Trends in ultrasensitive proteomics
AF Maarten Al telaar\textsuperscript{1,2} and Albert JR Heck\textsuperscript{1,2}

Here we review recent developments and trends in sample preparation, pre-fractionation, chromatography and mass spectrometry contributing towards the ultra-sensitive global analysis of proteins. Highly sensitive MS-based proteomics is not only beneficiary for the proteome analysis of single cells, an aim which is getting into reach, but also clearly relevant for the analysis of (a) subcellular organelles, (b) specific low-abundant cell-types such as adult stem cells, and (c) smaller but more homogeneous cell populations sorted or dissected from (diseased) tissue.

Addresses
\textsuperscript{1} Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
\textsuperscript{2} Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

Corresponding author: Heck, Albert JR (a.j.r.heck@uu.nl)

Introduction
Mass spectrometry (MS) is now the method of choice for identifying proteins [1,2], elucidating their posttranslational modifications [3,4], and reading out their functional interactions [5,6]. Therefore, MS has become an essential tool for biologists and biochemists in their efforts to understand the molecular mechanisms regulating cellular systems. On a more critical note covering the proteome in ‘comprehensive’ detail requires still rather large amounts of starting material, being cells, body fluid or tissue, hampering the analysis of smaller organelles or cellular subtypes. Thus, notwithstanding the major recent advances, innovations in MS based proteomics are still urgently needed [1]. These include improved dynamic range ($> 10^6$) to access low abundance components within the vicinity of high-abundance components, and sensitivity, to allow for the analysis of smaller sample sizes. Improving these two critical MS parameters will allow us to take full advantage of the ever increasing analysis speed of MS platforms and will enable even deeper analysis of complex samples to become routine.

One of the major goals of proteomic research is to be able to monitor all proteins in a particular biological system, such as a cell type or cellular subfraction/organelle. There are several different reasons why MS based proteomics needs to become more sensitive and comprehensive. First, targeting specific subcomponents of the cell or the proteome can both enhance the sensitivity of the analysis and contribute to the functional analysis of these proteins. Second, an increasing amount of data indicates that the behavior of cells in a population (in a culture, in an organ) cannot always be reliably approximated by the population average that results when cells are analyzed as a pool. Stochastic fluctuations in gene or protein expression, between cells of an otherwise identical group, can induce differences in their behavior having profound consequences for cell differentiation and responses to stimuli [7–9]. Therefore, upon external stimulus, neighboring cells or populations of cells may behave very differently and for example become activated, show no response at all or go into apoptosis. Averaging cellular responses, as generally done in current proteomics experiments, may thus dilute the sought responses or markers. Targeting only cells of interest has as additional benefit that the background signal may be reduced, enhancing the dynamic range achieved.

To achieve such targeted strategies at an ultrasensitive level, optimized methods are required in all sections of the MS based proteomics pipe-line (see Figure 1). It includes efficient subfractionation and purification of the cells and cellular components of interest, efficient methods to lyse the cells and digest the proteins without major sample losses, and finally fast and sensitive methods for separation, detection and identification of all possible peptides/proteins. Clearly, a combination of maximizing both the specificity of the sample preparation method and the protein detection is essential. Here we review some current trends in ultra-sensitive proteomics, highlighting different complementary approaches, in particular in cellular/organellar pre-fractionation and trends in chromatography and mass spectrometry targeted at improving sensitivity as summarized in Table 1.

Cellular pre-fractionation
Antibody-based purification methods for cellular populations are well established and have proven particularly useful for instance for the isolation of blood cell or immune cell populations. Two powerful cell purification methods that incorporate antibodies for specificity are fluorescence activated cell sorting (FACS) [10,11\textsuperscript{*},12] and immune magnetic separation [13]. FACS is a specialized
Towards ultrasensitive proteomics workflows. Samples containing inherently low amounts of material, such as cellular subpopulations within tissues or subcellular organelles require extensive (pre)fractionation. Subpopulations of cells can be enriched by techniques such as fluorescence activated cell sorting (FACS) or dissected from tissue using laser capture micro-dissection (LCM). To reduce complexity, sensitive nanoscale fractionation and separation at very high resolution (UPLC) are interfaced with advanced MS platforms, exhibiting high speed, sensitivity and resolution.

type of flow cytometry providing a method for sorting a heterogeneous mixture of biological cells into two or more containers, one cell at a time, based upon the specific light scattering and fluorescent characteristics of each cell. It involves passing cells through flow chambers at high rates (>20 000 cells/s) and using lasers to excite fluorescent tags (“fluorochromes”) that are usually attached to antibodies; different antibodies are tagged with different colors, enabling researchers to quantify molecules that define cell subtypes or reflect activation of specific pathways. With magnetic separation, the cells or subcomponents of interest must be labeled with a specific antibody. Both methods are capable of yielding hundreds up to millions of purified cells of a specific subtype. FACS-based cell purification methods have been already coupled to downstream MS based proteomics analysis for instance on purified human leukocyte populations, murine liver-cell populations, and mitochondria. Da Silva et al. [10] combined FACS sorting with proteomics analysis harvesting V-ATP-rich cells isolated from mice kidney and epidermis, generating and breeding EGFP-V-ATPase tagged mice. These V-ATP-rich cells are present in different tissues and have a dedicated function in proton transport acidifying the extracellular environment. Following collection of GFP-positive cells they were able to detect about 1500 proteins from about 200 000 FACS sorted cells. Luber et al. [14] used FACS sorting to purify dendritic cell (DC) subsets from mouse spleens using the expression of CD8α and CD4 surface molecules (Figure 2). Subsequently, they analyzed by mass spectrometry using label-free quantification more than 5000 proteins from a few micrograms of material. Therefore,

<table>
<thead>
<tr>
<th>Technique</th>
<th>Material</th>
<th>Quantity</th>
<th>Maximum # proteins identified</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>FACS – SDS</td>
<td>Dendritic cells</td>
<td>20 μg</td>
<td>6664</td>
<td>[14]</td>
</tr>
<tr>
<td>FACS – SDS</td>
<td>V-ATPase-rich cells</td>
<td>–</td>
<td>2297</td>
<td>[10]</td>
</tr>
<tr>
<td>FACS – HILIC</td>
<td>Stem cells</td>
<td>5000 cells</td>
<td>3775</td>
<td>[11]</td>
</tr>
<tr>
<td>LCM</td>
<td>Breast carcinoma tissue</td>
<td>300 ng/3000 cells</td>
<td>1003</td>
<td>[17]</td>
</tr>
<tr>
<td>LCM</td>
<td>Kidney glomeruli (K)</td>
<td>~10 000 cells</td>
<td>~2400</td>
<td>[19]</td>
</tr>
<tr>
<td>LCM – SAX</td>
<td>Breast and colon cancer tissue</td>
<td>5–7 μg/~20 000 cells</td>
<td>3600–4400</td>
<td>[20]</td>
</tr>
<tr>
<td>Single LC–MS 8 h gradient</td>
<td>HeLa cells</td>
<td>1 μg</td>
<td>2516</td>
<td>[33]</td>
</tr>
<tr>
<td>Single LC–MS 5 h gradient</td>
<td>HeLa cells</td>
<td>1 μg</td>
<td>2587</td>
<td>[29]</td>
</tr>
<tr>
<td>Single LC–MS 8 h gradient</td>
<td>HEK293 cells</td>
<td>2 μg</td>
<td>4622</td>
<td>[34]</td>
</tr>
<tr>
<td>Single LC–MS 4 h gradient</td>
<td>Yeast strain W303 MATα</td>
<td>4 μg</td>
<td>4084</td>
<td>[35]</td>
</tr>
<tr>
<td>Rare cell proteomic reactor</td>
<td>hESCs cells</td>
<td>50 000 cells</td>
<td>2281</td>
<td>[42]</td>
</tr>
<tr>
<td>Centrifugal proteomic reactor</td>
<td>ER and Golgi microsomal membranes</td>
<td>20 μg</td>
<td>955</td>
<td>[43]</td>
</tr>
<tr>
<td>AutoProteome system</td>
<td>Suprachiasmatic nucleus neurons</td>
<td>200 μg</td>
<td>1958</td>
<td>[44]</td>
</tr>
</tbody>
</table>
conventional DC (cDC) preparations from pooled spleens of 32 mice were used yielding more than $2.5 \times 10^6$ cDCs per subset. A major finding in this study was that differences exist in viral recognition upon subsets of dendritic cells, whereby the CD8α+ DCs largely lacked the receptors required to sense certain viruses. Following similar strategies Di Palma et al. collected and analyzed adult colon stem cells from a mouse strain in which green fluorescent protein (GFP) has been knocked into the Lgr5 locus. Lgr5 is a gene that is uniquely expressed in the stem cells of several adult tissues such as intestine, hair follicles and stomach [15]. Following, very high enrichment (>95%) of these GFP positive cells from mouse intestines, albeit in rather small quantities of around 5000 colon stem cells, they were still able to detect about 4000 proteins combining hydrophilic interaction liquid chromatography (HILIC) with regular reversed phase (RP) LC. Liu et al. [16] addressed the issue of different cell populations being present in organs such as the liver. Although liver is often seen as a relative homogenous organ, it contains multiple different cell types, which have been enriched for by a combination of collagenase-based density gradient centrifugation and magnetic activated cell sorting. This approach led to the isolation of four distinct types of liver cells, hepatocytes, hepatic stellate cells, Kupffer cells and liver sinusoidal endothelial cells, enabling cell type specific proteome profiling and preventing population averaging.

**Laser capture micro-dissection**

Proteomics methods still need to be further improved to become amendable for the analysis of (clinical) tissue samples, a task considerably hampered by tissue heterogeneity. Laser capture micro-dissection (LCM) can be used to selectively isolate target cells from their native tissue environment. However, so far the small number of cells that is typically procured by LCM severely limits proteome coverage. Umar et al. [17] combined LCM with nanoflow LC-FT-ICR MS analyzing protein digests of 3000 tumor cells from breast carcinoma tissue. A total of around 1000 proteins could be identified by matching LC-MS data to accurate mass and time (AMT) tag databases that were previously established for human breast cancer cell lines. In a similar approach, Dos Santos et al. [18] found tumor specific expression changes for 39 proteins in Human intrahepatic cholangiocarcinoma,
of which several could be linked to tumorigenic pathways. Waanders et al. [19] characterized kidney glomeruli isolated by laser capture micro-dissection to a depth of more than 2400 proteins. Moreover, they were able to identify over 2000 proteins from single pancreatic islets of Langerhans, containing 2000–4000 cells and could quantitatively compare the proteome of such single islets, treated with high or low glucose levels. Such results clearly indicate that direct proteomic analysis of functionally distinct cellular structures opens up new perspectives in physiology and pathology. Wisniewski et al. [20] combined filter-aided sample preparation (FASP) workflow with strong anion exchange (SAX) fractionation and LCM providing a very powerful method that even could be applied to the analysis of formalin fixed and paraffin embedded tissues. They reached a depth of 3600–4400 proteins per single LC MS/MS run comparing archival neoplastic and matched normal colonic mucosa cancer specimens for three patients observing the differential expression of 30 known colon cancer markers.

A severe challenge that remains to be addressed by clinical proteomics is intratumor heterogeneity. During tumor development individual cells compete for space and resources and at the same time cooperate to evade the immune system and progress into new areas. This evolution causes the appearance of distinct cell populations within the tumor, containing their own genetic and epigenetic identity and responding differently to treatment [21]. As these cell populations may look very similar in the tissue environment, targeting them specifically is difficult. One emerging technology to deal with this challenge in MS imaging, which, although struggling with sensitivity and identification power, is able to create distinct biomolecular profiles from such heterogeneous tumor regions [22,23].

Mass spectrometry

Significant advances have been in mass spectrometric instrumentation, from which proteomics research is benefiting immensely [24,25]. Hybrid instruments have been designed to combine the capabilities of the individual instruments, with a focus on faster and more sensitive analysis primarily of peptides by LC MS/MS. Several recently introduced MS platforms have improved ion inlet and transfer optics improving sensitivity, such as the S-lens in the Thermo Q Exactive, the Qjet in the AB Sciex TripleTof 5600 and the StepWave ion transfer in the Waters Synapt 2G-S. Furthermore, all these platforms have increased their sequencing speeds to such an extent that, with current dynamic range limits, virtually all observed eluting peaks can be targeted. The Q Exactive reaches an MS/MS speed of 12 Hz at a resolution of 12 500 at m/z 400, which comes down to approximately 1 s analysis time for a top 10 method, including the MS scan at 50 000 resolution [26]. This increased sequencing speed results in the identification of >2500 proteins from an undefined sample amount utilizing a single 90 min gradient. The 5600 platform of AB Sciex can even reach higher sequencing speeds as it can operate at 100 Hz. However, maximum identifications have been obtained at 20 Hz sequencing speed, with 25 000 resolution at m/z 186, identifying >1100 proteins from 200 ng of yeast in a single 85 min gradient [27]. Next, when dealing with limited sample amounts maximum information can be obtained when the majority of peptides can be efficiently fragmented. Besides improved identification rates through high resolution and mass accuracy of the fragment data, access to complementary techniques in a single analysis has increased sampling depth. Electron transfer dissociation (ETD) is complementary to collision induced dissociation (CID) in that it produces richer fragment spectra from highly charged peptides (>3+) and leaves PTMs intact on the peptide backbone. Sweney et al. [28] showed that, when both fragmentation methods are available on a single platform, the complementarity can be exploited to increase sequencing success rate by choosing either CID or ETD in a data dependent decision tree approach, based on precursor charge and m/z observed in the MS scan. Using this decision tree logic they identified 53 055 peptides in total, which was greater than using CID (38 293) or ETD (39 507) alone. Frese et al. [29] modified this decision tree logic to accommodate for fragment spectra readout in the Orbitrap analyzer for both ETD and CID (HCD) and Zhou et al. [30] showed that this approach aids in the identification depth of phosphopeptides.

An alternative approach to tackle the dynamic range issue and increase sensitivity is represented by single reaction monitoring (SRM) based approaches. As such approaches have been reviewed extensively recently, and do not provide an unbiased proteome coverage approach; we refer the interested reader to these reviews [31,32].

Chromatography

There are many ways to tackle the dynamic range issues in MS-based proteomics. These include the coupling of several stages/phases of chromatography to enhance peptide separation, which evidently goes at the expense of longer analysis time [25,33,34]. Moreover, care has to be taken that such an approach does not go hand-in-hand with cumulative sample loss, favoring somewhat online approaches. Additionally, new chromatographic approaches are evaluated that either use longer-columns, smaller inner diameters and/or alternative phase materials (e.g. SCX, HILIC, WAX, ERLIC) [33], all implemented to enhance the separation power and/or sensitivity. Illustrative of the state-of-the art in sensitivity in RP chromatography Kocher et al. [35] identified over 2500 proteins in a tryptic digest of 1 μg of lysate using a single 8 h gradient on a 50 cm x 75 μm column packed with 2 μm particles. Frese et al. [29] showed a similar depth of analysis (>2500 proteins) from 1 μg of material using single 5 h gradients
on a 35 cm × 50 μm column packed with 3 μm particles, in combination with decision tree guided peptide CID/HCD/ETD fragmentation, as described above. The Mann group reported the identification of almost 3000 proteins from yeast obtained using a 50 cm × 75 μm column, 1.8 μm particle sizes and a column heater to restrain the operating pressure, with an 8 h gradient run in triplicate [36]. Next they further increased the performance by transferring their LC setup to an EASY nLC 1000 UPLC system coupled to the new Q Exactive MS, combining increased separation power with increased sequencing speed [37**]. This setup identified over 4000 proteins in 4 h gradients ran in sixplex starting with 4 μg of material. This number of identified proteins comes close to the total number expressed and thus allowed system-wide yeast proteome analysis upon heat shock. Di Palma et al. showed that using a zwitterionic hydrophilic interaction liquid chromatography approach high resolution in separation could be achieved allowing the reasonable comprehensive proteome analysis of about 5000 FACS sorted adult stem cells [11*,38]. Masuda et al. [39] reported on a miniaturized LC–MS system with a high-recovery phosphopeptide enrichment protocol based on hydroxy acid-modified metal oxide chromatography (HAMMOC) that enabled them to detect over 1000 phosphorylation sites starting with only 10⁴ cells (1 μg starting material). Their miniaturized analytical column of 25 μm diameter provided a 3.6-fold improvement in sensitivity over the conventional 100 μm diameter column.

To go from several thousands of cells towards single cell proteomics, several of these developments will have to be combined and preferentially constructed into an online system to minimize sample losses. Individual steps in online sample preparation, fractionation and separation have already been reported. Waanders et al. introduced an on-line ‘replay’ system whereby the LC flow is initially split, and one portion is analyzed directly, while the other is diverted to a capture capillary, only analyzing undersampled features in the replay run [19*,40]. An automated online pressurized digestion system has been developed allowing the simultaneous introduction of the sample and the enzyme, circumventing sample handling steps [41]. The authors then combined this system with a modified form of replay, splitting the flow to have one portion digested with pepsin and the other portion analyzed as intact proteins, combining top-down and bottom-up proteomics in a single online setup [42]. Online chemical stable isotope labeling for quantification has been reported by Rajmakers et al. [43], where each sample was loaded onto a trap column followed by the loading of chemical reagents to accomplish full labeling, after which the next sample was loaded and treated. With this approach 3 samples can be compared quantitatively using dimethyl stable isotope labeling, while circumventing elaborate offline sample handling, minimizing sample losses, and thus making it very well suited for small sample volumes. The group of Figey has developed online proteomics platforms fully integrating pre-concentration, buffer exchange, reduction, alkylation, digestion and on-line 2D LC–MS/MS for the analysis of small sample amounts. With this system they targeted biological systems where sample availability is inherently low, such as ER and Golgi microsomal membranes in rat hepatic cells [44] and suprachiasmatic nucleus neurons from a single mouse [45], identifying 955 and 2131 proteins, respectively.

Figure 3

Mass cytometry (adapted from Ref. [49]). Cells are stained with epitope-specific antibodies conjugated to rare transition element isotope reporters, each with a different mass. Cells are nebulized into single-cell droplets, and an elemental mass spectrum is acquired for each by ICP-MS, providing system-wide views of immune signaling in healthy human hematopoiesis.
6 Omics

Conclusions and outlook
Genuine single cell analysis with MS so far only results in the detection of high abundant peptides, metabolites and lipids [46]. For the analysis of the protein content directly from single cells the complexity and huge dynamic range are still large hurdles to take. An encouraging piece of work in this respect was recently reported by Salehi-Reyhani et al. [47] describing a microfluidic antibody capture chip to determine protein copy numbers in single cells. They employed optical methods to isolate, trap and lyse a single cell, after which the protein of interest is captured by its antibody and measured by total internal reflection microscopy.

An exciting new development combining FACS and mass spectrometry is mass cytometry (Figure 3). In mass cytometry [48**,49] transition element isotopes not normally found in biological systems (e.g. lanthanides) are chelated to antibodies. Cells, with bound antibody isotope conjugates, are sprayed as single-cell droplets into an inductively coupled plasma mass spectrometer (ICP-MS) creating a quantifiable response profile. The ICP-MS based detection eliminates overlap between tags (as occurring in fluorescence detection), and allows a great variety and number of detectable markers, as sufficient isotopically different rare earth metals are available, certainly more than in traditional fluorescent based flow cytometry [50]. In an early application mass cytometry was used to unravel with high-resolution hematopoietic cellular subpopulations. Unfortunately, cells vaporized in mass cytometry cannot be recovered for further analysis, as with conventional flow cytometry. Moreover, both traditional as well as mass cytometry are dependent on the availability of good antibodies, which are in particular for posttranslational modified forms of proteins still in short supply.

In summary, ultrasensitive MS-based proteomics is making significant progress, made possible through improvements at different stages of the proteomics workflow. These advancements open up way to explore by proteomics important new biological questions such as the protein signatures of adult stem cells, tumor heterogeneity and ultimately stochastic fluctuations in protein expression.

Acknowledgements
This work was partly supported by PRIME-XS, grant agreement number 262067, funded by the European Union 7th Framework Programme; the Netherlands Proteomics Centre, embedded in the Netherlands Genomics Initiative and the Centre for Biomedical Genetics. We apologize for leaving out further important contributions.

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 analysis of MS imaging of myxoid sarcoma tissue, revealing intratumor heterogeneity and single tissue sections containing high-grade and low-grade tumors, supporting a concept of tumor progression through clonal selection.


38. System-wide perturbations in yeast are examined by single LC-MS analysis, combining high resolution LC separation with high resolution and accuracy MS detection.


40. Masuda T, Sugiyama N, Tomita M, Ishihama Y: Microscale phosphoproteome analysis of 10,000 cells from human cancer cell lines. Anal Chem 2011, 83:7698-7703. Analysis of a significant amount of phosphopeptides (1011) from as little as 1 μg of protein starting material, which is significantly lower than conventionally shown.


