tandem Mass Spectra

Anoop M. Mayampurath[1], Navdeep Jaitly[1], Samuel O. Purvine[1], Matthew E. Monroe[2], Kenneth J. Auberry[1], Joshua N. Adkins[2], and Richard D. Smith[2]
[1]Environmental & Molecular Sciences Laboratory, [2]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352

Overview

We present a new software tool for tandem MS analyses that:

- accurately calculates the monoisotopic mass and charge of high–resolution parent ions
- accurately operates regardless of the mass selected for fragmentation
- performs independent of instrument settings
- enables optimal selection of search mass tolerance for high mass accuracy experiments
- is open source and thus can be tailored to individual needs
- incorporates a SVM-based charge detection algorithm for analyzing low resolution tandem MS spectra
- creates multiple output data formats (.dta, .MGF)
- handles .RAW files and .mzXML formats
- compatible with SEQUEST, MASCOT, X!Tandem

Introduction

Peptide identification through tandem MS is a commonly used technique in proteomic-based research. Analyzing MS/MS fragmentation results from Thermo Fisher Scientific mass spectrometers currently involves using the program extract_msn (Figure 1) to create .dta file representations of spectra. Each .dta file contains a list of observed peaks in the MS/MS spectra and the corresponding parent ion charge and protonated mass, which extract_msn determines based on spectra characteristics.

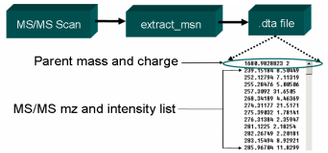


Figure 1: Thermo's software tool "extract_msn"

Problem: The instrument acquisition software can record the wrong parent monoisotopic mass (e.g., Figure 2). Thus, a .dta file is created with the monoisotopic mass +/-n amu from the correct mass (n being an integer).

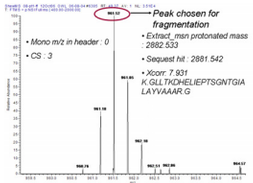
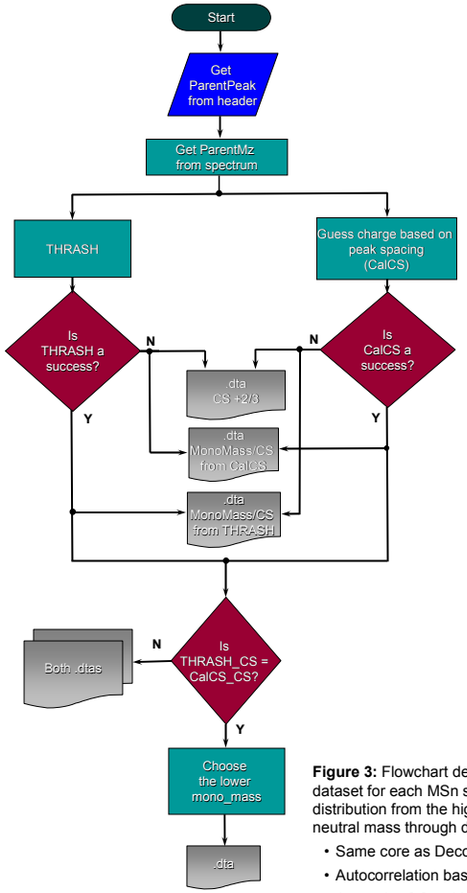


Figure 2: Example illustrating -1 amu difference between mono mass in dta and actual mono mass

Our solution: Use a combination of THRASH [1] and charge-based peak finding routines to deisotope the parent isotopic distribution.

Methods

High Resolution Precursor



Low Resolution Precursor

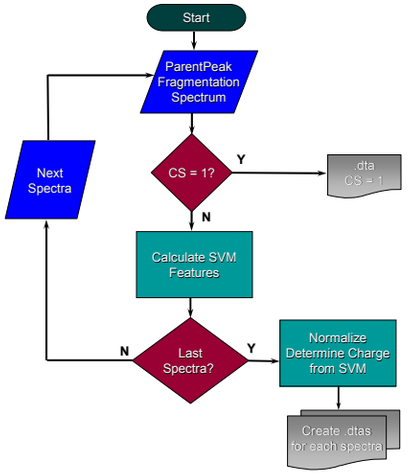


Figure 4 : Flowchart describing the DeconMSn algorithm for a low-resolution tandem MS dataset. DeconMSn overcomes the lack of peak profile information by incorporating a support-vector-machine-based charge detection algorithm that identifies the most likely charge of a parent ion.

- A trained support vector machine (SVM) [6] is used to assign a charge (+1, +2, +3, or +4) to a scan based on feature vector values
- Figure 5 shows the feature vector of 19 features based on [7] that is calculated for an example MS/MS scan
- Ambiguous spectra are assigned charge states +2 and +3

Figure 3: Flowchart describing the DeconMSn algorithm for high-resolution precursor dataset for each MSn spectrum. DeconMSn identifies the parent spectrum and m/z peak distribution from the high-resolution raw spectra, and then extracts the monoisotopic neutral mass through deisotoping, regardless of which mass was chosen for fragmentation

- Same core as Decon2LS [2] (written in C++ on .NET platform)
- Autocorrelation based peak fitting routines [3] are used for charge state detection.
- Averagine [4] and Mercury[5] used to create theoretical profiles which are matched to observed profiles
- CalCS determines charge by stepping from the parent peak by a distance calculated as 1.003/(cs) to determine the existence of a peak.

Parent_MZ	NET	Bscore2/3	Xscore2/3/ratio	Xscore2/3_Loss (H20, NH3, CO)	Peak Distribution [0:mz:2mz:3mz:4mz]	Fscore2/3
818.92	0.6	0.01/0.03	19794/0/0	….	0.4:0.5:0.02:0	9.1/5.2

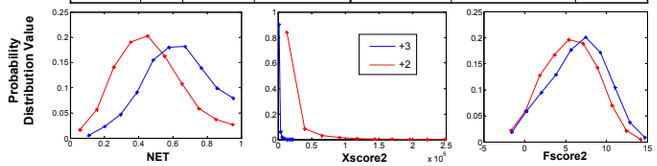
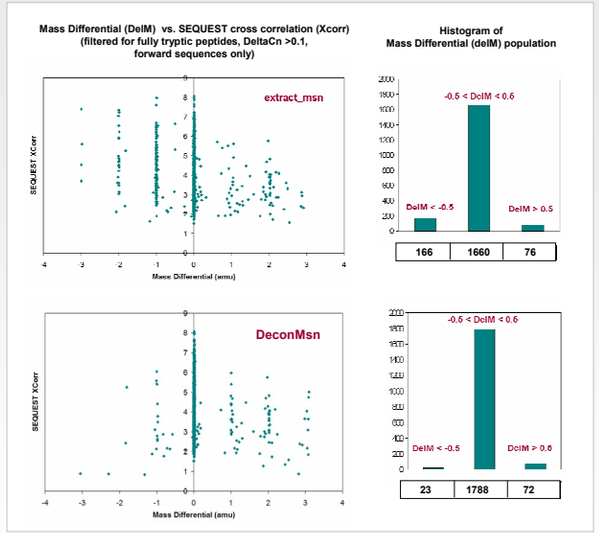


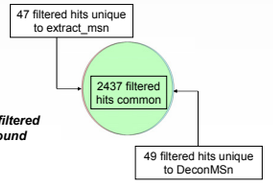
Figure 5 : Example probability distributions for selected feature vectors for each low-res MS/MS scan.

Results

High Resolution Precursor Spectrum



Low Resolution Precursor Spectrum



98% of all filtered peptides found

Conclusions

DeconMSn:

- Accurately determines the monoisotopic mass and charge of high resolution parent isotopic distributions
- Improves narrow mass tolerance-based SEQUEST searches
- Implements a SVM-based charge state detection algorithm to handle low resolution tandem mass spectra
- Incorporated as part of Decon2LS, which is available for download at **http://ncrr.pnl.gov/**

Acknowledgements

This research was supported by the National Institute of Allergy and Infectious Disease (NIAID) and the National Center for Research Resources (NCRR). Experimental portions of this research were performed at the Environmental and Molecular Sciences Laboratory (EMSL), a U.S. Department of Energy (DOE) national scientific user facility located at the Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract No. DE-AC05-76RLO 1830.

References

1. Horn, D.M., *et al.* Automated Reduction and Interpretation of High Resolution Electrospray Mass Spectra of Large Molecules. *J. Am. Soc. Mass Spectrom.* **2000,** *11*, 320-332.
2. Jaitly, N., *et al.* Open Source Tools for the Accurate Mass and Time (AMT) Tag Proteomics Pipeline, Proc. of ASMS **2006**, Seattle, WA.
3. Senko, M. W., *et al.* Automated assignment of charge states from resolved isotopic peaks for multiply charged ions. *J. Am. Soc. Mass Spectrom.* **1995,** *6*, 52–56.
4. Senko, M. W., *et al.* Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229–233.
5. Rockwood, A. L., *et al.* Rapid Calculation of Isotope Distributions. *Anal. Chem.* **1995**, *67*, 2699–2704.
6. Kanu S., *et. al. SVM And Kernel Methods Matlab Toolbox.* Perception Systemes et Information, INSA de Rouen, **2005** *France*
7. Klammer, A. A., *et al.* Peptide charge state determination for low-resolution tandem mass spectra. Proc. of 2005 IEEE Comp. Systems Bioinformatics. Conf. **2005** Stanford, CT.

Contact Information

Anoop M. Mayampurath
Environmental Molecular Sciences Laboratory
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
e-mail: anoop.mayampurath@pnl.gov
Phone: (509) 376-5267