

What is proteomics?

Andrew J McArdle*, Stephanie Menikou¹

This article has been accepted for publication in Archives of Disease in Childhood Education & Practice, 2020 following peer review, and the Version of Record can be accessed online at <https://doi.org/10.1136/archdischild-2019-317434>.

*Corresponding author: Section of Paediatric Infectious Disease, Imperial College, Medical School, St Mary's Hospital, Praed Street, London, W2 1NY, UK. E-mail: andrew.mcardle@gmail.com. Telephone: 0207 5943717

1. Section of Paediatric Infectious Disease, Imperial College, Medical School, St Mary's Hospital, Praed Street, London, W2 1NY, UK

Abstract

Proteomics is the study of a large number of proteins in biological systems. We aim to introduce the complex field to paediatricians and present some recent examples of applications to paediatric problems.

Various approaches have been used to study proteomes. The current mainstay is tandem mass spectrometry of enzymatically digested proteins ("bottom-up proteomics"), and we describe the experimental and computational approach further.

Proteomics can offer advantages over transcriptomics by giving direct information about proteins rather than RNA, however typically data is obtained at lower depth and the confident identification of mass spectra can be challenging. Proteomics frequently complements transcriptomics and other -omics.

Used effectively, proteomics offers promise to help answer important clinical and biological questions.

The term “proteome” was coined in a 1995 study of bacterial proteins.¹ In accordance with the common approach of -omics sciences, proteomics is the study of a large number of proteins in biological systems (e.g. cells, tissues, or organisms) at once.

The simplicity of this statement belies the complexity of the field, since an enormous range of basic questions can be asked regarding proteins: about their abundance, localisation, interactions, modification states and folding, through time, under different conditions and in different subjects. Regardless, to date, the majority of clinical proteomic studies focus on answering biological questions based on protein abundance and post-translational modifications.

TECHNIQUES

Early proteomic studies relied upon separating complex protein mixtures on 2D gels, and this approach still has a role.² Individual proteins can be resolved as spots and identified by time-consuming sequencing by Edman degradation or mass spectrometry (MS). Abundance changes can be recognised through changes in spot intensity.

Development of tandem MS and allied informatics allowed unfractionated protein mixtures to be analysed in one run. Presently tandem MS of digested proteins ("bottom-up proteomics") is the main approach,³ though others are used (figure 1a).

MASS SPECTROMETRY-BASED PROTEOMICS

Figure 1b illustrates the process of bottom-up proteomics which is described in more detail by Zhang et al.³ Proteins are extracted from a sample and digested (most often with trypsin). The digested peptides are separated by liquid chromatography and eluted over time and directly introduced into the mass spectrometer detector through the ionization interface. Successive rounds of peptide ions ("precursors") are subjected to a first round of MS (usually at high resolution) to measure their mass-charge ratio (m/z). Typically, the most abundant ions are then selected to be fragmented in turn and undergo a further round of MS.

Various bioinformatic approaches exist to identify peptides and proteins from the data.⁴ Most commonly, the fragment mass spectra are matched to possible peptides from a relevant protein database, controlling for chance matches. Proteins can be quantified by inspecting the intensity of their corresponding precursor signals.

There is increasing interest in top-down proteomics, in which complete proteins are analysed by tandem MS, with an intervening fragmentation step. There are advantages for detection of protein isoforms. However, there are challenges in dealing with complex protein mixtures, and it is most frequently applied to individual proteins. Interested readers are directed to a recent review.⁵

THE VALUE AND CHALLENGE OF PROTEOMICS

One may ask what value proteomics can add when the field of transcriptomics is very advanced. The great strength of proteomics is that it provides direct information about the structural, signaling and enzymatic building blocks of the body, and proteins comprise a majority of drug targets. In contrast, transcriptomics tells us about abundance of individual transcripts, but not directly about protein abundance or state. The correspondence between gene and protein expression levels is often poor; the two techniques provide complementary information and are frequently used together.⁶

Protein-based diagnostic tests are potentially more easily translatable into low-cost near-patient technologies. Such tests are already in wide use, for example pregnancy tests and rapid-diagnostic tests for malaria.

However, the confident and comprehensive identification and quantification of peptides and then proteins from mass spectra is not without challenge. Unlike sequencing technologies which directly

read nucleotides with confidence scores, further steps are necessary to assign fragment mass spectra to putative peptides and proteins. Further, where typical sequencing runs give tens of millions of reads per sample, typical proteomic studies obtain only tens of thousands of fragment spectra per sample.

APPLICATIONS

Diagnostic and prognostic markers

Diagnosis of sepsis and necrotising enterocolitis (NEC) on the neonatal unit remain challenging clinical problems. MS-based proteomic analysis of neonates' urine yielded seven proteins which were then validated in a separate cohort by enzyme-linked immunosorbent assay (ELISA) as protein biomarkers for diagnosis of sepsis, and prognosis and diagnosis of NEC.⁷ Areas under the curve in a validation cohort indicated good performance (all >0.95), with sensitivity and specificity for NEC vs. sepsis of 89% and 80% respectively.

There is increasing drive to target treatments to individual patients by deeper phenotyping (so-called "personalised" or "precision" medicine), especially in oncology. Jiang et al⁸ identified a large number of proteins differentially expressed between prednisolone-resistant and -sensitive acute lymphoblastic leukaemia cell lines. One of these, PCNA, was validated as a biomarker of prednisolone responsiveness in children, independent of subtype.

Understanding biology and disease

It is believed that cow's milk antigens enter maternal milk to drive cow's milk protein allergy in exclusively breastfed infants. However, antibody-based techniques can miss digested proteins and cross-reactivity cannot be excluded. Zhu et al⁹ recently identified 36 non-human proteins in human breastmilk, mostly of bovine origin, using bottom-up proteomics. A further study¹⁰ showed that cow's milk peptides in breastmilk peak at two-hours post ingestion, and that detected peptides are rich in proline which may have helped them resist digestion in the gut.

The factors which determine severity of malaria remain poorly understood. Reutersward et al¹¹ used an antibody bead array approach to characterise the serum proteome of children with malaria and healthy controls. 37 proteins were found to be differentially abundant in severe and non-severe malaria; these included acute phase reactants and proteins involved in cell migration/adhesion and tissue remodelling.

CLOSING REMARKS

In this brief review we have introduced proteomic principles and technologies, and a few examples of their application.

In common with all -omics, the rapidity of progress and application can be overstated, the volume of data leads to computational and statistical challenges, and techniques can be employed thoughtlessly or simply because they are in vogue.¹² Identifying biomarkers is not enough – the road to clinical application is long, as outlined in recent reviews and shown in figure 2.^{13,14} Helpfully, data sharing is increasingly mandated, allowing validation, reanalysis, repurposing and meta-analysis (e.g. PRIDE archive¹⁵).

Used effectively (and often complementing other "-omics" techniques), proteomics offers promise to help answer important clinical and biological questions.

Acknowledgments

N/A

Competing Interests

None.

Funding

AJM is funded through a Wellcome Trust/Imperial College London 4i PhD Fellowship. SM is funded through a Biomedical Research Centre Institute for Translational Medicine and Therapeutics grant

References

1. Wasinger VC, Cordwell SJ, Cerpa-Polkjak A *et al.* Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis*. 1995;16:1090–1094.
2. Oliveira BM, Coorsen JR and Martins-de-Souza D. 2DE: The Phoenix of Proteomics. *J. Proteomics*. 2014;104:140–150.
3. Zhang Y, Fonslow BR, Shan B *et al.* Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* 2013;113:2343–2394.
4. Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J. Proteomics*. 2010;73:2092–2123.
5. Gregorich ZR and Ge Y. Top-down proteomics in health and disease: challenges and opportunities. *Proteomics*. 2014;14:1195–1210.
6. Rogers S, Girolami M, Kolch W *et al.* Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models. *Bioinformatics*. 2008;24:2894–2900.
7. Sylvester KG, Ling XB, Liu GY *et al.* Urine Protein Biomarkers for the Diagnosis and Prognosis of Necrotizing Enterocolitis in Infants. *J. Pediatr.* 2014;164:607-612.e7.
8. Jiang N, Kham SKY, Koh GS *et al.* Identification of prognostic protein biomarkers in childhood acute lymphoblastic leukemia (ALL). *J. Proteomics*. 2011;74:843–857.
9. Zhu J, Garrigues L, Toorn *et al.* Discovery and Quantification of Nonhuman Proteins in Human Milk. *J. Proteome Res.* 2018;18:225–238.
10. Picariello G, De Cicco M, Nocerino R *et al.* Excretion of Dietary Cow's Milk Derived Peptides Into Breast Milk. *Front. Nutr.* 2019;6:10.3389/fnut.2019.00025.
11. Reuterswärd P, Bergström S, Orikiiriza J *et al.* Levels of human proteins in plasma associated with acute paediatric malaria. *Malar. J.* 2018;17:426.
12. Ioannidis JPA. A roadmap for successful applications of clinical proteomics. *Proteomics Clin. Appl.* 2011;5:241–247.
13. Zhang Z and Chan DW. The Road from Discovery to Clinical Diagnostics: Lessons Learned from the First FDA-Cleared In Vitro Diagnostic Multivariate Index Assay of Proteomic Biomarkers. *Cancer Epidemiol. Prev. Biomark.* 2010;19:2995–2999.
14. Parker CE and Borchers CH. Mass spectrometry based biomarker discovery, verification, and validation – Quality assurance and control of protein biomarker assays. *Mol. Oncol.* 2014;8:840–858.
15. Vizcaíno JA, Csordas A, del-Toro N *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 2016;44:D447–D456.

FIGURES

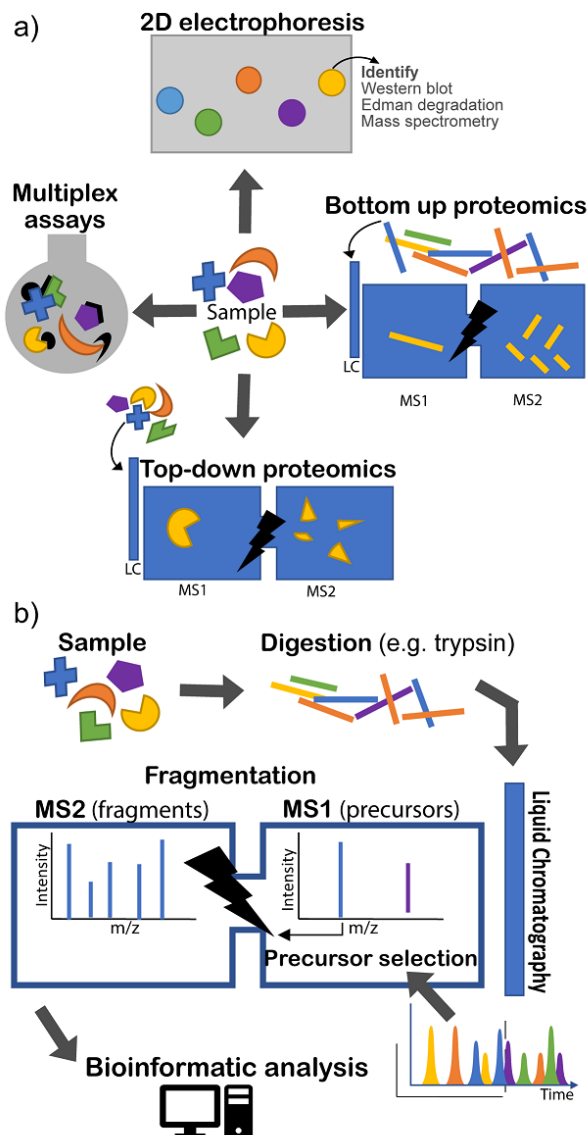


Figure 1. Proteomic techniques illustrated. (a) Four types of approach are illustrated. 2D electrophoresis separates proteins on a gel by electrochemical means. With subsequent staining, protein “spots” can be recognised and compared between samples. Identification of proteins requires other tools, such as Western blotting or MS. Top-down and bottom-up proteomics both rely upon tandem (multiple rounds of) MS. The bottom-up approach analyses peptides produced by enzymatic digest of proteins. The top-down approach analyses intact proteins directly. In both cases a second round of MS is undertaken following fragmentation (MS2). Finally, various multiplex assay technologies exist for quantifying panels of proteins in solution simultaneously (e.g. Luminex and Mesoscale Discovery). LC: liquid chromatography, MS: mass spectrometry. (b) The bottom-up proteomic approach is illustrated in more detail. A sample is digested, usually by trypsin. The various peptides in the sample are eluted over time through a liquid chromatography column and ionised into the mass spectrometer. The mass spectrometer regularly analyses the eluted peptide species (MS1) and selects a number of precursors to be sequentially fragmented and analysed (MS2). Bioinformatic analysis is required to identify peptides from their precursor and fragment mass spectra (MS1 and MS2).

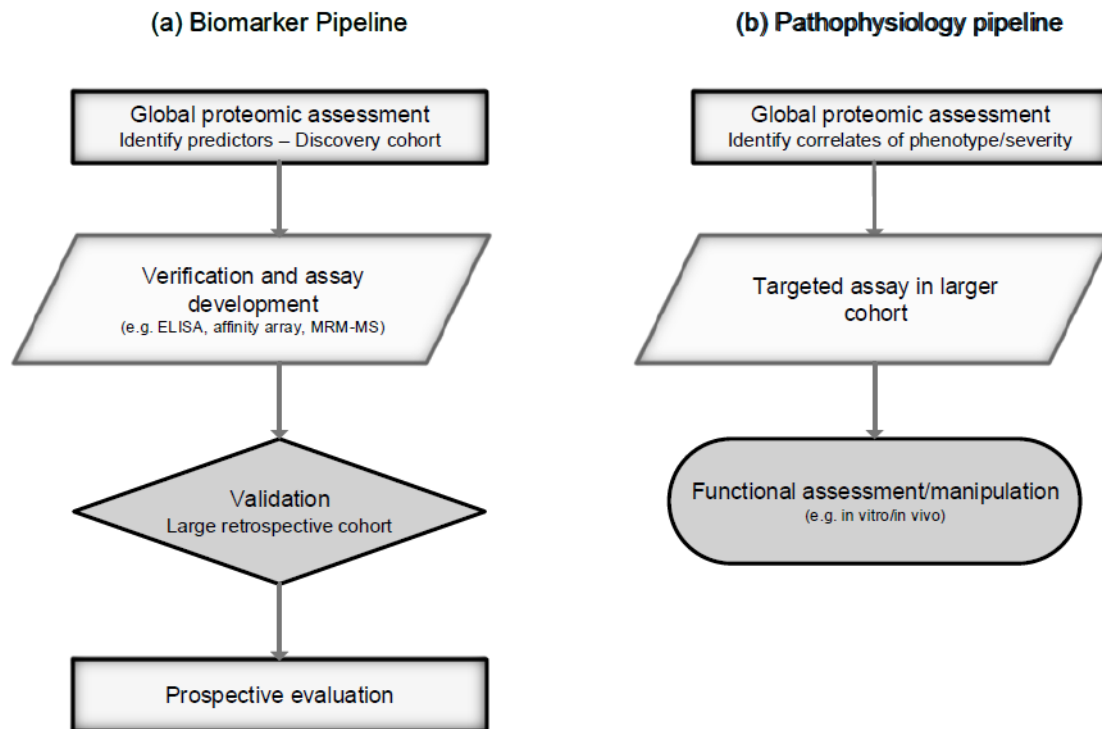


Figure 2. Proteomics can be used to identify biomarkers and understand pathophysiology. Multiple steps are required to ensure validity, and for biomarkers, clinical utility. Outline processes are illustrated. Further detail on biomarker discovery is given by Parker and Borchers (2014).