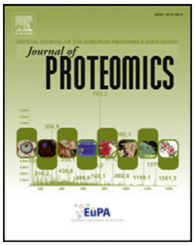
JOURNAL OF PROTEOMICS 105 (2014) 31–45

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot



Improving the identification rate of data independent label-free quantitative proteomics experiments on non-model crops: A case study on apple fruit☆

Kim Buts[a,*], Servaas Michielssens[b,c], Maarten L.A.T.M. Hertog[a], Eisuke Hayakawa[d], Jan Cordewener[e], Antoine H.P. America[e], Bart M. Nicolai[a,f], Sebastien C. Carpentier[g,h,**]

[a]BIOSYST–MeBioS, KU Leuven, Belgium
[b]Quantum Chemistry and Physical Chemistry Section, KU Leuven, Belgium
[c]Computational Biomolecular Dynamics Group, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
[d]Research Group of Functional Genomics and Proteomics, KU Leuven, Belgium
[e]Plant Research International, Wageningen, Netherlands
[f]Flanders Centre of Postharvest Technology, Leuven, Belgium
[g]SYBIOMA, KU Leuven, Belgium
[h]BIOSYST — Crop Biotechnics, KU Leuven, Belgium

ARTICLE INFO

Available online 22 February 2014

Keywords:
Non-model crop
Plant proteomics
Data independent acquisition
Label free quantification

ABSTRACT

Complex peptide extracts from non-model crops are troublesome for proper identification and quantification. To increase the identification rate of label free DIA experiments of Braeburn apple a new workflow was developed where a DDA database was constructed and linked to the DIA data. At a first level, parent masses found in DIA were searched in the DDA database based on their mass to charge ratio and retention time; at a second level, masses of fragmentation ions were compared for each of the linked spectrum. Following this workflow, a tenfold increase of peptides was identified from a single DIA run. As proof of principle, the designed workflow was applied to determine the changes during a storage experiment, achieving a two-fold identification increase in the number of significant peptides. The corresponding protein families were divided into nine clusters, representing different time profiles of changes in abundances during storage. Up-regulated protein families already show a glimpse of important pathways affecting aging during long-term storage, such as ethylene synthesis, and responses to abiotic stresses and their influence on the central metabolism.

Biological significance

Proteomics research on non-model crops causes additional difficulties in identifying the peptides present in, often complex, samples. This work proposes a new workflow to retrieve

---

☆ This article is part of a Special Issue entitled: Proteomics of non-model organisms.

* *Correspondence to:* K. Buts, BIOSYST–MeBioS, KU Leuven, Willem de Croylaan 42, Leuven (Heverlee) B-3001, Belgium.
** *Correspondence to:* S. Carpentier, BIOSYST — Crop Biotechnics, KU Leuven, Willem de Croylaan 42, Leuven (Heverlee) B-3001, Belgium. Fax: +32 16 32 19 93.
E-mail addresses: kim.buts@kuleuven.be (K. Buts), sebastien.carpentier@biw.kuleuven.be (S.C. Carpentier).

http://dx.doi.org/10.1016/j.jprot.2014.02.015

more identifications from a set of quantitative data, based on linking DIA and DDA data at two consecutive levels. As proof of principle, a storage experiment on Braeburn apple resulted in twice as much identified storage related peptides. Important proteins involved in central metabolism and stress are significantly up-regulated after long term storage.

This article is part of a Special Issue entitled: Proteomics of non-model organisms.

1. Introduction

Since the birth of 'proteomics' [1], workflow improvements have been made, ranging from protein extraction up to data analysis. Although plant proteomics isn't in its infancy anymore, many hurdles still need to be taken [2–4], especially when working with non-model crops [5,6]. Often, the genome is not sequenced and protein databases are poorly structurally and functionally annotated. In addition, ploidy, heterozygosis, alternative splicing and posttranslational modifications enlarge the gap between the genome and the proteome.

Shotgun proteomics based on liquid chromatography coupled to mass spectrometry, is under continuing development. Parallel to the improvement of the mass spectrometers new proteomic strategies and workflows are further perfectionized. The key steps of a peptide-based shotgun proteomic method are (*i*) digestion of the extracted proteins, (*ii*) LC–MS/MS analysis, followed by (*iii*) computational analysis of the obtained spectra to deduce the peptide sequences [7–9]. Identification of the present peptides, and of their corresponding proteins, is not only based on database searching [10–14] and de novo sequencing of unassigned spectra [15–17], but increasingly also on spectral library searching [18–22], giving rise to the use of spectral archives [23]. Typically, LC–MS/MS analysis starts with an MS survey scan where peptide precursor masses are measured, followed by an MS/MS scan for fragmentation of the selected precursor ion. Because the selection of precursor ions relies on, typically, intensity and charge information gathered in the MS cycle this type of LC–MS/MS is called data-dependent acquisition (DDA). Even though it is a versatile and commonly used technique, it has several shortcomings inherent to the serial nature of the MS and MS/MS cycles. Not only is there a bias towards the more abundant peptides but also no MS scans can be obtained while fragmentation is being performed in the MS/MS scan in most of the current mass spectrometers. In complex biological samples, co-eluting peaks can lead to chimeric spectra, reduced reproducibility and loss of information about less abundant peptides in DDA analysis. During recent years, mass spectrometers have developed towards higher mass accuracy, resolving power, sensitivity, scan speed, reproducibility and lower detection limits [9,24]. For example, the use of hybrid LTQ-Orbitrap devices [25,26], high energy C-trap dissociation [27], parallel reaction monitoring [28], coupling of a quadrupole mass filter to an Orbitrap analyzer [29,30] and combining quadrupole, Orbitrap and ion trap mass analysis [31,32], all contributed to improvements in proteomics experiments, and in particular towards better peptide identifications and quantifications.

Protein quantification can be performed using one of the available labeling approaches [33–37] that mostly rely on measuring the relative protein abundance, although also absolute quantification is possible [38]. Label-free techniques [39,40] rely either on spectral counting [41] or on the use of averaged, normalized ion intensities [42]. Although label based protocols in general reach higher sensitivity than label-free protocols, the compatibility with complicated experimental set-ups, ease of use and lower cost still favor the application of label-free shotgun proteomics experiments [43].

In order to identify and quantify peptides from multiple complex samples in a rapid, consistent, reproducible, accurate and sensitive way, data independent acquisition (DIA) protocols have been developed for label-free shotgun proteomics.

In order to generate both precursor and product ions in a single run, Waters Corporation developed a new data independent MS approach, UPLC/MS^E [44]. To separate complex peptide mixtures in several fractions, 2D ultra performance liquid chromatography (UPLC) is used [45]. This step is followed by a mass spectrometry run where an alternating energy level is applied to the collision cell. At low energy, accurate precursor masses are obtained, while fragmentation spectra of all parent masses are taken at high collision energy. So in theory, only one analytical run is needed to obtain quantitative and qualitative information on the sample. However, in the case of complex whole tissue extracts from non-model organisms, the hybrid MS^E fragmentation spectra usually lead to poor protein identifications. Therefore, data independent (MS^E) and data dependent acquisition (DDA) runs should be combined to acquire all necessary information. An advantage from the MS^E approach is the availability of fragmentation information, which can be used together with the parent mass and retention time in order to link this data to the DDA data.

Compared to the DDA data, DIA spectra are noisy because of the reduced precursor selectivity. Therefore AB SCIEX uses a high-resolution MS/MS time of flight instrument to cycle repeatedly through a number of pre-set sequential precursor isolation windows (swaths) in order to retrieve fragment ion spectra from the detected precursor ions [46]. Egertson et al. [47] on the other hand presented a multiplexing MS/MS strategy, in which they used five separate 4-m/z-wide isolation windows per spectrum. Afterwards the data was demultiplexed into the different isolation windows, resulting in data with a high sampling frequency and high precursor selectivity.

Another development is the use of ion mobility (IM) [48–50] in which structural properties of the parent ion are used to give the different parent ions a different kinetic energy before fragmentation. The different mobility of the parent ions enables to calculate the origin of the MS/MS fragments and leads to a deeper proteome coverage and a more confident peptide identification. When ion mobility is used in combination with an MS^E workflow, it is called high definition MS^E (HDMS^E). Due to transmission loss and detector saturation, both lower and higher intensity ions might be missed during HDMS^E analysis. So, even though IM