Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Developing top down proteomics to maximize proteome and sequence coverage from cells and tissues

Dorothy R Ahlf[1], Paul M Thomas[2] and Neil L Kelleher[2,3]

Mass spectrometry based proteomics generally seeks to identify and characterize protein molecules with high accuracy and throughput. Recent speed and quality improvements to the independent steps of integrated platforms have removed many limitations to the robust implementation of top down proteomics (TDP) for proteins below 70 kDa. Improved intact protein separations coupled to high-performance instruments have increased the quality and number of protein and proteoform identifications. To date, TDP applications have shown >1000 protein identifications, expanding to an average of ~3–4 more proteoforms for each protein detected. In the near future, increased fractionation power, new mass spectrometers and improvements in proteoform scoring will combine to accelerate the application and impact of TDP to this century's biomedical problems.

Addresses
[1] Department of Chemistry and Biochemistry and the Harper Cancer Institute, University of Notre Dame, Notre Dame, IN, United States
[2] Department of Chemistry, and the Proteomics Center of Excellence, Northwestern University, Evanston, IL, United States
[3] Department of Molecular Biosciences, and the Proteomics Center of Excellence, Northwestern University, Evanston, IL, United States

Corresponding author: Kelleher, Neil L (n-kelleher@northwestern.edu)

Current Opinion in Chemical Biology 2013, **17**:787–794

This review comes from a themed issue **Analytical techniques**

Edited by **Milos V Novotny** and **Robert T Kennedy**

For a complete overview see the Issue and the Editorial

Available online 27th August 2013

1367-5931/$ – see front matter,

http://dx.doi.org/10.1016/j.cbpa.2013.07.028

Introduction

Proteomics: from inception to enduring goals

The analysis of proteins has undergone a major revolution over the past 20 years from the earliest days of amino acid analysis and Edman sequencing to today's sophisticated mass spectrometry platforms. The successes of the human genome project have inspired similar efforts within the context of the proteome and have thus led the rapid development of high-throughput methods for proteomics [1,2]. Characterizing the chemical state of these proteins provides valuable biological information. The complexity of proteomics, a 'global cellular view', arises when all combinatorial patterns are taken into account across a variety of cell types. To date, bottom-up proteomics has proven ineffective to detect combinatorial proteomics, unless the modifications are co-located on one peptide.

In many regards, the human proteome is more complex than its genome. Each somatic cell in the human body encodes the same genetic information in $\sim 3 \times 10^9$ base-pairs of DNA. However, the human proteome cannot be defined this trivially. The proteoform content of a cell changes with cell type, over time and in response to external stressors. While the human genome contains just over 20 000 protein-expressing genes, RNA processing alone increases the number of possible base sequences to perhaps >100 000 in most cells. Finally, proteins may also be highly modified with differential combinatorial patterns of post-translational modifications (PTMs) [3,4]. Extensive studies of singly, highly modified proteins (e.g. histones) show that though these multitudes of modification combinations are possible, only a limited number modified forms are observed [5–7].

A word on language and protein databases

During the development of mass spectrometry-based proteomics, many new terms have entered the scientific vernacular. One sequence translated from a gene in the Universal Protein Resource, or UniProt, is selected as the 'canonical sequence', and variations to the base amino acid sequence are referred to as isoforms. However, this term fails to capture the complexity of highly post-translationally modified proteins that may also have base sequence changes. As different isoforms may be modified differently from each other, it is important to have language to differentiate the level at which one is speaking, analogous to the levels of protein higher order structure. The term 'proteoform' encapsulates the combinatorial combination of a set of modifications on a particular UniProt isoform (stably identified with a hyphen and then an integer, e.g. *-1 for the canonical*, *-2*, *-3* and so on) [8••]. The proteoform term includes all site specific features such as coding single nucleotide polymorphisms, mutations, or PTMs that map to the same gene. One isoform may have many different possible proteoforms. Note also that the UniProt KnowledgeBase is a gene-centric database, and, if used precisely with database search engines, can provide better clarity on the lingering issue of protein inference for bottom up; top down technology achieves gene-specific identification for proteins and thus has no such inference problem.

Mass spectrometry methods for proteomics: top down and bottom up

From the earliest days of proteomics (even before it was termed as such) two main types of mass spectrometric

analysis were performed. The primary method for protein identification is bottom-up, where peptides, generated from enzymatic proteolysis of proteins, are analyzed in a mass spectrometer [9,10]. To increase dynamic range, many groups have employed polyacrylamide gel electrophoresis (SDS-PAGE), either in one dimension, separating by molecular weight, or in two dimensions with a primary isoelectric focusing component. As excising proteins from a gel is labor intensive, many groups have preferentially turned to on-column separation techniques such as Multidimensional Protein Identification Technology (MudPIT) or other separation strategies [11,12]. Digestion of proteins requires the researcher to infer the identity of a protein from smaller peptides in a robust, relatively easy, and rapid fashion. Further analytical techniques have been based around this method to give quantification and identify modified proteins by class [13]. However, a major limitation of these enrichment protocols is their potential to alter observed stoichiometry. Rarely do the peptides detected provide information covering the entire protein because certain peptides may not be detected (particularly true for low abundance proteins). Finally, as with many scientific methods generating 'big data', researchers continue to optimize the most correct statistical methods of reporting identifications and false discovery rates [14–16].

To complement the speed and sensitivity of bottom-up proteomics, top-down proteomics introduces intact proteins into the mass spectrometer and then fragments whole protein ions directly [17•]. When the complete intact protein is present and measured at high mass accuracy, 100% sequence coverage is obtained and PTM combinations are preserved, leading to precise identification and characterization of specific genes, isoforms and proteoforms. However, due to inherent difficulties in both the separation and detection of intact proteins, there is low proteome coverage per injection compared with peptide-based analyses [18]. Also, the cost of mass spectrometers required to obtain high mass accuracy measurements is prohibitive to many groups. Moving forward, benchtop style instruments will bring this capability to more research groups than in past years [19–21]. With this and further development on high-throughput methods for intact proteins, the barriers to implementation of the top-down approach will drop substantially over the coming years [22,23••]. The full platform recently developed by the Kelleher lab combines all the elements discussed in the following sections to obtain high proteome coverage (Figure 1). For this reason, it will serve as the focus of this perspective, along with selected other methods discussed in the sections below.

A platform for top down proteomics on a high throughput basis

Mass-based fractionation of intact proteins

Once protein samples have been obtained from many different available methods, the next downstream step can be a mass-based separation. This approach allows the researcher to sequester proteins into similar ranges of molecular weight and apply a few adjustments to downstream analytical methods for low (>30 kDa), medium (30–70 kDa), and high (>70 kDa) mass proteins [24]. Many previous researchers had attempted to use mass-based separation for intact proteins, with limited success [3,25]. A special gel band elution device can be used, but few papers exist due to its low recovery of intact proteins [4].

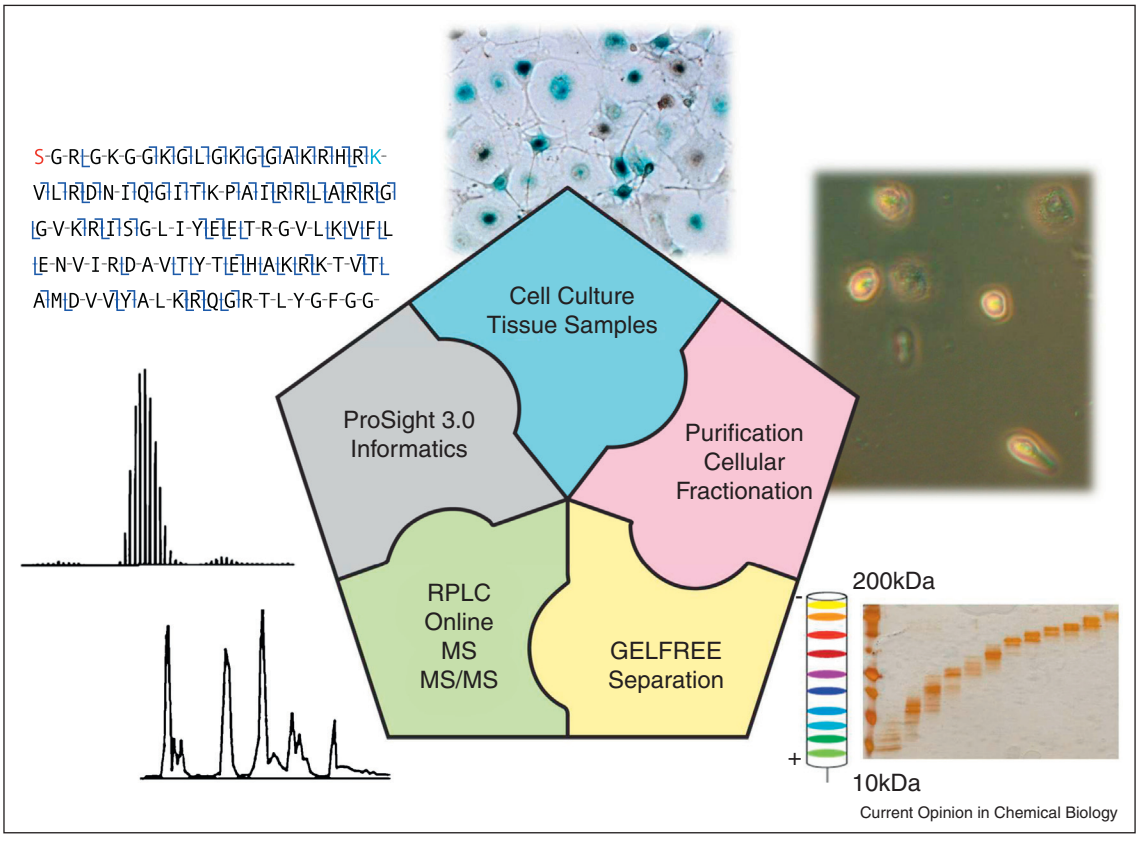
Tube gel electrophoresis overview and theory

Tube gel electrophoresis operates upon the same separation principles of SDS-PAGE gel electrophoresis; however, in the Gel Elution Liquid-based Fractionation Entrapment Electrophoresis (GELFrEE) device and other similar devices, proteins elute through the gel and into solution (Figure 2). Tube gel separation, therefore, gives higher sample recovery and is amenable to other separations either before or afterwards. Depending on the cross-sectional area of the separation tube, much greater sample amounts can be separated than in a single lane of a SDS-PAGE slab gel. Similar to gel electrophoresis, the separation can be optimized for an expected mass range by changing the degree of gel crosslinking. Each time-based fraction harvested correlates to a specific expected mass range which one may optimize with standard proteins and lysates for reproducible results [26–29]. Some highly hydrophobic proteins can be maintained in solution with surfactants present (even integral membrane proteins with up to ~8 transmembrane domains). GELFrEE allows the researcher to obtain protein fractions in a time-based manner, although the sample harvesting is currently manual [28•,30]. Since the publication of the initial paper in Analytical Chemistry, this technology has been commercialized as the GELFREE 8100 Fractionation System. Each particular sample may present unique challenges; yet the GELFrEE device allows many parameters to be optimized such as stacking gel length, loading amount, and collection time. Many different types of protein sample have been coupled to this separation platform due to the ease of use and its similarity to SDS-PAGE [26,31,32].

Reversed-phase liquid chromatography (RPLC) and online separations

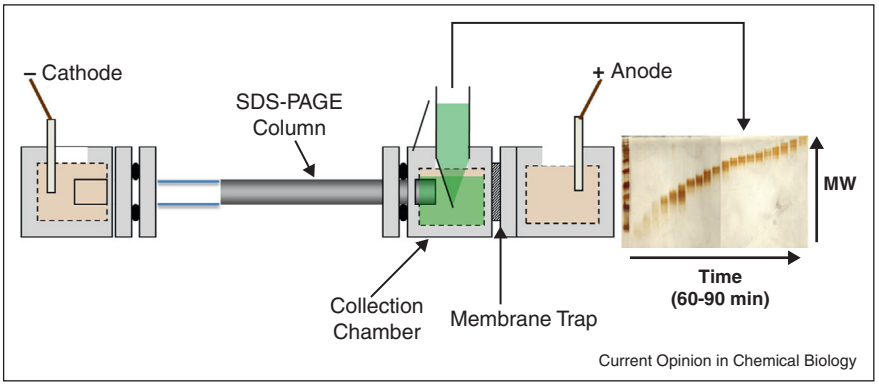
Liquid chromatography (LC) is among the most popular method of separation for peptides and intact proteins. Reverse phase liquid chromatography, RPLC, in particular is among the most common separation before mass spectrometry. This technique separates proteins based on hydrophobicity, with the most hydrophilic molecules eluting first. In large part due to the popularity of this technique, a wide range of materials are available and numbers are continuing to grow. In addition, even though challenges still exist for nanocapillary-based RPLC of whole proteins, many research groups are using this for

Figure 1



A schematic depiction of one work and data flow used for high-throughput top-down proteomics. Total protein content is quantified after cell purification or subcellular fractionation and loaded onto a GELFREE column (see Figure 2) for molecular weight-based separation. Protein fractions of increasing molecular weight are processed to remove SDS before injection onto reversed phase columns for online MS or MS/MS analysis. LC–MS/MS files are processed with ProSight and accompanying software for high-throughput protein identification and characterization.

Figure 2



Gel Eluted Liquid Fractionation Entrapment Electrophoresis (GELFrEE). The device consists of a resolving gel which can be cast analogous to an SDS-PAGE gel depending on the separation desired. Fractions are manually removed from the collection chamber in a time-based manner. A portion of each fraction can be removed and optionally visualized on a traditional SDS-PAGE gel (at right) to assess separation performance. This mass-based separation allows instrument parameters to be selected for acquisition of optimal top down MS and MS/MS data sets.

advantageous separations in sample-limited situations. Materials used include C4, C5, and C8 for smaller proteins, while polymeric media such as PLRP-S have been utilized to a much wider range [26,31,33,34•,35]. Fenselau and colleagues in particular have used C8 media for separating bacterial proteins up to 20 kDa [36]. Using PLRP-S as a portion of the platform Tran *et al.* identified 1043 proteins from the proteome of a human cell line [23••]. Patrie *et al.* have been using a superficially porous medium with success for HeLa cells and other endeavors [37•]. Other chromatographic media continues to be developed for top down proteomics (TDP) and will allow improvements to RPLC, particularly >70 kDa where chromatographic resolution tends to suffer for complex mixtures.

Alternate online separations

An option used by many researchers for additional separation power, as the separation is orthologous to RPLC, is isoelectric focusing (IEF). This separation occurs based on isoelectric point, along a pH gradient generated by small molecules called carrier ampholytes. Wherever the protein is placed along a pH gradient, as an electric field is applied, the protein will move toward the oppositely charged electrode until it has reaches an uncharged state. Proteins focus into sharp bands based on their individual isoelectric points. Several commercial IEF devices exist, but few examples of true high-throughput TDPs have been shown on these systems [38–42]. A custom IEF system has also been coupled to GELFrEE analysis, and though the system requires many manual steps, the high recovery and lack of conductive mixing are major assets [23••,38,43].

Capillary electrophoresis separates proteins according to their size to charge ratio, and can achieve vary narrow protein elution profiles (~10 s) [44,45]. Several groups have coupled this technique to hybrid mass analyzers for intact protein analysis [46–48]. In addition, many different groups are working on adapting separation techniques such as strong cation exchange or others that have proven effective at the peptide level to the intact protein level. For the time being, RPLC remains the default separation for intact proteins, coupled to mass spectrometry.

Mass analyzers for top down proteomics

To quickly and robustly characterize complex samples for TDP, even ones with multiple dimensions of separation, hybrid Fourier transform (FT) mass spectrometers have become the most common instruments. For TDPs, these instruments use a separate mass filter before high mass accuracy measurement in either an ICR cell or an Orbitrap. High mass accuracy is particularly vital for characterization of the human proteome, as it provides confidence to mass shifts that may be of biological or artifactual origin [35,49,50]. In addition, high mass accuracy allowed confident identification and classification of multiple bacterial species [36,51]. Even similar mass shifts, such as acetylation (42.0106 Da), may be differentiated from trimethylation (42.0470 Da) with high mass accuracy instruments (Figure 3).

Before intact protein mass spectrometry was readily accessible, researchers required customized instrumentation to challenge the limitations of protein analysis. These instruments have one or more mass filters before the ICR cell, making them hybrid instruments, and were mostly tailored for direct infusion of purified proteins or protein fractions. These instruments have been used with great effect for the analysis of purified single proteins. Han *et al.* showed fragmentation of a 200 kDa protein using a 6 Tesla FT-ICR instrument, and Valeja *et al.* achieved baseline resolution for a 148 kDa protein [52,53•]. Several commercial hybrid instruments make use of quadrupoles or ion traps before a FT-ICR or a FT-Orbitrap. The ion trap–Orbitrap pair, a hybrid arrangement from ThermoFisher, allows both mass filtering and fragmentation within the ion trap. The hybrid has been used for both bottom-up and top-down proteomic analyses to great effect as the scans may be performed in the ion trap for quicker scan speeds, or the FT for high resolution. These sophisticated instruments have become a bridge between the work of highly customized instrumentation labs and those labs newer to the field of TDP. On the basis of the Kingdon trap, this instrument has proven useful for intact protein analysis. Improved versions of the Orbitrap, such as the Orbitrap Elite, show strong capabilities for detection and fragmentation of proteins (even antibodies up to 160 kDa) [54]. Many researchers are employing the Orbitrap for both top down and bottom-up experiments, as all fragmentation modes can be achieved on an LC time scale. These include electron transfer dissociation, collision induced dissociation and Higher-energy C-trap dissociation (HCD) fragmentation. Expanding the capabilities of mass spectrometers to analyze larger intact proteins can lead to a very high level of characterization.

Data processing and informatics

The first software designed for top down mass spectrometry data was ProSightPTM, but since its initial commercialization as ProSightPC, other software has been developed. MascotTD (a.k.a. Big Mascot) supports analysis up to 110 kDa, while the standard Mascot supports up to 16 kDa. Similar to ProSight, MascotTD allows the identification of different proteoforms and isoforms, but does not address the database tailoring that best captures the value of data obtained during top down fragmentation. MSAlign+ is based on spectral alignment and allows the identification of unexpected PTMs with dynamic programming [55,56]. The Precursor Ion Independent Top-Down Algorithm, uses the fragmentation data to match a protein from a predicted gene [57]. With a

Figure 3

An overview of some mass analyzers used for top down proteomics analysis. From custom instruments (top left) that were used to pioneer the technique to hybrid FT-ICR systems (top right), FTMS continues to be the workhorse mass spectrometers for top down proteomics. Recently however, several research groups have shown the capability of the Orbitrap line as a promising option for many labs (bottom).

gene match made, the intact mass is used to map observed shifts from the gene-predicted mass.

TDP data processing typically uses data with high mass accuracy at both the intact and fragmentation scan level, so the software must take this into account during both analysis and scoring. The lengthy sequences of whole proteins, compared with 5–20 amino acid long peptides, can cause database sizes to grow exponentially and a larger number of fragments must be assessed for matches. With one of the enduring goals of TDP being characterization of PTMs, large combinations of PTMs can be annotated and considered within databases. Therefore, all of the considerations of true proteome level data (isoform and proteoform-resolved) can be considered. Much improved from its early days of single protein searching, tools such as ProSight have become a suite of software tools for processing this type of data and enabling the automated correlation of complex tandem MS data sets with multiply-modified proteoforms.

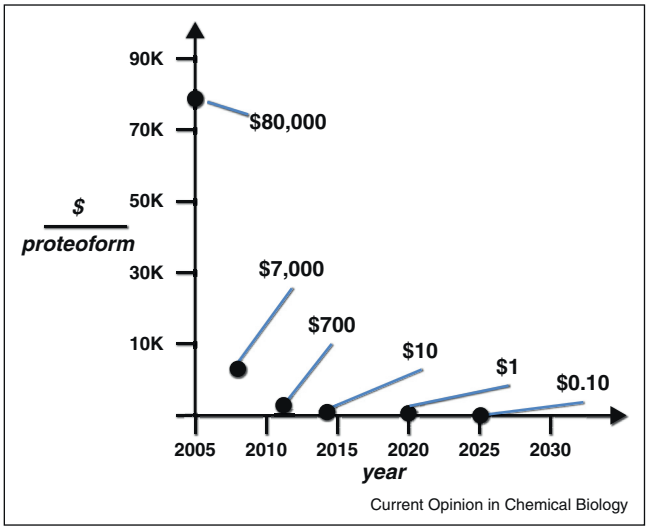
Conclusions

Each of the sections above outlines an area where researchers are advancing TDP. Concerted efforts have improved the ease of use, throughput, identification, and characterization power of TDP. Each stage is important to the overall platform and will no doubt be the subject of more work over the coming years. However, as more researchers turn to proteomics for precise answers and find relatively low value peptide lists (particularly in discovery/validation of protein biomarkers), top-down proteomics is poised to grow as a major, complementary method for the future. The protocols and work outlined above seek to expand proteomics based on whole proteins and enable efficient, accessible analyses based on highly confident and gene-specific data. This foundation will help expand top-down proteomics to complement information from other studies in translational and basic research.

Future prospects

TDP can now characterize hundreds of proteoforms per day, but use of the technique is underused and

Figure 4



A conceptual comparison of how proteomics measurements (in $/proteoform) might be reduced in the future from the same type of public/private sector construct used in the Human Genome Project. Cost per measurement can fall with investment and focused effort. Figures are approximate and projections are not based on advanced modeling (adapted from [8••]).

underdeveloped when compared with the multitude of laboratories and authors who contributed to the development of other methods [58,59]. Several areas of TDP are poised to alter dramatically in the following few years. Returning to targeted analysis of complex proteins via native or 'supercharged' ESI (e.g. for antibodies and high molecular weight therapeutic proteins), represents a great challenge that many researchers are motivated to solve [60–62]. Other research areas such as protein biopharmaceuticals, allergens, or toxins could benefit from full characterization. This could remove ambiguity, and replace concepts such as 'biosimilarity,' or methods resulting in large numbers of consumer or industry complaints [63–66]. Additional potential exists in combining top-down and middle-down (e.g. heavy and light chain fragments) approaches to obtain a great depth of identification and characterization of therapeutic antibodies.

To meet these needs, TDP methods will still need advancements. Coupling approaches such as subcellular fractionation methods are straightforward, with non-incremental advances such as electrospray supercharging offering disruptive possibilities for the future. Being able to meet the demands for intact measurement of samples limited by clinical availability will assist TDP in reaching more interested groups. Also, strong connections must be made between proteoforms detected and the mechanisms underlying complex molecular mechanisms and even human disease. In recent articles about the role of TDP in the Human Proteome Project, the potential to tie whole protein mass spectrometry to cataloguing human proteoforms in tissues, cell types, and fluids is rising [2,8••]. One paper calls for tying a disruptive reduction in the cost of proteomics to achieve a $1/proteoform price point (Figure 4), a >1000 fold increase in the efficiency of current analyses [8••].

Acknowledgements

The authors would like to thank their colleagues for their valuable input, as well as NIH Grant GM 067193 for the support.

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Aebersold R, Bader GD, Edwards AM, Van Eyk JE, Kussmann M, Qin J, Omenn GS: **The biology/disease-driven human proteome project (B/D-HPP): enabling protein research for the life sciences community**. *J Proteome Res* 2013, **12**:23-27.

2. Paik Y-K, Jeong S-K, Omenn GS, Uhlen M, Hanash S, Cho SY, Lee H-J, Na K, Choi E-Y, Yan F *et al.*: **The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome**. *Nat Biotechnol* 2012, **30**:221-223.

3. Du Y, Meng F, Patrie S, Miller L: **Improved molecular weight-based processing of intact proteins for interrogation by quadrupole-enhanced FT MS/MS**. *J Proteome Res* 2004, **3**:801-806.

4. Tomiyama S: **Experimental autoimmune labyrinthitis: assessment of molecular size of autoantigens in fractions of inner ear proteins eluted on the Mini Whole Gel Eluter**. *Acta Otolaryngol* 2002, **122**:692-697.

5. Garcia B, Pesavento J, Mizzen C, Kelleher N: **Pervasive combinatorial modification of histone H3 in human cells**. *Nat Methods* 2007, **4**:487-489.

6. Hake SB, Garcia BA, Duncan EM, Kauer M, Dellaire G, Shabanowitz J, Bazett-Jones DP, Allis CD, Hunt DF: **Expression patterns and post-translational modifications associated with mammalian histone H3 variants**. *J Biol Chem* 2006, **281**:559-568.

7. Jiang L, Smith JN, Anderson SL, Ma P, Mizzen CA, Kelleher NL: **Global assessment of combinatorial post-translational modification of core histones in yeast using contemporary mass spectrometry. LYS4 trimethylation correlates with degree of acetylation on the same H3 tail**. *J Biol Chem* 2007, **282**:27923-27934.

8. •• Kelleher NL: **A cell-based approach to the Human Proteome Project**. *J Am Soc Mass Spectrom* 2012 http://dx.doi.org/10.1007/s13361-012-0469-9.
This article provides insight into the basis for a cell-based version of the Human Proteome Project, and the milestones achievable with such a project to define cell and protein heterogeneity within the human body.

9. Aebersold R, Mann M: **Mass spectrometry-based proteomics**. *Nature* 2003, **422**:198-207.

10. Ho Y, Gruhler A, Heilbut A, Bader G, Moore L: **Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry**. *Nature* 2002, **415**:2-5.

11. Washburn MP, Wolters D, Yates JR: **Large-scale analysis of the yeast proteome by multidimensional protein identification technology**. *Nat Biotechnol* 2001, **19**:242-247.

12. Wolters DA, Washburn MP, Yates JR: **An automated multidimensional protein identification technology for shotgun proteomics**. *Anal Chem* 2001, **73**:5683-5690.

13. Schmidt SR, Schweikart F, Andersson ME: **Current methods for phosphoprotein isolation and enrichment**. *J Chromatogr B* 2007, **849**:154-162.

14. Gupta N, Bandeira N, Keich U, Pevzner PA: **Target-decoy approach and false discovery rate: when things may go wrong**. *J Am Soc Mass Spectrom* 2011, **22**:1111-1120.

15. Goloborodko AA, Gorshkov MV, Good DM, Zubarev RA: **Sequence scrambling in shotgun proteomics is negligible**. *J Am Soc Mass Spectrom* 2011, **22**:1121-1124.

16. Bell AW, Deutsch EW, Au CE, Kearney RE, Beavis R, Sechi S, Nilsson T, Bergeron JJM, Group HTSW: **A HUPO test sample study reveals common problems in mass spectrometry-based proteomics**. *Nat Methods* 2009, **6**:423-430.

17. • Kelleher NL: **Top-down proteomics**. *Anal Chem* 2004, **76**:197A-203A.
This review provides a basis for Top-down Proteomics as well as a primer on many of the overall concepts.

18. Compton PD, Zamdborg L, Thomas PM, Kelleher NL: **On the scalability and requirements of whole protein mass spectrometry**. *Anal Chem* 2011, **83**:6868-6874.

19. Scigelova M, Makarov A: **Orbitrap Mass Analyzer—overview and applications in proteomics**. *Proteomics* 2006, **6**:16-21.

20. Olsen JV, Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, Lange O, Remes P, Taylor D, Splendore M *et al.*: **A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed**. *Mol Cell Proteom* 2009, **8**:2759-2769.

21. Makarov A, Denisov E, Lange O: **Performance evaluation of a high-field Orbitrap mass analyzer**. *J Am Soc Mass Spectrom* 2009, **20**:1391-1396.

22. Garcia BA: **What does the future hold for Top Down mass spectrometry?** *J Am Soc Mass Spectrom* 2010, **21**:193-202.

23. •• Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M *et al.*: **Mapping intact protein isoforms in discovery mode using top-down proteomics**. *Nature* 2011, **480**:254-258.
This article highlights some major developments in TDP, using the overall platform to identify more than 1000 proteins and approximately three times more proteoforms.

24. Harris L, Churchward M, Butt R, Coorssen JR: **Assessing detection methods for gel-based proteomic analyses**. *J Proteome Res* 2007, **6**:1418-1425.

25. Singh R, Shasany a K, Aggarwal A, Sinha S, Sisodia BS, Khanuja SPS, Misra R: **Low molecular weight proteins of outer membrane of *Salmonella typhimurium* are immunogenic in Salmonella induced reactive arthritis revealed by proteomics**. *Clin Exp Immunol* 2007, **148**:486-493.

26. Lee JE, Kellie JF, Tran JC, Tipton JD, Catherman AD, Thomas HM, Ahlf DR, Durbin KR, Vellaichamy A, Ntai I *et al.*: **A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome**. *J Am Soc Mass Spectrom* 2009, **20**:2183-2191.

27. Kellie J, Catherman A, Durbin K, Tran J: **Robust analysis of the yeast proteome under 50 kDa by molecular-mass-based fractionation and top-down mass spectrometry**. *Anal Chem* 2012, **84**:209-215.

28. • Tran JC, Doucette A: **Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation**. *Anal Chem* 2008, **80**:1568-1573.
This article describes a custom mass-based separations device, now commercialized, which provided a critical piece for high-throughput TDP.

29. Kellie JF, Tran JC, Lee JE, Ahlf DR, Thomas HM, Ntai I, Catherman AD, Durbin KR, Zamdborg L, Vellaichamy A *et al.*: **The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput**. *Mol BioSyst* 2010, **6**:1532-1539.

30. Tran JC, Doucette A: **Multiplexed size separation of intact proteins in solution phase for mass spectrometry**. *Anal Chem* 2009, **81**:6201-6209.

31. Vellaichamy A, Tran JC, Catherman AD, Lee JE, Kellie JF, Sweet SMM, Zamdborg L, Thomas PM, Ahlf DR, Durbin KR: **Size-sorting combined with improved nanocapillary-LC–MS for identification of intact proteins up to 80 kDa**. *Anal Chem* 2010, **82**:1234-1244.

32. Botelho D, Wall M, Vieira D, Fitzsimmons S, Liu F, Doucette A: **Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separation**. *J Proteome Res* 2010, **9**:2863-2870.

33. Doucette A, Tran J, Wall M: **Intact proteome fractionation strategies compatible with mass spectrometry**. *Rev Proteomics* 2011:8.

34. • Tipton JD, Tran JC, Catherman AD, Ahlf DR, Durbin KR, Lee JE, Kellie JF, Kelleher NL, Hendrickson CL, Marshall AG: **Nano-LC FTICR tandem mass spectrometry for top-down proteomics: routine baseline unit mass resolution of whole cell lysate proteins up to 72 kDa**. *Anal Chem* 2012, **84**:2111-2117.
This article highlights capability for high-throughput TDP up to 72 kDa on a routine basis.

35. Boyne MMT, Garcia BBA, Li M, Zamdborg L, Wenger CD, Babai S, Kelleher NL: **Tandem mass spectrometry with ultrahigh mass accuracy clarifies peptide identification by database retrieval**. *J Proteome Res* 2009, **8**:374-379.

36. Wynne C, Edwards NJ, Fenselau C: **Phyloproteomic classification of unsequenced organisms by top-down identification of bacterial proteins using capLC–MS/MS on an Orbitrap**. *Proteomics* 2010, **10**:3631-3643.

37. • Roth MJ, Plymire DA, Chang AN, Kim J, Maresh EM, Larson SE, Patrie SM: **Sensitive and reproducible intact mass analysis of complex protein mixtures with superficially porous capillary reversed-phase liquid chromatography mass spectrometry**. *Anal Chem* 2011, **83**:9586-9592.
Superficially porous media is highlighted as an option for the separation of intact proteins. High reproducibility, sensitivity, and resolution were all achieved using this media to identify 343 proteins from HeLa cell lysates.

38. Tran JC, Wall MJ, Doucette AA: **Evaluation of a solution isoelectric focusing protocol as an alternative to ion exchange chromatography for charge-based proteome prefractionation**. *J Chromatogr B* 2009, **877**:807-813.

39. Ouvry-Patat SA, Torres MP, Quek H-H, Gelfand CA, O'Mullan P, Nissum M, Schroeder GK, Han J, Elliott M, Dryhurst D *et al.*: **Free-flow electrophoresis for top-down proteomics by Fourier transform ion cyclotron resonance mass spectrometry**. *Proteomics* 2008, **8**:2798-2808.

40. De Godoy LMF, Olsen JV, Cox J, Nielsen ML, Hubner NC, Fröhlich F, Walther TC, Mann M: **Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast**. *Nature* 2008, **455**:1251-1254.

41. O'Cualain RDM, Hyde JE, Sims PFG: **A protein-centric approach for the identification of folate enzymes from the malarial parasite, *Plasmodium falciparum*, using OFFGEL™ solution-based isoelectric focussing and mass spectrometry**. *Malaria J* 2010, **9**:286.

42. Scruggs SB, Reisdorph R, Armstrong ML, Warren CM, Reisdorph N, Solaro RJ, Buttrick PM: **A novel, in-solution separation of endogenous cardiac sarcomeric proteins and identification of distinct charged variants of regulatory light chain**. *Mol Cell Proteom* 2010, **9**:1804-1818.

43. Tran JC, Doucette AA: **Rapid and effective focusing in a carrier ampholyte solution isoelectric focusing system: a proteome prefractionation tool**. *J Proteome Res* 2008, **7**:1761-1766.

44. Li Y, Champion MM, Sun L, Champion PAD, Wojcik R, Dovichi NJ: **Capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry as an alternative proteomics platform to ultraperformance liquid chromatography–electrospray ionization-tandem mass spectrometry for samples of intermediate complexity**. *Anal Chem* 2012, **84**:1617-1622.

45. Sun L, Knierman MD, Zhu G, Dovichi NJ: **Fast top-down intact protein characterization with capillary zone electrophoresis–electrospray ionization tandem mass spectrometry**. *Anal Chem* 2013, **85**:5989-5995.

46. Haselberg R, Brinks V, Hawe A, De Jong GJ, Somsen GW: **Capillary electrophoresis–mass spectrometry using noncovalently coated capillaries for the analysis of biopharmaceuticals**. *Anal Bioanal Chem* 2011, **400**:295-303.

47. Haselberg R, De Jong GJ, Somsen GW: **Capillary electrophoresis-mass spectrometry for the analysis of intact proteins**. *J Chromatogr A* 2007, **1159**:81-109.

48. Taichrib A, Pelzing M, Pellegrino C, Rossi M, Neusüss C: **High resolution TOF MS coupled to CE for the analysis of isotopically resolved intact proteins**. *J Proteomics* 2011, **74**:958-966.

49. Mann M, Kelleher NL: **Precision proteomics: the case for high resolution and high mass accuracy**. *Proc Natl Acad Sci U S A* 2008, **105**:18132-18138.

50. Gnad F, De Godoy LMF, Cox J, Neuhauser N, Ren S, Olsen JV, Mann M: **High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast**. *Proteomics* 2009, **9**:4642-4652.

51. Demirev Pa, Fenselau C: **Mass spectrometry for rapid characterization of microorganisms**. *Annu Rev Anal Chem (Palo Alto, Calif.)* 2008, **1**:71-93.

52. Han X, Jin M, Breuker K, McLafferty FW: **Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons**. *Science* 2006, **314**:109-112.

53. • Valeja SG, Kaiser NK, Xian F, Hendrickson CL, Rouse JC, Marshall AG: **Unit mass baseline resolution for an intact 148 kDa therapeutic by mass spectrometry**. *Anal Chem* 2011, **83**:8391-8395.

This article demonstrates the capabilities of high resolution mass spectrometry to achieve baseline resolution of isotopic distributions even at high mass.

54. Mao Y, Valeja SG, Rouse JC, Hendrickson CL, Marshall AG: **Top-down structural analysis of an intact monoclonal antibody by electron capture dissociation-Fourier transform ion cyclotron resonance-mass spectrometry**. *Anal Chem* 2013, **85**:4239-4246.

55. Liu X, Sirotkin Y, Shen Y, Anderson G: **Protein identification using top-down spectra**. *Mol Cell Proteomics* 2012, **11**:1-34.

56. Ansong C, Wu S, Meng D, Liu X, Brewer HM, Deatherage Kaiser BL, Nakayasu ES, Cort JR, Pevzner P, Smith RD *et al.*: **Top-down proteomics reveals a unique protein S-thiolation switch in *Salmonella typhimurium* in response to infection-like conditions**. *Proc Natl Acad Sci U S A* 2013, **110**:10153-10158.

57. Tsai YS, Scherl A, Shaw JL, MacKay CL, Shaffer SA, Langridge-Smith PRR, Goodlett DR: **Precursor ion independent algorithm for top-down shotgun proteomics**. *J Am Soc Mass Spectrom* 2009, **20**:2154-2166.

58. Mallick P, Kuster B: **Proteomics: a pragmatic perspective**. *Nat Biotechnol* 2010, **28**:695-709.

59. Schirle M, Bantscheff M, Kuster B: **Mass spectrometry-based proteomics in preclinical drug discovery**. *Chem Biol* 2012, **19**:72-84.

60. Ayaz-Guner S, Zhang J, Li L, Walker JW, Ge Y: **In vivo phosphorylation site mapping in mouse cardiac troponin I by high resolution top-down electron capture dissociation mass spectrometry: Ser22/23 are the only sites basally phosphorylated**. *Biochemistry* 2009, **48**:8161-8170.

61. Yin S, Loo JA: **Top-down mass spectrometry of supercharged native protein–ligand complexes**. *Int J Mass Spectrom* 2011, **300**:118-122.

62. Shaw JB, Li W, Holden DD, Zhang Y, Griep-Raming J, Fellers RT, Early BP, Thomas PM, Kelleher NL, Brodbelt JS: **Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation**. *J Am Chem Soc* 2013 http://dx.doi.org/10.1021/ja4029654.

63. Ahmed I, Kaspar B, Sharma U: **Biosimilars: impact of biologic product life cycle and European experience on the regulatory trajectory in the United States**. *Clin Ther* 2012, **34**:400-419.

64. Sherwood S, Rush D: **Defining cellular senescence in IMR-90 cells: a flow cytometric analysis**. *Proc Natl Acad Sci U S A* 1988, **85**:9086-9090.

65. Hebling CM, McFarland MA, Callahan JH, Ross MM: **Global proteomic screening of protein allergens and advanced glycation endproducts in thermally processed peanuts**. *J Agric Food Chem* 2012 http://dx.doi.org/10.1021/jf303554t.

66. Savaryn JP, Catherman AD, Thomas PM, Abecassis MM, Kelleher NL: **The emergence of top-down proteomics in clinical research**. *Genome Med* 2013, **5**:53.