BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 21 2013, pages 2804–2805
doi:10.1093/bioinformatics/btt465

Data and text mining

Advance Access publication September 5, 2013

LC-IMS-MS Feature Finder: detecting multidimensional liquid chromatography, ion mobility and mass spectrometry features in complex datasets

Kevin L. Crowell[1], Gordon W. Slysz[1], Erin S. Baker[1], Brian L. LaMarche[2], Matthew E. Monroe[1], Yehia M. Ibrahim[1], Samuel H. Payne[1], Gordon A. Anderson[1] and Richard D. Smith[1,*]

[1]Biological Sciences Division, Pacific Northwest National Laboratory, 999 Battelle Boulevard, Richland, WA 99352, USA and [2]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, 999 Battelle Boulevard, Richland, WA 99352, USA

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: The addition of ion mobility spectrometry to liquid chromatography-mass spectrometry experiments requires new, or updated, software tools to facilitate data processing.

Results: We introduce a command line software application LC-IMS-MS Feature Finder that searches for molecular ion signatures in multidimensional liquid chromatography-ion mobility spectrometry-mass spectrometry (LC-IMS-MS) data by clustering deisotoped peaks with similar monoisotopic mass, charge state, LC elution time and ion mobility drift time values. The software application includes an algorithm for detecting and quantifying co-eluting chemical species, including species that exist in multiple conformations that may have been separated in the IMS dimension.

Availability: LC-IMS-MS Feature Finder is available as a command-line tool for download at http://omics.pnl.gov/software/LC-IMS-MS_Feature_Finder.php. The Microsoft.NET Framework 4.0 is required to run the software. All other dependencies are included with the software package. Usage of this software is limited to non-profit research to use (see README).

Contact: rds@pnnl.gov

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on June 5, 2013; revised on July 16, 2013; accepted on August 6, 2013

1 INTRODUCTION

In conventional proteomics, liquid chromatography (LC) separations are generally coupled with mass spectrometry (MS) to increase analytical sensitivity and proteome coverage (Washburn *et al.*, 2001). More recently, ion mobility spectrometry (IMS) has been added as an extra dimension of separation to further enhance sensitivity (Baker *et al.*, 2010; Lee *et al.*, 2002; Liu *et al.*, 2004). Although several software programs are available for deisotoping and/or grouping related peaks (Bertsch *et al.*, 2011; Deutsch *et al.*, 2010; Monroe *et al.*, 2007) to aid in the identification of species (typically peptides or metabolites), none are currently able to process LC-IMS-MS data. In this regard, we introduce LC-IMS-MS Feature Finder, a software application that enables researchers to discover possible molecular ion signatures in LC-IMS-MS data by reporting characteristic features such as monoisotopic mass, charge state, LC elution time and IMS drift time for use by downstream identification software (Lamarche *et al.*, 2013).

2 IMPLEMENTATION

LC-IMS-MS Feature Finder is a command line software application for clustering deisotoped peaks in LC-IMS-MS data. The workflow for this software application is schematically depicted in Figure 1. The preferred input file for the software application is a DeconTools (Jaitly *et al.*, 2009; Slysz *et al.*, 2010) (http://omics.pnl.gov/software/DeconTools.php) 'isos' file format. This comma-separated value file contains information about deisotoped features from individual spectra for a single instrument analysis, including monoisotopic mass, m/z, charge state, LC elution time and IMS drift time. A raw instrument data file is also required. Currently, LC-IMS-MS Feature Finder only supports instrument files in Unified Ion Mobility Frame (Beagley *et al.*, 2009) file format, which is a SQLite database format created at Pacific Northwest National Laboratory similar to the YAFMS (Shah *et al.*, 2010) file format. We note that files from commercially available IMS instruments are unsupported at this time because of their unreleased format specification. However, once corresponding application programming interfaces become available for commercial file formats, the LC-IMS-MS Feature Finder will be updated to support them.

In LC-IMS-MS data, multiple charges states of the same species have distinct IMS drift times. To take advantage of this fact, the LC-IMS-MS Feature Finder partitions the data by charge state to allow it to be processed independently. By partitioning the data and using the Microsoft.NET 4.0 Task Parallel Library (TPL), multiple data partitions are processed on separate computing cores. The TPL automatically detects the number of available computing cores and decides how to use them to process the data in parallel. Processing the data in a parallel fashion during the most computationally intensive steps significantly decreases the overall runtime. Moreover, by using the TPL, parallelization is not hard coded in the software or part of the user-specified input; instead, it is adaptable to any computing core structure. This adaptability is especially important, as the IMS dimension adds an extra level of complexity and significantly increases the computational burden compared with conventional LC-MS data processing.

*To whom correspondence should be addressed.

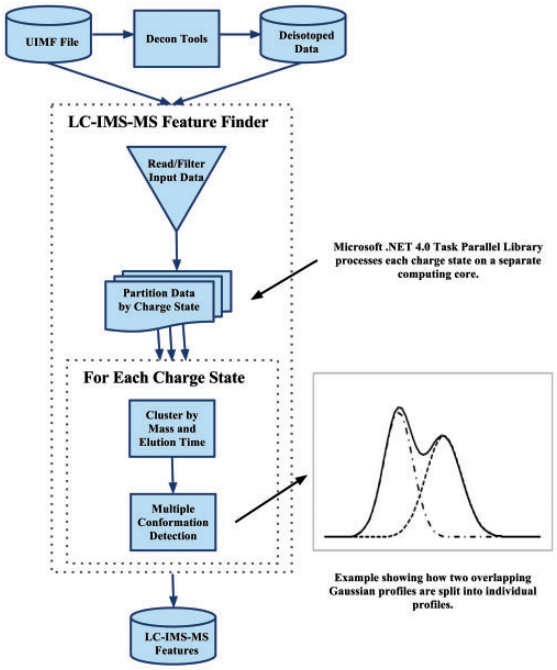

Fig. 1. A flow chart of the LC-IMS-MS Feature Finder software application. The input data are the raw spectra in UIMF format and deisotoped data. The deisotoped data are clustered by monoisotopic mass and LC elution time and tested for multiple conformations to create LC-IMS-MS Features

An important component of LC-IMS-MS Feature Finder is the ability to detect multiple conformations of a chemical species. The algorithm is particularly useful when (i) a single molecular gas-phase ion species exists in multiple structural conformations that have distinct IMS drift times and (ii) two different molecular ion species co-elute in the IMS dimension. In these cases, the intensity profiles are frequently not base-line resolved but instead form a set of overlapping peaks, which hinders the ability to accurately quantify each individual intensity profile. To ensure the most accurate representation of these profiles, LC-IMS-MS Feature Finder returns to the raw data file to collect raw intensity values. Returning to the raw data is essential to get the most accurate representation of these profiles. The mixture of profiles is split into distinct intensity profiles that will be quantified and reported separately in the output file. The individual intensity profiles reported are resolved using both the LC and IMS dimensions, therefore creating distinct LC-IMS-MS Features. Future versions of LC-IMS-MS Feature Finder will explore the possibility of resolving which of the two cases give rise to the multiple conformations.

The software application outputs a tab-separated value text file that contains associated information about each molecular ion signature detected: monoisotopic mass, m/z, charge state, LC elution time, IMS drift time and abundance. The final monoisotopic mass and m/z values are weighted averages (based on intensity) of the monoisotopic mass and m/z values of each deisotoped feature associated with a given signature. The LC elution time and IMS drift time values are the time points at which the signature exhibits the highest intensity. The output file can be used by downstream software to cross-reference the detected molecular ion species with a database of known ions to aid in identification of the ions contained in the dataset. More details about the software input, output and configurable parameters are available in the supplementary information.

3 CONCLUSIONS

LC-IMS-MS Feature Finder affords a stand-alone console application that uses raw LC-IMS-MS data and deisotoped features contained in the raw data to discover molecular ion signatures. Results in the form of a flat text file can readily be imported into Excel or other data analysis tools for further analysis, as well as associated with a database of known molecular ion species by using other downstream software applications. LC-IMS-MS Feature Finder has been integrated into the accurate mass and time (AMT) tag data processing pipeline for LC-IMS-MS datasets at Pacific Northwest National Laboratory (Crowell *et al.*, 2013). The utility of LC-IMS-MS Feature Finder extends to a broad range of studies, including, but not limited to, proteomics, metabolomics and lipidomics.

Funding: National Institute of General Medical Sciences (8 P41 GM103493-10), the Environmental Research Genome Sciences Program under the Pan-omics project, the National Institute of Health (R01ES022190) and by the U.S. Department of Energy (DOE)/BER Research Genome Sciences Program. 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 in Richland, Washington. PNNL is operated for the DOE by Battelle under Contract DE-AC05-76RLO1830.

Conflict of Interest: none declared.

REFERENCES

Baker,E.S. *et al.* (2010) An LC-IMS-MS platform providing increased dynamic range for high-throughput proteomic studies. *J. Proteome Res.*, **9**, 997–1006.

Beagley,N. *et al.* (2009) Increasing the efficiency of data storage and analysis using indexed compression. *e-Science, 2009. e-Science'09 Fifth IEEE International Conference on IEEE*. Oxford, UK, pp. 66–71.

Bertsch,A. *et al.* (2011) OpenMS and TOPP: open source software for LC-MS data analysis. *Methods Mol. Biol.*, **696**, 353–367.

Crowell,K.L. *et al.* (2013) Increasing confidence of LC-MS identifications by utilizing Ion mobility spectrometry. *Int. J. Mass Spectrom.* [Epub ahead of print, doi.org/10.1016/j.ijms., June 28, 2013].

Deutsch,E.W. *et al.* (2010) A guided tour of the Trans-Proteomic Pipeline. *Proteomics*, **10**, 1150–1159.

Jaitly,N. *et al.* (2009) Decon2LS: an open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics*, **10**, 87.

LaMarche,B.L. *et al.* (2013) MultiAlign: a multiple LC-MS analysis tool for targeted omics analysis. *BMC Bioinformatics*, **14**, 49.

Lee,Y.J. *et al.* (2002) Development of high-throughput liquid chromatography injected ion mobility quadrupole time-of-flight techniques for analysis of complex peptide mixtures. *J. Chromatogr. B*, **782**, 343–351.

Liu,X. *et al.* (2004) Development of high throughput dispersive LC-ion mobility-TOFMS techniques for analysing the human plasma proteome. *Brief. Funct. Genomic Proteomic*, **3**, 177–186.

Monroe,M.E. *et al.* (2007) VIPER: an advanced software package to support high-throughput LC-MS pepide identification. *Bioinformatics*, **23**, 2021–2023.

Shah,A.R. *et al.* (2010) An efficient data format for mass spectrometry-based proteomics. *J. Am. Soc. Mass Spectrom.*, **21**, 1784–1788.

Slysz,G.W. *et al.* (2010) *The DeconTools Framework: An Application Programming Interface Enabling Flexibility in Accurate Mass and Time Tag Workflows for Proteomics and Metabolomics*. ASMS, Salt Lake City, UT.

Washburn,M.P. *et al.* (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.*, **19**, 242–247.