Published in final edited form as:
Expert Rev Proteomics. 2014 December ; 11(6): 649–651. doi:10.1586/14789450.2014.976559.

Deep and quantitative top-down proteomics in clinical and translational research

Neil L Kelleher, **Paul M Thomas**,
Department of Molecular Biosciences, Northwestern University, 2145 N. Sheridan Rd, Evanston, IL, 60208, USA

Ioanna Ntai,
Department of Chemistry, Northwestern University, 2145 N. Sheridan Rd, Evanston, IL, 60208, USA

Philip D Compton, and
Department of Chemistry, Northwestern University, 2145 N. Sheridan Rd, Evanston, IL, 60208, USA

Richard D LeDuc
Proteomics Center of Excellence, Northwestern University, 2145 N. Sheridan Rd, Evanston, IL, 60208, USA

Abstract

It has long been understood that it is proteins, expressed and post-translationally modified, that are the primary regulators of both the fate and the function of cells. The ability to measure differences in the expression of the constellation of unique protein forms (proteoforms) with complete molecular specificity has the potential to sharply improve the return on investment for mass spectrometry-based proteomics in translational research and clinical diagnostics.

Keywords

clinical; proteomics; quantitative; top-down; translational

While cells within the human body may share the same genome, it is the proteins that serve as downstream effector molecules that perform enzymatic reactions, regulate cellular processes and, in general, give rise to the organism's phenotype. The past decade has witnessed a remarkable evolution in proteomics research as it transformed from a technique practiced by a specialized community into a thriving field of science [1,2]. Most forms of proteomics rely on tandem mass spectrometry in an attempt to *identify* the gene from which

Author for correspondence: Northwestern University, 2145 N. Sheridan Rd, Evanston, IL, 60208, USA, Tel.: +1 847 467 4362, n-kelleher@northwestern.edu.

Financial & competing interests disclosure
The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript.
No writing assistance was utilized in the production of this manuscript.

the protein derives, *characterize* the alterations relative to the reference sequence and all covalently bound moieties attached to the sequence (i.e. a specific proteoform [3,4]) and *quantify* the relative abundance of the protein when comparing two or more biological states.

Most practitioners utilize the now well-developed methods of bottom-up shotgun proteomics in which proteins are enzymatically digested with a protease, such as trypsin, and the resulting small peptides (<30 amino acids in length) are analyzed by mass spectrometry [5]. These bottom-up proteomic techniques introduce a 'peptide to protein' inference problem that complicates the identification and quantitation steps in even well-organized proteomic studies [6]. Ambiguities in protein inference carried forward into quantitation, we postulate, decrease the chances of biomarker discovery and validation. In contrast, top-down proteomics eliminates the use of proteases during sample preparation [7], and instead measures the intact protein directly, and then fragments the protein for identification and characterization. By measuring and quantifying whole proteins, we achieve more confident identification and better characterization of individual proteoforms, and as a result, a deeper understanding of the biological processes that control human health and disease.

The proteoform hypothesis

The completed human genome sequence revealed a much smaller complement of genes than was originally anticipated, suggesting that a major source of complexity within our bodies arises from variations of protein molecules and not solely from gene/protein expression. Protein variation can arise from changes in the genome (e.g. allelic variants from coding polymorphisms or mutation), from alternative RNA splicing, from *in vivo* proteolysis (e.g. signal/transit peptide cleavage) and from any number of diverse post-translational modifications. It is the accumulation of all of these events that define a specific proteoform and govern its biological function [4]. Therefore, proteoform-resolved measurements offer a clarified view of transcription, translation and post-translational events that underlie complex phenotypes [8]. Many candidate assays for clinical diagnostics rely on imperfect ELISAs, mRNA transcript measurements or the analysis of *in vitro* enzymatically generated peptides; these measurement modalities have returned advances [9], but indirectly reflect the presence or actions of proteoforms. A clear way to understand clinically relevant differences at the protein level between biological states is to measure the *differences in the expressed proteoforms* between those states.

Our proposition, that intact proteoforms represent a powerful class of molecules for use as biomarkers of disease states, is referred to as the 'proteoform hypothesis'. The word 'powerful' is used in the statistical sense; power is the ability to detect a true difference between two or more populations when such a difference is present. The proteoform hypothesis therefore states that proteoforms have the greatest ability to differentiate biologically real differences in samples of complex material; the presence or absence of cancers, the onset of disease, the classification of cell types or the differentiation of two or more biological states. In fact, recent findings suggest that mRNA abundances are only weakly correlated to protein expression levels [10]. Likewise, *in vitro* enzymatically generated peptides offer only a small piece of the puzzle; we posit that measuring intact

proteoforms will deliver increased value and return on investment in clinical research, provided the technology is available and robust.

A label-free platform for differential measurement of proteoforms

The new platform we highlight here [11] largely separates proteoform identification and characterization from relative quantitation. To implement the approach, high-performance mass spectrometers are currently needed to measure intact proteoforms. We use Fourier Transform (Orbitrap) mass spectrometers with Automatic Gain Control for label-free top-down quantitation [11]. Furthermore, to achieve sufficient peak capacity for complex samples, orthogonal phases of separation are employed; typically, we use a molecular weight-based separation (GELFrEE) followed by liquid chromatography (hydrophobicity-based) that is directly coupled to the mass spectrometer (i.e. liquid chromatography–mass spectrometry). This analysis relies on intact mass profiling of proteoforms from multiple liquid chromatography–mass spectrometry runs, and then calculating a proteoform intensity based on the sum of all the relevant isotopic peak heights over the elution time of the proteoform. Using a statistical model, we quantify individual proteoforms within nested technical and biological replicates. We are then able to estimate the relative differences in proteoform expression observed between two or more clinically relevant states with statistical confidence.

A top-down proteomics experiment, operated in discovery mode, can now detect thousands of proteoforms derived from over a thousand unambiguously identified genes [12]! With the advent of label-free relative quantitation, differences in the relative abundance of over a thousand proteoforms can be tested, even if some quantified masses lack identifications [11,13]. The value proposition of this new approach for biomarker discovery appears high, as it allows the deepest analysis of relative proteoform expression in the low mass proteome yet reported. To the extent the positive outlook projected here proves true, the value of proteoforms will be felt by improving the return on investment on clinical/translational research for protein-based biomarkers. However, proteoform-aware, targeted assays are also inexpensive to deploy relative to other technologies – and this principle is already being proven by the availability of new clinical assays.

Current clinical & translational applications of top-down proteomics

The use of top-down mass spectrometry to measure clinically relevant proteoforms is not new. A recent review describes a half dozen prescient examples [8]. An increasing number of labs working at the interface of technology and human health are showing the emergent use of top-down proteomics in translational research [14–17].

The revolution in molecular analysis is also beginning to make headway in the clinic. In 2013, Bruker was granted 510(k) clearance to use its MALDI BioTyper CA system for the identification of Gram-negative bacteria cultured from patients [18]. In this system, a microbial colony is directly spotted onto a MALDI plate and the most abundant proteins in the sample are profiled in a top-down manner using their intact mass. After comparison of the protein profile with a library, a bacterial identification can be made at the genus, species and even sub-species levels. This new capability not only changes the business calculus

regarding development of specific antibiotics but it may also reduce the problem of antibiotic resistance. By operating at the proteoform level (i.e. capturing sequence variation of entire ribosomal proteins), this assay is made economically viable and functional.

Large clinical laboratories have also started to use intact proteoform profiling information (i.e. without tandem mass spectrometry) to assist in the diagnosis of disease. Major examples of the use of intact protein profiling (in combination with DNA sequencing) are the detection of transthyretin sequence changes within plasma to help diagnose hereditary amyloidosis [19], the analysis of hemoglobin variants within erythrocytes to diagnose blood disorders such as thalassemia [20] and the measurement of insulin-like growth factor, a protein whose serum levels are indicative of certain growth abnormalities [21].

Future outlook & conclusions

While the value of measuring intact proteoforms in disease diagnosis is being realized, there is still untapped potential in using proteoform analysis within the clinic. Most of the examples described above only utilize the highly accurate mass of specific intact proteoforms; greater confidence in diagnosis will arise when both identification and characterization of proteoforms are incorporated in the assays. Mass spectrometers need to be further developed for routine intact protein analysis experiments; most mass spectrometers are developed with peptide analyses in mind. Increased sensitivity will allow for reduced sample size. Both the required sample amounts and the degree of sample handling before analysis must decrease when working with precious clinical samples, such as biopsies. Finally, the process from mass spectrometry data collection to diagnosis needs to be automated. The comprehensive analysis of proteoforms has the potential to revolutionize the molecular understanding of health and disease, but only with further development can this innovation be brought to fruition. As the value and efficiency of proteoform analysis comes more clearly into view, we expect the number of proteoform-resolved diagnostics to expand in the years ahead.

Acknowledgement

This study was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM067193. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Biographies

Paul M Thomas

Ioanna Ntai

Philip D Compton

Richard D LeDuc

References

1. Schiess R, Wollscheid B, Aebersold R. Targeted proteomic strategy for clinical biomarker discovery. Mol Oncol. 2009; 3(1):33–44. [PubMed: 19383365]
2. Walther TC, Mann M. Mass spectrometry-based proteomics in cell biology. J Cell Biol. 2010; 190(4):491–500. [PubMed: 20733050]
3. LeDuc RD, Fellers RT, Early BP, et al. The C-score: a Bayesian framework to sharply improve proteoform scoring in high-throughput top down proteomics. J Proteome Res. 2014; 13(7):3231–3240. [PubMed: 24922115]
4. Smith LM, Kelleher NL. Proteomics, C.T.D., Proteoform: a single term describing protein complexity. Nat Methods. 2013; 10(3):186–187. [PubMed: 23443629]
5. Pandey A, Mann M. Proteomics to study genes and genomes. Nature. 2000; 405(6788):837–846. [PubMed: 10866210]
6. Shteynberg D, Deutsch EW, Lam H, et al. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol Cell Proteomics. 2011; 10(12):M111.007690. [PubMed: 21876204]
7. Siuti N, Kelleher NL. Decoding protein modifications using top-down mass spectrometry. Nat Methods. 2007; 4(10):817–821. [PubMed: 17901871]
8. Savaryn JP, Catherman AD, Thomas PM, et al. The emergence of top-down proteomics in clinical research. Genome Med. 2013; 5:53. [PubMed: 23806018]
9. Li XJ, Hayward C, Fong PY, et al. A blood-based proteomic classifier for the molecular characterization of pulmonary nodules. Sci Transl Med. 2013; 5(207):207ra142.
10. Khan Z, Ford MJ, Cusanovich DA, et al. Primate transcript and protein expression levels evolve under compensatory selection pressures. Science. 2013; 342(6162):1100–1104. [PubMed: 24136357]

11. Ntai I, Kim K, Fellers RT, et al. Applying label-free quantitation to top down proteomics. Anal Chem. 2014; 86(10):4961–4968. [PubMed: 24807621]
12. Catherman AD, Durbin KR, Ahlf DR, et al. Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. Mol Cell Proteomics. 2013; 12(12): 3465–3473. [PubMed: 24023390]
13. Wu S, Brown JN, Tolic N, et al. Quantitative analysis of human salivary gland-derived intact proteome using top-down mass spectrometry. Proteomics. 2014; 14(10):1211–1222. [PubMed: 24591407]
14. Kellie JF, Higgs RE, Ryder JW, et al. Quantitative measurement of intact alpha-synuclein proteoforms from post-mortem control and Parkinson's disease brain tissue by intact protein mass spectrometry. Sci Rep. 2014; 4:5797. [PubMed: 25052239]
15. Barnidge DR, Dasari S, Botz CM, et al. Using mass spectrometry to monitor monoclonal immunoglobulins in patients with a monoclonal gammopathy. J Proteome Res. 2014; 13(3):1419–1427. [PubMed: 24467232]
16. Iavarone F, Melis M, Platania G, et al. Characterization of salivary proteins of schizophrenic and bipolar disorder patients by top-down proteomics. J Proteomics. 2014; 103:15–22. [PubMed: 24690516]
17. Edwards RL, Griffiths P, Bunch J, Cooper HJ. Compound heterozygotes and beta-thalassemia: top-down mass spectrometry for detection of hemoglobinopathies. Proteomics. 2014; 14(10):1232–1238. [PubMed: 24482221]
18. [Last accessed 10 August 2014]] 510(k) Substantial equivalence determination decision summary. 2014. www.accessdata.fda.gov/cdrh_docs/reviews/K130831.pdf
19. Amyloidosis, transthyretin-associated familial, reflex. Blood. 2014 www.mayomedicallaboratories.com/test-catalog/Overview/83674.
20. Thalassemia and hemoglobinopathy evaluation. 2014 www.mayomedicallaboratories.com/test-catalog/Overview/84158.
21. IGF-I, LC/MS. 2014 www.questdiagnostics.com/testcenter/TestDetail.action?ntc=16293.