JOURNAL OF PROTEOMICS 74 (2011) 2204–2209

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Technical note

Deconvolution of overlapping isotopic clusters improves quantification of stable isotope–labeled peptides

Salvatore Cappadona[1], *Javier Muñoz*[1], *Wim P.E. Spee*, *Teck Y. Low*, *Shabaz Mohammed*, *Bas van Breukelen*[*], *Albert J.R. Heck*[*]

Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Centre for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 28 March 2011
Accepted 24 April 2011
Available online 15 May 2011

Keywords:
Proteomics
Mass spectrometry
Bioinformatics
Quantification
Stable isotope labeling
Overlap of isotopic clusters
Dimethyl labeling

ABSTRACT

High-resolution mass spectrometry and the use of stable isotopes have greatly improved our ability to quantify proteomes. Typically, the relative abundance of peptides is estimated by identifying the isotopic clusters and by comparing the peak intensities of peptide pairs. However, when the mass shift between the labeled peptides is small, there can be the possibility for overlap of the isotopic clusters which will hamper quantification accuracy with a typical upwards bias for the heavier peptide. Here, we investigated the impact of the overlapping peak issue with respect to dimethyl based quantification and we confirmed there can be need for correction. In addition, we present a tool that can correct overlapping issues when they arise which is based on modeling isotopic distributions. We demonstrate that our approach leads to improved accuracy and precision of protein quantification.

1. Introduction

Quantitative proteomics has emerged as a powerful technique to address questions regarding the molecular mechanisms that regulate biological processes [1]. A common strategy to compare the difference in proteins abundance between two or more samples is stable isotope dilution, followed by LC/MS/MS [2]. In such an approach, distinct yet similar samples are first labeled using stable isotopes, such as ^{2}H, ^{13}C, ^{18}O and ^{15}N, and then mixed in equal amounts followed by a classic proteomics workflow [3]. The MS intensities of the light and heavy peaks accurately reflect the relative abundance of peptides, and therefore proteins, in the samples. Although potentially straightforward, this quantitative strategy can be hampered by the overlap of isotopic clusters of light and heavy peaks, which occurs whenever the mass shift between the peptide pairs is smaller than their isotopic envelope. This issue affects, to different extents, most isotopic labeling techniques, but is still largely disregarded by the software commonly used for quantification. At this moment, only a few correction tools have been proposed to deconvolute overlapping isotopic distributions and most typically address a specific isotopic labeling. IEMM [4], for instance, proposes a method called isotopic envelope mixture modeling to

* *Corresponding authors at: Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Centre for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.*
E-mail addresses: b.vanbreukelen@uu.nl (B. van Breukelen), a.j.r.heck@uu.nl (A.J.R. Heck).
[1] These authors contributed equally to this work.

1874-3919/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2011.04.022

overcome overlapping in ^{18}O labeling, where the isotopic peaks are shifted by 2 and 4 Da. Q3 [5] uses a Poisson approximation to predict the isotopic distribution for acrylamide labeling, where the shift is a multiple of 3 Da, depending on the number of cysteines present in the peptides. More recently, an algorithm has been proposed that uses quadratic equations to resolve overlapping peaks occurring with a 4 Da shift when mTRAQ labeling is used [6].

In this work we systematically investigated the theoretical quantification error caused by the overlapping issue. First, the averagine model [7] was used to estimate peptide isotopic distributions and to model the predictable quantification error at increasing masses and differential expressions. The analysis confirmed a significant trend toward overestimation of the heavier isotopologue that becomes significantly apparent when the peptide mass approaches 3 kDa.

Then, to illustrate the need to correct for isotope peaks overlap, a simple strategy was devised to deconvolute overlapping peptides. The algorithm predicts the isotopic distribution of peptides based on their sequence, as identified by fragmentation spectra. Our approach can be used in the form of a post processing tool that can correct the results obtained by quantitation software tools that evaluate peptide ratios based on the elution profiles of the monoisotopic peak of their precursor ions, such as MSQuant [8] and XPRESS [9]. Our strategy showed more accurate peptide ratios and resulted in improved accuracy and precision of protein quantification.

2. Materials and methods

2.1. Sample preparation

HeLa cells were harvested by centrifugation at 2,500 g for 10 min at 4 °C. Cell lysis was performed in a buffer containing 8 M urea and 2 M thiourea in 25 mM ammonium bicarbonate at pH 8.2 with protease and phosphatase inhibitors. The cell suspension was later subjected to ultra-sonication with the input power at 60 W in 30 cycles for one minute. Subsequently, cell debris was removed by centrifugation at 1,000 g for 10 min at 4 °C. Proteins (50 μg) were then reduced, alkylated and digested for 4 h with Lys-C. The mixture was then diluted 4-fold to 2 M urea and digested overnight with trypsin. Finally, sample was acidified with formic acid to a final concentration of 5%.

2.2. Dimethyl labeling

Trypsinized peptides were equally divided into 3 pools and subsequently desalted using Sep-Pak C18 cartridges (Waters Corporation, Milford, MA). Peptides were then labeled with stable isotope dimethyl labeling as described previously [10]. Efficiency of incorporation of each label was individually checked by LC-MS/MS before being mixed in a 10:1:1 (Light: Intermediate: Heavy) ratio.

2.3. Liquid chromatography and tandem mass spectrometry

For mass-spectrometric analysis, peptides were separated with a C18 column (3 μm, 200 Å, 50 μm × 40 cm) (Dr Maisch, Ammerbuch-Entringen) arranged in a vented-column configuration [11]. A 5 hours gradient consisting of: 0–10 min, 0% B at 5.0 μl/min for sample loading; 10.1-240 min, 10% to 23% B at 0.10 μl/min; 240.1-277 min, 23% to 50% B at 0.10 μl/min; 277.1-279 min, 50% to 100% B at 0.10 μl/min; 279.1 – 281.5 min, maintained at 100% B, 281.51 -282.0 min, 0% B at 0.10 μl/min; 282.1 – 299.0 min, 0% B at 0.1 μl/min was used for separation while eluted peptides were introduced by nano-electrospray into an LTQ Orbitrap Velos (Thermo Fisher, Bremen). Mass-spectrometric duty-cycle consists of a high resolution (30,000 FHMW) survey scan in the Orbitrap. The ten most intense precursors were fragmented using a data-dependent decision tree utilizing HCD (essentially beam type CID), ETD-IT and ETD-FT [12]. In brief, doubly charged peptides were subjected to HCD fragmentation and higher charged peptides were fragmented using ETD. The normalized collision energy for HCD was set to 35%. ETD reaction time was set to 50 ms for doubly charged precursors.

2.4. Database search

Raw files were converted to peak lists using Proteome Discoverer 1.2 (Thermo Scientific). For protein identification, .mgf files were searched against a concatenated forward-decoy IPI human database (Version 3.52, 148,400 protein sequences), including all the frequent contaminants observed in MS, using Mascot 2.3 (Matrix Science) search engine. Trypsin/P was selected, allowing 2 missed cleavages. Precursor mass tolerance was initially set at 50 ppm [13], while fragment mass tolerance was set at 0.6 Da for ETD-IT fragmentation and to 0.05 Da for HCD and ETD-FT fragmentation. Carbamidomethylation of cysteines was set as fixed modification whereas oxidation of methionines, dimethylation, dimethylation:2 H(4) and dimethylation:2 H(4)13 C(2) of K and N-term were set as variable modifications. The resulting dat files were exported and filtered for a 1% false discovery rate at peptide level using in-house developed software "Rockerbox" (version 1.2.6) [14].

2.5. Peptide and protein quantification

Peptide and protein quantification were performed using a dimethyl-adapted version of MSQuant [15]. All peptide ratios were manually validated to assure proper peak area integration and minimize, in this way, quantification errors caused by retention time shifts due to the deuterated forms of the dimethyl labeling.

2.6. Prediction of the quantification error

The quantification error caused by the overlapping issue was estimated theoretically as a function of mass and labeling ratio. The model amino acid *averagine*, with molecular formula $C_{4.9384}H_{7.7583}N_{1.3577}O_{1.4773}S_{0.0417}$ and average molecular mass of 111.1254 Da was used to estimate isotopic distributions at different molecular masses, as described by Senko *et al* [7].

2.7. Correction of quantitative results

Overlapping Peaks Finder, an in-house written Java program, was used to correct quantitative data and is available at