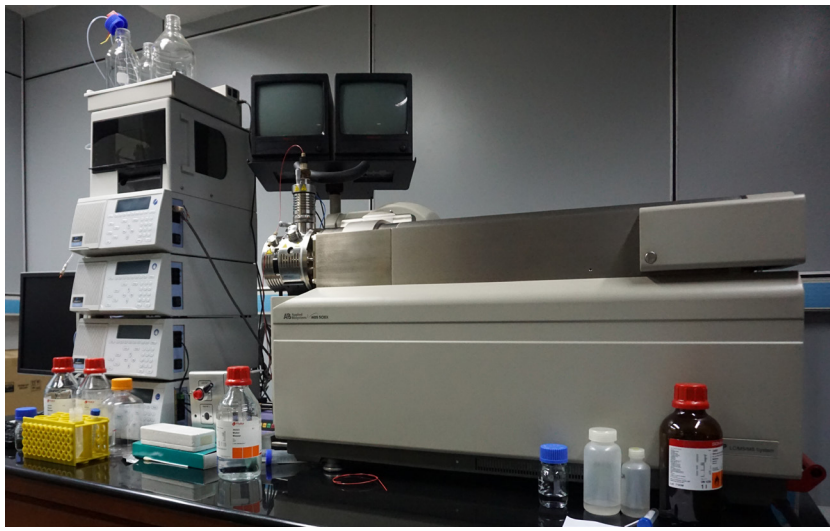
doi: 10.1038/nindia.2015.116



# Next generation proteomics tools

Srikanth Rapole[1], Mahesh J. Kulkarni[2], Suman S. Thakur[3] & Shantanu Sengupta[4]

*Srikanth Rapole*

*Mahesh J. Kulkarni*

*Suman S. Thakur*

*Shantanu Sengupta*

In the early era of proteomics, conventional tools such as two dimensional gel electrophoresis to separate a large number of proteins, spot picking, tryptic digestion of individual spots and MALDI-TOF-MS based identification were quite popular. However, they were tedious, time consuming, lacked throughput and quantitative ability.

The quantitative ability of mass spectrometry was realised with the development of nano liquid chromatography-based separation and high resolution mass spectrometry. Subsequently, shot gun proteomics and semi quantitative labeling technologies such *as isobaric tags for relative and absolute quantitation* (iTRAQ) and stable isotope labeling by amino acids in cell culture (SILAC) were developed.

The multiplexing ability of iTRAQ allowed discovery of candidate biomarkers including tissue biomarkers, serum biomarkers and drug resistance markers despite the limitations of variability in labelling efficiencies and quantitation at MS/MS level. SILAC is another mass spectrometry based quantitative approach predominantly compatible with cell culture system, which eventually became a powerful tool in quantitative biology. Label free approaches like $MS^E$ (MS at elevated energy) facilitate untargeted quantitation.

## Quantitation methods

Historically, triple-quadrupole based mass spectrometers were used for quantitation because of their high sensitivity and scan speed. These instruments facilitated development of quantitative approaches such as multiple reaction monitoring (MRM) or selected reaction monitoring (SRM), which recently have made a mark in the area of proteomics for biomarker validation. In MRM, a specific precursor and fragment ion are monitored for quantitation. MRM is highly reproducible and provides absolute concentration if stable isotope-labeled internal standards are included in the workflows. MRM based targeted quantification is becoming quite popular in the proteomics community, as this approach is able to replace expensive antibody-based quantification like Western blotting and ELISA.

[1]Proteomics Lab, National Centre for Cell Science, Pune, India (rsrikanth@nccs.res.in). [2]Proteomics Facility, Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune (mj.kulkarni@ncl.res.in). [3]Proteomics and Cell Signaling, CSIR-Centre for Cellular & Molecular Biology Hyderabad, Andhra Pradesh, India (sst@ccmb.res.in). [4]CSIR-Institute of Genomics and Integrative Biology, New Delhi (shantanus@igib.res.in).

Several recent studies have used this approach for targeted quantitation. One such example is quantification of cancer related proteins in body fluids using targeted protemics[1]. The sensitivity, selectivity and scan speed of triple quadrupole mass spectrometers have been incorporated into high resolution mass spectrometers such as QTOF and Orbitraps, for identification and quantification of the proteome. ABI5600, Q-Exactive HF and Xevo G2 XS QTOF are some of the instruments that can perform the dual functions of identification and quantitation.

MRM performed on QTOFs and Orbitraps are called pseudo MRM or high resolution MRM (HR-MRM). Sometimes it is also referred to as parallel reaction monitoring (PRM). Unlike MRM, in PRM it is not possible to monitor the specific fragment ion during acquisition. Post mass spectral acquisition, extracted ion chromatograms (XIC) for selected ions are used for quantitation. The high scan speed facilitates development of sequential window acquisition of all theoretical mass spectra (SWATH).

In this approach, a spectral library is created by information dependent acquisition (IDA), later the instrument is specifically tuned for the selection of precursor ions from an overlapping window of 25 m/z spread over a precursor mass range of 400-1250 m/z window 25 m/z wide. Peptides are quantitated by targeted data extraction of SWATH-MS data.

## Post-translational modifications

Post-translational modifications are vital for regulating a number of cellular processes and cellular control mechanism. High resolution accurate mass spectrometers like TOFs and Orbitraps also facilitate better characterisation of PTMs. Various mass spectrometric strategies like MS/MS, neutral loss and electron transfer dissociation (ETD) are being used for precise characterisation of PTMs. However, the quantitation of PTMs heavily relies on the fragment ion library. Thus construction of fragment ion library for synthetically modified peptides becomes a prerequisite for quantification of PTMs. Once the library for the modified peptide is established, the PTMs can be quantified by either MRM, PRM or SWATH. One of the inevitable consequences of post-translational modifications is generation of various 'proteoforms'. They also arise as a result of genetic variations, mutations and splicing.

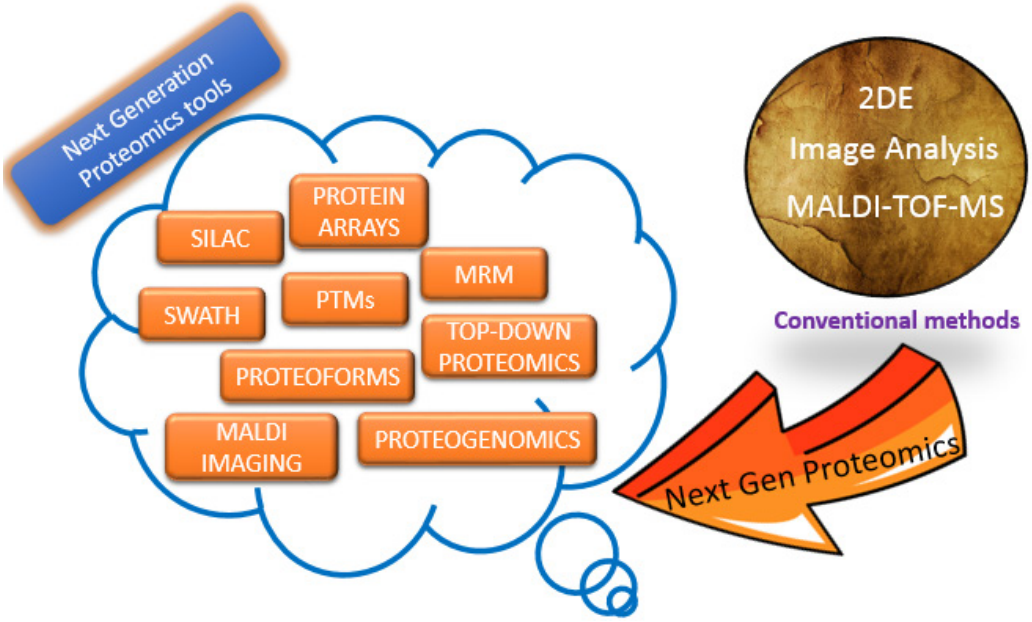
Proteoforms are important since they can be differentially expressed in disease conditions and activate different pathways leading to completely different disease physiology. Proteoforms play an important role in biological processes and could be potentially used as biomarkers. One of the popular examples of such proteoforms is HbA1c, or glycated haemoglobin, used as a diagnostic marker to assess glycaemic status over preceding 100-120 days.

There is an increasing need to discriminate and quantify such isoforms with accuracy and efficiency. But highly homologous amino acid sequences and a great variation in their cellular concentrations pose a major challenge. However, advancements in mass spectrometry-based techniques have now enabled the identification and quantification of these protein isoforms.

## Bottom-up or top-down?

A major roadblock in identification of proteoforms using the popular bottom-up approach is that very few peptides of an intact protein can be detected unambiguously. In most cases, the unique peptide of an isoform may be missed or may remain undetected. Thus, the top-down approach can be useful as it provides more information about the single intact protein.

Top-down proteomics is one of the classical techniques in mass spectrometry with great potential. It is not as popular as the bottom-up approach due to its complexity – it is still in the developing phase.




In top-down proteomics, the focus is to get an intact protein mass and characterise it by fragmentation of different charge states. Top-down proteomics has great benefits, especially to detect degradation products, sequence variants and simultaneous detection of multiple post-translational modifications[2].

Top-down proteomics experiments have been performed using different modes CID, HCD, ECD, ETD for different proteins, non-covalent complexes and subunit complexes using a variety of new generation instruments such as modified FT-ICR MS, Orbitrap, Q-TOF with ion mobility and triple quadruple. Top-down proteomics will play an important role in clinical and translational research and also in identifying unique protein forms or proteoforms. However, this method is less widely used due to the lack of MS compatible methods and instruments.

## Proteogenomics

Proteogenomics is another upcoming discipline that utilises proteomics, genomics and transcriptomics information to find, assign and confirm the functional gene, as individually they don't provide wholesome information about a system[3]. Genomics and transcriptomics experiments help make customised protein sequence database.

*De novo* sequencing is also very helpful to find novel peptide sequences and novel genes by searching against databases. Notably novel peptides are identified using shotgun proteomics through customised databases that are absent in reference protein sequence databases. Proteomics is playing an important role in providing protein level evidence on the basis of expressed proteins and finally enriching gene information.

Proteogenomics has great importance in clinical research and biomarker discovery. Onco-proteogenomics is becoming especially popular in cancer research to find tumour specific signature peptides. As normal shot gun proteomics does not provide information regarding point mutation, unusual splice variants and gene fusion, proteogenomics will be helpful in this regard. Proteogenomics has the potential to link DNA, RNA and protein expression information in the perspective of central dogma.

## MALDI imaging

Matrix assisted laser desorption ionisation imaging mass spectrometry (MALDI-IMS) is emerging as a powerful tool to explore the molecular content of tissues within their morphological context. It allows direct measurement of proteins, peptides, metabolites, lipids and drugs from tissue sections.

Distribution of detected compounds can be seen as an image. Being a label free approach, MALDI-IMS is increasingly being recognised in the field of biomarker discovery, especially tissue-based research. The technical developments in MALDI imaging acquisition and data analysis are facilitating this approach for better understanding of molecular changes associated with the progression of disease.

## Protein microarrays

The protein microarray platform is another gel free approach emerging as a powerful tool to study thousands of proteins simultaneously. Protein arrays are miniaturised 2D arrays generally printed on functionalised glass slides comprising immobilised proteins of interest which can be analysed in a high throughput manner[4].

There are several protein microarray formats including tissue arrays, reverse-phase arrays, capture arrays and lectin arrays that have advanced in recent years. These are being successfully applied in various fields including protein-protein interaction studies, immunological profiling, biomarker discovery and vaccine development. Rapid advances in next generation proteomics tools are delivering meaningful biological insights in modern biology.

## References

1. Huttenhain, R. *et al.* Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci. Transl. Med.* **4**, 142ra94 (2012).
2. Tran, J. C. *et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254-258 (2011).
3. Faulkner, S. *et al.* Proteogenomics: emergence and promise. *Cell Mol. Life Sci.* **72**, 953-957 (2015).
4. Chandra, H. *et al.* Protein microarrays and novel detection platforms. *Expert Rev. Proteomics* **8**, 61-79 (2011).