

Microorganism Identification by MS/MS Typing Using Spectral Correlation Methods

Jane Razumovskaya; Sergey Kurnosenko; V. Sergey Rakov, Appravu K. Sundaram; Constance Murphy; Berk Oktem; Sue Shanbhag; Vladimir M. Doroshenko

Science & Engineering Services Inc., Columbia, MD



Science & Engineering Services, Inc.

INTRODUCTION

The masses of biological molecules reflect their composition and can be used for their detection and characterization. MS/MS typing using AP-MALDI Quadrupole Ion Trap MS/MS is a targeted proteomics method to rapidly detect multiple microorganisms in a complex biological sample. Detection of microorganisms is performed by analyzing gas-phase fragmentation of their unique species-specific peptide biomarkers. The biomarkers are unique (species-specific) peptides, whose MS/MS spectra were experimentally obtained and recorded into a Reference Library as etalon spectra.

Database Search Methods

Database Searches are proteome-wide methods to detect the presence of a particular peptide followed by organism detection in the sample, based on the best matching between experimentally measured spectrum and a range of candidate theoretical spectra derived from the peptides available in a sequence database.

- > **Pros**
 - > Allows to detect a wide range of microorganisms available in sequence database.

- > **Cons**
 - > Microorganism of interest must be present in the sequence database.
 - > Lacks detection sensitivity (requires good spectral quality, frequently high sample concentration, predictable fragmentation pattern dominated by 'b/y' ions).
 - > Does not easily utilize peptide fragmentation specifics (fragmentation methods, instruments, peptides).

MS/MS typing

MS/MS typing is a targeted proteomics method which performs peptide detection by comparing a previously recorded experimental (MS/MS) etalon spectrum of a species specific biomarker peptide to the current MS/MS spectrum.

- > **Pros**
 - > High sensitivity of detection (noise tolerant, can perform identifications on lower sample concentrations)
 - > Utilizes peptide fragmentation specifics
 - > Is not restricted to peptides
- > **Cons**
 - > Requires the presence of species-specific biomarker etalon spectrum in the Reference Library (MS/MS spectrum of biomarker peptide must be pre-recorded under similar conditions to the following detection experiment).

GOALS

- > Assess the expected FP (False Positive) rates for MS/MS typing method based on unrelated spectra and related spectra concepts.
- > Assess the FP rate for MASCOT for selected dataset.
- > Perform comparison between a Database Search engine (MASCOT) and an MS/MS typing method (using Spectral Correlation) for different sample concentrations.
- > Perform comparison between MASCOT and MS/MS typing on the same dataset, focusing on success rate of detection, as a function of sample concentration.

Materials and Methods

Experimental Materials:

- > Dataset includes 5 peptide mixture resulted from ovalbumin tryptic digest.
- > 9 different concentrations of the mixture were prepared following procedure presented in ASMS 2007 poster ThPJ 157 and analyzed using Thermo LCQ-DecaXP MS equipped with APMALDI source; the amounts of ovalbumin in each concentration are (in µg/ml):

co(PBS)	10	50	100	250	500	750	1000	1500
---------	----	----	-----	-----	-----	-----	------	------
- > 1 µl of each solution was deposited on MALDI plate and digested with trypsin (ThPJ 157). Thus, the amount of protein for MS and MS/MS analysis was of the order of ng/sample.
- > The analysis was repeated 10 times for each concentration.

Computational Materials

- > Database Analysis was performed using MASCOT v2.1.
- > MS/MS typing was performed by computing spectral correlation coefficients between two spectra.
- > Efficiency of detection was determined as a percent of peptides assigned correctly by each method, for each concentration.
- > The FP rate for correlation coefficient (assuming spectra of unrelated peptides) was computed using comparisons a) between each etalon spectrum and the spectra of other peptides, b) between each etalon spectrum and blank spectra (0 sample concentration).
- > The maximal FP rate for correlation coefficient was modeled by maximizing correlation coefficient between the etalon spectrum and set of hypothetical spectra of peptides of matching precursor mass found in the database (with exclusion of proline peak effect).

RESULTS

Experimental and theoretical etalons

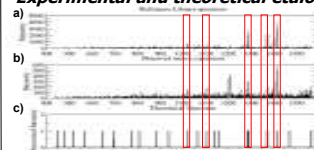


Figure1.

- a) Reference Library spectrum (etalon) recorded to represent a peptide.
- b) Noisy spectrum of that same peptide, which does not give an identification in database search
- c) 'b/y' fragmentation of the peptide used in common database searches.

Modeling Spectral Correlation FP rate

$$FP\ rate = \frac{FP}{FP + TN} * 100\%; \text{Corresponds to } (1 - \text{Specificity})$$

$$TP\ rate = \frac{TP}{TP + FN} * 100\%; \text{Corresponds to Sensitivity}$$

- > Assuming unrelated (physically independent) spectra (absence of any physical reasons to expect dependency or common features)

- > **Measured FP rate:** By calculating correlation between MS/MS spectra, for each peptide with the other 4 (unrelated) peptides [Figure2a, black line].
- > **Theoretical FP rate:** Based on asymptotically normal distribution of correlation coefficient, assuming physically independent spectra. [Figure2a, blue line].

- > Assuming related spectra (the two spectra may exhibit partial dependency/common features)

- > **"Worst case" FP-rate:** Maximal FP-rate for the correlation between MS/MS spectra of each peptide and the database peptides selected by all "incorrect" database search assignments (on the order of 3000 peptides/examined spectrum). [Figure2b]

Spectral Correlation FP rate

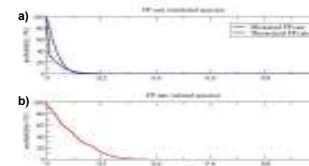


Figure2. a) Unrelated spectra: Measured and Theoretical FP rates of Spectral correlation are similar. b) Related spectra: FP rate based on maximal possible correlation coefficient between two spectra.

- > **Correlation coefficient cutoff for less than 1% FP identifications, selected based on the analysis for both unrelated spectra and related spectra is 0.35.**

Database Search Identifications

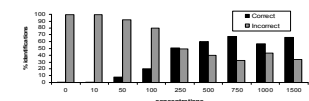


Figure3. Percent of correct and incorrect top-hit Mascot identifications (without score cutoff) as a function of sample concentration.

Database Search FP rate

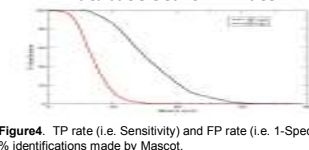


Figure4. TP rate (i.e. Sensitivity) and FP rate (i.e. 1-Specificity) vs % identifications made by Mascot.

- > **Mascot top score of 30 shows less than 1% FP identifications in the given dataset. The selected score of 30 is used for a cutoff in concentration comparison.**

MS/MS typing versus Database Search

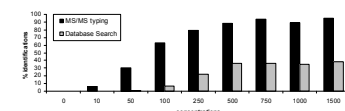


Figure5. Percent of identifications made at different sample concentrations by Spectral Correlation (with correlation coefficient cutoff of 0.35) and Mascot (with score cutoff of 30).

For concentrations between 500 and 1500 ng/sample, the probability of detection stays the same, showing that above 500 ng/sample, the increased sample concentration does not affect the detection (well above limit of detection). At these concentrations MS/MS typing shows the average of 92% probability of detection (92 out of 100 samples will be detected with this concentration range), while Mascot performed in average 36% probability of detection.

As the sample concentration is reduced, MS/MS typing is capable of detecting peptides in 79% and 63% of the tests for 250 and 100 ng/sample respectively. At the 50 ng/sample, current MS/MS typing capability is 30% of the tests, which is higher than Mascot's at 250 (22%) ng/sample.

CONCLUSIONS

- > This study outlines methodology for assignment of a probabilistic measure of organism identification by MS/MS.
- > MS/MS typing offers possibility of detection at lower concentration levels such as 50-100ng of protein per sample in context of current sample preparation protocol.
- > While database search will always be necessary for detection of unknown organisms without specific biomarkers, MS/MS typing offers a significant increase in speed and sensitivity in targeted proteomics for detection of specific bio-agents.

ACKNOWLEDGEMENTS

This work is funded by a contract from the US Army RDECOM Acquisition Center under the Joint PEO for Chemical and Biological Defense.

55th ASMS Conference, June 3-7, 2007, Indianapolis, IN