Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Advances in proteomics for production strain analysis

Andrew Landels[1,3], Caroline Evans[1,3], Josselin Noirel[2] and Phillip C Wright[1]

Proteomics is the large-scale study and analysis of proteins, directed to analysing protein function in a cellular context. Since the vast majority of the processes occurring in a living cell rely on protein activity, proteomics offer a unique vantage point from which researchers can dissect, characterise, understand and manipulate biological systems. When developing a production strain, proteomics offers a versatile toolkit of analytical techniques. In this commentary, we highlight a number of recent developments in this field using three industrially relevant case studies: targeted proteomic analysis of heterologous pathways in *Escherichia coli*, biofuel production in *Synechocystis* PCC6803 and proteomic investigations of lignocellulose degradation. We conclude by discussing future developments in proteomics that will impact upon metabolic engineering and process monitoring of bio-producer strains.

Addresses
[1] Department of Chemical and Biological Engineering, University of Sheffield, S1 3JD, UK
[2] Chaire de bioinformatique, Laboratory 'Genomics, Bioinformatics and Applications', Conservatoire national des arts et métiers, 75003 Paris, France

Corresponding author: Wright, Phillip C (p.c.wright@sheffield.ac.uk)

[3] These authors contributed equally to the manuscript.

Current Opinion in Biotechnology 2015, **35**:111–117

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

Edited by **Uwe T Bornscheuer** and **Alex Toftgaard Nielsen**

http://dx.doi.org/10.1016/j.copbio.2015.05.001

Introduction

Chemical biotechnology is a field directed to harnessing living organisms as cellular factories, for bio-based production of small molecules and polymers [1,2]. These biological production systems are less well understood than traditional chemical engineering processes due to their inherent complexity. As a result, advanced molecular techniques like proteomics are required to engineer more efficient processes and develop new applications.

First defined in 1995 as a portmanteau of 'protein' and 'genomics', proteomics is the large-scale study of proteins within a cell, tissue or organism [3]. It is a rapidly evolving field focused on identification and characterisation of these proteins and their proteoforms (isoforms and post translational modification (PTM) variants). Quantitative methods in proteomics have enabled comparative analysis of protein expression profiles, typically providing 'snapshots' of cells and proteins in different stages of bio-production. Recent studies have also measured protein turnover by determining rates of protein synthesis and degradation. These techniques offer a means to gain information on mechanisms of bio-production for purposes of optimisation and process monitoring. To date proteomics has found application to well characterised strains such as *Escherichia coli* [4], emergent bio-producer strains like the cyanobacteria *Synechocystis* PCC6803 (herein referred to as *Synechocystis*) [5•]; as well as meta-proteomic analysis of mixed microbial communities [6].

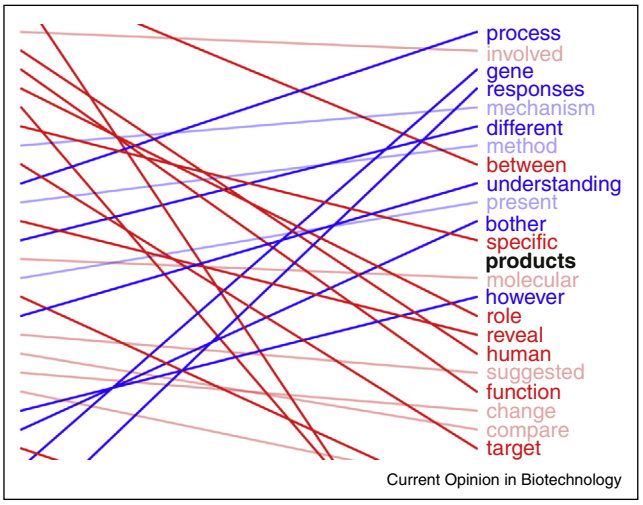
Bibliometric analysis (see supplementary material) of recent proteomics publications has highlighted a couple of key trends in producer strain studies: Proteomics in producer strain analysis tends to focus much more on understanding mechanisms and responses, or suggesting molecular pathways, indicating that in general production analysis is lagging behind the general trend toward targeted proteomics (Figure 1). We cover the topic of targeted proteomics in more detail below and highlight a small number of cutting edge studies in our first case study.

In this commentary, we present a typical approach for conducting a proteomics experiment, highlighting key terms and concepts. We then outline novel proteomics approaches using post-2012 examples, focusing on three industrially relevant case studies: a method-specific approach, a strain-specific approach and a process-specific approach, concluding with a discussion of the impact of recent developments in the field.

Proteomic analysis pipeline

High-throughput proteomic methods commonly used in biotechnology approaches utilise the 'shotgun' or bottom-up technique [7•], where the proteome is digested into peptides that are typically 5–14 amino acids long. Whole proteins or larger polypeptides can also be analysed (top-down, middle down respectively), but this strategy has several technical issues detailed elsewhere [8••]. A digested proteome is complex, containing thousands of peptides with varying abundances. The mix requires

Figure 1



A selection from a rank-plot of the 200 most frequently used words in abstracts of production-strain proteomics publications, ranked by frequency. Words in blue are higher-ranked in production-strain proteomics than in proteomics in general; words in red are lower ranked, words in black have not changed relative position. Faded words have changed rank by 5 places or fewer and words in bold-face are only present in the production-strain list. The solid line indicates the change in rank. This figure presents a snapshot of the full list, which is available in the supplementary material.

fractionation, typically using offline high performance liquid chromatography (HPLC) or in solution isoelectric focusing, which splits the single sample into lower complexity fractions. Doing this collects together peptides with similar features — such as hydrophobicity, charge state or isoelectric point — and significantly improves the quality of the final data. Samples are then subject to nano-flow reverse phase HPLC, coupled directly to mass spectrometer. This process is referred to as MS–MS or MS^2.

The mass spectrometer (MS) initially scans the masses and intensities of all eluting peptides from the HPLC, on a scale of seconds to milliseconds, this is an MS 'survey scan'. Eluting peptides are then selected for fragmentation from the survey scan, either in a data dependent (DDA) or data independent (DIA) acquisition mode. DDA targets a specific peak from the survey scan for further analysis, whilst DIA fragments all ions from the survey scan simultaneously. The data is then analysed computationally to identify and characterise the proteome. Two detailed reviews provide further information, Altelaar *et al.* [9••] provide an overview, whilst Zhang *et al.* [8••] cover the topic more comprehensively.

Approaches in proteomics

Proteomic approaches can be subdivided into discovery and targeted modes, for characterisation of the proteome and analysis of an identified subset of proteins respectively. Their application and relevance are outlined in the case studies. Examples of workflows, gel and non-gel based, together with their major benefits and drawbacks and examples of their application to bio-producer strains are outlined (Table 1). Classical proteomics employs two-dimensional electrophoresis (2DE) for analysis of expression profile, where protein identification requires MS as a second step. Gel free quantification methods achieve protein identification and relative quantification using MS with significant advantages over 2DE [10]. Protein and peptide labeling methods, metabolic (e.g. SILAC) or chemical labels (e.g. iTRAQ) have been widely employed in proteomics but 'label-free' methods are increasingly gaining in popularity [11,12]. To date, no direct comparison of all these techniques has been reported (Table 1). Technique selection is dependent on the biological context, number of samples to be processed and compared.

Targeted proteomic approaches are directed to the detection and the precise quantification of specific subset of proteins of interest. This complements discovery proteomics and applications include verification of candidate proteins and process monitoring [13]. Quantification is based on detection and measurement of proteotypic peptides that represent the protein, based on unique amino acid sequence. Specificity and sensitivity are both conferred via 'reaction monitoring' for the presence of (co-eluting) fragment ions, linking precursor and product transition information. Application of high-resolution mass measurement and acquisition of full fragment ion spectra have enabled developments, including higher throughput and specificity conferred by parallel reaction monitoring (PRM) as recently reviewed [14].

Inclusion of stable isotope forms of reference proteotypic peptides, at known concentrations, enables absolute quantification. QconCATs (concatenated proteotypic peptide sequences), are custom designed recombinant proteins, which can be metabolically labeled, purified and tryptically digested, to provide a set of standards for absolute quantification of multiple proteins in parallel [15•]. Label free approaches are popular due to limited sample pre-processing requirements prior to analysis compared to label based methodologies. Examples include Intensity-Based Absolute Quantification (iBAQ) and Absolute Protein Expression (APEX), which have been compared for different sample types and MS platforms [16–18].

Case study: targeted proteomics for process optimisation

A key area of proteomic application is assessment and modelling of heterologous pathways. Whilst assessing how an inserted pathway is affecting the proteomic background provides useful information on how the organism is responding; for pathway engineering purposes it is often more informative to assess either the pathway proteins directly, or a specific subset of the proteome