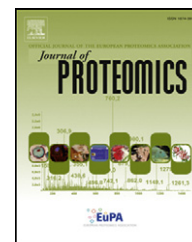


Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot

EasyProt — An easy-to-use graphical platform for proteomics data analysis

Florent Gluck^{a,b,*}, Christine Hoogland^{a,b,d,1}, Paola Antinori^{a,b}, Xavier Robin^b, Frederic Nikitin^d, Anne Zufferey^b, Carla Pasquarello^c, Vanessa Fétaud^{b,c}, Loïc Dayon^{b,2}, Markus Müller^d, Frederique Lisacek^d, Laurent Geiser^{a,b,d}, Denis Hochstrasser^{a,b,e}, Jean-Charles Sanchez^b, Alexander Scherl^{a,b,c}

^aSwiss Center for Applied Human Toxicology (SCAHT), Geneva, Switzerland

^bBiomedical Proteomics Research Group (BPRG), Department of Human Protein Sciences, University of Geneva, Geneva, Switzerland

^cProteomics Core Facility, University of Geneva, Geneva, Switzerland

^dProteome Informatics Group (PIG), Swiss Institute of Bioinformatics, Geneva, Switzerland

^eClinical Proteomics Laboratory, Department of Genetics and Laboratory Medicine, Geneva University Hospitals, Geneva, Switzerland

ARTICLE INFO

Article history:

Received 12 September 2012

Accepted 18 December 2012

Available online 28 December 2012

Keywords:

Protein identification

Protein quantification

Label-free

Isobaric labeling

Post-translational modifications

Mass spectrometry

ABSTRACT

High throughput protein identification and quantification analysis based on mass spectrometry are fundamental steps in most proteomics projects. Here, we present EasyProt (available at <http://easyprot.unige.ch>), a new platform for mass spectrometry data processing, protein identification, quantification and unexpected post-translational modification characterization. EasyProt provides a fully integrated graphical experience to perform a large part of the proteomic data analysis workflow. Our goal was to develop a software platform that would fulfill the needs of scientists in the field, while emphasizing ease-of-use for non-bioinformatician users. Protein identification is based on OLAV scoring schemes and protein quantification is implemented for both, isobaric labeling and label-free methods. Additional features are available, such as peak list processing, isotopic correction, spectra filtering, charge-state deconvolution and spectra merging. To illustrate the EasyProt platform, we present two identification and quantification workflows based on isobaric tagging and label-free methods.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In the field of proteomics, a wide number of software exists [1] to perform the analysis of mass spectrometry-based proteomics data. The number of software available can be overwhelming and even confusing in regard to which one to use. Which tool would be most appropriate to solve a given problem? There is no obvious answer to this simple question. With the

omnipresence of the Internet, several portals, such as ExPASy (<http://expasy.org>) for instance, are now featuring a vast array of tools that can potentially be used. However, the selection remains difficult, especially given the lack of fully integrated and easy-to-use software available to perform the entire proteomic data analysis workflow. Given the diversity of available software and their specificities, there is a need for an integrated, easy-to-use software platform. Unfortunately,

* Corresponding author at: Biomedical Proteomics Research Group, Department of Human Protein Sciences, Faculty of Medicine, University of Geneva, Rue Michel Servet 1, 1211 Geneva 4, Switzerland. Tel.: +41 22 379 41 69; fax: +41 22 379 55 02.

E-mail address: florent.gluck@unige.ch (F. Gluck).

¹ Present address: Queensland Institute of Medical Research, Brisbane, Australia.

² Present address: Proteomics and Metabolomics Core, Nestlé Institute of Health Sciences, Lausanne, Switzerland.

creating a full-fledged pipeline by connecting the different pieces of software together is a daunting task for anyone but programmers, considering such a project certainly involves programming. Developing graphical interfaces to existing command line tools is a non-trivial and time consuming task which requires extensive programming skills.

Software such as the Trans-Proteomic Pipeline [2], the Computational Proteomics Analysis System [3], the Systems Biology Experiment Analysis Management System (<http://www.sbeams.org/Proteomics/>) or the Virtual Expert Mass Spectrometrist [4] are obvious candidates, but while being greatly customizable, they generally require nontrivial configuration work as well as various external dependencies to work properly. EasyProt on the other hand is a fully integrated solution in which underlying technologies (such as the search engine) and their complexities are integrated in the platform and thus not visible to the end user. Compared with other web-based software, EasyProt offers a more modern and dynamic web interface as well some unique features such as on-the-fly false discovery rate (FDR) computation and fully integrated isobaric and label-free quantification processing.

Our goal was to develop a software platform that would fulfill the needs of scientists in the field, while emphasizing ease-of-use for non-bioinformatician users. To accomplish this objective, we worked in close collaboration with researchers from proteomic laboratories, starting from a basic protein identification workflow, while incrementally adding new features over time, such as exports, visualizers and quantification pipelines. The EasyProt platform covers the whole workflow from proprietary data file formats produced by mass spectrometers, to identification and quantification results, ready to be analyzed by researchers and scientists with various backgrounds. EasyProt is implemented in the Java language and is structured around two distinct graphical applications.

The first one, EasyprotConv, is a standalone desktop application to process mass spectrometers' proprietary data formats. EasyprotConv features peak list processing such as precursor isotope correction [5], spectra filtering, charge state deconvolution and low collision energy — higher collision energy spectra merging for isobaric relative peptide quantification with Orbitrap-hybrid instruments (LTQ-OT) [6]. Through the use of Superhirn [7], EasyprotConv performs label-free [8] processing, such as peak detection, liquid chromatography alignment and feature map normalization.

The second component, EasyProt, is a multi-user web application implementing peptide and protein identification through Olav [9,10], unexpected post-translational modification identification with Popitam [11], isobaric quantification (TMT [12] and iTRAQ [13]) with IsoQuant and Isobar [14], label-free quantification, and several viewers and exports. A particularity of EasyProt resides in how one can set the false discovery rate [15] on-the-fly, after the identification search has completed rather than at submission time. When dealing with multiple identification searches spread across multiple files (e.g., various peptide fractionation methods), EasyProt transparently merges all results from several searches into a single result that can then be exported.

Both EasyProt and EasyprotConv are freely available to academic institutions at the following web address: <http://easyprot.unige.ch>. This website also features video tutorials

showing how to use EasyProt to perform common tasks such as protein identification and quantification.

To illustrate the EasyProt platform, two identification and quantification workflows are presented. The first one is based on a labeled quantification method by isobaric tagging, while the second one is based on a label-free quantification approach. Both quantification workflows were conducted and validated against the same samples from ProteoRed Multicentric Proteomic Experiment 2009 PME5 (<http://www.proteored.org>).

2. Material and methods

2.1. Sample preparation

Our laboratory participated in the ProteoRed multicentric experiment 2009 PME5 initiative and received two identical complex protein mixtures, labeled A and B, containing each 100 µg of total protein. The mixture consisted of a single ion-exchange chromatographic fraction of a soluble *Escherichia coli* digest. Four mammalian proteins, cytochrome C, apomyoglobin, aldolase and serum albumin, were spiked at different concentration levels into samples A and B. For the TMT sample preparation, off-gel electrophoresis (OGE) [16] separation was performed in order to decrease the mixture's complexity. Using TMT 6-plex tags, three replicates of each group, A and B, were prepared leading to 12 sample fractions. They were then analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS) on a LTQ-Orbitrap XL. In case of the label-free sample preparation, samples of group A and B were also analyzed by LC-MS/MS on a LTQ-Orbitrap XL. Gas-phase fractionation (GPF) [17] was used to decrease sample's complexity and a total of eight fractions were obtained. The experimental procedure is covered in greater details in the Supplementary material document.

2.2. Data file preparation

For the isobaric quantification experiment, a total of 12 RAW files were produced. EasyprotConv was used to create peak lists into the MGF file format. The option to combine peaks from CID and HCD fragments within the TMT 6-plex reporter mass window was selected. EasyprotConv was used to merge the resulting twelve MGF files into a single MGF file. In regard to the label-free quantification experiment, a total of eight RAW files were obtained from the GPF analysis. Each group, A and B, was defined by four RAW files, corresponding to four technical replicates per group. EasyprotConv was used to generate peak lists into the MGF file format with the isotopic correction option enabled in order to correct precursor ions masses. The label-free option was selected to proceed to the alignment of the analyses as well as the generation of the normalized master map of all chromatographic features. EasyprotConv was used again to merge the resulting eight MGF files into a single MGF file.

2.3. Protein identification

As described in the ProteoRed multicentric experiment's guidelines, the PME5 database, as a FASTA file, was downloaded from

Download English Version:

<https://daneshyari.com/en/article/7638242>

Download Persian Version:

<https://daneshyari.com/article/7638242>

[Daneshyari.com](https://daneshyari.com)