

# NIH Public Access

**Author Manuscript** 

*Comput Biol Chem.* Author manuscript; available in PMC 2009 June 1.

Published in final edited form as: *Comput Biol Chem.* 2008 June ; 32(3): 215–217.

# MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/ MS) features

Matthew E. Monroe<sup>a</sup>, Jason L. Shaw<sup>a</sup>, Don S. Daly<sup>a</sup>, Joshua N. Adkins<sup>a</sup>, and Richard D. Smith<sup>a<sup>\*</sup></sup>

aPacific Northwest National Laboratory<sup>†</sup>, Richland, WA 99354 USA

# Abstract

Quantitative analysis of liquid chromatography (LC)- mass spectrometry (MS) and tandem mass spectrometry (MS/MS) data is essential to many proteomics studies. We have developed MASIC to accurately measure peptide abundances and LC elution times in LC-MS/MS analyses. This software program uses an efficient processing algorithm to quickly generate mass specific selected ion chromatograms from a dataset and provides an interactive browser that allows users to examine individual chromatograms with a variety of options.

## Keywords

Liquid chromatography-tandem mass spectrometry (LC-MS/MS); Quantitation; Proteomics; Quantification; Selected ion chromatogram

# 1. Introduction

Proteomic studies frequently depend on quantitative information to gain useful biological insights. Several software packages, including ASAPRatio (Li et al., 2003) and RelEx (MacCoss et al., 2003), have been developed to generate quantitative information, generally in the context of relative abundances of isotope-labeled species from liquid chromatography-tandem mass spectrometry (LC-MS/MS) analyses. This approach is limited by the resolution of the mass spectrometer; therefore analysis of isotopic pairs is best accomplished by using high-mass-measurement-accuracy MS (Qian et al., 2005b). Isotopic pairs processing is efficiently managed in-house by an advanced software package called VIPER (Monroe et al., 2007) that is uniquely tailored to the advantages of high resolution and high mass accuracy LC-MS. For characterizing species detected in low-resolution MS/MS analyses, we created the software package MASIC (<u>MS/MS Automated Selected Ion Chromatogram Generator</u>) to enable "label-free" quantitation.

<sup>\*</sup>Corresponding Author at: Pacific Northwest National Laboratory, PO Box 999, MS: K8-98, Richland, WA 99352, USA. Tel.: 1 509 376 0723; fax: 1 509 376 2303.

Availability: MASIC may be downloaded free of charge at http://ncrr.pnl.gov/software/

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# 2. Materials and Methods

MASIC evaluates peptide abundances and elution times in an LC-MS/MS analysis and provides the user with a flexible interface for viewing this information. The program first analyzes LC-MS/MS data to generate descriptions for each MS and MS/MS spectrum in terms of scan acquisition time, scan type, total ion intensity, base peak intensity, etc. This information is then used to generate "selected ion chromatograms" (SICs) for each species chosen for MS/ MS fragmentation. Chromatographic data such as peak width, height, area, and apex elution time are computed for each selected ion chromatogram. The peak apex is a more accurate measure of elution time than the scan numbers associated with fragmentation spectra, which allows for more precise alignment to LC-MS features in the AMT tag approach (Zimmer et al., 2006). MASIC also generates text files that summarize SIC and MS(/MS) scan information, allowing the user to generate "total ion chromatograms" (TIC) or "base peak intensity" (BPI) chromatogram plots using either a standard graphing program or the PNNL developed software package called Decon2LS (available free of charge at http://ncrr.pnl.gov/software/). The data generated are also written to an .Xml file that can be read by an accompanying executable called MASIC Browser. Both MASIC and MASIC Browser include graphical user interfaces (GUIs) along with Visual Basic .NET DLLs that allow for command line or code-based automation. The compiled executables and the source code are available for download free of charge at http://ncrr.pnl.gov/software/.

#### 2.1 Data Processing

MASIC recognizes .Raw files from Thermo Fisher mass spectrometers, mzXML files (http://sashimi.sourceforge.net/software\_glossolalia.html), mzData files (http://www.psidev.info/), and pairs of CDF and Mascot generic format (MGF) files that can be generated for Agilent ion trap mass spectrometers. Each of these formats contains similar information about the raw data, which allows MASIC to determine base peak intensity and *m*/*z*, total ion intensity, scan (LC elution) time, scan type, and in the case of MS/MS spectra, parent ion *m*/*z*.

MASIC includes a basic compression algorithm to deal with high-resolution data. If more than two consecutive points have zero intensity, the entire range is specified simply by the starting and ending m/z. The program also groups all points within an m/z range specified by the user; example bin sizes are 0.02 m/z for MS spectra and 0.2 m/z for MS/MS spectra. These steps reduce the amount of data used without a significant decrease in data quality, which improves processing rates, eases memory requirements, and reduces output file sizes.

After calculating a noise threshold for each MS parent ion spectrum, MASIC creates the SICs. First, a list of parent ions—m/z values chosen for fragmentation—is compiled, which may be supplemented by custom values provided by the user. These values are grouped into small m/z ranges (e.g., 0.05 m/z), and one SIC is created for each range by searching the MS1 spectra for data within a user-defined tolerance of the median m/z value within the range. In each mass spectrum, the sum of the intensity values for the data within the m/z search window defines that spectrum's contribution to the SIC. If a spectrum has no data in the m/z window, then the minimum positive intensity across the entire spectrum is taken as its contribution.

Following SIC generation, MASIC analyzes each SIC, which represents a unique LC elution time interval and m/z range, for useful information. First, the software finds the largest peak (that includes the scan number of the original m/z) in the SIC and applies Butterworth data smoothing, as implemented by the XPRESS software package (Han et al., 2001). The peak's signal-to-noise ratio, full-width-at-half-maximum, and area are calculated using the unsmoothed data, while the smoothed data are used to compute several statistical measures of peak shape. These measures include the Kolmogorov-Smirnov statistic (K-S stat) to estimate

Comput Biol Chem. Author manuscript; available in PMC 2009 June 1.

peak skew and the computation of the statistical moments, which provide peak area (zeroth moment), center of mass (1<sup>st</sup> central moment), peak standard deviation (2<sup>nd</sup> moment), and peak skew (derived from the 3<sup>rd</sup> moment).

MASIC also searches for spectra with similar parent ion m/z values and elution times. The fragmentation patterns of potentially similar spectra are compared using a Pearson correlation. If the correlation value exceeds a threshold, then the relationship between the spectra is recorded and the peak apex elution time of the spectrum with the largest peak area is used to define the elution time of the other related spectra.

A 3.2 GHz Pentium 4 computer was used to benchmark the performance of MASIC when processing Thermo Fisher .Raw files generated by LCQ, LTQ, and LTQ-FT analyses of a standard peptide mixture (a tryptic digest of a mixture of 12 proteins modeled after the mixture suggested by Purvine et al., 2004). As shown in Table 1, for separations with ~ $10^4$  spectra, the time spent loading and analyzing the file was usually less than 10 minutes, and often less than 5 minutes.

#### 2.2 MASIC Browser

After the data file has been analyzed and chromatograms created, the MASIC Browser allows the user to examine the chromatograms individually, as shown in Figure 1 and Figure 2. The list of SICs can be sorted by various criteria, including m/z, peak intensity, peak width, peak area, and signal-to-noise ratio. The SICs may also be filtered by intensity, signal-to-noise ratio, or m/z. The Butterworth-smoothed data are shown along with the SIC data by default; however, these data can be either removed from the display or replaced by data processed by Savitzky-Golay smoothing. The user can navigate the display window by zooming and panning to view adjacent regions of the chromatogram. Two floating cursors are provided to help the user visually evaluate the SIC. The initial cursor positions indicate the scan number at the peak apex of the SIC and the scan number of the MS/MS fragmentation event that provided the m/z value used to generate the SIC. The Browser also allows the user to annotate the plot with arrows and text.

### 3. Summary

MASIC quickly analyzes raw data from LC-tandem mass spectrometry studies and generates dataset-wide chromatograms and selected ion chromatograms for MS parent ions chosen for MS/MS fragmentation and/or for user-defined m/z values. The results can be visualized using MASIC Browser and are summarized numerically in text files. The abundance and elution time information computed for individual MS features enables more accurate generation of reference databases of mass and (elution) time tags as utilized by the AMT tag approach. The abundance values can also be used to compare peptides between two samples to compute peptide peak area ratios, as demonstrated in a proteomic study analyzing human plasma (Qian et al., 2005a).

#### Acknowledgements

Portions of this research were supported by the U. S. Department of Energy Office of Biological and Environmental Research Genomes:GtL Program, the NIH National Center for Research Resources (Grant RR018522), and the National Institute of Allergy and Infectious Diseases (NIH/DHHS through interagency agreement Y1-AI-4894-01).

# References

Han DK, Eng J, Zhou H, Aebersold R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat. Biotechnol 2001;19:946–951. [PubMed: 11581660]

Comput Biol Chem. Author manuscript; available in PMC 2009 June 1.

Monroe et al.

- Li XJ, Zhang H, Ranish JA, Aebersold R. Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. Analytical Chemistry 2003;75:6648–6657. [PubMed: 14640741]
- MacCoss MJ, Wu CC, Liu H, Sadygov R, Yates JR 3rd. A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. Analytical Chemistry 2003;75:6912–6921. [PubMed: 14670053]
- Monroe ME, Tolić N, Jaitly N, Shaw JL, Adkins JN, Smith RD. VIPER: an advanced software package to support high-throughput LC-MS peptide identification. Bioinformatics. 2007
- Purvine S, Picone AF, Kolker E. Standard mixtures for proteome studies. OMICS 2004;8:79–92. [PubMed: 15107238]
- Qian WJ, Jacobs JM, Camp DG, Monroe ME, Moore RJ, Gritsenko MA, Calvano SE, Lowry SF, Xiao W, Moldawer LL, Davis RW, Tompkins RG, Smith RD. Comparative proteome analyses of human plasma following *in vivo* lipopolysaccharide administration using multidimensional separations coupled with tandem mass spectrometry. Proteomics 2005;5:572–584. [PubMed: 15627965]
- Qian WJ, Monroe ME, Liu T, Jacobs JM, Anderson GA, Shen Y, Moore RJ, Anderson DJ, Zhang R, Calvano SE, Lowry SF, Xiao W, Moldawer LL, Davis RW, Tompkins RG, Camp DG, Smith RD. Quantitative proteome analysis of human plasma following *in vivo* lipopolysaccharide administration using <sup>16</sup>O/<sup>18</sup>O labeling and the accurate mass and time tag approach. Mol. Cell. Proteomics 2005;4:700–709. [PubMed: 15753121]
- Zimmer JS, Monroe ME, Qian WJ, Smith RD. Advances in proteomics data analysis and display using an accurate mass and time tag approach. Mass Spectrom. Rev 2006;25:450–482. [PubMed: 16429408]

MASIC Browser	
<u>File E</u> dit <u>H</u> elp	
Select File D:\QC_Shew_07_02_SICs.xml	
Parent Ion SIC to View Scan 2201 (428.19 m/z) Scan 4101 (575.44 m/z) Scan 4636 (575.33 m/z) Scan 4363 (575.44 m/z) Scan 6171 (566.65 m/z) Scan 10638 (1171.14 m/z) Scan 10001 (1171.23 m/z) Scan 10430 (1171.36 m/z) Scan 6535 (780.09 m/z) Scan 6733 (749.97 m/z) Scan 6955 (749.78 m/z)	Navigation <u>Previous</u> <u>Jump to Scan</u> <u>Next</u> <u>Auto Step</u> 150 ✓ Move forward
Sort Order Sort Descending	SIC Filters   MS/MS Results Filters   SIC Type Filter   All SIC's
✓Show Baseline Corrected StatsIndex: 5681m/z: 780.09Scan at apex: 6390Peak StDev: 63.4Center of mass: 6420Peak Skew: 1.131Intensity: 2.005E+06Peak KSStat: 0.9394Area: 2.850E+08PwHM: 64FWHM: 64StDev: 4.156E+04Noise level: 4.156E+04Noise StDev: 4.57E+04Points used: 964Noise Mode Used:	✓ Fix X Range       1250       scans         Fix Y Range       5000000       counts         Minimum S/N       3         Minimum Intensity       1000000       counts         Filter by m/z       550       m/z         ±       0.2       m/z
Smoothing Options Show Smoothed Data O Do Not Show Smoothed Data O Use Butterworth Smooth Filter Order Use Savitzky Golay Smooth Filter Order Minimum Peak Width (points)	SIC Peak Finding         Recompute SIC Peak Stats         Subtract baseline when computing Intensity and Area         Redo SIC Peak         Redo SIC Peak         Finding For All Data

#### Figure 1.

MASIC Browser interface. The user can browse the list of selected ion chromatograms, view the statistics for each, and filter the displayed data. As each chromatogram is selected, it is plotted in a separate window (see Figure 2).

Monroe et al.



#### Figure 2.

Selected Ion Chromatogram plot in MASIC Browser. Both the raw data (line with points) and the smoothed data (no points) are plotted. The user can navigate the chromatogram, zooming, panning, and annotating as desired.

Author Manusci	NIH-PA /	uthor Manuscript	NIH-PA A	uscript	PA Author Man	NIH-I
Proce	essing henchmarks for MAS	Ta IC using a 3.2 GHz Pen	<b>ble 1</b> tium 4 computer			
Thermo Fisher Mass Spectrometer	Separation Time (minutes)	Total Spectra	MS/MS Spectra	Finnigan RAW File Size (MB)	File Load Time (minutes)	SIC Processing Time (minutes)
LCQ	100	2,779	1,851	10	0.1	0.2
LTQ	100	19,298	15,845	74	0.5	2.9
LTQ-FT	100	10,656	6,905	534	2.4	1.7
LTQ	600	133,677	93,131	875	7.4	74
LTQ	720	150,034	110,071	1067	9.2	78

**NIH-PA** Author Manuscript