International Journal of Mass Spectrometry 427 (2018) 91–99

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

Full Length Article

An algorithm to correct saturated mass spectrometry ion abundances for enhanced quantitation and mass accuracy in omic studies

Aivett Bilbao, Bryson C. Gibbons, Gordon W. Slysz, Kevin L. Crowell, Matthew E. Monroe, Yehia M. Ibrahim, Richard D. Smith, Samuel H. Payne, Erin S. Baker*

Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, USA

ARTICLE INFO

Article history:
Received 19 May 2017
Received in revised form
16 September 2017
Accepted 3 November 2017
Available online 6 November 2017

Keywords:
Mass spectrometry
Detector saturation
Analog-to-digital converter saturation
Saturation correction
Isotopic envelope
Isotopic ratios
Quantitation

ABSTRACT

The mass accuracy and peak intensity of ions detected by mass spectrometry (MS) measurements are essential to facilitate compound identification and quantitation. However, high concentration species can yield erroneous results if their ion intensities reach beyond the limits of the detection system, leading to distorted and non-ideal detector response (e.g. saturation), and largely precluding the calculation of accurate *m*/*z* and intensity values. Here we present an open source computational method to correct peaks above a defined intensity (saturated) threshold determined by the MS instrumentation such as the analog-to-digital converters or time-to-digital converters used in conjunction with time-of-flight MS. In this method, the isotopic envelope for each observed ion above the saturation threshold is compared to its expected theoretical isotopic distribution. The most intense isotopic peak for which saturation does not occur is then utilized to re-calculate the precursor *m*/*z* and correct the intensity, resulting in both higher mass accuracy and greater dynamic range. The benefits of this approach were evaluated with proteomic and lipidomic datasets of varying complexities. After correcting the high concentration species, reduced mass errors and enhanced dynamic range were observed for both simple and complex omic samples. Specifically, the mass error dropped by more than 50% in most cases for highly saturated species and dynamic range increased by 1–2 orders of magnitude for peptides in a blood serum sample.

1. Introduction

Current mass spectrometers provide molecular measurements (i.e. *m*/*z*) with both high mass accuracy and resolving power for concentrations typically ranging from 3 to 4 orders of magnitude in a single mass spectrum depending upon instrumental details. These measurements are extremely important for identifying molecules occurring in complex samples and determining how they change under varying biological and environmental conditions. Two important characteristics for confident molecular identifications from mass spectrometry (MS) are the isotopic peaks (isotopic envelope) for the ion observed and tandem (MS/MS) measurements of fragment ion species. Together these observations yield information about the molecular mass, observed charge state, elemental composition and structural arrangement for the potential molecule. However, both mass and quantitation errors occur in MS measurements for numerous reasons and understanding their source is necessary to ensure that incorrect conclusions are not made. Mass error has been discussed in detail in several publications [1–5] and is associated with all MS measurements for reasons related to both the MS platform type and non-ideal performance through the instrument (e.g. slight deviations in voltage and pressure). Fundamental limitations also influence mass error when insufficient ion statistics result in the inability to accurately define a peak. Quantitation errors are associated with ion statistics and limits of the MS detector dynamic range, which impair accurate measurements of both the high and low concentration species in a sample. If the concentration of a molecule is too low, ions can be lost while traveling through the instrument, resulting in no detection or not enough ion accumulation to create a well-defined peak. High concentration species also cause issues if their signal intensity exceeds the detector capacity saturating one or more of the isotopic peaks. Furthermore, if the isotopic envelope contains one or more saturated peaks, the natural distribution is distorted and results in incorrect quantitative readings [6]. These incorrect intensity values returned by the deisotoping algorithms are not accurate representations of the actual ion abundances regardless of whether the maximum intensity, area or volume is used. Detector

* Corresponding author at: 902 Battelle Blvd., P.O. Box 999, MSIN K8-98, Richland, WA 99352, USA.
E-mail address: erin.baker@pnnl.gov (E.S. Baker).

https://doi.org/10.1016/j.ijms.2017.11.003
1387-3806/

saturation is commonly observed in analog-to-digital converters (ADC) or time-to-digital converters used in conjunction with time-of-flight (TOF) mass spectrometers [2,6–8]. However, saturation has also been observed in other MS instrumentation such as triple quadrupoles [9,10] and trap based [11] instruments, and its extent usually depends on the detector design. In addition, the incorporation of on-line separations, such as liquid chromatography (LC) and ion mobility spectrometry (IMS) also result in saturation due to the concentrating effect of the separations.

Practical approaches for avoiding saturation include optimizing sample concentrations and instrument operating settings in order to keep the abundances below the known saturation levels of the platform [8,12]. However, this precludes covering the large dynamic range needed in many applications such as that occurring in blood plasma and environmental samples. Over the last decade, several other approaches have been implemented to extend dynamic range and reduce saturation effects. A majority of these advances have been based on hardware or instrumental techniques applied during MS data acquisition to avoid or minimize saturating the detection systems [9,13,14]. However, as hardware has improved, a handful of post-acquisition software approaches have also been utilized or suggested to correct the saturated MS data. For example, in targeted MS analysis with selected reaction monitoring, an algorithm called SignalFinder was integrated into the SCIEX proprietary software MultiQuant and utilized for saturation correction in data from triple quadrupole MS systems [10]. In the case of untargeted MS-studies, to our knowledge, no proprietary or open source software is available to perform saturation correction post-acquisition. Nevertheless, strategies to correct the saturated peaks using isotopic distributions have been suggested in the literature. For instance, theoretical isotopic distributions can be compared against the most intense observed isotopic envelopes in the data to detect saturation [7]. Furthermore, the intensity and mass shift errors of a saturated ion peak can be corrected by using information from the adjacent (e.g. second or third C13) isotopic peak that is not saturated [2]. Several methods addressing the computationally demanding calculations for theoretical isotopic distributions have been reported [15–22] and reviewed in particular for large biomolecules [23]. Among the cited examples, methods based on utilizing the molecular formula of a known molecular species are preferred and known to be most accurate. However, the elemental composition of ionized molecular species are not always known and, indeed, establishing their identity is often a key aspect of the measurements. In proteomic analyses, methods utilizing known elemental compositions to determine the monoisotopic mass based on the makeup of amino acids provide a good estimation [21,23].

Considering these initial ideas, we implemented a computational method to first detect all possible saturated peaks in a spectrum by flagging those with an intensity above a defined ADC threshold (e.g., 70%), which is close to the ADC capacity of the utilized instrument. The saturated peaks were then corrected using the unsaturated peaks in the isotopic envelope and theoretical isotopic models based upon assumed elemental compositions of the analytes (e.g. peptides). The utility of this approach was then evaluated by analyzing datasets of varying complexity and determining their dynamic range, mass error and quantitation accuracy before and after correction.

2. Materials and methods

2.1. Software algorithm

The software algorithm described in this manuscript first identifies saturated peaks in the mass spectrum. Since a 256 channel 8-bit ADC digitizer was utilized in the MS studies, the maximum signal possible per pixel was 256. The total number of pixels per accumulated scan was then considered, and a peak at or above a threshold of 70% of this number was determined to be saturated. Each flagged peak was then further investigated by utilizing the unsaturated peaks from the ion's isotopic envelope detected in a mass spectrum. To correct the saturated peaks, the intensity values of the peaks of the isotopic envelope $\{A_1, A_2 \ldots A_n\}$ and the intensity values of the peaks of a theoretical isotopic envelope $\{T_1, T_2 \ldots T_n\}$ were required, where n is the number of isotopic peaks.

For the peptide analyses, the software implementation reported here used the so-called averagine method to create an approximate peptide molecular formula from the given monoisotopic mass [21]. Since the purpose of the algorithm is to report an accurate intensity value for each observed isotopic envelope instead of attempting to correct every saturated value in the profile mass spectrum or ion distribution, a peak centroiding method was applied to the isotopic peaks for both the observed and theoretical isotopic envelopes prior to comparison. Intensity values of saturated isotopic peaks were corrected by assuming that, when unsaturated, the relative intensities of individual peaks of the observed isotopic envelope will match the relative intensity of the isotopic peaks of the theoretical isotopic envelope. In other words, the intensity ratio for each isotope in the form observed/theoretical, is constant for all isotopic peaks. The correction of each saturated peak was accomplished using the formula:

$$A_x = T_x \times A_u \div T_u$$

where A is the group of observed isotopic peaks, T is the group of theoretical isotopic peaks, *x* is the index of the isotopic peak being corrected, and *u* is the index of the most intense observed isotopic peak that is not saturated. The algorithm then iterates over the observed isotopic peaks to find the most intense isotopic peak (A_u) that is not saturated since it has both an accurate abundance and the best signal quality. The current version of the algorithm requires defining the saturation level of the detector, which is characteristic for a given mass spectrometer. Using the index of the selected peak, the relative intensity of the corresponding peak (T_u) of the theoretical isotopic envelope is determined. For each saturated isotopic peak (A_x), the new intensity value is computed by multiplying the relative intensity of the corresponding theoretical peak (T_x) by the defined ratio ($A_u \div T_u$). Similarly, the new *m*/*z* value of the saturated isotopic peak is computed by back calculating from the *m*/*z* of the unsaturated isotopic peak.

The described method for saturation correction was implemented as a command-line application in C# and it has been integrated into the data processing pipeline for LC-IMS–MS datasets at the Pacific Northwest National Laboratory. Raw MS files in the Unified Ion Mobility Frame (UIMF) file format [24] were processed by DeconTools [25,26], where the saturation correction was integrated. R software (v3.3.1 × 64) was then used for visualization and comparison of the results from the DeconTools output. This software is freely available and can be downloaded from http://omics.pnl.gov/software/decontools-decon2ls. The source code is available at https://github.com/PNNL-Comp-Mass-Spec/DeconTools. Saturation correction was enabled in DeconTools by setting the "ScanBasedWorkflowType" parameter to uimf_saturation_repair. See, for example, the parameter file SampleParameterFileIMS.xml in the repository on GitHub.

2.2. Instrumentation

All data files were acquired with an IMS-MS platform previously described [27,28]. Briefly, the IMS instrument couples a 1-m IM drift cell with an Agilent 6224 TOF MS (Agilent Technologies,