JOURNAL OF PROTEOMICS 118 (2015) 112–129

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Development of a computational framework for the analysis of protein correlation profiling and spatial proteomics experiments☆

Nichollas E. Scott[a,*], Lyda M. Brown[a], Anders R. Kristensen[b], Leonard J. Foster[a,*]

[a]Centre for High-throughput Biology, University of British Columbia, Vancouver V6T 1Z4, British Columbia, Canada
[b]Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver V5Z 4S6, British Columbia, Canada

ARTICLE INFO

Available online 20 November 2014

Keyword:
Protein correlation profiling
SILAC
Matlab
Spatial proteomics

ABSTRACT

Standard approaches to studying an interactome do not easily allow conditional experiments but in recent years numerous groups have demonstrated the potential for co-fractionation/co-migration based approaches to assess an interactome at a similar sensitivity and specificity yet significantly lower cost and higher speed than traditional approaches. Unfortunately, there is as yet no implementation of the bioinformatics tools required to robustly analyze co-fractionation data in a way that can also integrate the valuable information contained in biological replicates. Here we have developed a freely available, integrated bioinformatics solution for the analysis of protein correlation profiling SILAC data. This modular solution allows the deconvolution of protein chromatograms into individual Gaussian curves enabling the use of these chromatography features to align replicates and assemble a consensus map of features observed across replicates; the chromatograms and individual curves are then used to quantify changes in protein interactions and construct the interactome. We have applied this workflow to the analysis of HeLa cells infected with a *Salmonella enterica* serovar Typhimurium infection model where we can identify specific interactions that are affected by the infection. These bioinformatics tools simplify the analysis of co-fractionation/co-migration data to the point where there is no specialized knowledge required to measure an interactome in this way.

Biological Significance
We describe a set of software tools for the bioinformatics analysis of co-migration/co-fractionation data that integrates the results from multiple replicates to generate an interactome, including the impact on individual interactions of any external perturbation. This article is part of a Special Issue entitled: Protein dynamics in health and disease. Guest Editors: Pierre Thibault and Anne-Claude Gingras.

Abbreviations: FBS, fetal bovine serum; PBS, phosphate buffer saline; r.c.f., relative centrifugal force; PCP, Protein correlation profiling; SEC, size exclusion chromatography; SILAC, stable isotope labeling by amino acids in cell culture; TPR, true positive rate; FPR, false positive rate; PPI, protein–protein interaction; TP, true positive; FP, false positive; TN, true negative; FN, false negative; M/L, medium over light; H/L, heavy over light

☆ This article is part of a Special Issue entitled: Protein dynamics in health and disease. Guest Editors: Pierre Thibault and Anne-Claude Gingras.

* *Corresponding authors at:* Department of Biochemistry and Molecular Biology, University of British Columbia, 2125 East Mall, Vancouver V6T 1Z4, British Columbia, Canada. Tel.: +1 604 822 8311.
E-mail addresses: nichollas.e.scott@gmail.com (N.E. Scott), foster@chibi.ubc.ca (L.J. Foster).

http://dx.doi.org/10.1016/j.jprot.2014.10.024
1874-3919/

1. Introduction

The coordination of proteins into high-order arrangements is a ubiquitous phenomenon within all domains of life [1,2]. These associations, which typically form in an ordered manner [3] via energy dependent [4,5] and independent (self-assembly) processes [6], act to enhance the functional capabilities of the proteome through the formation of protein complexes. Within biological systems, protein complexes provide a means by which multi-stepped procedures can be accomplished more efficiently through spatial and temporal confinement of both enzymes and substrates [7,8]. This confinement leads to an enhancement in the efficiency of multi-stepped reactions [7,8], as well as a means to coordinate processes between subunits irrespective of the transcriptional and translational regulation, which, as has been noted, can vary between protein complex subunits [9,10]. This increase in efficiency and coordination, in conjunction with the enhancement in protein stability, modulation of binding properties and as a means for specialization/evolvability have all been attributed to the widespread implementation of protein–protein binding and protein complex formation in biological systems [7,8]. Consistent with these benefits, protein associations/complexes are thought to be widespread within all proteomes, with the majority of proteins assumed or shown to be involved with at least one protein complex [1,2,11,12].

Due to the ubiquitous nature of protein complexes and their fundamental role in protein function, numerous efforts have aimed to enhance our understanding of the interaction landscape of both model systems [1,2,11,13,14] and higher eukaryotes [15–17]. These efforts have provided a wealth of information about the composition of and interactions within protein complexes, providing an essential resource for the scientific community [18,19] and novel insight into the organization of biological systems as a whole [8,20]. The data from these studies have revealed the presence of modular but interconnected networks [21,22], referred to as protein–protein interaction networks (PPI) or interactomes, in which the topology, connectivity and temporal state all influence the network [23]. Importantly, the modular nature of PPI [22,24] appears to directly result in the robustness, versatility and error-tolerance of biological systems [21]. The modular nature of the PPI bestows these properties in four ways:1) by allowing the re-use of protein subunits between related but functionally discrete complexes, such as in the case of the RPB10 subunit of RNA polymerases I, II, and III [25]; 2) by enabling the ability to augment the properties of specific protein complexes via the exchange/addition of unique subunits, such as the altered antigen processing of the proteasome on binding of the proteasome activator of P28 (composed of PA28α, PA28β) [26–28]; 3) by tolerating the potential for duplication/specialization of a subunit which can occur uncoupled to all other subunits of a given complex or component of the proteome [8,23]; and 4) by enabling potential redundancy of duplicated subunits [29]. This modular nature of the proteome has widespread ramifications, especially during the elucidation of interactomes since the composition of complexes can differ dramatically between biological conditions of interest.

Traditional approaches to studying interactomes have typically employed targeted strategies such as protein tagging or protein-fragment complementation assays to assess protein interactions [30–32]. These approaches, although powerful, have limited scalability due to being largely incompatible with sample multiplexing and the requirement of protein tagging through genetic manipulation. These factors result in a limitation in the number/combination of constructs that can be assessed in a single experiment. Furthermore the requirement that constructs behavior similarly to the native untagged counterparts is not guaranteed with the type of tag, the level of expression and the protein being manipulated all affecting the success of these approaches [33]. Although these concerns can be minimized through careful control of the experimental design, such as the expression of constructs at or near native levels [1,2,34,35] and the use of parallelization/automation to manage libraries of increasing complexity [36,37], sample multiplexing at a system wide level still represents a significant challenge. Recently, we and others demonstrated that co-fractionation/co-migration [15,16,38] and the principle of "guilt by association" [39] can be used to measure interactomes much faster, yet with the same sensitivity and specificity, as more conventional approaches. These approaches have the added advantages that they do not require protein tagging and that they can be multiplexed through the use of metabolic labeling [40] approaches.

The protein correlation profiling SILAC (PCP–SILAC) approach to monitor an interactome is an extension of the methodology pioneered for organelle proteomics [41,42]. This approach involves the use of SILAC to construct chromatographic elution profiles (typically from size exclusion chromatography) for all proteins separated, under native conditions, from a complex mixture. Proteins interactions are then elucidated based on similarities in specific features of the chromatograms of two proteins [16]. Importantly, the use of SILAC enables the measurement of two experimental conditions within the same experiment, making this technique uniquely suited to monitoring changes in protein interactions. However, as co-migration approaches to study interactomes are still in their infancy and due to the volume of data generated within a single SEC–PCP–SILAC experiment, the bioinformatics treatment of such datasets presents several challenges. These challenges include: alignment of the fractionation scale between replicates, curve deconvolution and controlling for over-fitting, statistical assessment of false positive rates and combining ratio measurements among replicates.

Here we build on the initial approaches used to derive interactome information out of PCP–SILAC data to develop a unified and more user-friendly set of tools to enable the rapid alignment, processing, visualization and assignment of protein interactions from PCP–SILAC datasets. As a case study, we apply these tools to a recently collected PCP–SILAC dataset comparing the protein complexes from uninfected HeLa cells versus cells infected with *Salmonella enterica* serovar Typhimurium, a system of key interest to our lab as a systems biology model [43,44]. The primary motivation of this work, however, was to overcome many of the informatics challenges involved in going from raw PCP–SILAC data to an interactome with knowledge of which proteins and complexes are affected by an external stimulus. Within this work