Technical Note

Data Independent Acquisition analysis in ProHits 4.0

Guomin Liu [a], James D.R. Knight [a], Jian Ping Zhang [a], Chih-Chiang Tsou [b,c], Jian Wang [d,e], Jean-Philippe Lambert [a], Brett Larsen [a], Mike Tyers [f], Brian Raught [g], Nuno Bandeira [d,e,h], Alexey I. Nesvizhskii [b,c], Hyungwon Choi [i,j], Anne-Claude Gingras [a,k,*]

[a] Centre for Systems Biology, Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, Canada
[b] Department of Pathology, University of Michigan, Ann Arbor, USA
[c] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, USA
[d] Center for Computational Mass Spectrometry, University of California, San Diego, La Jolla, California, USA
[e] Department of Computer Science and Engineering, University of California, San Diego, La Jolla, California, USA
[f] Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, Québec, Canada
[g] Princess Margaret Cancer Institute, Department of Medical Biophysics, University of Toronto, Ontario, Canada
[h] Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California, USA
[i] Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore
[j] Institute of Molecular and Cell Biology, Agency for Science, Technology, and Research, Singapore, Singapore
[k] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

ARTICLE INFO

ABSTRACT

Affinity purification coupled with mass spectrometry (AP-MS) is a powerful technique for the identification and quantification of physical interactions. AP-MS requires careful experimental design, appropriate control selection and quantitative workflows to successfully identify bona fide interactors amongst a large background of contaminants. We previously introduced ProHits, a Laboratory Information Management System for interaction proteomics, which tracks all samples in a mass spectrometry facility, initiates database searches and provides visualization tools for spectral counting-based AP-MS approaches. More recently, we implemented Significance Analysis of INTeractome (SAINT) within ProHits to provide scoring of interactions based on spectral counts. Here, we provide an update to ProHits to support Data Independent Acquisition (DIA) with identification software (DIA-Umpire and MSPLIT-DIA), quantification tools (through DIA-Umpire, or externally via targeted extraction), and assessment of quantitative enrichment (through mapDIA) and scoring of interactions (through SAINT-intensity). With additional improvements, notably support of the iProphet pipeline, facilitated deposition into ProteomeXchange repositories and enhanced export and viewing functions, ProHits 4.0 offers a comprehensive suite of tools to facilitate affinity proteomics studies.

Significance: It remains challenging to score, annotate and analyze proteomics data in a transparent manner. ProHits was previously introduced as a LIMS to enable storing, tracking and analysis of standard AP-MS data. In this revised version, we expand ProHits to include integration with a number of identification and quantification tools based on Data-Independent Acquisition (DIA). ProHits 4.0 also facilitates data deposition into public repositories, and the transfer of data to new visualization tools.

© 2016 Elsevier B.V. All rights reserved.

Mass spectrometry (MS) data can be a challenge to manage, analyze and subsequently present to an audience in simple, intuitive ways. MS data must be archived, searched, scored, compared and visualized, often requiring a variety of unconnected, non-standardized software tools. These challenges are now compounded by the growing use of data-independent acquisition (DIA) methods [1–9], and more specialized applications such as interaction proteomics [10–14].

In 2010, we introduced an open source Laboratory Information Management System (LIMS) called ProHits [15,16], designed to handle data generated in a proteomics facility, but which also provided specialized tools for the analysis of AP-MS, a popular proteomics application. ProHits is installed on a LINUX server (behind a firewall); via a web interface, authorized users can access search engines, analysis tools and multiple data visualization options. Through its Data Management module, ProHits automatically backs-up all data acquired in a mass spectrometry facility, converts vendor specific files to common formats (through ProteoWizard [17] or vendor software) and facilitates database searching using both free, open source tools (Comet [18], MSGF+ [19], X!Tandem [20]) and the commercial search engine
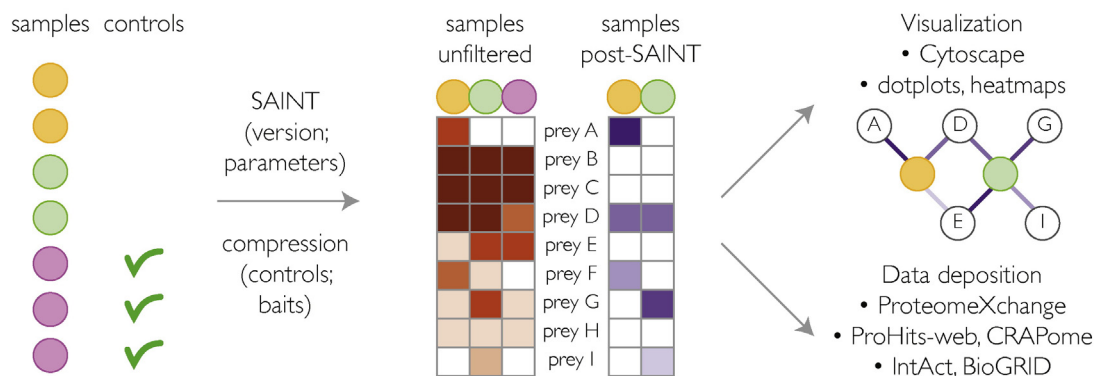
a
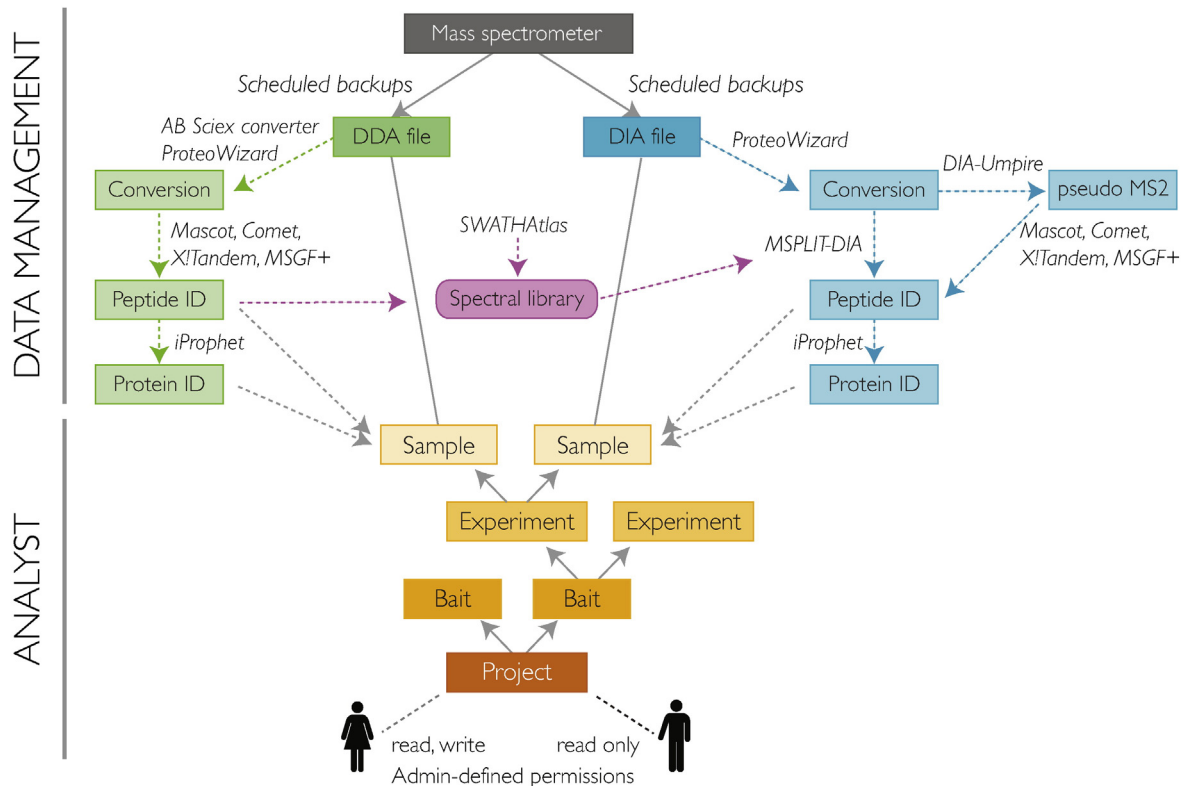
b

Fig. 1. ProHits organization and protein interaction scoring. A) The ProHits system consists of two principal modules, a Data Management module and an Analyst module. All mass spectrometers in a facility can be connected to ProHits: scheduled backups of the mass spectrometry data are performed followed by file conversion and database searches. DIA identification is supported by DIA-Umpire and the spectral matching tool MSPLIT-DIA. Peptide and protein identification results are parsed to a Sample defined in the Analyst module. The Samples are defined in a Project → Bait → Experiment → Sample hierarchy. Permissions for different projects are assigned to users in an Admin section. B) Schematic workflow for protein interaction analysis using SAINT through ProHits. Within a project, a user defines which samples should be analyzed and specifies which of those are the controls. The SAINT version (SAINTexpress or standard SAINT) is selected, alongside optional parameters and sample compression level. SAINT uses the quantitative matrix to derive the probability of interactions. Post analysis with SAINT, the data can be visualized or deposited in repositories from ProHits itself.

Mascot [21]. Search results can be evaluated using the PeptideProphet [22] and ProteinProphet [23] components of the Trans Proteomics Pipeline (TPP [24]). The current version of ProHits also enables the use of iProphet [25] for combining the identification results from multiple search engines into a single output (Fig. 1a; green boxes).

To enable sample tracking, an "Analyst" module helps users organize their data into projects to which different user permissions can be assigned (Fig. 1a). Due to the strong focus of our team on affinity proteomics, the flow within each project is organized according to a "bait" protein. The bait is defined by its gene name, species, protein accession number and epitope tag(s), as appropriate. Note however, that the

system can be used for any type of enrichment approach (e.g. with a nucleic acid or chemical compound as a bait), or even for more generic profiling by simply considering the "bait" level as part of an organizational hierarchy (there are 4 layers: "project" → "bait" → "experiment" → "sample"). Once a "bait" is entered, "experiments" can be associated to it, and annotated with text-based protocols (accessible via drop-down menus), controlled vocabularies, and free text notes. Under the "experiment" hierarchy, "samples" can be created (Fig. 1a; orange boxes). Each sample is associated with a unique identifier: creating "samples" first in the Analyst module and naming the files on the mass spectrometry acquisition computer following this