JOURNAL OF PROTEOMICS 72 (2009) 761–770

available at www.sciencedirect.com

www.elsevier.com/locate/jprot

On the acquisition of +1 charge states during high-throughput proteomics: Implications on reproducibility, number and confidence of protein identifications

Victor J. Nesatyy, Ksenia Groh, Holger Nestler, Marc J.-F. Suter*

Eawag - Swiss Federal Institute of Aquatic Science and Technology, Ueberlandstrasse 133, Duebendorf, Switzerland

ARTICLE INFO

Article history:
Received 22 December 2008
Accepted 11 March 2009

Keywords:
Proteomics
High-throughput
MudPIT
Charge state
Reproducibility

ABSTRACT

Modern high-throughput methods for the proteome analysis are gradually replacing more traditional 2D gel-based techniques. Almost immediately after the introduction of high-throughput proteomics techniques in 2001, reproducibility of the results became an issue. Extensive discussion in the literature led to the conclusion that certain "undersampling" exhibited during measurements could be due to the stochastic nature of the data-dependent sampling, routinely used with current mass spectrometry equipment. At the same time, the effect of the acquisition of different charge states on the reproducibility and confidence of protein identifications, to the best of our knowledge, has never been properly evaluated. There exists the frequently voiced yet hardly documented opinion that +1 charge states should be rejected during data-dependent acquisition. The work presented here shows that inclusion of the +1 charge state in the data-dependent acquisition protocols can indeed lead to improved proteome coverage, reproducibility, and the confidence of protein identifications by high-throughput proteomics. It was also shown that contrary to the established opinion, gas-phase dissociation of singly charged peptide species results in rich fragmentation patterns containing both b- and y-ions allowing for successful and confident peptide identification.

1. Introduction

Modern high-throughput methods for the analysis of a proteome such as the Multidimensional Protein Identification Technology (MudPIT) [1] and Gel interfaced with Liquid Chromatography (GeLC) [2,3] are gradually replacing more traditional 2D gel-based techniques. Today their application ranges from environmental [4,5] to medical research [6]. The success of these LC-based approaches is primarily based on their ability to consistently outperform 2D gels with regard to the number of protein identifications, their ability to work with mixtures of wider concentration range, and high-throughput capability. For example, a traditional 2D gel protocol requires at least 3 days to separate and visualize a complex protein mixture, with additional days needed for spot picking, in-gel digestion and MS analysis for protein identification. The same task can be completed in two days by the above mentioned high-throughput techniques. It was also repeatedly shown by various research groups that the number of protein identifications produced by GeLC and MudPIT is 5 to 10 times higher than that produced by 2D gels [7,8]. Thus the main point to be discussed is not the superiority of high-throughput techniques, but the quality, consistency and reproducibility of the data obtained by these methods.

Almost immediately after the introduction of high-throughput proteomics techniques in 2001, reproducibility of the results became an issue. To date, the majority of the high-throughput data acquired was obtained using MudPIT [9–13]. In one of the first publications devoted to this topic Durr et al. showed that 7 to 9 repetitive MudPIT measurements with the

* *Corresponding author.*
E-mail addresses: victor.nesati@epfl.ch (V.J. Nesatyy), suter@eawag.ch (M.J.-F. Suter).

1874-3919/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2009.03.003

ion trap equipment available at that time were needed in order to reach 95% confidence of analytical completeness of the proteome of the luminal endothelial cell membranes [9]. Liu et al. addressed the same question by presenting a statistical model capable of accurately predicting the number of repetitive MudPIT measurements required for saturated sampling of complex mixtures [10]. Using previously published protein abundance data, they showed that the soluble protein fraction from yeast cell lysates required 9 repetitive MudPIT measurements to achieve 95% saturation in protein identifications [10]. At the same time, an analysis by Kislinger et al. of 4 subcellular fractions from a heart tissue isolate, performed on the same ion trap equipment (LCQ DECA XP, ThermoFisher Scientific), showed that virtually complete saturation with regards to the total number of identified proteins could be achieved after only 5 individual MudPIT analyses [13].

Extensive discussion in the literature led to the conclusion that certain "undersampling" exhibited by single MudPIT measurements could be due to the stochastic nature of the data-dependent sampling, where the mass spectrometer sequentially selects individual peptide precursor ions for fragmentation in the order of their intensities [10,13,14]. In general, data-dependent acquisition is biased towards detection of higher intensity signals, originating from higher abundance proteins, thus leaving the lower intensity peptide signals, produced by less abundant proteins, undetected [14]. This problem is caused by the limited duty cycle and restricted dynamic range of the current MS instrumentation, resulting in the inability to analyze every precursor ion eluting from the chromatographic column [15]. Moreover, the simultaneous presence of coeluting compounds in the ion source, competing for protons often results in ion suppression effects mostly affecting less abundant species [14]. It appears that repeated measurements of the same sample are able to partially overcome this stochastic "undersampling" problem and thus achieve better proteome coverage.

At the same time, the effect that the acquisition of different charge states has on the reproducibility and confidence of protein identifications has never properly been addressed. There exists the frequently voiced opinion that +1 charge states should be rejected during data-dependent acquisition. It has been repeatedly stated that singly charged signals can lead to false positive identifications because they are unlikely to arise from a tryptic peptide but may rather be artifacts caused by chemical noise [16]. It has then been further argued that the removal of +1 charge states from the analysis will allow spending more time on the analysis of ions potentially yielding useful data and thus reducing the cycle time [16]. The same idea was expressed in a couple of protocols published online [17,18]. However, the original MudPIT publications and follow-up reports using the same technique always included MS/MS data from singly charged species [1,19].

The work presented here shows that fragmentation of singly charged peptide species in scan-dependent MS/MS experiments on the Orbitrap mass spectrometer can provide a basis for reliable and confident peptide identification, leading to improved reproducibility and proteome coverage in high-throughput proteomics. The experiments were performed on protein extracts from the two widely used model organisms, the green alga *Chlamydomonas reinhardtii* and zebrafish (*Danio rerio*).

2. Materials and methods

2.1. *Zebrafish maintenance and breeding*

The zebrafish (WIK strain) were maintained according to recommended procedures [20]. Briefly, fish were reared in a recirculating flow-through system filled with tap water treated with active carbon and UV light. The room was maintained at 28 °C and 14 h/10 h light/dark cycle. The diet consisted of live food *Artemia* nauplia and dry vitamin flakes. For breeding, about 30 reproductively active adults were kept in 75 L glass aquaria with active filtering. The eggs were collected every morning 2 h after the lights-on. The eggs were washed and raised in E3 medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$). Before protein extraction, larvae were euthanized in MS222 and devitellogenization was carried out according to Link et al. [21].

2.2. *Algae cultivation*

C. reinhardtii strain CC-125, obtained from the Chlamydomonas Genetics Center (Duke University, NC), was cultured autotrophically in the liquid medium as described by Le Faucheur et al. [22]. Precultures in late exponential phase were diluted to 5×10^5 cells/ml in fresh medium and grown for 6 h on a rotary shaker (90 rpm) at 25 °C and constant illumination of 120 μmol/m^2·s photons which corresponds to photosynthetically active radiation. The cells were then harvested and further processed for protein extraction.

2.3. *Protein extraction*

Algae were placed in a lysis buffer (0.1% CHAPS, 25 mM KCl, 25 mM Tris–HCl, 25 mM $MgCl_2$, pH 7.5) and disrupted by three 10 s sonication bursts. Zebrafish proteins were extracted from devitellogenized 3-day-old zebrafish embryos placed in the same lysis buffer and homogenized with a pestle. In both cases, cell debris were removed by centrifugation and proteins were precipitated from the supernatant using methanol/chloroform. The precipitated proteins were then redissolved in resolubilization buffer (9 M urea, 2 M thiourea, 50 mM Tris–HCl) to a concentration of 1 mg/ml (measured by Bradford assay), reduced, alkylated and digested as described previously [1,19]. Peptides were loaded onto a triphasic MudPIT column (C18-SCX-C18) made in-house from a 100 μm ID fused-silica capillary drawn to a fine tip using a Sutter-2000 needle puller, desalted and eluted using the standard MudPIT protocol.

2.4. *Mass spectrometry and data processing*

The LTQ-Orbitrap was mass calibrated using a polyLys mixture and operated at 1.5 kV spray voltage in positive ion mode. To promote formation of singly charged peptide species the tube lens was set to 250 V and the ion transfer capillary temperature to 300 °C. One full scan FT mass spectrum (400–2000 *m/z*, resolution of 60 000) was followed by seven data-dependent MS/MS scans acquired in the linear ion trap with normalized collision energy (setting of 35%) and continuously repeated throughout each MudPIT step. Dynamic exclusion