

# Optimization of a LC-FTMS proteomics pipeline for high throughput and confident peptide identifications

Ronald J. Moore, Aleksey V. Tolmachev, Anil K. Shukla, Therese R.W. Clauss, Rui Zhang, David J. Anderson, Karl K. Weitz, Brianne O. Petritis, and Richard D. Smith  
Pacific Northwest National Laboratory, Richland, WA



Pacific Northwest  
NATIONAL LABORATORY

## Overview

High resolution FT-MS instruments (FT-ICRs and Orbitraps), coupled to HPLC, constitute the core of our proteomics data production facility. Here, we report on the quality control and system optimization procedure that ensures high quality and reproducibility and assists in upgrades and enhancements of the high resolution MS platforms and chromatography systems.

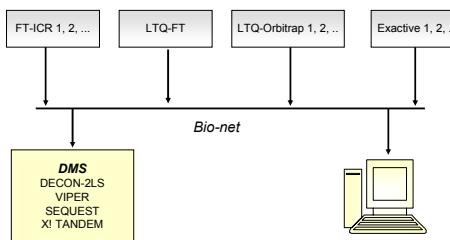


Fig. 1. High resolution LC-MS platforms connected via network for automated data processing and QC

## Introduction

The high throughput proteomics facility at PNNL uses custom built automated HPLC systems that are coupled to several high resolution FTMS instruments, including an 11T FTICR developed in-house, a customized Bruker 9T FTICR, a Thermo LTQ-FT, three Thermo LTQ-Orbitraps, and Thermo's Exactive Orbitrap. In addition to a routine 24/7 work load to analyze samples from a diverse range of organisms, technological improvements are periodically implemented to improve throughput and/or data quality. The optimization process involves an automated procedure for extensively characterizing quality control (QC) samples used to monitor instrument performance. This procedure assesses instrument performance by evaluating data quality for biological studies and provides metrics for assessing instrument optimizations and upgrades.

## Methods

The QC procedure is based on the accurate mass and time (AMT) tag approach developed in our laboratory [1]. Using this approach with our standard LC-MS parameters, analysis of our QC sample (*Shewanella oneidensis* tryptic digest) typically results in a set of ~3000 high resolution, high accuracy mass spectra. For rapid turn around time and thorough evaluation of instrument performance, a fully automated software package, ProDAT, was developed to process this QC data and provide a number of metrics for assessment of system performance. This application uses a subset of data from a previously established *S. oneidensis* peptide database that contains >40,000 AMT tags.

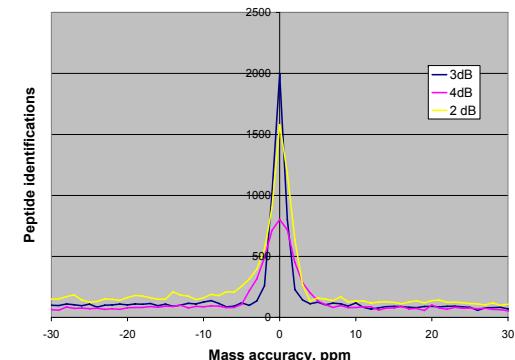


Fig. 2. Mass accuracy histogram for QC peptides identified in LC-MS with 11T FT-ICR, using 3 different settings for the ion excitation power attenuation, dB.

Each LC-MS data set is recalibrated [2], to determine the external calibration quality and make coverage results less sensitive to the instrument calibration. The results include a list of detected LC features and a corresponding list of peptides found in the AMT tag database. A statistical analysis of the collection of high resolution mass spectra obtained during an LC separation produces the mass accuracy and precision. Identification confidence is assessed based on the mass accuracy histogram base line level [3]. Experimental LC elution times are aligned with normalized elution time (NET) data from the AMT tag database. The alignment quality is characterized in terms of the NET precision, calculated as the baseline width of the histogram of NET deviations.

## Results

A summary table is generated that includes characteristics of the peptide coverage and quality of the data (i.e., the number of elution features and number of identified peptides). Also reported are mass measurement accuracy and precision, chromatographic elution time precision, and the false discovery rate (FDR). The table also includes key instrument settings, e.g., excitation power, trapping voltage and external accumulation time for custom FT-ICR instruments. This provides a basis for optimization of multiple instrument settings based on a series of LC-FTMS measurements. Additionally, sample concentrations and AGC targets are optimized to maximize coverage and confidence of identifications. As a result, the procedure provided stable operation of all LC-FTMS proteomics platforms, with overall performance gradually improving. The QC analyses generate ~10,000 peptide identifications per LC separation, with higher than 90% identification confidence (i.e. FDR < 0.1).

## Monitoring Performance of High Resolution LC-MS Platforms: LTQ-Orbitrap XL

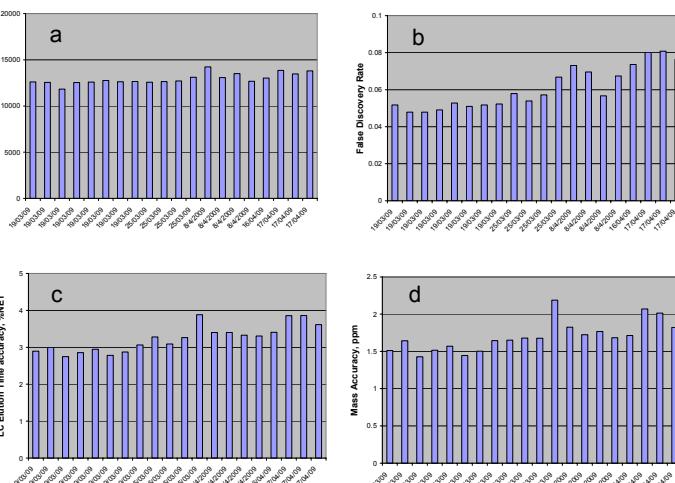


Fig. 3. QC results for LTQ-Orbitrap-XL. Reproducible coverage, ~14,000 unique peptide identifications, was observed over ~1 month time span (a). However, the identification confidence drifted to higher FDR values (b). This was explained by deterioration of the LC elution time accuracy (c) and a slow increase in the mass measurement errors (d). Accordingly, service of LC columns and AGC target recalibration were done to address the issues and restore optimal performance.

## Example Application with Exactive Orbitrap

The Exactive is a benchtop Orbitrap instrument that was recently optimized for application studies using the AMT tag approach. The following results were obtained using a series of *Shewanella* QC LC-MS measurements, with 3 different AGC targets: balanced ( $1 \times 10^6$ ), high Dynamic Range ( $3 \times 10^6$ ), and a low AGC setting ( $3 \times 10^5$ ).

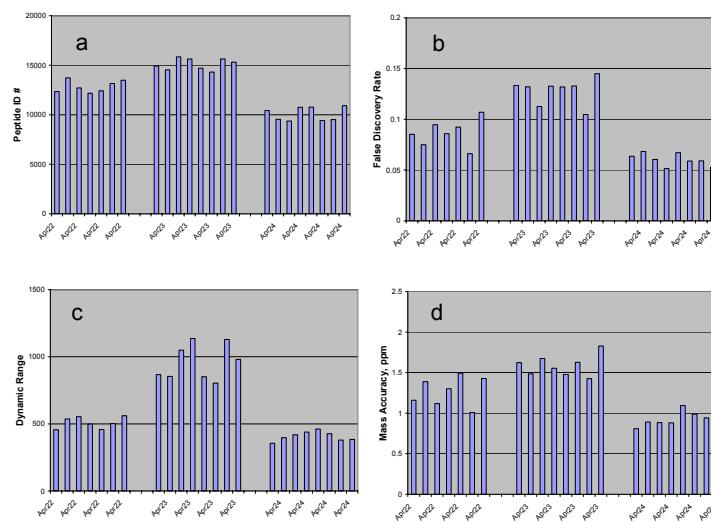


Fig. 4. QC results for Exactive. AGC target settings are balanced ( $1 \times 10^6$ , Apr22), high Dynamic Range ( $3 \times 10^6$ , Apr23), and a low setting ( $3 \times 10^5$ , Apr24). The peptide coverage (a) varied from ~10,000 to 15,000. The high AGC target (3E6) produced highest dynamic range, as reported by ProDAT (b). However, the best FDR values were obtained with the lowest AGC target (c). This is explained by mass measurement accuracy markedly improving with the low ion population setting (d). As a result, a range of operation modes is formulated for application studies, aimed at either highest possible coverage/sensitivity, or maximized identification confidence.

## References

- Smith, R. D., Anderson, G. A., Lipton, M. S., Paša-Tolić, L., Shen, Y., Conrads, T. P., Veenstra, T. D. and Useth, H. R. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* 2002, 2: 513–523.
- A. V. Tolmachev, M. E. Monroe, N. Jaitly, V. A. Petyuk, J. N. Adkins, and R. D. Smith. Mass measurement accuracy in analyses of highly complex mixtures based upon multidimensional recalibration. *Anal. Chem.* 2006, 78: 8374–8385.
- A. V. Tolmachev, M. E. Monroe, S. O. Purvine, R. J. Moore, N. Jaitly, J. N. Adkins, G. A. Anderson and R. D. Smith. Characterization of strategies for obtaining confident identifications in bottom-up proteomics measurements using hybrid FT MS instruments. *Anal. Chem.* 2008, 80: 8514–8525.

## Conclusions

- High resolution LC-MS platforms require regular monitoring using rapid and reliable QC procedures
- ProDAT software automatically processes LC-MS data and generates major performance characteristics
- QC results acquired over 4 years provide a comparison of up to 7 high resolution platforms
- The QC procedure provides an approach for fine-tuning and optimizing all steps involved in LC-MS measurements, including choosing sample concentration, LC configuration, electrospray conditions, ion source, and mass analyzer settings.
- As a result of the gradual improvement, reproducibly for up to ~15,000 peptide identifications in a single LC-MS measurement is possible with ~1.5 ppm mass accuracy (Exactive, FWHM over one LC-MS).

## Acknowledgements

This work was funded by the National Center for Research Resources (RR 018522), the National Institute of Allergy and Infectious Diseases (NIH/DHS) through interagency agreement Y1-AI-4894-01, the National Institute of General Medical Sciences (NIGMS, R01 GM063883), and the U. S. Department of Energy Office of Biological and Environmental Research (DOE/BER).

Samples were analyzed using capabilities developed under the support of the NIH National Center for Research Resources and DOE/BER. Significant portions of the work were performed in the Environmental Molecular Science Laboratory, a DOE/BER national scientific user facility at Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is operated for the DOE by Battelle under contract DE-AC05-76RLO-1830.

**CONTACT:** R. J. Moore  
Biological Sciences Division, K8-98  
Pacific Northwest National Laboratory  
P.O. Box 999, Richland, WA 99352  
E-mail: ronald.moore@pnl.gov