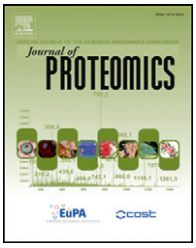
JOURNAL OF PROTEOMICS 75 (2012) 5516–5522

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot



Technical note

MilQuant: A free, generic software tool for isobaric tagging-based quantitation

Xiao Zou[a,1], *Minzhi Zhao*[b,1], *Hongyan Shen*[a], *Xuyang Zhao*[a], *Yuanpeng Tong*[a], *Qingsong Wang*[a], *Shicheng Wei*[b,*], *Jianguo Ji*[a,**]

[a]*State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing 100871, China*
[b]*Center for Biomedical Materials and Tissue Engineering, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China*

ARTICLE INFO

Article history:
Received 23 March 2012
Accepted 29 June 2012
Available online 10 July 2012

Keywords:
Isobaric tagging
iTRAQ
MilQuant
Public software
Quantitative proteomics
TMT

ABSTRACT

Isobaric tagging techniques such as iTRAQ and TMT are widely used in quantitative proteomics and especially useful for samples that demand *in vitro* labeling. Due to diversity in choices of MS acquisition approaches, identification algorithms, and relative abundance deduction strategies, researchers are faced with a plethora of possibilities when it comes to data analysis. However, the lack of generic and flexible software tool often makes it cumbersome for researchers to perform the analysis entirely as desired. In this paper, we present MilQuant, mzXML-based isobaric labeling quantitator, a pipeline of freely available programs that supports native acquisition files produced by all mass spectrometer types and collection approaches currently used in isobaric tagging based MS data collection. Moreover, aside from effective normalization and abundance ratio deduction algorithms, MilQuant exports various intermediate results along each step of the pipeline, making it easy for researchers to customize the analysis. The functionality of MilQuant was demonstrated by four distinct datasets from different laboratories. The compatibility and extendibility of MilQuant makes it a generic and flexible tool that can serve as a full solution to data analysis of isobaric tagging-based quantitation.

Abbreviations: ETD, Electron transfer dissociation; HCD, High energy collision dissociation; IT, Ion trap; iTRAQ, Isobaric tags for relative and absolute quantitation; MS, Mass spectrometry; PQD, Pulsed Q dissociation; PSM, Peptide-spectrum match; SILAC, Stable isotope labeling by amino acids in cell culture; TMT, Tandem mass tags.

* *Correspondence to:* S. Wei, Center for Biomaterials and Medical Devices, School of Stomatology, Peking University, Beijing 100081, China. Tel./fax: +86 10 82195780.
** *Correspondence to:* J. Ji, P.O. Box 31, State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing 100871, China. Tel.: +86 10 62755470; fax: +86 10 62751526.
E-mail addresses: marinzou@gmail.com (X. Zou), xgt1986627@163.com (M. Zhao), hyshen1985@163.com (H. Shen), zhao-xu-yang@163.com (X. Zhao), tongwang98@yahoo.com.cn (Y. Tong), wangqingsong@pku.edu.cn (Q. Wang), sc-wei@pku.edu.cn (S. Wei), jijg@pku.edu.cn (J. Ji).
[1] These authors contributed equally to this work.

1874-3919/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2012.06.028

1. Introduction

Advances in instrumentation and techniques over the past decade have helped proteomics mature into the stage of quantitative profiling of proteomes. A variety of quantitation approaches have been developed and applied in an increasing number of research fields [1], among which stable isotope-based techniques have been especially favored over gel-based quantitation due to its high throughput, sensitivity, and reproducibility. Researches conducted in cell cultures have benefited greatly from metabolic labeling approaches such as stable isotope labeling by amino acids in cell culture (SILAC) [2], whereas *in vitro* labeling [3], also applicable in cell culture studies, is required for specific sample types, especially clinical samples such as serum, saliva, and cerebrospinal fluid. Moreover, while most stable isotope-based quantitation approaches are capable of simultaneous comparison of up to three samples, isobaric tags for relative and absolute quantitation (iTRAQ) [4] and tandem mass tags (TMT) [5], two commercially available isobaric tagging techniques, offer higher multiplexing capacity of up to eight multiplex channels. Altogether, these advantages made isobaric tagging based techniques popular choices for relative proteome quantitation.

With isobaric tagging techniques, relative quantities of peptides become apparent when the parts of tagging reagents fall off from these precursor peptide ions, producing reporter ions that appear in the tandem MS spectra at multiplexed m/z and with varying intensities, which serves as the basis for the deduction of relative protein abundances. Traditionally, TOF has been the mass analyzer of choice for production and acquisition of reporter ions, while the usage of ion trap (IT) is limited by the "one-third rule" [6]. However, recent development of novel dissociation methods, including pulsed Q dissociation (PQD) [7], electron transfer dissociation (ETD) [8] and high energy collision dissociation (HCD) [9], has also enabled ion trap (IT) and Orbitrap for such tasks. As a result, the diversity in mass spectrometry (MS) collection approaches and native acquisition files formats is greatly increased. Added with the diversity in choices of database searching, intensity normalization, and relative abundance computation methods, this could result in a great many variations on the theme of data analysis. Among the available tools in this field, a lot of them were developed before the advent of novel dissociation approaches and thus not directly compatible, whereas most other software tools, such as Mascot and Proteome Discoverer, works in an integrated fashion that reports only the ultimate relative abundance results, making it difficult for researchers to tailor the entire analysis protocol to their needs and to compare results from different instrumentation platforms.

In this paper, we present MilQuant (mzXML-based isobaric labeling quantitator), an open source software that is compatible with all mainstream mass spectrometers, tandem MS dissociation approaches, and database search algorithms that have been reported in isobaric tagging-based quantitation researches. It also offers a panel of effective normalization and relative abundance combination methods. Furthermore, the analysis with MilQuant is highly flexible pipeline, with each step generating a handful of output files to facilitate researchers to customize their downstream handling. Finally, we demonstrate the merits of MilQuant with four datasets collected with different approaches.

2. Technical basics of MilQuant

Since the identification of peptide and protein is the prerequisite for their quantitation, a typical quantitation workflow for isobaric tagging data usually import peptide and protein identities and associate them with peptide relative abundances indicated by the intensities of reporter ion peaks extracted from tandem MS spectra. These intensities will then undergo a series of transformations and integration into the final forms of relative peptide and protein abundance ratios [3].

Specifically, the MilQuant pipeline consists of the six following steps (Fig. 1): (i) reporter ion peak extraction, (ii) importing and optional filtering of protein and peptide hits, (iii) mapping spectral ratios to corresponding peptide sequences, (iv) isotopic impurity correction, normalization for inequality in amounts of samples for labeling, (v) an optional VSN normalization, and (vi) combination into peptide and protein abundance ratios.

MilQuant was written in Python and R programming languages. The source code is freely available at http://

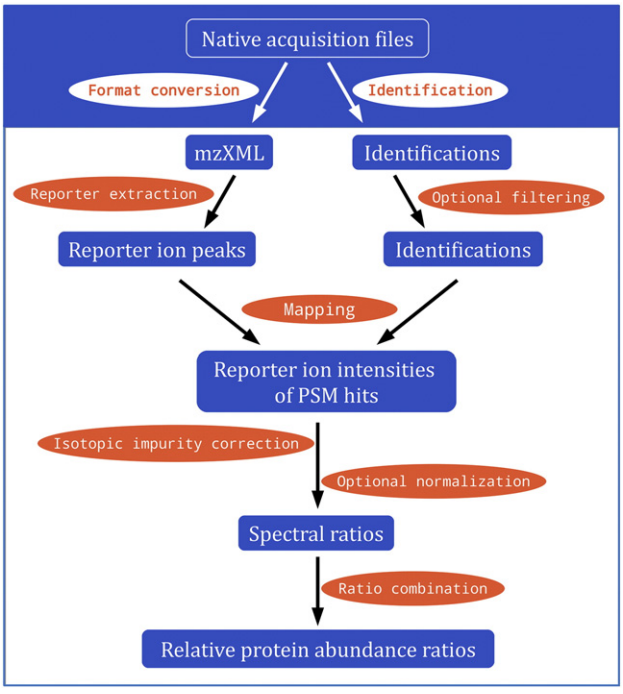


Fig. 1 – MilQuant's workflow. Preparations before MilQuant include conversion of native acquisition files and identification. Afterwards, MilQuant starts by extracting reporter ion peaks and mapping them to identified spectra and peptide sequences. Intensities of reporter ion peaks from qualified spectra are then corrected for isotopic impurities and normalized. Finally, peak intensities or their relative ratios of distinct spectra are combined into the relative abundance ratios of the corresponding peptides and proteins. PSM: peptide–spectrum match.