Automated diagnosis of LC-MS/MS performance

Hua Xu1,* and Michael A. Freitas2,*

1Proteomics and Informatics Services Facility, Research Resources Center, University of Illinois at Chicago, Chicago, IL 60612 and 2Department of Molecular Virology, Immunology and Medical Genetics, the Ohio State University, Columbus, OH 43210, USA

Received on December 1, 2008; revised on February 21, 2009; accepted on March 15, 2009

ABSTRACT

Summary: We report a software scheme for automated diagnosis of liquid chromatography tandem mass spectrometry (LC-MS/MS) system performance. The proposed software scheme provides a robust framework for establishing automated diagnosis of LC-MS/MS system performance for a variety of instruments and experiments. This schematic consists of four main software components: (i) data conversion, (ii) peptide identification, (iii) LC retention time analysis and (iv) system performance evaluation. The implementation of a standard approach for assessing LC-MS/MS system performance enables researchers to apply reliable metrics to assess their workflows performance over different batch experiments. Furthermore, the results from system diagnosis can provide feedback to the workflow to stop batch analysis if system performance falls below prescribed thresholds. A basic implementation of the approach based on the MassMatrix database search and LC retention time analysis programs is presented.

Availability: An open source implementation of the LC-MS/MS system diagnosis software based on the MassMatrix database search program is freely available to non-commercial users and can be downloaded at www.massmatrix.net.

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) is one of the most commonly used methods to identify and characterize peptides and proteins in complex samples (Aebersold and Mann, 2003). The experimental MS/MS spectra are most commonly transformed by use of database search software into peptide and protein identifications (Sadygov et al., 2004). In some cases, the LC retention times for peptides are also used as complementary information for peptide identification and characterization (Shinoda et al., 2008). Standard protein digests or peptide mixtures are routinely used to calibrate the LC-MS/MS system. Results from standards are manually inspected and/or searched by peptide identification software to determine system performance. In an environment where system performance requires constant monitoring, automated diagnosis of system performance would be highly desirable to allow for rigorous quality assessment and control. The need for ‘rules and prerequisites’ to determine the ‘technical and biological reproducibility, comparability and statistical robustness of the data’ is of clear importance in establishing quality control in clinical/translational proteomics experiments (Mischak et al., 2007).

This application note describes an approach for on-the-fly automated diagnosis of LC-MS/MS system performance. The scheme is designed to evaluate the quality of the experimental LC separation and MS instrument performance for standard protein mixtures. A decision whether or not to process a batch based on instrument performance can be made based on the results of the system performance diagnosis. These results also provide metrics to assess performance of the proteomic workflow over time. The proposed scheme was designed such that it can be incorporated into most standard proteomics workflows and can provide automated system diagnosis. A basic implementation of the approach for protein identification by LC-MS/MS is demonstrated.

2 DESCRIPTION OF APPROACH AND IMPLEMENTATION

The flow diagram for the proposed software scheme describing automated LC-MS/MS system diagnosis is shown in Figure 1a. The scheme contains four main software components: data conversion, peptide identification, peptide retention time analysis and statistical result evaluation. In this approach, the raw data file for a reference protein standard is acquired on the mass spectrometer and converted to an appropriate data format (e.g., mzXML, mzData, etc.) by the data conversion module. Data conversion programs are freely available online and are also provided by the mass spectrometer manufacturers. The converted data file is then submitted to peptide/protein identification software, the most common of which are database search algorithms (Sadygov et al., 2004). The identified peptides are further evaluated by a peptide retention module (Xu et al., 2008). The results from peptide/protein ID and peptide retention time analysis are then input into a statistical result evaluation module to determine overall LC-MS/MS system performance. The results from the final system performance evaluation are then used to determine if a batch of experiments can proceed or abort. By evaluation of both the MS/MS and LC retention times, problems with both the LC and MS can be diagnosed.

An implementation of this scheme based on the MassMatrix database search program and MassMatrix LC retention time analysis is described herein. MassMatrix was chosen because it is freely
available and capable of performing the database search and peptide LC retention time analysis (Xu et al., 2008). Furthermore, the MassMatrix search engine has been systematically validated against other database search programs (Xu and Freitas, 2007, 2008). The flow diagram for the implementation is shown in Figure 1b. The current implementation was designed to diagnosis data from LCQ, LTQ, LTQ/Orbitrap and LTQ-ETCR mass spectrometers (Thermo Fisher, CA, USA). The diagnosis of other mass spectrometers only differs in the process of converting raw data files to mzXML data files. The diagnosis program can be easily configured to use different data conversion programs for those mass spectrometers. Much of our proteomic experiments are performed with limited samples, which requires that system performance be established prior to batch sample runs.

In our automated system diagnosis approach, the following parameters for the MassMatrix software modules are required at time of configuration: (i) the standard protein database; (ii) the enzyme used during protein digestion and maximum number of missed cleavages; and (iii) mass accuracies for precursor and product ions. This configuration process is done by use of a configuration program in the system diagnosis program package. Once a standard configuration is created, LC-MS/MS system diagnosis is then performed without supervision. To establish strict quality process control a protein standard is placed first and last in the sample queue and after every 10th sample to assess system performance for the entire batch. Once a protein standard is run, a post-run script is initiated to diagnosis system performance. During this process the raw data file from the mass spectrometer is converted to an mzXML file by use of ReAdW (http://tools.proteomecenter.org/ReAdW.php). The mzXML file is then searched with MassMatrix to obtain both the peptide identification and the predicted relative retention time reports. These processes are automated by use of Python scripts in the system diagnosis program package (provided as Supplementary Material). The Python scripts also parse the search results from MassMatrix in html format to obtain values from the LC-MS/MS diagnosis variables.

Poor performance of a mass spectrometer for LC-MS/MS normally causes low sensitivity or low fragmentation efficiency resulting in fewer high-quality peptide MS/MS spectra. Therefore, the mass spectrometer performance can be evaluated by two diagnosis variables from the database search results of non-complex protein standard samples: (i) the number of high-quality peptide matches (p.p. score > 8.0 and p.p. score > 3.0 in MassMatrix), and (ii) the sequence coverage of the protein standards based on those high-quality peptide matches. Poor performance of the LC system results in insufficient or no separation of the peptides. This results in a poor correlation between observed and predicted peptide retention times based on the peptide’s calculated hydrophobicity. This quality of the LC system performance can be evaluated with two system diagnosis variables obtained from the retention time analysis in MassMatrix: (i) the coefficient of determination of the retention time prediction model, and (ii) the percentage of peptides with observed retention times within the 99% confidence bands of the predicted retention times (Xu et al., 2008). The system diagnosis program also provides an assessment of mass error for the mass spectrometer by calculating the mean value and 95% confidence interval of mass errors for high-quality peptide matches. A systematic mass shift significantly larger than the configured critical value issues a warning in the diagnosis report and suggests a poor calibration of the mass spectrometer.

The whole diagnosis process is run without human supervision and the status of the system is reported in five ranks: poor, fair, good, very good and excellent based on the set of criteria for the diagnosis variables. The criteria for the diagnosis variables can be configured in the program. Example settings of diagnosis criteria for system diagnosis of a LCQ mass spectrometer by use of the tryptic digest of bovine histones and a LTQ-FT ICR mass spectrometer by use of the tryptic digest of a bovine serum albumin (BSA) protein standard are shown in Supplementary Table 1. System performance that falls outside the range of ‘good’ to ‘excellent’ results in termination of the batch experiment and requires that the analyst manually verify the quality of the data files in that batch. For Thermo Fisher’s instruments, the Xcalibur data acquisition system is setup to run the system diagnosis program post acquisition. The script is run synchronously and can block the queue from proceeding if the standard falls outside the established criteria. In addition, future versions will also include additional modules and setup instructions for different manufacturers as well as email notifications to alert users of system performance problems. Continuous monitoring of system performance criteria is useful in setting thresholds for routine mass spectrometer maintenance/calibration and replacement/reconditioning of chromatographic columns. We have included an open-source version of the diagnosis program for evaluation. The program is modular and will allow researchers with different criteria to easily adapt the approach for the necessary data conversion, protein ID and evaluation metrics of their choice.

![Flow diagram for the implementation of MassMatrix database search program.](Image)

**Fig. 1.** (a) The scheme for automated LC-MS/MS system diagnosis, (b) the implementation of the scheme based on MassMatrix database search program.
Automated diagnosis of LC-MS/MS performance

**Funding:** Ohio State University, National Institutes of Health (CA107106, RR023647, CA101956), V Foundation (AACR Translational Cancer Research Grant); Leukemia & Lymphoma Society (SCOR).

**Conflict of Interest:** none declared.

**REFERENCES**


