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

Strategies for obtaining confident identifications in high coverage, high throughput LC-MS proteomics measurements using hybrid FT instruments

Aleksey V. Tolmachev, Matthew E. Monroe, Ronald J. Moore, Samuel O. Purvine, Joshua N. Adkins, Gordon A. Anderson and Richard D. Smith
Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA

LTQ-Orbitrap LC-MS and IT MS/MS results are combined in a high throughput, confident proteomics strategy

Overview

- A new proteomics strategy explores the hybrid capabilities of the LTQ Orbitrap
- High resolution LC-MS data were combined with the Linear Ion Trap (LIT) MS/MS results obtained in the same LC-MS(/MS) experiment
- Three alternative approaches were evaluated for coverage and confidence of identifications
- Peptide false discovery events were nearly eliminated with a new hybrid approach that combined high resolution MS and LIT MS/MS data and applied elution time constraints from a previously generated accurate mass and time (AMT) tag database

Introduction

Hybrid FTMS instruments, such as the LTQ-FT and LTQ-Orbitrap, are capable of generating fast duty cycle linear ion trap MS/MS in parallel with high resolution mass spectra without compromising the overall throughput of measurements. We explore strategies for high throughput, high coverage proteomics measurements using the hybrid FTMS instruments.

Our accurate mass and time tag (AMT tag) strategy (**Fig. 1**) typically enables identification of thousands of peptides in a single LC-FTMS analysis by comparing accurate molecular mass and LC elution time information from the analysis to a reference database. An alternative strategy uses linear ion trap (low resolution) MS/MS identifications obtained using SEQUEST that is refined using an accurate precursor mass filter (APMF).

These two approaches were compared with a new hybrid approach that combines AMT tag identifications with MS/MS-derived peptide identifications.

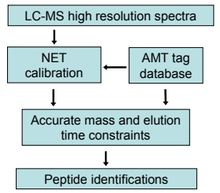

Fig. 1 AMT tag data processing chart

Methods

A *Shewanella oneidensis* global cell lysate tryptic digest solution was chosen for this study. The LTQ Orbitrap XL MS instrument was coupled to a 75 µm i.d. capillary LC column using 100 min LC separations, providing 2940 high resolution, high accuracy full MS spectra (400-2000 *m/z*, resolution of 100,000 each) along with lower resolution linear ion trap (LIT) MS/MS spectra for the 6 most abundant precursor ions, obtained in the linear ion trap in parallel with acquisition of the high resolution spectra. The LC-MS (MS/MS) dataset used here contained approximately 18,000 LIT MS/MS spectra.

Approach 1: AMT tag

High resolution LC-MS data matched to AMT tags

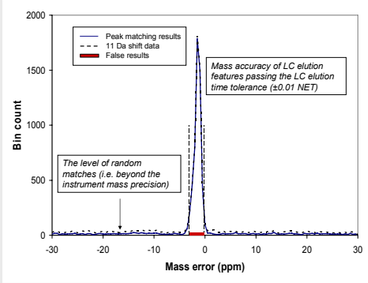

Fig. 2 Mass accuracy histogram of LC-MS features matched to AMT tags

AMT tag data processing (**Fig. 1**), was done as described previously [1, 2]. The AMT tag database for S. oneidensis included approximately 20K peptide AMT tags, filtered according to the peptide tag quality metrics. **Fig. 2** shows the mass accuracy histogram for LC-MS features matched to AMT tags; the vertical "Bin count" axis shows a number of matched LC-MS features per 0.5 ppm bin. Such a histogram can be used to estimate the false discovery rate (FDR), i.e. a percentage of random, or false, matches [1]. FDR ≈ 2% was obtained using a ratio of the shaded area to the peak area, which agreed with a control FDR estimation based on the decoy AMT tag database generated using 11Da shift [3], dashed curve in Fig. 2.

Approach 2: APMF

MS/MS identifications, filtered using the high resolution precursor masses

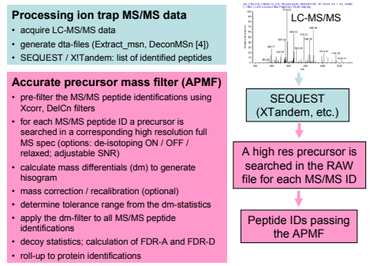

Fig. 3 APMF tag data processing chart

Fig. 4 Peptides identified in LC-MS and LC-MS/MS measurements

The APMF analysis provided 3577 peptide identifications compared to 5180 peptides found using the AMT tag approach, with 2659 of the peptides identified by the both approaches. 918 peptides unique to APMF can be also found by the AMT tag approach using an increased set of tags, ~100K, and relaxed constraints, which also produced increased FDR values, up to 20%.

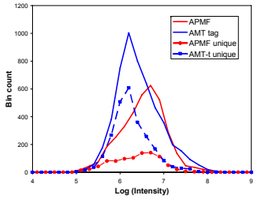

Fig. 5 Histogram of ion intensities

2521 peptides detected only by the AMT tag approach have on average lower intensities, indicating that the MS/MS based coverage is limited by under-sampling, i.e. insufficient number of MS/MS spectra. This issue will be addressed in further work.

Approach 3: PMTF

IT MS/MS (APMF), filtered using AMT tag elution times

The precursor mass and time filter (PMTF) approach combines AMT tag identifications with MS/MS-derived peptide identifications. In this approach, the low resolution MS/MS identifications are first filtered using the APMF, then matched to an AMT tag database in which they are additionally filtered using LC elution times.

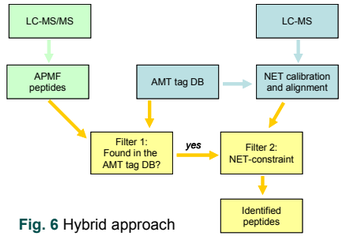

Fig. 6 Hybrid approach

LC-MS elution time data are aligned to AMT tag LC NET values for an improved confidence due to the LC elution time constraint. As a result of multiple filters, an increased number of AMT tags can be used (e.g., here 50K), with FDR values below 0.1%

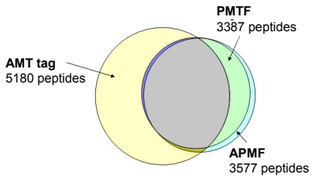

Fig. 7 Peptides identified with the three alternative strategies

3387 high confidence (FDR ≈ 0.07%) peptides identified with the PMTF approach include 3301 peptides also found by the AMT tag approach and 2702 peptides found by the APMF approach. 685 peptides identified with the PMTF approach and not reported in the AMT tag results can be attributed to a larger database used, 50K vs. 20K AMT tags. 276 APMF peptides not reported by the PMTF approach did not pass the AMT tag filter and/or the LC elution time filter.

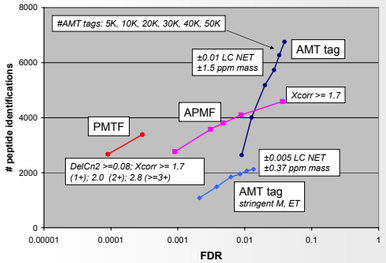

Fig. 8 Peptide identification coverage vs. FDR

Stringent filters used for the AMT tag approach reduced peptide coverage to <~2000 with improved FDR <1% (0.01 in absolute units in Fig. 8). Variable tolerances for SEQUEST scores, LC NET and mass accuracy produced the coverage vs. FDR variations for each of the three strategies. See [5] for a systematic approach for coverage vs. confidence optimization of AMT tag results.

Strategy	AMT tag	APMF	PMTF
# peptides found with regular database	5180	3577	3387
# peptides, composite decoy database (decoys subtracted)	5159	3507	3337
# decoy peptides: N_{decoy}	144	11	1
FDR from histogram areas: FDR_A	2.0%	0.27%	0.07%
FDR based on N_{decoy}: FDR_D	2.8%	0.31%	0.03%
# proteins with regular database	1177	978	977
# proteins with composite database, corrected	1173	967	969
# decoy proteins	133	11	1
# proteins with FDR < 10%	774 (2)	978 (1)	977 (1)
# proteins with FDR < 1%	432 (4)	571 (2)	977 (1)
# proteins with FDR < 0.1%	278 (6)	571 (2)	565 (2)
# proteins with FDR < 0.01%	231 (7)	385 (3)	565 (2)

Peptide and protein identifications obtained in a single high throughput LC-Orbitrap measurement are summarized in the above table. The four bottom rows show protein counts for specified levels of the probability of false identification, "protein FDR"; numbers in parentheses show the minimum count of peptides per one protein identification. The PMTF approach produced most confident peptide identifications, with one or fewer false peptides. Using 2 or more peptides per protein ID, it was possible to obtain 565 proteins with the protein FDR < 0.01%

Conclusions

- Three alternative strategies have been considered for bottom-up proteomics using hybrid instruments such as LTQ-FT and LTQ-Orbitrap: AMT tag for high resolution data, APMF for MS/MS IDs filtered using high-res precursor masses, and a new hybrid approach that employs the AMT tag database in processing combined MS and MS/MS results
- Peptide FDR estimated using the decoy and mass accuracy histogram approaches produced consistent results in the FDR range ~0.1% to >~10%
- The AMT tag approach produced the highest peptide coverage, >5000 IDs in a single LC-MS, with FDR >~1%
- The MS/MS-based accurate precursor mass filter (APMF) approach produced <~1% peptide level FDR; the coverage was reduced vs. AMT tag results, <~4000 peptides (i.e. 3577 peptides in the sample dataset used here)
- The new PMTF approach produced a marked peptide ID confidence improvement, FDR < 0.001, for 3387 peptide IDs obtained, which resulted in an increased quantity of confidently identified proteins

Acknowledgements

Portions of this research were supported by the U.S. Department of Energy Office of Biological and Environmental Research, the National Center for Research Resources (RR 018522), the National Institute of Allergy and Infectious Diseases (NIH/DHHS through interagency agreement Y1-AI-4894-01), and the National Institute of General Medical Sciences (NIGMS, R01 GM063883). The authors gratefully acknowledge the *Shewanella* Federation and DOE Genomics:GtL program for the standard *Shewanella* sample. Work was performed in the Environmental Molecular Science Laboratory, a DOE national scientific user facility located on the campus of Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.

References

1. Tolmachev AV, Monroe ME, Jaitly N, Petyuk VA, Adkins JN, and Smith RD. Mass measurement accuracy in analyses of highly complex mixtures based upon multidimensional recalibration. *Anal. Chem.* **2006**, *78*: 8374-8385.
2. Monroe ME, Tolić N, Jaitly N, Shaw JL, Adkins JN, and Smith RD. VIPER: an advanced software package to support high-throughput LC-MS peptide identification. *Bioinformatics* **2007**, *23*:2021-2023.
3. Petyuk VA, Qian W-J, Chin MH, Wang H, Livesay EA, Monroe ME, Adkins JN, Jaitly N, Anderson DJ, Camp DG, Smith DJ, and Smith RD. Spatial mapping of protein abundances in the mouse brain by voxelation integrated with high-throughput liquid chromatography–mass spectrometry. *Genome Res.* **2007**, *17*: 328-336.
4. Mayampurath AM, Jaitly N, Purvine SO, Monroe ME, Auberry KJ, Adkins JN, and Smith RD. DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. *Bioinformatics* **2008**, *24*:1021-1023.
5. Jaitly N, Adkins JN, Dabney AR, Monroe ME, Norbeck AD, Mottaz HM, Lipton MS, Anderson GA, and Smith RD. A Statistical Approach to Quantifying Uncertainties in the AMT Tag Pipeline. ASMS, ThOA pm.

Contact Information
Aleksey Tolmachev, PhD
Biological Sciences Division, K8-98
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
e-mail: tolmachev@pnl.gov