JOURNAL OF PROTEOMICS 91 (2013) 375–384

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Technical note

Large precursor tolerance database search — A simple approach for estimation of the amount of spectra with precursor mass shifts in proteomic data☆

Rueyhung Roc Weng[a,1], *Lichieh Julie Chu*[a,b,1], *Hung-Wei Shu*[a], *Timothy H. Wu*[c], *Mengchieh Claire Chen*[a,2], *Yuwei Chang*[a], *Yihsuan Shannon Tsai*[c,f,3], *Michael C. Wilson*[f], *Yeou-Guang Tsay*[d,e], *David R. Goodlett*[f], *Wailap Victor Ng*[a,b,c,g,*]

[a]*Institute of Biotechnology in Medicine, National Yang Ming University, Taipei, Taiwan, ROC*
[b]*Molecular Medicine Research Center, Chang Gung University, Taoyuan, Taiwan, ROC*
[c]*Institute of Biomedical Informatics, National Yang Ming University, Taipei, Taiwan, ROC*
[d]*Proteomics Research Center, National Yang Ming University, Taipei, Taiwan, ROC*
[e]*Institute of Biochemistry and Molecular Biology, National Yang Ming University, Taipei, Taiwan, ROC*
[f]*University of Maryland, School of Pharmacy, Baltimore, MD 21201, USA*
[g]*Department of Biotechnology and Laboratory Science in Medicine, National Yang Ming University, Taipei, Taiwan, ROC*

ARTICLE INFO

Article history:
Received 25 March 2013
Accepted 27 July 2013
Available online 8 August 2013

Keywords:
LTQ-Orbitrap
Mass shift
PeptideProphet
Precursor mass tolerance

ABSTRACT

Mass measurement and precursor mass assignment are independent processes in proteomic data acquisition. Due to misassignments to C-13 peak, or for other reasons, extensive precursor mass shifts (i.e., deviations of the measured from calculated precursor neutral masses) in LC–MS/MS data obtained with the high-accuracy LTQ-Orbitrap mass spectrometers have been reported in previous studies. Although computational methods for post-acquisition reassignment to monoisotopic mass have been developed to curate the MS/MS spectra prior to database search, a simpler method for estimating the fraction of spectra with precursor mass shift so as to determine whether the data require curation remains desirable. Here, we provide the evidence that an easy approach, which applies a large precursor tolerance (2.1 Da or higher) in *SEQUEST* search against a forward and decoy protein sequence database and then filters

Abbreviation: p, *PeptideProphet* probability; H23, Human lung adenocarcinoma cell line H23; CL1-5, Human lung adenocarcinoma cell line CL1-5.

* *Corresponding author at:* Institute of Biotechnology in Medicine, National Yang Ming University, 155 Linong Street, Section 2, Taipei 112, Taiwan, ROC. Tel.: +886 2 2826 7321; fax: +886 2 2826 4092.
E-mail addresses: wvng@ym.edu.tw, vng7321@hotmail.com (W.V. Ng).
[1] Rueyhung Roc Weng and Lichieh Julie Chu contributed equally to this work.
[2] Current address of Mengchieh Claire Chen: Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA.
[3] Current address of Yihsuan Shannon Tsai: Curriculum in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, NC 27599, USA.

1874-3919/$ – see front matter
http://dx.doi.org/10.1016/j.jprot.2013.07.030

the data with *PeptideProphet* peptide identification probability ($p \geq 0.9$), could detect most of the MS/MS spectra containing inaccurate precursor masses. Furthermore, through the implementation of artificial mass shifts on 4000 randomly selected MS/MS spectra, which originally had accurate precursor mass assigned by the mass spectrometers, we demonstrated that the accuracy of the precursor mass has almost negligible influence on the efficacy and fidelity of peptide identification.

Biological significance

Integral precursor mass shift is a known problem and thus proteomic data should be handled and analyzed properly to avoid losing important protein identification and/or quantification information. A quick and easy approach for estimating the number of MS/MS spectra with inaccurate precursor mass assignments would be helpful for evaluating the performance of the instrument, determining whether the data requires curation prior to database search or should be searched with specific search parameter(s). Here we demonstrated most of the MS/MS spectra with inaccurate mass assignments (integral or non-integral changes) that could be easily identified by database search with large precursor tolerance windows.

1. Introduction

Mass spectrometry (MS), due to its diverse applications in qualitative and quantitative proteomic analysis, has become an essential tool frequently applied in basic and biomedical researches [1–12]. The advancements in instrumentation [13–17], labeling chemistry [4–6,18], bioinformatics [19–25], etc. have expedited proteomic research in the past decade and a large number of proteins can be identified and quantified at high efficiency and with high reliability.

Despite high-end mass spectrometers could perform highly accurate measurements of the peptide ion mass/charge ratios (m/z) in proteome analysis, deviations of the measured from calculated precursor neutral masses in LTQ-Orbitrap data have been reported in previous studies [26–29]. This phenomenon is not too surprising because each peptide ion is present in the form of an isotopic cluster, and the precursor m/z measurement and monoisotopic mass assignment are two related but separate processes during data acquisition. The inaccuracy of the recorded precursor masses in the raw data may result from computational misassignment of the precursor mass to a heavier isotopic peak instead of the ideal monoisotopic peak during isotopic-cluster deconvolution or posttranslational modifications (e.g., +1 Da shift in deamidation of asparagine and glutamine to aspartate and glutamate) or other reasons. To take advantage of high-accuracy precursor mass measurement in data analysis, software such as the *extract_msn* (Thermo Scientific, San Jose, CA), *DeconMSn*, *Bulleye*, *YADA* and a Perl script by Scherl et al. (2008) has been developed for post-acquisition curation of the misassigned precursor masses in raw data prior to database searching to enhance peptide identification [26–29]. For instance, after the curation of the precursor mass in a *Pseudomonas aeruginosa* dataset, approximately 10% more peptides were identified, and the false discovery rate (FDR) was reduced from 1 to 0.2% [27].

A simple approach for estimating the fraction of MS/MS spectra with precursor mass shifts is useful not only for monitoring the accuracy of mass assignments in mass spectrometry and improvements of mass spectrometers (i.e., data acquisition hardware and software) but also helpful for revealing the successful rate in precursor mass curation. Database searching with a large tolerance window is apparently a possible approach for the identification of spectra with precursor mass shifts since it has been used in the analysis of low mass accuracy data. However, a major concern of this approach is that this may lead to the loss of search sensitivity since each peptide match must now compete against a larger pool of candidate peptides, each of which has a chance to randomly score higher than the correct peptide [30]. To demonstrate the feasibility of using a large precursor tolerance window in *SEQUEST* searches to monitor most, if not all, of the precursor mass shifts, we examined the FDRs and compared the conserved and conflicted peptide identifications in 11 different search windows, ranging from 5 ppm to very large 15.1 Da, of the proteomic data from a prokaryote *Halobacterium salinarum* strain NRC-1 and human lung adenocarcinoma cancer cell lines (CL1-5 [31] and H23 (ATCC, Manassas, VA)) obtained with two models of LTQ-Orbitrap mass spectrometers (i.e., the earlier installed "XL" and the relatively newer model "Discovery"). In addition, we also analyzed the influence of artificially implemented integral and non-integral precursor mass shifts in peptide identification with a large precursor tolerance window of 2.1 Da. Details of our analyses and the influence of inaccurate precursor mass in peptide identification are discussed in detail below.

2. Materials and methods

2.1. Proteome samples

H. salinarum strain NRC-1 (ATCC 700922) cell lysate samples were prepared from cells grown in CM^+ medium [32,33] at 37 °C as previously described [34]. Human lung adenocarcinoma cell lines H23 (ATCC, Manassas, VA) and CL1-5 [31] were cultured in RPMI-1640 medium at 37 °C and 5% CO_2 to 80–90% confluence in 75 cm^2 flasks (Corning, Corning, NY) and then lysed in cell lysis buffer containing a protease inhibitor cocktail (Sigma, St. Louis, MO) [35]. Approximately 100 μg of protein were digested with 2 μg sequencing grade modified trypsin (Promega, Madison, WI) in a total volume of 500 μl of