Signal enhanced proteomics: a biological perspective on dissecting the functional organisation of cell proteomes
Dalila Bensaddek¹, Armel Nicolas² and Angus I Lamond

Proteomes are highly dynamic and can respond rapidly to environmental and cellular signals. Within cells, proteins often form distinct pools with different functions and properties. However, in quantitative proteomics studies it is common to measure averaged values for proteins that do not reflect variations that may occur between different protein isoforms, different subcellular compartments, or in cells at different cell cycle stages and so on. Here we review experimental approaches that can be used to enhance the signal from specific pools of protein that may otherwise be obscured through averaging across protein populations. This signal enhancement can help to reveal functions associated with specific protein pools, providing insight into the regulation of cellular processes. We review different strategies for proteomic signal enhancement, with a focus on the analysis of protein pools in different subcellular locations. We describe how MS-based proteome analyses can be combined with a general physico-chemical cell fractionation procedure that can be applied to many cultured cell lines.

Address
Laboratory of Quantitative Proteomics, Centre for Gene Regulation and Expression, School of Life Sciences, University of Dundee, Dundee, DD1 5EH, UK.

Corresponding author: Lamond, Angus I. (a.i.lamond@dundee.ac.uk)
¹ Present address: Bruker UK Limited, Banner Lane, Coventry, CV4 9EH, UK.
² Present address: DC Biosciences, James Lindsay Place, Dundee Technopole, Dundee, DD1 5JJ, UK.

Current Opinion in Chemical Biology 2019, 48:114–122
This review comes from a themed issue on Omics
Edited by Ileana M Cristea and Kathryn S Lilley

https://doi.org/10.1016/j.cbpa.2018.10.011
1367-5931/© 2018 Elsevier Ltd. All rights reserved.

Introduction
The field of proteomics has recently seen tremendous progress and technical advances that have improved the efficiency of protein detection at multiple levels, including experimental design, sample preparation workflows, LC–MS instrumentation and in silico analysis. As a result, it is now possible not only to identify a large proportion of a steady state cell proteome in a single experiment, either with, or without, fractionation [1,2] but also to describe additional proteome dimensions, such as protein turnover rates, cell cycle-specific changes, post-translational modifications and subcellular localization and so on.

From a biological perspective, a limitation of most shotgun proteomics experiments is that protein extracts are typically prepared from heterogeneous populations of cells, for example either from tissues, or whole organisms, or from unsynchronised cell lines with cells at different cell cycle stages. The resulting quantitative data represent an averaged value across all of the pools and populations that each protein belongs to. However, to characterise a cell’s proteome in more detail, particularly to gain insights into biological regulatory mechanisms, it is important not only to include quantitation of protein expression levels, but also to resolve protein groups into single isoforms (the so called ‘isoform inference’ problem associated with bottom-up proteomics). Furthermore, information from additional proteome ‘dimensions’ for example describing the subcellular distribution of proteins and cross-correlating this with data on post-translational modifications (PTMs), protein complexes, rates of protein synthesis and turnover, can provide valuable insights into regulatory mechanisms.

This combined analysis approach has been referred to as either ‘Next Generation Proteomics’, or, ‘multidimensional proteomics’ [3]. A major advantage of the multidimensional characterization of cell proteomes is the ability to mine the resulting data to establish correlations between different protein properties, for example linking the subcellular location of a protein with either a specific isoform, or post-translational modification state and so on. This can generate useful hypotheses regarding the functional significance of such correlations that subsequently can be evaluated directly in follow-on experiments.

In this manuscript we will review examples of methods that make it possible to enhance proteomic signals and thereby detect protein-level changes that would not have been detected in standard one-dimensional analyses.
Signal enhanced proteomics versus classical proteomics

Most proteomics approaches have tended to provide measurements that describe an averaged view, or steady state proteome. Protein expression levels are measured from the combined analysis of different subpopulations of protein molecules extracted from homogenized cell or tissue extracts. In turn, the extracts are generated from pools of cells, which typically include cells at different stages of the cell cycle, and may include cells that have shown different response levels to external stimuli and so on. While this is useful in providing general information on the proteome and its remodelling, in disease for example, the averaging that occurs at different levels misses a lot of biological detail and can obscure the detection of changes in proteins that occur specifically in subsets of the global proteome (Figure 1).

To address this, several enrichment strategies at either cellular, or subcellular levels, can be applied that add value and are useful for linking proteomics information more effectively with cell biology. For example, immunologists have long used surface markers to label and sort different subpopulations of immune cells, using Fluorescence Activated Cell Sorting (FACS). With recent improvements in sensitivity and throughput for MS analysis, it has become technically feasible and cost effective to combine FACS sorting with proteomics analysis, which helps to target protein detection in specific cell subsets. Recently, Ly et al. [4*], extended this approach to using FACS also to isolate cell subpopulations defined by immunolabelling intracellular, rather than cell surface antigens. Using this strategy, termed PRIMMUS (PRoteomic analysis of Intracellular iMMUnolabelled cell Subsets), Ly et al. were able to separate interphase and mitotic cells, and also able to isolate populations of FACS-sorted cells enriched for specific mitotic subphases in sufficient quantities for detailed MS-based proteomic analysis as shown schematically in Figure 1. This demonstrates that even minor subsets of cells in a population, which exist only transiently, can be isolated and protein responses detected that would otherwise be obscured in the bulk analysis of the proteome in extracts prepared from mixtures of cells at different cell cycle stages.

Isoform-specific proteomics

In higher eukaryotes, many genes encode two or more protein isoforms, which behave as distinct pools of related proteins and that may differ in terms of subcellular localizations, interaction partners and function. For example, alternative splicing of pre-mRNA transcripts is commonplace and this can generate multiple mRNAs from the same gene, which in turn gives rise to proteins that differ in structure and function. Isoforms can also arise via differential protein processing, for example cleaving the original translation product into shorter forms. In other cases, protein isoforms can arise from expression of closely related, duplicated genes.

Whatever the mechanism, a common feature of closely related protein isoforms is that they usually share extensive regions of protein sequence identity and consequently have many shared peptides. The corollary is that many peptides that are identified in the typical bottom-up MS-based proteomics workflows cannot reliably be assigned to only one specific protein species for quantitation (Figure 2). If the structure of different protein isoforms is known experimentally, or predicted from genomic and transcriptomic studies, it can be possible to identify peptides that are either unique to

Figure 1

Schematic comparison of averaged proteomics analysis and signal enhanced proteomics. When a population of cells in culture is subjected to external stimuli, cells may respond differently resulting in a heterogeneous population of cells (with different properties). In a classical proteomics experiment, cell extracts are prepared from the mixed population resulting in averaged protein signals. In enhanced signal proteomics, the two population of cells are separated before proteomics analysis avoiding the dilution of the signal.
Isoform-specific proteomics. A schematic representation of how a single gene can encode two protein isoforms as a result of alternative splicing. These proteins isoforms have a largely shared sequence but differ in some segments, exemplified here by the blue and red peptide sequences. The two protein isoforms may have different properties, such as subcellular localizations and can only be resolved if subcellular fractionation is used before the proteomics analysis.

a specific isoform, or shared between different isoforms. This information can then be used to identify and quantify the expression of the respective protein isoforms, based on the levels of each peptide detected by MS analysis.

The sequence identities of all of the possible protein isoforms that may be expressed in different cell types and organisms are not always known in advance, however. One way of addressing the isoform problem experimentally, without relying on accurate genome annotations, is
to include a size fractionation step for protein separation in the workflow. For example, size fractionation can be carried out either by using SDS-PAGE, or size exclusion chromatography (SEC), to fractionate proteins before digestion. This will help to enrich for distinct protein pools where isoforms differ in size, which will in turn reduce the isoform inference problem in subsequent MS analysis. Computational approaches can also be used to try to distinguish protein isoforms. For example, Ahmad et al. [5] used a candidate approach, combined with subcellular fractionation, to detect protein isoforms that showed differential behaviour in separate subcellular compartments. Here, average values for protein intensity are first calculated using all of the peptides detected from a given gene, irrespective of isoforms. Next, the potential protein coding region is subdivided along its length and protein intensity is re-calculated, either using the peptides from the respective amino terminal, middle and carboxy-terminal ‘thirds’ of the protein sequence (‘rule of thirds’ approach), or else protein intensity calculated using groups of adjacent peptides, moving sequentially along the protein length from the amino to carboxy terminus. The principle is that if a protein isoform has a region of peptide sequence that is not present in other isoforms (e.g. resulting from inclusion of a differentially spliced exon), that may result in a protein intensity value for the isoform that differs from the average value obtained using all of the peptides matched to the protein group. Using this unbiased approach, Ahmad et al. [5] detected candidate protein isoforms via the analysis of peptide subsets and showed that the expression of some of these isoforms differed between the respective cytoplasmic and nuclear compartments of cultured human cell lines.

The subcellular proteome

While most proteomic analyses have studied whole cell extracts, avoiding the issue of subcellular localization of proteins, some studies have focussed on analysing the proteome of specific, purified organelles, albeit without regard to the distribution of proteins in other subcellular locations. More recently, approaches have been developed to tackle the spatial dimension of the global proteome. In general, these global proteome localization approaches can be divided into two groups: first, targeted studies, which attempt to isolate biologically defined compartments (organelles), using fractionation methods yielding relatively pure fractions; second, global studies where the fractionation method generates multiple fractions, of different composition, for example using compartment-specific characteristics, such as density, solubility to detergents, and so on as the basis for fractionation. Subsequently, the cross-gradient profile of identified proteins is used to group them and assign proteins to compartments based on co-fractionation markers. Examples of methods from the first group include the classic nucleoli extraction protocol, which has been used in the ‘spatial proteomics’ workflow (Figure 3) [6] as well as methods using detergents of increasing strength to target compartment proteomes, based on their accessibility and solubility [7]. Methods which belong in the second group include Protein Correlation Profiling (PCP) [8] and Localization of Organelle Proteins by Isotope Tagging (LOPIT) [9], both of which rely on statistical methods to unravel the pattern of distribution of compartment proteomes in mixed populations separated using a gradient fractionation approach. The LOPIT workflow has been used successfully to investigate how the proteome, at steady state, is partitioned between multiple organelles and compartments. LOPIT has taken advantage of isobaric labelling, such as iTRAQ [10] and TMT [11], which allows simultaneous analysis of up to ten samples in a single MS run. Isobaric labelling is particularly well suited for analysis of fractionation experiments, because physically combining fractions early in the workflow makes the analysis internally controlled and improves data quality by reducing the problem of missing values. Hyperplexed LOPIT [12] leverages new technological development both in the TMT-technology and MS instrumentation, allowing more accurate quantification of an increased range of reporter tags. In addition to higher multiplexing capabilities, HyperLOPIT features improved subcellular fractionation protocols, which aim to preserve as many suborganelles as possible [13].

In addition to characterising the spatial proteome in a static manner, that is describing the organelle compartment proteome at steady state, it is also important to understand the dynamic remodelling of the subcellular distribution of the proteome, which may occur within a cell undergoing a response to stimuli, such as stress, or viral infection. This will typically result not only in changes in protein abundance, but also see protein translocation from one compartment to another. Recently, Cristea and co-workers extended the methods described above to study virus-induced spatial cell remodeling; they combined label free and isobaric labelling to measure the abundance levels for host and viral proteins as well as their localization throughout the time course of human cytomegalovirus (HCMV) infection [14,15].

Furthermore, the extent of proteome relocalization can be affected by the cell genotype. An early example of this was provided by a comparison of the response to stress induced by DNA damage in human HCT116 cells that were either wild type, or null, for the tumour suppressor p53 [16]. Using a MS-based proteomics approach, combined with subcellular fractionation, the distribution of the proteome between the nucleus and cytoplasm was compared before and after DNA damage induced by etoposide treatment in both p53 +/+ and p53 −/− HCT116 cells. This showed that there were few p53-dependent differences in proteome localization under
normal cell growth conditions, but clear differences after induction of DNA damage, particularly affecting the ability of ribosomal proteins to accumulate in nucleoli. This study illustrates how the unbiased proteomic analysis of part of the role of p53 in the DNA damage response was only revealed after linking MS-based proteome measurements with subcellular fractionation.

**Protein turnover analysis**

Early biochemical studies of protein turnover relied on detecting the incorporation of radiolabeled amino acids into newly translated proteins [17]. Typically, proteins were labelled with $^{35}$S methionine and pulse-chase experiments were used to determine their rate of degradation after adding drugs to block protein synthesis. Nowadays, MS-based proteomics allows the measurement of turnover rates for large numbers of proteins simultaneously, for example by using pulse labelling experiments combined with stable isotope labelling by amino acids in cell culture (SILAC), where labelling is carried out with amino acids incorporating heavy stable isotopes. The underlying principle of pulse SILAC is to metabolically label proteins with heavy isotope substituted amino acids and then to quantify how the isotope-labelled protein population changes across different time points (Figure 4). We and others have previously used pulse SILAC to study protein synthesis and turnover [7,18–24].
An alternative method to pulse SILAC labelling is metabolic labelling using bioorthogonal amino acids [25], such as azidohomoalanine (AHA) which is incorporated into newly synthesised proteins instead of methionine [26]. AHA contains an azide group, enabling capture of newly synthesised proteins via click chemistry [25]. This, combined with SILAC, enables relatively short pulse times [27,28].
Most recently protein turnover has been studied in high throughput by MS-based proteomics which combined SILAC and TMT labelling [29**]. Not only did this approach address the problem of missing data at pulsed time points, it also allowed different proteoforms to be resolved by providing peptide-level measurements of turnover rates. Another example of peptide level turnover data was recently reported by Ly et al. in immortalised human breast epithelial cells [30]. By using pulse-SILAC and cellular assays, Ly et al. studied the activation of v-Src tyrosine kinase activity in untransformed MCF10A cells. This resulted in rapid oncogenic transformation of the cells, which showed major phenotypic changes within 48–72 hours, affecting their morphology, motility and invasiveness. Over this time course, the expression and/or turnover levels of only ~3% of proteins was found to change. Furthermore, since many of the transformation-responsive proteins were relatively low abundance, oncogenic transformation affected only ~1.5% of the total protein molecules in the MCF10A proteome.

It has been shown that the turnover rates of specific proteins can vary for separate pools of the same protein located in different subcellular compartments (Figure 4). For example, this was identified for ribosomal proteins in HeLa cells using a combination of pulsed stable isotope labelling with SILAC and fluorescence microscopy [21,31]. This showed that there is a much higher rate of ribosomal protein turnover in HeLa cell nuclei than in the cytoplasm. Newly translated ribosomal proteins are immediately imported into the nucleus, ready for assembly into nascent ribosomal subunits in the nucleolus. If not bound to rRNA, however, free ribosomal proteins in this nuclear pool are rapidly degraded. It was observed that ribosomal protein stability was dramatically increased upon assembly into ribosome subunits and export to the cytoplasm [31].

Another example showing how different pools of the same proteins can exhibit differential turnover rates was provided in a study on the assembly of RNA polymerase II complexes by Boulon et al. [32]. By using pulsed SILAC analysis to analyse protein turnover rates, combined with subcellular fractionation, they studied the assembly of RNA polymerase II, which occurs predominantly within the cytoplasm. RNA polymerase II is then transported into the nucleus after it is assembled, thereby preventing partially assembled and potentially non-functional subpolymerase complexes having an opportunity to compete for binding to gene promoters. The pulse-SILAC data showed that, similar to the situation with ribosomal proteins, protein turnover rates for subunits of the large polymerase complexes are higher in the cell compartment where assembly takes place (in this case the cytoplasm), as opposed to the compartment where the complex functions (in this case the nucleus, NB the reverse roles of the same compartments for ribosome subunits).

Further systematic proteome level analyses of the relation between protein turnover levels and subcellular localization were carried out using unbiased MS approaches in U2OS cells by Larance et al. [7]. This study systematically compared protein turnover levels in the respective nuclear, cytosolic, membrane and cytoskeletal compartments. It also revealed an important feedback mechanism, whereby inhibition of protein degradation by the proteasome resulted in a rapid inhibition of new protein translation, mediated by induced phosphorylation of eIF2alpha. Importantly, all of these studies together show that protein half-life values based only on analyses of whole cell extracts provide average values that can mask the existence of pools of protein with different properties.

Concluding remarks
In this manuscript we have highlighted some of the practical issues involved in integrating data from MS-based proteomic studies with functional studies in cell biology. In particular, we have focussed on how proteomics can be used to study subcellular localization and to identify pools of protein and distinct protein isoforms that can exhibit differences in structure and properties (e.g. turnover rate, interaction partners and PTMs etc.) in different subcellular compartments. This information can be critical for understanding biological regulatory and response mechanisms, but is often lost or obscured in proteomic studies because of the effect of cell and protein population averaging when whole cell or tissue extracts are analysed.

Subcellular fractionation approaches [9,12**,13*,14**,15*] can be conveniently applied in a multi-dimensional proteomics strategy to improve the functional analysis of cell proteomes [15*] and in the characterization of spatial remodelling following a perturbation such as viral infection [14**,15*].

As shown for the role of p53 in the cellular response to DNA damage [16] characterising how subcellular proteome dynamics is affected by genotype will also be important and can now be analysed more systematically in human cells thanks to the availability of genome engineering technology with CRISPR/Cas9.

We anticipate that in future more detailed studies examining the composition and dynamic remodelling of organelle proteomes at a multidimensional level will help to reveal new insights into the specific protein complexes and functional pools of proteins and isoforms that participate in cell regulatory mechanisms and metabolism.

Acknowledgements
This work was funded by awards from the Wellcome Trust[073980/Z/03/BR and 098503/Z/12/Z and a BBSRC sLoLa grant BR/K003801/1. Imaging and image analysis was supported by two Wellcome Trust Strategic Awards [099945/01/Z and 095931/Z/11/Z] and an MRC Next generation Optical Microscopy Award (MR/K015869/1).
Signal enhanced proteomics: a biological perspective on dissecting the functional organisation of cell proteomes

Bensaddek, Nicolas and Lamond 121

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

● of special interest
◆ of outstanding interest


The authors used FACS to isolate cell subpopulations, specific mitotic subphases, by immunlabelling intracellular, rather than cell surface antigens. This required optimization of cell permeabilization protocols to allow efficient antibody labelling without causing significant loss of intracellular proteins.


The authors describe here a substantial improvement or redevelopment of their localization of organelle proteins by isotope tagging (LOPIT) into a complete pipeline, including the creation of an open-source infrastructure for the analysis of quantitative MS-based spatial proteomics data (http://bioconductor.org/packages/pRoloct http://bioconductor.org/packages/pRoloct and an accompanying interactive visualization framework (http://bioconductor.org/packages/pRoloctGU http://bioconductor.org/packages/pRoloctGU).


HyperLOPIT was used to describe the subcellular proteome of a mouse pluripotent stem cell population, providing localization data for more than 5000 proteins with unprecedented spatial resolution. The authors describe several dimensions of the proteome including protein complexes and functional networks, and make their data available to the community for interactive exploration.


This cell-wide study of organelles in primary fibroblasts throughout the time course of human cytomegalovirus (HCMV) infection. Through a combination of label-free, isobaric-labeling and machine learning, the authors described changes in the spatial localization of ~4000 host and 190 viral proteins. This study provides a comprehensive resource for understanding host and virus biology during HCMV pathogenesis. For example, by measuring dynamic spatial reorganization of proteome, the authors were able to show that the translocation of MYO1A from the plasma membrane to the viral assembly complex is necessary for efficient HCMV replication.

15. Jean Beltran PM, Cook KC, Cristea IM: Exploring and exploiting proteome proteomics during viral infection. J Virol 2017, 91. The viral life cycle requires viral components to be active at the proper time and targeted to specific subcellular locations within the host cell. The authors review different mechanisms used by viruses to repurpose cell organelles, and the various tools that can be used for studying dynamic remodelling of the proteome as a result of viral infection.


The authors describe a new method for studying protein turnover by combining dynamic SILAC and TMT-labeling of ten pulse time-points in a single experiment. They report high proteome coverage without missing values (at pulse times) and measure turnover rates with peptide level resolution, thus distinguishing various proteoforms.


Pulse SILAC was used to measure changes in protein turnover over the course of oncogenic transformation. The authors showed that Src-induced transformation changed the expression and/or turnover levels of ~3% of proteins, affecting ~1.5% of the total protein molecules in the cell. This is the first comprehensive study describing changes in protein expression and/or turnover rates upon oncogenic transformation. The data are provided via the EPD database (www.peptracker.com/epd).
