Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

The promise of targeted proteomics for quantitative network biology

Masaki Matsumoto and Keiichi I Nakayama

Proteomics is a powerful tool for obtaining information on a large number of proteins with regard to their expression levels, interactions with other molecules, and posttranslational modifications. Whereas nontargeted, discovery proteomics uncovers differences in the proteomic landscape under different conditions, targeted proteomics has been developed to overcome the limitations of this approach with regard to quantitation. In addition to technical advances in instruments and informatics tools, the advent of the synthetic proteome composed of synthetic peptides or recombinant proteins has advanced the adoption of targeted proteomics across a wide range of research fields. Targeted proteomics can now be applied to measurement of the dynamics of any proteins of interest under a variety of conditions as well as to estimation of the absolute abundance or stoichiometry of proteins in a given network. Multiplexed targeted proteomics assays of high reproducibility and accuracy can provide insight at the quantitative level into entire networks that govern biological phenomena or diseases. Such assays will establish a new paradigm for data-driven science.

Address
Department of Molecular and Cellular Biology and Division of Proteomics, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan

Corresponding authors: Matsumoto, Masaki (masakim@bioreg.kyushu-u.ac.jp), Nakayama, Keiichi I (nakayak1@bioreg.kyushu-u.ac.jp)

Current Opinion in Biotechnology 2018, **54**:88–97

This review comes from a themed issue on **Analytical biotechnology**

Edited by **Hiroshi Shimizu** and **Fumio Matsuda**

https://doi.org/10.1016/j.copbio.2018.02.014

Introduction

The completion of the human genome sequencing project engendered the development of 'omics' technologies that now allow comprehensive and quantitative measurement of various biological molecules — including nucleic acids, proteins, lipids, carbohydrates, and other metabolites — and thereby provide important insight into the functioning of biological systems under both physiological and pathological conditions. In particular, the development of next-generation sequencing (NGS) has provided a powerful tool for uncovering the molecular signatures associated with various biological phenomena or disease states. Although NGS-based studies shed light on how the genome operates to support a living organism [1], they do not provide direct information as to how components encoded by the genome drive cellular activities. Biochemical reactions that underlie physiological or pathological phenomena are directly executed or governed by proteins that play enzymatic, structural, or regulatory roles. The goal of proteomics is to provide an overall view of how proteins perform these functions [2,3].

The development of proteomics has depended in large part on technical advances in mass spectrometry (MS) as well as in various biochemical methodologies and computational tools [2,4,5]. Shotgun proteomics, also known as bottom-up proteomics, is a widely adopted approach in which peptides generated by proteolytic digestion of a protein sample are analyzed by tandem MS coupled with liquid chromatography (LC). There are two basic types of shotgun proteomics: discovery (nontargeted) and targeted. Discovery proteomics is based on a data-dependent acquisition (DDA) mode in which the most intense precursor ions detected in the MS spectrum are automatically selected for acquisition of MS/MS spectra. This approach is able to identify a large number of peptides without any prior knowledge. As performed with a modern LC–MS instrument, it allows the identification of most proteins of *Escherichia coli* or yeast in a single-shot (without prefractionation) experiment [6,7•]. Pioneering studies also provided an in-depth analysis of the human proteome to an extent almost equivalent to that of transcriptome analysis [8,9]. Furthermore, massive efforts based on the discovery approach to obtain data from human tissues, body fluids, and cell lines have provided the basis for a draft map of the human proteome [10•,11••] that aims to reveal the overall structure of the whole-body proteome in humans. In addition to simple enumeration of proteins, through the incorporation of various technologies including metabolic stable-isotope labeling [12,13], chemical labeling with isotopic [14] or isobaric [15] tags, or label-free methods [16–18], discovery proteomics allows comparison of the amounts of a large number of proteins among biological samples. The maturation of discovery proteomics has thus expanded its applicability to addressing key questions in biology and medicine [19••,20,21,22].

Many proteins identified by discovery proteomics as being of potential importance in a phenomenon of

interest await validation in a more quantitative manner and under various conditions. An increasing need for reliable and high-throughput methods for accurate protein quantification has led to the development of targeted proteomics. This approach typically relies on multiple reaction monitoring (MRM), also known as selected reaction monitoring (SRM), which had previously been adopted for the quantitation of small compounds [23] and peptides [24–28]. In this review, we will focus on recent technical advances in proteomics, most notably in targeted proteomics, and on the application of proteomics approaches to complex biochemical systems such as cancer metabolism.

Targeted proteomics: a tool for hypothesis testing

The finding of answers to key questions in biology and medicine will require the development of quantitative assays for specific proteins and the robust measurement of such proteins in different biological samples [29]. The nontargeted proteomics approach is not suitable for this purpose because of several technical limitations. The acquisition of MS/MS spectra in the DDA mode depends on the signal intensity of the peptide precursor ions observed, which results in the preferential detection of high-abundance proteins. Furthermore, in the case of highly complex mixtures, the selection of precursor ions for acquisition of MS/MS spectra becomes random, with a consequent loss of repeatability of peptide identification. This bias of detection toward highly abundant proteins that results from the stochastic nature of DDA is thus an obstacle to consistent detection of a given protein, with the incompleteness of the proteome data hampering verification of biological hypotheses.

Targeted proteomics based on MRM has been expected to solve this problem of data inconsistency [24,25,27,30]. MRM is performed on a triple-stage quadrupole MS instrument, in which the precursor ions of peptides are selected by the first quadrupole (Q1) and then fragmented in the second quadrupole (Q2) by collision-induced dissociation (CID), with the resulting product ions being further selected by the third quadrupole (Q3). The pair of mass/charge (m/z) values selected by Q1 and Q3 is termed the 'MRM transition.' Each MRM transition serves as a specific filter that allows transmission of only the selected ion to the detector and consequent generation of a chromatographic trace for the targeted ion. A set of MRM transitions determining a peptide signature together with the peptide elution time in high-performance liquid chromatography (HPLC) is referred to as an MRM assay. Although MRM is limited with regard to the number of peptides that can be measured concomitantly, it has a wider dynamic range with a lower limit of detection and it provides more robust quantification in comparison with the DDA approach, thus allowing the detection of relatively low-abundance proteins in a single-shot LC–MS experiment. With the use of internal standards that are typically isotope-labeled peptides, targeted proteomics is suitable for the precise and absolute quantification of proteins [27,28,31,32], which is essential for mathematical modeling of biochemical reactions, stoichiometric theory of protein–protein interactions, and integration and comparison of data obtained at different times and places.

Development of large-scale targeted assays

Despite the substantial advantages of targeted proteomics, this approach has yet to be widely adopted even by specialized proteomics laboratories because of the labor-intensive and time-consuming process of MRM assay development [27]. The most crucial step of targeted proteomics is the selection of optimal peptides in terms of sensitivity and specificity. Such peptides are referred to as 'proteotypic peptides (PTPs)' if they are found only in a given protein. The accumulation of data from DDA-based experiments in public repositories [33] has provided a primary resource for PTPs inferred from the detection frequency across projects (Figure 1, path a). After candidate PTPs have been selected, their sequences are converted into sets of MRM transitions with the use of informatics tools such as Skyline [34]. These MRM transitions are then subjected to empirical validation by determination of specificity and signal intensity together with the HPLC retention time of each peptide. This step is a bottleneck in MRM assay development because it requires peptides of predetermined purity and concentration. A synthetic peptide library is a practical tool for such validation because of the high level of throughput and flexibility of peptide synthesis (Figure 1, path b) [35••]. Several large-scale efforts to develop MRM assays with synthetic peptide libraries have indeed been successful for proteomes of relatively low complexity such as those of *Saccharomyces cerevisiae* [36], *Mycobacterium tuberculosis* [37], and *Streptococcus pyogenes* [38]. In such cases, a nearly complete identification of expressed proteins with a sufficient number of peptides is a prerequisite for generation of a synthetic peptide library. In the case of mammals such as human and mouse, although massive amounts of data have been deposited in repositories [39], they are still insufficient to generate a high level of sequence coverage for the entire proteome. To mitigate this situation, researchers have developed informatics tools based on the accumulated DDA data for the prediction of PTPs [40]. Kusebauch *et al.* [41••] thus developed human SRMAtlas, a targeted assay resource for the human proteome that relies on a synthetic peptide library based on the combination of empirically scored PTPs (Figure 1, path a) and PTPs predicted by informatics (Figure 1, path c). An analogous project, designated ProteomeTools, provides a peptide library with a richer content including variants, posttranslational modifications, and nontryptic peptides as well as a larger number of tryptic peptides [42••]. MS/MS measurement of tryptic peptides by five different fragmentation methods with