Published in final edited form as:
Anal Chem. 2018 January 02; 90(1): 110–127. doi:10.1021/acs.analchem.7b04747.

Top-down Proteomics: Ready for Prime Time?

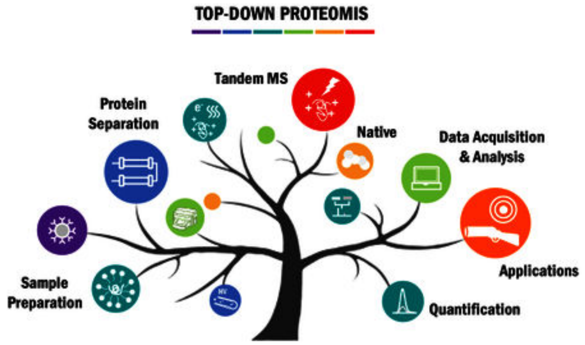
Bifan Chen[1], **Kyle A. Brown**[1], **Ziqing Lin**[2,3], and **Ying Ge**[1,2,3,*]

[1]Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States

[2]Department of Cell and Regenerative Biology, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States

[3]Human Proteomics Program, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States

TOC



INTRODUCTION

In the post-genomics era, the study of proteins is critical for understanding cellular functions at the molecular level.[1–5] Beyond the genetic code, the human proteome is greatly diversified by various genetic variations, alternative splicing of RNA transcripts, and post-translational modifications (PTMs).[6,7] In 2013, the term "proteoform" was designated to describe "all of the different molecular forms in which the protein product of a single gene can be found",[6] clearing up the confusion in nomenclature and joining research efforts to develop methodologies for proteoform characterization. Top-down proteomics, which analyzes intact proteins without digestion, has proved to be a premier technology for global and comprehensive analysis of proteoforms.[4,8,9] The top-down approach retains intact protein mass information, providing a "bird's-eye" view of the proteome and allowing for identification of novel proteoforms, in-depth sequence characterization, and quantification of disease-associated PTMs.[4,8,9]

[*]Corresponding Author: ying.ge@wisc.edu.

The authors declare no competing financial interest.

Although some technical challenges remain, development over the past five years has expanded top-down proteomics from a mostly targeted approach to one capable of proteoform-profiling across multiple samples.[9] Now, thousands of proteoforms can be identified, characterized, and quantified using the high throughput top-down proteomics approach.[10,11] Moreover, developments in instrumentation and fragmentation have positioned top-down proteomics in the fast lane for future progression.[12] With the successful development of commercial high-resolution mass spectrometers such as the solariX XR Fourier transform ion cyclotron resonance (FT-ICR) (Bruker),[13] Orbitraps (Thermo) and quadrupole time-of-flight (Q-TOFs) (Bruker and Waters), excellent instruments are now widely available in academic and industrial labs for top-down proteomics. Moreover, the 21 Tesla FT-ICR mass spectrometers equipped with state-of-the-art fragmentation features at the Pacific Northwest National Laboratory and National High Magnetic Field Laboratory have demonstrated unprecedented resolving power, acquisition rate, and flexible tandem MS (MS/MS) capabilities, providing enormous potential for top-down proteomics practitioners to probe complicated proteomics applications.[14,15] Reciprocally, the methodological and technological gains from developing large-scale and high-throughput workflow have further empowered targeted analysis, from which top-down proteomics roots and thrives. As a result, interest in top-down MS has grown considerably and many studies have already underscored the potential of top-down proteomics for unraveling disease mechanisms and discovering novel biomarkers.[4,8,16,17]

Clearly, top-down proteomics has gained remarkable space in the proteomics landscape over the last few years. It is no longer a specialized method, and has become a solid, established technique in the proteomics field. Recently, the growing top-down proteomics community has gained momentum through the creation of the Consortium for Top-down Proteomics (http://www.topdownproteomics.org/).

A number of recent reviews have already given an overview of the technical requirements for top-down proteomics and delineate the history and fundamentals of the field as well as its application to biomedical research.[4,8,9,18–20] With a special emphasis on publications in the past two years (2015–2017), this review examines recent technological trends and developments in the areas of intact protein sample preparation, separation, MS/MS, acquisition strategies, data analysis, native MS, and quantitation from the perspectives of the authors. We also highlight recent applications for global and targeted top-down proteomics applications, and conclude with outlooks on the field.

SAMPLE PREPARATION STRATEGIES

Often overlooked, sample preparation remains one of the most important and challenging aspects in top-down proteomics. Although MS is a sensitive analytical technique, isotope and charge state distributions of protein ions produced by electrospray ionization (ESI), spreads the signal of a single species over a large *m/z* range. Thus, signal suppression from salt adducts (i.e. Na^+, K^+), detergents, or even coexisting protein species can greatly hamper a top-down experiment. In the section below, common methods for extracting intact proteins, replacing/removing buffer components incompatible with MS, and approaches to decrease sample complexity and to enrich low-abundance proteins will be discussed.

Traditionally, physical methods for lysing samples (e.g. homogenization and sonication), are performed using a mixture of Good's buffers, salts, reducing agents, and protease and phosphatase inhibitors to extract cellular components while avoiding proteoform alteration or degradation.[4,21] However, these conditions necessitate post-extraction work-up to remove or replace non-volatile salts that suppress MS signal by forming adducts to protein ions and increase the chemical noise.[22,23] Molecular-weight cutoff (MWCO) filters provide an easy method for exchanging protein samples into MS compatible conditions such as volatile ammonium salt buffers or low concentration, acidic solutions for downstream analysis.[24] Additionally, reversed-phase chromatography (RPC), a regular component in top-down proteomics workflows, desalts samples prior to the MS analysis.[25,26] Although proper desalting is generally critical for reliable MS data, recent developments in small emitter tips have pushed the boundaries of salt inclusion, making MS detection of protein ions more tolerant of non-volatile salt contamination, even with conditions mimicking physiological conditions (25 mM Tris, 150 mM KCl).[27]

Detergents are common buffer additives that facilitate cell permeabilization and aid in hydrophobic membrane proteins solublization.[28–30] In particular, anionic detergents such as sodium dodecyl sulfate (SDS) exhibit excellent protein solubility; however, they pose a challenge for downstream MS analysis by causing complete signal suppression at levels as low as 0.01%.[31] Protein precipitation, usually with chloroform/methanol or cold acetone, removes detergents and other MS incompatible contaminants.[32–34] These methods are straightforward and add little cost to the experiment, but unfortunately, loss of proteins, difficulty in re-solubilizing the protein pellet, and incomplete removal of contaminants result in low reproducibility in the down-stream analysis. In addition, hydrophobic membrane proteins generally require high levels of formic acid to dissolve the protein pellet, which can result in artificial protein formylation.[33,35] Although alternative filtration devices have been employed to deplete and remove SDS, non-specific binding and/or incomplete removal limits their widespread utility.[36,37] Recent advancements by Kachuk *et al.* demonstrated near-complete front-end removal of SDS using a transmembrane electrophoresis.[38] Similarly an online matrix removal device developed by Kim *et al.* showed a 2- to 10-fold increase in signal after just 5 min.[38,39] Although anionic detergent removal using cleavable linkers (i.e. acid labile detergents) between the hydrophobic tail and hydrophilic head have shown promise in bottom-up workflows, these methods have yet to be translated into top-down proteomics.[40] On the other hand, nonionic detergents, such as n-Dodecyl β-D-Maltoside (DDM), are also commonly used to solubilize protein, and generally exhibit lower signal suppression.[24,41,42] Often times, these mild, non-denaturing detergents do not perturb protein-protein interactions, making them an indispensable tools in the area of native top-down MS.[41,43] Technological advancements in detergent removal as well as alternative extraction strategies (i.e. organic sovlents[44,45]) have made membrane proteomics an intriguing area for further investigation and development.

Given the complexity and dynamic range of the proteome, enrichment strategies represent an important area of development. Organelle fractionation by differential centrifugation is widely used to isolate and enrich specific sub-proteomes.[46] Nuclear, mitochondrial, membrane, and cytosolic fractions can be collected for global protein analysis and deep proteoform coverage of organelle specific targets.[47] One common example is histone proteoform analysis, which has recently gained importance for its role in understanding epigenetic regulation.[48]

In other cases, low abundant proteoforms (e.g. phosphoproteins) are enriched by taking advantage of their unique physicochemical properties. Recently the phosphoproteomics field has benefited from technologies such as functionalized superparamagnetic nanoparticles and microspheres that enrich phosphoproteins for downstream top-down proteomics characterization.[49–52] In 2017, Chen *et al.* demonstrated a strategy coupling functionalized cobalt ferrite nanoparticles to targeted liquid chromatography (LC)-MS/MS for phosphoprotein characterization (Figure 1a).[52] This method demonstrated a significant enrichment of spiked-in phosphorylated β-casein from 0.5% to 94% in a complex tissue lysate, showing the high specificity and efficiency of the functionalized nanoparticles. More importantly, low-abundance endogenous phosphoproteins from a complex tissue lysate were greatly enriched, which allowed for confident identification and phosphorylation site localization (Figure 1b-d).[52] Although overwhelmed by the signals of co-eluting non-phosphorylated proteins prior to enrichment, the phosphorylated ions became dominant after enrichment for subsequent MS/MS experiments.[52] Therefore, this strategy serves as a promising tool for probing novel phosphorylation sites for a more comprehensive understanding of their role in biological systems. Additionally, when targeted protein and protein complexes are of particular interest, co-immunoprecipitation or affinity purification are commonly used for comprehensive proteoform coverage and quantitation as demonstrated by a number of studies.[16,17,53–58]

INTACT PROTEIN SEPARATION

The complexity of the proteome requires the fractionation of intact proteins prior to top-down MS analysis.[59] The under-developed front-end separation of intact proteins has long been a major obstacle for further advancing top-down methodology.[8,60] Thus, recently, significant efforts have been dedicated to the development of chromatographic and electrophoretic separation strategies to address intact protein separation challenges.[60–69] Numerically, the intact proteome appears to be a much simpler mixture than its corresponding peptide digests.[59] In practice, however, protein-level fractionation and separation are daunting tasks due to diverse physicochemical properties (i.e. size, charge, and hydrophobicity) and the wide dynamic range of the proteome.[8,59] In this section, we will discuss recent developments in chromatographic methods, especially in denaturing RPC, non-denaturing ion exchange chromatography (IEX), and hydrophobic interaction chromatography (HIC) which can be directly coupled to MS, as well as electrophoretic methods and multidimensional separations strategies.

RPC is the most prevalent method for online LC-MS analysis of intact protein. Unfortunately, because of complex physicochemical properties of protein mixtures and poor recovery of intact proteins, the peak capacity and resolution of intact protein separation largely lag behind that of peptide separation in RPC.[59,60] Depending on the column and gradient length, the peak capacity obtained in the one-dimensional RPC separation of intact proteins is usually below 100.[62] Recently, Shen *et al.* reported the use of long RPC columns (>1 m) and long gradient time (>600 min) with short alkyl (C1–C4) bonded phases to achieve peak capacities greater than 400 (similar to peptide level RPC separation) for complex intact protein mixture up to 50 kDa.[68] They concluded that increasing column length results in the most dramatic improvement in their intact protein separation. Interestingly, they also notice minimal influence of pore size between 200 Å to 400 Å,[68] although many others found 1000 Å generally provides more accessible pore volume for intact protein.[62] Using a capillary column 4 cm long packed with sub-micron particles (0.47 μm; C18 bonded phase), Wu *et al.* demonstrated a slip-flow[65] nanocapillary RPC separation that has peak capacities up to 750 with *E. coli* lysate in a 60 min gradient.[70] However, both methods were conducted under high pressure, 14k psi for the long column approach,[68] and 9k psi for the slip-flow approach.[70] The requirement of ultrahigh pressure LC systems has limited the accessibility to these methods. Alternatively, monolithic columns have shown great promise in intact protein separations,[71] and offer several advantages such as high permeability, low backpressure, fast mass transfer, and better recovery.[72] For example, Simone *et al.* developed **γ**-ray-induced polymethacrylate-based monolithic capillary columns that can provide peak capacity over 1000 in a 240 min gradient with the addition of 0.1% TFA and column heating at 60 °C.[73] Owing to the low backpressure, analysis time can be as short as five min using a higher flow rate, while still offering a peak capacity of 190.[73] Recent developments in longer columns, reduced stationary phase particle size, and monolithic materials have all shown potential to improve intact protein RPC.

Other than RPC, online LC-MS with non-denaturing conditions coupled to top-down analysis was also recently advanced to address pharmaceutical needs to characterize protein aggregates and conjugates.[67,74–76] Online size exclusion chromatography (SEC)-MS,[74] ion exchange chromatography (IEX)-MS,[75,76] hydrophobic interaction chromatography (HIC)-MS,[67] were developed with analytes detected by top-down native-ESI MS yielding lower charge states. Muneeruddin *et al.* recently demonstrated the used of IEX with an increasing concentration of salt gradient (ammonium acetate) coupled online to MS and MS/MS to obtain meaningful structural information on protein conjugates, such as PEGylated and glycosylated interferon β−1a.[75,76] In contrast to the increasing salt gradient in IEX, a decreasing salt gradient (e.g. ammonium sulfate) is often used for HIC. As a non-denaturing separation method, HIC offers high sensitivity to the hydrophobic surface of proteins. While recent work has showed that ammonium tartrate does not interfere with downstream MS analysis after a quick desalting step,[64] direct online LC-MS analysis with HIC was not realized until the use of more-hydrophobic HIC materials and ammonium acetate, as demonstrated by Chen *et al.*[67] The HIC-MS remains sensitive to minor modifications and holds a potential for analyzing antibody-drug conjugates in an online manner.[67] The SEC-MS, IEX-MS, and HIC-MS methods share the common feature of utilizing non-denaturing conditions in the form of volatile salts with proper selection of the stationary phase to maintain the efficacy of the chromatographic mode, preserving the noncovalent interactions. Thus, the ability of these native separation methods to couple online with MS will lead to new avenues for the analysis of intact complexes.

Although LC methods are preferred for direct coupling with MS, capillary electrophoresis techniques, including capillary zone electrophoresis (CZE) and capillary isoelectric focusing (CIEF), have proved to be useful alternatives.[77,78] Recent technical advancement of the sheath-flow and sheathless interface have encouraged more effective use of CZE coupled with top-down MS.[63,79–83] Compared to traditional LC techniques, CZE offers several advantages in sensitivity, resolution, and speed.[84] However, the small sample loading amount have limited the number of total identifications from a protein mixture by CZE alone.[80]

While different chromatographic and electrophoretic methods effectively target a subset of the proteome based on specific physicochemical properties, no single separation strategy can comprehensively resolve the intact proteome. Therefore, a multidimensional separation strategy combining orthogonal modes is highly desirable and necessary to achieve a deeper proteome coverage. After implementing RPC as a pre-fractionation step and optimizing fragmentation, Zhao *et al.* identified 180 proteins and 580 proteoforms from yeast in their CZE-MS/MS study in 2016.[85] Similarly, by adding an offline high pH RPC fractionation step, Wang *et al.* doubled the identifications (163 proteins and 328 proteoforms) when compared to the 1D low pH RPC-MS/MS experiment from *E. coli*.[86] The power of additional dimensions for deep proteome coverage was further showcased by the three-dimensional LC-MS approach incorporating offline IEX (first dimension), HIC (second dimension), and online RPC (third dimension) demonstrated by Valeja *et al.*[66] A total of 640 proteins were identified from a single fraction from IEX (out of 35) as compared to 47 in the 2D approach. Moreover, only minimal improvement in protein separation and identifications with simply prolonging the gradients in RPC in the 2D approach. Therefore, this 3DLC (IEX-HIC-RPC) method greatly outperformed the conventional 2DLC (IEX-RPC) approach demonstrating the great potential for effective separation of intact proteins to achieve deep proteome coverage in top-down proteomics.[66]

Despite the improved resolving power of these multidimensional separation methods, high molecular weight (MW) proteins were not detected with these methods due to a variety of challenges associated with MS analysis of large proteins.[87] First, low MW proteins interfere with the detection of high MW proteins in the mass spectrometer. Second, high MW proteins inherently exhibit lower MS signal due to wider distributions of isotopes and charge states.[87] To address these challenges, practitioners of top-down MS have turned to size-based fractionation prior to MS analysis to alleviate spectral interference from low MW proteins with high MW proteins. This has been well demonstrated in the use of gel-eluted liquid fraction entrapment electrophoresis (GELFrEE).[61,62,88,89] Although effective recovery of intact protein from polyacrylamide gels has been challenging, recent attempt by Takemori *et al.* shows progresses to utilize dissolvable gel for top-down analysis.[90]

Recently, without a protein precipitation step, Cai *et al.* developed a MS-compatible serial-SEC (sSEC) strategy that enabled detection of proteins up to 223 kDa using a high-resolution Q-TOF mass spectrometer (Figure 2a-b).[69] By employing a two-dimensional sSEC-RPC approach, more than 4000 unique proteoforms were detected with a 15-fold increase of the proteins above 60 kDa, which greatly outperformed one-dimensional RPC, especially on high MW proteins.[69] Importantly, many of these high MW proteins undetected in one-dimensional RPC, possessed important PTMs (e.g. phosphorylation) (Figure 2c). Therefore, size-based fractionation prior to other dimensions of separation will continue to play an important role in improving coverage of the intact proteome. Despite its effectiveness, the large amount of sample requirement and labor-intensive offline sample handling steps have limited MDLC in practice. Although online 2DLC for top-down proteomics has been developed for histone separation,[91] wide adaptation requires further improvement to the robustness of online MDLC hardware and the proper selection of orthogonal separation methods.

TOP-DOWN TANDEM MS

One of the key merits in top-down proteomics is the prospect of comprehensively interrogating all PTMs within the whole protein sequence from a “bird’s-eye” view.[4,8] Efficient and extensive MS/MS fragmentation techniques are the essential prerequisites to characterize proteoforms. Energy-based dissociation processes such as collision-induced dissociation (CID), also referred to as collisionally activated dissociation (CAD); and higher-energy collisional dissociation (HCD),[92] remain the most robust methods and have enabled many large-scale and high-throughput top-down proteomics studies.[88,89,93] These types of energy transfer processes predominantly cleave C-N amide bonds of the protein backbone, leading to abundant *b*- and *y*-type product ions. Generally, the tendency of cleaving labile PTMs (e.g. phosphorylation) has limited the utility of CID in PTM localization and structural examination. However, intact protein ions in the gas phase appear less susceptible to cleavage of labile PTMs during CID or HCD, possibly due to the high-order structure.[94]

In contrast, electron-based methods such as electron transfer dissociation (ETD) and electron capture dissociation (ECD) generally preserve labile PTMs, producing primarily *c*- and $z^\bullet$-type ions subsequent to the cleavage of N-C_a bonds.[95,96] In particular, targeted top-down MS with ECD and ETD represents a powerful method to comprehensively characterize biologically relevant proteins, especially those with labile PTMs, as demonstrated on multiple cardiac and skeletal muscle proteins and others.[97–101] For instance, Peng *et al.* recently identified and characterized a previously unknown phosphoprotein known as enigma homolog isoform 2 from Z-disc by ECD, opening up future functional investigation of this novel phosphoprotein.[97] To address the need of pharmaceutical analysis, characterization of intact or reduced heterogeneous immunoglobulins G (IgG) with electron-based top-down MS/MS techniques continues to improve.[102–105] Benefited from the advancement of the high-resolution MS instrument, Fornelli *et al.* recently showed top-down analysis of intact IgGs by ETD, achieving comparable ~30% sequence coverage with 2-fold less transient summed and 6-fold less acquisition time when compared to the previous study.[102,106]

Moreover, significant progress has made in developing hybrid dissociation methods such as electron-transfer dissociation/higher-energy collisional dissociation (EThcD)[107] and activated ion electron transfer dissociation (AI-ETD),[108,109] with improved fragmentation for intact proteins. Brunner *et al.* demonstrated that with the proper setting of the collisional energy, EThcD and electron-transfer dissociation/collision-induced dissociation (ETciD) increased the appearance of *b*-/*y*- type ions, while retaining *c*-/ $z^\bullet$- type ions that contain phosphorylation site information.[107] Overall, EThcD and ETciD provide more extensive fragmentation than CID/HCD/ETD alone over a broad range of charge states, and retain labile phosphorylation, making these methods suitable for top-down analysis. While AI-

ECD methods have been developed and yielded improved ECD product ions, it is largely limited to FT-ICR instruments.[110] By using infrared photon bombardment simultaneously with ETD reaction, Riley *et al.* showcased the improvement of AI-ETD on intact proteins characterization over HCD and ETD, particularly on low charge state precursor ions, which extends the utility of electron-based fragmentation in virtually any trap-type instruments in the top-down proteomics field.[108]

Another MS/MS method that recently comes into the spotlight is 193 nm ultra-violet photodissociation (UVPD) pioneered by the Brodbelt group.[12,90,111–116] The high energy deposition of the photons on intact proteins results in a wide array of fragment ions: in addition to *b*-/*y*- and *c*-/ $z^{\bullet}$- product ions, *a*-/*x*- from the cleavage of C_a-C bonds and even *d*-,*v*-, and *w*- ions from secondary fragmentation are produced.[12] The rich fragmentation along the protein backbone and the preservation of labile modifications lead to a promising utility of UVPD in pinpointing mutation in sequence variants,[111] unnatural amino acid incorporation,[112] and sites of modification[115] or ligand binding[113] on intact proteins. Because UVPD has minimal dependency on precursor charge states, it not only dissociates low charge state complexes produced by native mass spectrometry into subunits for probing structural topology, but also efficiently fragments monomeric subunits for secondary structural characterization (Figure 3).[90] For streptavidin tetramers, in contrast to CID, UVPD led to symmetric charge portioning pathways, providing insights into subunit organization and higher order structure (Figure 3a-f).[90] Interestingly, while the resulting monomeric units carried similar charges to that during CID, it remained folded during UVPD as demonstrated by the relative fragment abundances (Figure 3g-i).[90] Other than targeted analysis of a single protein, large-scale proteoform profiling that compares UVPD to HCD, was recently demonstrated.[116] The results showed that while HCD still provided almost twice as much proteoform identification on a chromatographic time scale, UVPD enable more confident proteoform characterization, indicating the complementary nature of UVPD and HCD.[116] Furthermore, proton transfer reaction (PTR)[117] that focuses multiple charge states of intact proteins into a single lower charge state can be coupled with UVPD as demonstrated by Holden *et al.*[114] The combined ion signals and the relatively efficient UVPD fragmentation on lower charge state ions, allow for higher-quality tandem mass spectra.[114] However, applying this strategy in LC/MS analysis remains challenging because of the time requirement to perform PTR and UVPD in a sequential manner.

As an alternative, front-end ETD, an electrical discharge-based reagent ion source that produces stable and abundant ETD reagent ions, provides flexibilities for ion-ion reactions in the later stage for intact protein characterizaiton.[118] Using sequential front-end ETD and PTR to disperse overlapping fragmentation ions over a broader *m/z* range, Anderson *et al.* showed sequence overage up to 81% for intact histone proteoforms in online LC-MS experiments.[119]

To better implement these fragmentation techniques in high-throughput top-down proteomics on chromatographic time scales, higher quality tandem mass spectra with improved signal-to-noise (*S/N*) and mass accuracy in a single scan is necessary. Under denaturing mode, intact proteins are often highly charged, and the number of charges generally increases proportionally to the molecular weight of the protein.[87] However, a trapping type cell can only accommodate a fixed number of charges during one scanning event. This implies that highly charged ions of large intact proteins or exceedingly large number of fragment ions, will result in limited number of detectable ions, and consequently lower *S/N*. Accumulation of multiple scans can mitigate the space charge effect, but the trade-off is the decrease in duty cycle during online LC-MS/MS analysis. Recognizing this compromise, Riley *et al*. presented a high capacity design of the precursor ion storage in the high pressure cell on the Orbitrap Fusion Lumos.[120] This new scheme allows nearly 3-fold more precursor charges to be accumulated (up to 1,000,000 charges or more), which ultimately leads to *S/N* improvement without sacrificing spectral acquisition speed during ETD experiments for intact proteins.[120] Instead of increasing cell capacity of storing more ions, Holden *et al*. reported a method of overfilling the ion trap with precursor ions and subsequently implementing resonance ejection of the undissociated precursor ions after UVPD, which resulted in improved *S/N*, higher resolution, and lower mass error for the fragment ions.[121]

Various MS/MS techniques have been developed and are steadily evolving to achieve improved identification and characterization of proteoforms at higher acquisition rate and a higher *S/N*. Now, nearly complete fragmentation at single residue level of a 30 kDa intact protein has been demonstrated.[15,111] We envision that the arsenal of top-down MS/MS methods continues to expand in the next few years, particularly on a chromatographic time scale with higher *S/N*.

TOP-DOWN DATA ACQUISITION STRATEGIES

Top-down proteomics in early years directly adopted highly developed bottom-up data acquisition strategies implemented in the commercial instruments. In general, most top-down proteomics experiments are conducted in TopN data-dependent acquisition (DDA), during which the most intense precursor ions from the MS scan are isolated in a relatively large *m/z* window (up to 15 *m/z*) for MS/MS.[89] However, data acquisition is uniquely challenging in top-down experiments. In this section, we will explain some special considerations on this subject and further discuss the strategies developed to address these.

First, ESI of intact proteins generates a wide range of charge states, and precursors of abundant proteoforms in different charge states are selected and fragmented multiple times, often yielding redundant information. Because of this, fragmentation of co-eluting and low-abundance proteoforms might be hindered or missed entirely. Autopilot, an online control system that features online spectral deconvolution, project-wise exclusion list, intelligent MS/MS (HCD and ETD) based on past fragmentation events, combination of product ions, and real-time restricted database searching, has provided deeper and higher-confident identifications.[20] Nonetheless, in their study, DDA still provided 50 unique protein identifications that were not found during Autopilot acquisition.[20] Therefore, recently introduced and more advanced DDA that can select only the most abundant charge states while putting other redundant charge states on an exclusion list by deconvoluting and assigning charge states on-the-fly further improves intact protein identification and characterization.[48]

Second, although large isolation window (e.g. 15 *m/z*) allows for efficient precursor isolation for identification, it might hamper the precise characterization of individual proteoforms. Fornelli *et al.*, recently demonstrated the use of 3 *m/z* isolation window to improve specificity as well as to alleviate space charge effect, and identifed 439 proteoforms in the 30–60 kDa range from human fibroblasts.[10] Using a small isolation window, 0.6 *m/z*, Zhou *et al.* unraveled the complexity of histone PTMs in an online fashion and identified novel modifications including tyrosine bromination and glutathionylation in histone subunits from mouse brain combining a state-of-art high-resolution instrument and a new bioinformatics workflow.[48] With advances on mass spectrometers, narrow and efficient isolation will continue to be improved.

Finally, since many top-down studies rely heavily on Fourier transform trap-type instruments (Orbitrap and FT-ICR), the balance among ion accumulation, acquisition time, *S/N*, and mass accuracy during a scanning event critically influences the data quality. Fornelli *et al.* demonstrated improved top-down analysis of higher molecular species (up to 60 kDa) with a benchtop quadrupole-orbitrap.[10] With a shorter transient in the MS^1 scan to eliminate the accumulation of unnecessary noise before the second isotopic beat,[122] enhanced *S/N* of large protein ions were achieved while maintaining high accuracy of the average masses (e.g. 2.5 ppm for a 41 kDa protein).[10] Leveraging the high resolving power, mass accuracy, sensitivity, and spectral acquisition rate offered by the 21 Tesla FT-ICR mass spectrometer, Anderson *et al.* recently demonstrated a large-scale top-down LC-MS/MS study on human colorectal cancer cells.[123] Instead of summing multiple transients over a longer period, the installation of a multiple storage device (MSD, external quadrupole ion trap) between the Velos ion trap and the ICR cell enables multiple fills of precursor ions or fragment ions in the MSD before high-resolution detection in the ICR cell.[123,124] This strategy not only improves *S/N*, but also boosts the duty cycle in a chromatographic time scale. This study identified 684 unique protein entries and over 3200 proteoforms in 40 LC-MS/MS runs by front-end ETD and CID, including 372 isotopically resolved proteoforms with MW above 30 kDa.[123]

TOP-DOWN DATA ANALYSIS

Many fragment ions (in multiple charge states) are produced by top-down MS/MS, resulting in incredibly complex spectra with overlapping charge states. This, combined with the presence of unexpected and combinatorial PTMs imposes a tremendous computational challenge for top-down data analysis. Generally, the basic analysis of top-down proteomics data consists of isotopic peak picking and deconvolution from the spectra, identification and characterization through database search, validation, and visualization. Over the years, various groups have contributed to each step, pushing the bioinformatics boundaries in top-down proteomics.[125–127] New software packages such as MASH Suite Pro[128], Informed-Proteomics[11], and ProteinGoggle 2.0,[129] as well as new search algorithms such as pTop[130], TopPIC (from MS-Align+),[131] among others,[132,133] were spotlighted in the past two years.

The resulting complexity of intact protein (MS^1 and MS^2) spectra produced by ESI, necessitates deconvolution algorithms to combine all charge states and determine monoisotopic masses. The first algorithm for deconvolution of high-resolution mass spectra, THRASH, uses a subtractive peak finding method with charge determination and a monoisotopic mass determining strategy based on the averagine model.[134] This influential method is still widely used through adaptation and improvement, which have led to the development of Decon2LS,[135] DeconMSn,[136] Xtract,[137] MASH Suite,[138] as well as MS-Deconv,[125,126] an alternative combinatorial algorithm. In 2015, Marty *et al.* proposed UniDec, a Bayesian framework to separate mass and charge dimensions for fast deconvolution of ion mobility MS data, which could also be potentially applied to top-down data analysis.[139] Last year, Sun *et al.* introduced a novel algorithm, pTop with a pre-processing module pParseTD, which greatly improves the accuracy of precursor detection.[130] For the human histone data set, pTop recalled 22% more correct precursors than Xtract with 30 % less exported masses.[130] The improvement was facilitated by a machine-learning module, pParseTD, which evaluates and characterizes candidate isotopic clusters through features obtained from experimental isotopic distribution, LC profiles, and the same protein ion in different charge states. Similarly, taking advantage of the isotopic clusters in different charge states and LC-MS profiles, ProMex in the recently published Informed-Proteomics package, generates a list of deconvoluted isotopic masses from LC-MS data by aggregating signals from different charge states over a LC elution time span and examining the aggregated isotopomer envelope (Figure 4a).[11] Compared to other MS^1 feature-detection algorithm, such as ICR-2LS (http://omics.pnl.gov/software/icr-2ls) and MS-Deconv, ProMex accurately and reproducibly detected a significantly higher number of LC-MS features across all ten replicates of ovarian tumor sample, which has led to comparable coefficients of variation to label-free bottom-up studies.[11]

After deconvolution, a database search strategy is commonly used for protein identification and proteoform characterization. Given the likelihood of observing combinatorial or even unexpected PTMs and proteolytic truncations in a top-down workflow, strategies such as extended database search (e.g. ProteinGoggle,[129,140] ProSightPC absolute mass search), blind PTM search (e.g. MS-Align+[141], PIITA[142]) or the combination of the two (ProSightPC biomarker search mode, MS-Align-E[143]) are mainly applied to interrogate proteoforms.[144] Among these, ProSight PC and MS-Align+ (recently renamed to TopPIC[131] with improvements) remain to be two of the most popular software tools.

The recent pTop uses a sequence-tag based strategy and a scoring approach to greatly accelerate the database searching process, with the introduction of indexes of both sequence tags and combinatorial modifications, leading to 10- to 100-fold decrease in time required than MS-Align+ on complex data sets.[130] Nevertheless, truncated proteoforms are not yet considered during searches in the initial version of pTop. On the other hand, TopMG, a mass graph-based approach, enables the identification of proteoforms with terminal truncations, although it requires prolonged running time.[145] With dramatically faster running speed and higher numbers of identified proteins compared to ProsightPC (V3.0) and MSAlign+, MSPathFinder, a command-line algorithm, plays an important role in the recent Informed-Proteomics package (Figure 4b-d).[11] It utilizes a graph-based approach and a *de novo* sequencing algorithm based on sequence tags to improve search efficiency.[11] However, MSPathFinder requires pre-specified PTMs, and therefore does not discover unknown PTMs.

In 2016, MASH Suite Pro was introduced as a comprehensive package for large-scale top-down proteomics, which offers deconvolution (THRASH and MS-Deconv), identifications (MS-Align+), in-house developed quantitation and characterization of proteoforms (Figure 4e-i).[128] It greatly simplifies and speeds up the interpretation of high-resolution MS and MS/MS data with a user friendly and customizable interface.[128] Furthermore, with very few other software capable of manual validation, MASH Suite Pro provides useful visualization tool to correct mis-assignment of charge state and isotopic distributions, facilitating accurate and reliable proteoform characterization. In this regard, LcMsSpectator in the Informed-Proteomics package also provides an interactive results viewer to simplify the inspection of MS^1 precursor match and MS^2 fragment match.[11,48]

Statistical significance of the identification and characterization of proteoforms is evaluated by a number of parameters. *p*-value and *E*-value measures how good the fragmentation data matches the identified sequence (the lower the value, the less likely the match is random by chance), although computation of these values slightly varies in different software.[141] In general, a false-discovery-rate of 1% governs the confidence level of the identification in large-scale proteoform studies with a target/decoy approach.[10,89] Furthermore, introduction of C-score has improved identification and characterization of proteoforms, particularly in a high throughput fashion.[146] However, it is not suitable for unknown modifications.[126] Recently, also based on Bayesian models, MIScore is described by Kou *et al.* for proteoform characterization. MIScore provides a simple and fast method to identify and localize up to two unknown modifications in proteoforms.[126] Ongoing development and adaptation of these scoring validations among laboratories will improve workflows and provide more consistency within the community.

Over the years, a plethora of informatics tools for top-down proteomics have been developed which greatly streamlined the data analysis process. Although challenges remain (e.g. identification of larger proteins, utilization of internal fragments, etc.), [69,147] collaboration based on many open-source platforms provides tremendous opportunities for top-down bioinformatics to further advance.

NATIVE TOP-DOWN PROTEOMICS

The improvement of non-denaturing separation, fragmentation techniques, as well as instrument advance on transferring and detecting high *m/z* ions have brought native mass spectrometry and top-down proteomics together. In addition to the native separation methods described previously (e.g. SEC-MS, IEX-MS, and HIC-MS),[67,74–76] clear native GELFrEE (CN-GELFrEE) and native CZE were recently developed for separation of noncovalent assemblies of biomolecules.[148,149] Through a native and multistage (e.g. MS^3) MS approach, the protein complexes, monomer units, and monomer fragments can be readily detected from the fractions.[150] To facilitate identification and characterization of protein assemblies, a computational framework with scoring metrics was developed by Skinner *et al* (Figure 5a).[151] By utilizing CN-GELFrEE and the newly developed informatic framework, Melani *et al.* were able to access the large protein complexes from the king cobra venom proteome. [152] They identified two of the largest venom glycoprotein complexes, homodimeric $_{L}$-amino acid oxidase (~130 kDa) with different number of N-glycan moieties and the multichain toxin cobra venom factor (~147 kDa) by HCD (Figure 5b-e).[152] More recently, Belov *et al*. demonstrated a proof-of-principle use of native sheathless CZE coupled online to MS and MS/MS (HCD) to analyze standard protein complexes, monoclonal antibody aggregation, and *E. coli* ribosomal extract at femtomole levels.[149] With the identification of several non-covalent protein-protein and protein-metal ion ribosomal protein complexes, this approach shows potential to characterize protein interactions in complex biological samples with low sample consumption and high sensitivity.[149] Furthermore, the recent investigation of gas phase fragmentation propensity of HCD during denature and native top-down MS of by Haverland *et al*., will provide insights for developing tailored scoring metrics to improve identification of native complexes in the future.[153]

Other than identifying high MW protein assemblies from a mixture, targeted analysis of a single native complex by various fragmentation techniques can reveal non-covalent binding regions and other structural information. [10,154–157] Using a 15T FT-ICR mass spectrometer, Li *et al*. demonstrated that ECD of a 158 kDa aldolase tetramer under native conditions provides rich fragments in the region that ECD and CAD failed to access to under denatured conditions.[154] Nonetheless, depending on the specific structure of the protein complex, ECD sometimes provides minimal fragmentation information. Native top-down with 30 eV electron ionization dissociation (EID), as an alternative, preserves ligand-protein or protein-protein interactions and delivers much more structural information compared to ECD in certain cases, as demonstrated by a purified apo-superoxide dismutase dimmer (Figure 5f-g).[10] Providing comparable fragmentation to UVPD, EID has proved to be a useful technique as it requires no additional instrument modification in FT-ICR instruments. Similarly using a targeted native top-down approach, in 2017, Schneeberger and Breuker successfully probed the binding sites of a model transactivation response RNA to a trans-activating peptides at the single–residue level and captured time-resolved information on the stoichiometry of the complex using CAD.[155] Other than using electron and collisional energy to dissociate protein complexes and fragment monomeric units, UVPD, as previously mentioned, can also provide structural characterization for protein complexes by retaining noncovalent interaction upon dissociation.[12] For instance, UVPD on native dihydrofolate reductase complexes enabled the assessment of the binding regions of cofactor NADPH and inhibitor methotrexate.[113] Besides these fragmentation techniques, surface-induced dissociation (SID) also offers insights on structural topology with symmetric charge partitioning.[156]

The native top-down approach has expanded the toolbox and will continue to enable the study of larger complexes in the proteome beyond the normal detection limit of denaturing top-down proteomics. With the further development in non-denaturing separation strategies, we expect online analysis and characterization of native complexes to become an increasingly important area in top-down proteomics.

QUANTITATIVE TOP-DOWN PROTEOMICS

To further interrogate how proteoforms are associated with complex disease phenotypes, the acquirement of quantitative measurement in addition to qualitative information becomes critical. Over the recent years, top-down quantitation has expanded from examination of

PTM changes of a few targeted proteoforms to investigation of expression changes in large-scale proteoform studies.[93]

Within the same top-down mass or tandem mass spectra, relative quantification of unmodified and modified proteoforms using intensity ratios was first demonstrated feasible by Pesavento *et al.*, and used in many cardiac biomarker studies and others.[53,55,98,158,159] This method is extremely powerful to globally examine histones as shown in recent studies using high-resolution MS, providing extensive elucidation and relative quantification of histone variants and modifications in different samples.[160–162] The relative quantification of proteoforms abundance relies on the idea that the physiochemical properties of intact proteins in the gas phase during electrospray ionization is less affected by small modifications or variations than that of peptides,[158] however, it is worth noting that different modifications might have an impact on ionization, fragmentation, and therefore the accuracy of the quantification.[163,164]

In the realm of quantifying proteoform changes across multiple samples in large-scale biomarker discovery experiments, measuring relative ratio of proteoforms has limited applicability. Incorporation of labeling techniques such as stable isotope labeling by amino acids (SILAC),[165] isobaric and pseudoisobaric tags,[166–168] as well as NeuCode SILAC,[169] have shown some potential. For instance, the intact-mass strategy developed by the Smith group utilizes NeuCode SILAC to determine lysine count for the elucidation of proteoform families.[170,171] However, variation in labeling efficiency and reproducibility might lead to spreading (lower *S/N*) and asymmetry of protein ion signals.[172] In addition, dependence on precursor ion isolation, instrument resolution, and specialized data analysis software, have limited the adaptation of these labeling approaches.[173] To overcome some of these limitations, in 2016, Quijada *et al*. introduced a cost-effective partial metabolic labeling strategy coupling to top-down proteomics for living organisms, tunable intact mass protein mass increases (TIPMI), by spiking ^{13}C-glucose or ^{2}H-water into laboratory feedstocks.[174] Compared to SILAC, this method results in a more symmetric and predictable mass shift for heavy and light peak pairs, as the percentage of ^{13}C (or ^{2}H) incorporation in a protein deviates less than the percentage of lysine or arginine.[174] Although this labeling approach provides outstanding precision and accuracy, potential overlapping of the unlabeled proteoform peaks and the heavy peaks, and software requirement for picking labeled pairs still need to be addressed for large-scale quantitation. Therefore, the more accessible label-free techniques remain appealing in top-down quantitation proteomics.

The advantages of label-free quantitation such as simplicity and throughput has been demonstrated previously.[175,176] Recently, Ntai *et al*. applied and extended this idea to a larger scale platform with hierarchical linear model for statistical analysis, which quantified over 800 proteoforms with more than 100 significantly different between two yeast samples.[177] Using this quantitation pipeline and the Autopilot acquisition strategy, recently Durbin *et al*. achieved identification and quantification of thousands of proteoforms under 30 kDa in two human fibroblasts states.[93] In addition, the new Informed-Proteomics package improves quantification with the enhanced detection of LC-MS features by ProMex and identification with MSPathFinder, resulting in a significant increase in the number of differentially expressed proteoforms detected when compared to a previous study on the same samples.

[11,178] In short, these top-down label-free strategies and others all generally include the following steps: automatic isotopic peak picking, integrating, normalization, stringent proteoform binning and filtering, and statistical analysis.[11,177,179] Another simple label-free quantitation method, spectrum counting, was recently evaluated by Geis-Asteggiante *et al.* by comparing to other label-free methods (e.g. AUC).[173] Although spectrum counting offers comparable (slightly less) accuracy, it is a simple and robust preliminary screening method for putative differentially expressed targets.[173] Moreover, to ensure the quality of the label-free quantitative experiment, large sample size, and multiple biological and technical replicates are often necessary.

It is important to recognize that quantification is most accurate only within the linear response as implied in many quantitation works.[173,175,177] Protein signals in MS that fall close to or out of the upper and lower ends of the curve will result in over- or under-estimation of the true expression changes. Unfortunately, for intact proteins, the dynamic range and the instrument response greatly vary among instrument types, and often times, sensitivity differs for each protein.[180–182] Nevertheless, with advancement of instrumentation and optimization of workflows, top-down label-free quantitation holds great potential in translational and clinical research.

APPLICATIONS

Top-down proteomics is uniquely equipped to identify and characterize proteoforms without prior knowledge. Recently it has been widely applied in discovery research especially to study biological systems that have complex PTM combinations and large number of variants, for instance, histones and venom proteins.[152,183,184] More importantly, utilizing clinical samples (e.g. blood, tissues, cerebrospinal fluid, saliva), an increasing number of top-down applications have shown promise in the fields of cancer, neurodegenerative, and cardiovascular diseases recently.[56,97,178,179,185–193] Owing to the potential for providing a precise and comprehensive view of all proteoforms, top-down proteomics is thought to be more directly connected to complex disease phenotypes.[19] Thus, this powerful method is particularly valuable in understanding disease mechanisms and discovering potential biomarkers as well as therapeutic targets.[4,20] Here, we highlight a few recent and relevant examples in global profiling and targeted analysis of proteoforms in this fast growing field to demonstrate the value of studying proteoforms in biological and biomedical research.

Global profiling.

As previously discussed, improved sample prefractionation methods, LC-MS/MS workflow, and label-free quantitation of top-down proteomics have led to the ability to profile samples from healthy to diseased states.[93] Global top-down proteomics in discovery mode allows for the examination of potential biomarkers at the proteoform level. Cheon *et al.* quantitatively compared the low-MW proteome (<30 kDa) of human plasma from healthy control and colorectal cancer patients.[186] In addition to identifying proteoforms containing various PTMs (e.g. acetylation, S-glutathionylation, and *O*-glycosylations, etc), they detected previously-unreported coexisting pyroglutamylation at Gln24 and S-cysteinylation at Cys29 on Apolipoprotein A-II (8804.48 Da).[186] By using label-free quantification, they have identified 17 low MW proteoforms that showed >1.5-fold changes from the colorectal cancer samples, some of which were verified by western blot analysis and were consistent with the literature.[186]

Recently the Kelleher group comparatively analyzed peripheral blood mononuclear cells from kidney and liver transplant recipients in two different studies.[187,188] For the study on kidney transplant recipients, the top-down label-free approach with GELFrEE fractionation and nanocapillary LC-MS/MS identified 2905 proteoforms corresponding to 344 proteins and preliminarily suggested changes between the two patient groups with transplant excellent and with acute rejection (Figure 6a).[187] For the more recent liver transplant studies, among the differentially expressed proteoforms, those associated with chemokine/cytokine signaling and cytoskeletal regulation appeared most significant.[188] For instance, platelet factor 4 (PF4/CXCL4) were found more abundant in transplant excellent than acute rejection samples, which have shown potential protective effect by facilitating blockage of Th17 differentiation in other transplantation models.[188] Importantly, out of the 15 proteoforms of PF4/CXCL4, resulting mainly from N-terminal proteolytic processing, only 3 were differed with statistical significance (Figure 6b-f).[188] These results indicate the possibility that only specific proteoforms are related to certain disease phenotypes, and the label-free top-down analysis at the proteoform level provides an indispensable tool for investigating this hypothesis.

Other than clinical blood samples, studies of tumor tissues from patient-derived mouse xenograft models of basal and luminal B human breast cancer, WHIM2 and WHIM16, were also recently demonstrated by two independent top-down studies.[11,178] Ntai *et al.* quantified 982 proteoforms from 358 proteins (<30 kDa), and differentiate PTM changes such as phosphorylation level alternation of α-endosulfine between WHIM2 and WHIM16 samples using label-free top-down quantitation.[178] Although the bottom-up approach on the same samples yielded eight-times more identifications, it overlooked PTM changes and relative expression of heterozygous alleles in the proteoform level.[178] Using the same samples but with improved top-down workflow, Park *et al.* demonstrated the detection of over 7000 differentially expressed proteoforms.[11] Within the 3207 identified proteoforms, 1636 proteoforms were found with adjusted P value < 0.01 and fold change > 2 in two subtypes of breast type tumor samples.[11] These differentially expressed entities were found about 10 times more than what Ntai *et al.* reported with much shorter instrument acquisition time.[11,178] As much as top-down proteomics has advanced in global proteoform profiling and quantification in clinical research, these potential biomarkers found in blood and tumor tissue samples require further follow-up targeted analysis and validation, as many other global proteomics experiments do.

Targeted analysis.

While large-scale global proteoform profiling in discovery mode is powerful, the hypothesis-driven targeted proteomics approach that detects and quantifies specific proteoforms from a sub-proteome or a purified system with high sensitivity and reproducibility has often offered insights in the understanding of underlying molecular mechanisms of normal and disease-associated cellular events, and provided potential utility in clinical diagnostics.[4,194–196] For instance, in addition to the detection of novel truncated and glycosylated proteoform, Gafvels *et al*. revealed that levels of glycosylated apolipoprotein A-I (ApoA-1) distinguished significantly from the serum sample of the diabetic to non-diabetic patients using a semi-quantitative top-down LC-MS approach, suggesting glycated ApoA-1 may sever as a potential biomarker for diabetes.[197]

To improve selectivity and specificity of the proteoform characterization, affinity purification is often coupled with top-down analysis, as demonstrated by the Ge group and others.[16,17,53–58] Elucidation of potential molecular mechanisms related to proteoforms greatly benefits from this immunoaffinity top-down MS method. Recently, Carel *et al*. used a top-down targeted approach with other techniques to decipher the role of o-mycoloylation in targeting the outer membrane proteins (OMPs) to the mycomembrane in bacteria (Figure 7).[198] Subsequent to the expression and affinity purification of recombinant OMPs, PorA, ProH, PorB, and Por C of bacteria *C. glutamicum* (Figure 7a), top-down LC-MS/MS analysis of the OMPs associated with mycoloyl-arabinogalactan-peptidoglycan (mAGP) complex and secreted in the extracellular medium revealed well-conserved PTMs (Figure 7b-f), including *O*-mycoloylatoin, pyroglutaminatoin, and N-formylation.[198] Among these modifications, in particular, *O*-mycoloylatoin was only found in the mAGP-associated proteoforms, indicating that the presence of mycoloyl residues is essential for targeting OMPs to the mycolic acid-containing lipid bilayer.[198] The top-down approach plays a particularly important role in this case, because these OMPs from *C. glutamicum* are relatively hydrophobic with the presence of PTMs containing C32–C36 mycolyl residues and lack arginine and lysine residues for generating adequate tryptic peptides. Similarly, in another microbiology study on *M. tuberculosis*, using affinity purification and top-down MS, Parra *et al*. uncovered a complex repertoire of nearly 130 proteoforms of virulence associated 19 kDa lipoglycoprotein antigen (LpqH) resulting from various combination of lipidation, glycosylation, and proteolytic truncations.[199] In addition to the examination of molecular diversity, the identification of a novel phosphorylation on unprocessed LpqH further provided insights to microbial biogenesis.

As an example of the application in translational research, applying relative quantification in a targeted approach, Gregorich *et al*. studied sarcopenia (muscle loss with aging) by assessing age-related changes in the myosin regulatory light chain (RLC) proteoforms from the fast-twitch skeletal muscles of rats (Figure 8a-d).[200] RLC is a critical protein involved in the modulation of muscle contractility, and their results revealed a significant progressive decline in the RLC phosphorylation with increasing age.[200] Top-down MS/MS identified a previously unreported bis-phosphorylated proteoform of fast skeletal RLC, and further localized the sites of decreasing phosphorylation to Ser14/15 in addition to a N-terminal trimethylation.[200] Subsequent mechanical analysis of the single muscle fiber from rats of different ages revealed that the decrease in RLC phosphorylation is responsible for altered mechanical function in aged muscle fibers, such as significant decreases in maximal force, the Ca^{2+}-sensitivity of force, loaded shortening velocity, and power output (Figure 8e-j).[200] Suggesting RLC phosphorylation as a therapeutic target of sarcopenia muscle dysfunction, this study and others demonstrate the great potential of top-down targeted proteomics for quantifying proteoform changes across biological samples and elucidating disease pathology at a molecular level. [97,200,201]

CONCLUDING REMARKS

Top-down proteomics has gained tremendous momentum in the past few years especially after the establishment of Consortium for Top-down Proteomics (http://www.topdownproteomics.org/). Instrument developments and technical advancement in the areas discussed above (sample preparation, separation, tandem MS, data acquisition and analysis, quantitation), have equipped the top-down approach with wider dynamic range, higher sensitivity, and better PTM characterization to elucidate proteoforms in a deeper fashion than ever before. Thousands of proteoforms can now be identified, quantified, and characterized in a top-down study.[10,11,93] Nevertheless, efforts still needed to achieve a complete coverage of all proteoforms.

Despite the significant advancement in top-down hardware and software, some on-going challenges remain. For example, solubilization of hydrophobic proteins, fractionation or separation of proteins with good recovery, detection and characterization of high MW and low-abundance proteins, localization of (labile) PTMs in an accurate and high-throughput manner during LC-MS/MS among others, remain to be further addressed by novel methodologies. In parallel with the technical advancement, we anticipate accelerated and much-needed development on data analysis and bioinformatics tools to meet the growing demands of top-down practitioners, and to fully capitalize on the state-of-the-art instruments and top-down MS workflows.

Of the recent trends, many technological developments have focused on pushing the limit of large-scale proteoform profiling. Indeed, quantitative global profiling can result in specific proteoform targets and pathways that can potentially be linked to disease phenotypes in translational research. However, targeted analysis remains powerful and necessary, as it often leads to untangling of underlying molecular mechanisms of disease-associated cellular events. We expect a renaissance and resurgence of the targeted approach to focus on molecular specificity. Recent attempts to use multiple reaction monitoring to quantify intact proteins or proteoforms exemplify the need for further development of top-down targeted analysis.[180,181]

Is top-down proteomics ready for prime time? The significant advances in the field in all the areas discussed above have shown how well the community is marching towards this goal. With the development of high-resolution instrumentation and the maturation of the techniques, the age when top-down proteomics was specialized in a few laboratories, has long passed. Centered around proteoforms, top-down proteomics attracts numerous mass spectrometrists and biomedical researchers alike with its power to decipher the complex proteoforms and provide a holistic view on the molecular level, which might better reflect disease phenotypes. Nevertheless, the prime time comes with both opportunities and challenges. As the Consortium for Top-Down Proteomics continues to provide the fostering ground for education, collaboration, and development of comprehensive analysis of proteoforms, the field continues to progress and grow. We optimistically envision that more and more laboratories will adopt, practice, and advance top-down proteomics to answer important biological questions and decipher underlying disease mechanisms.

ACKNOWLEDGEMENT

The authors acknowledge support by National Institute of Health (NIH) R01 Grants (GM117058, HL109810, HL096971) and S10 OD018475. The authors also thank Ruixiang Sun, and Zhijie Wu for helpful discussions, and Trisha Tucholski and Zachery Gregorich for critical reading of the manuscript.

Biographies

Bifan Chen is a chemistry Ph.D. candidate in the research group of Professor Ying Ge at the University of Wisconsin-Madison. He received his B.A. degree in Chemistry and Asian Studies from St. Olaf College in 2014, and he anticipates his Ph.D. in Analytical Chemistry in 2019. His research focuses on the development of new methodologies for top-down proteomics including separation, phosphoprotein enrichment, and online tandem MS workflows.

Kyle A. Brown is a graduate student in the Department of Chemistry at the University of Wisconsin-Madison. In 2015, He earned his B.S. degree in Chemistry and B.A. in Mathematics from the University of North Carolina at Chapel Hill. His current research, under the mentorship of Professor Ying Ge, focuses on the development of top-down methodologies for membrane protein characterization.

Ziqing Lin is a postdoctoral research associate in Professor Ying Ge's research group at University of Wisconsin-Madison. He received his B.S. and M.S. in Chemistry from Tsinghua University, and acquired his Ph.D. in Biomedical Engineering at Purdue University. His research focuses on quantitative top-down proteomics, PTM characterization, and instrumentation. He is also in charge of user projects for the Human Proteomics Program in School of Medicine and Public Health.

Ying Ge is an Associate Professor of Chemistry and Cell and Regenerative Biology. Ge earned her B.S. degree at Peking (Beijing) University, and received her Ph.D. at Cornell University under the guidance of Professor Fred McLafferty in 2002. Her current research program focuses on developing novel strategies to address the challenges in top-down proteomics and understanding the molecular mechanisms underlying heart failure and cardiac regeneration via systems biology approaches.

REFERENCES

(1). Yates JR; Ruse CI; Nakorchevsky A Annu. Rev. Biomed. Eng 2009, 11, 49–79. [PubMed: 19400705]

(2). Pandey A; Mann M Nature 2000, 405, 837–846. [PubMed: 10866210]

(3). Altelaar AFM; Munoz J; Heck AJ R. Nat. Rev. Genet 2013, 14, 35–48.

(4). Gregorich ZR; Ge Y Proteomics 2014, 14, 1195–1210. [PubMed: 24723472]

(5). Aebersold R; Mann M Nature 2016, 537, 347–355. [PubMed: 27629641]

(6). Smith LM; Kelleher NL; Down C. f. T.; Proteomics. Nat. Methods 2013, 10, 186–187. [PubMed: 23443629]

(7). Patrie SM; Roth MJ; Zhang JM Proteomic and Metabolomic Approaches to Biomarker Discovery 2013, 313–332.

(8). Cai WX; Tucholski TM; Gregorich ZR; Ge Y Expert Rev. Proteomics 2016, 13, 717–730. [PubMed: 27448560]

(9). Toby TK; Fornelli L; Kelleher NL Annu. Rev. Anal. Chem 2016, 9, 499–519.

(10). Fornelli L; Durbin KR; Fellers RT; Early BP; Greer JB; LeDuc RD; Compton PD; Kelleher NL J. Proteome Res 2017, 16, 609–618. [PubMed: 28152595]

(11). Park J; Piehowski PD; Wilkins C; Zhou M; Mendoza J; Fujimoto GM; Gibbons BC; Shaw JB; Shen Y; Shukla AK; Moore RJ; Liu T; Petyuk VA; Tolic N; Pasa-Tolic L; Smith RD; Payne SH; Kim S Nat. Methods 2017, 14, 909–914. [PubMed: 28783154]

(12). Brodbelt JS Anal. Chem 2016, 88, 30–51. [PubMed: 26630359]

(13). Nikolaev EN; Boldin IA; Jertz R; Baykut GJ Am. Soc. Mass. Spectrom 2011, 22, 1125–1133.

(14). Shaw JB; Lin TY; Leach FE; Tolmachev AV; Tolic N; Robinson EW; Koppenaal DW; Pasa-Tolic LJ Am. Soc. Mass. Spectrom 2016, 27, 1929–1936.

(15). Weisbrod CR; Kaiser NK; Syka JEP; Early L; Mullen C; Dunyach JJ; English AM; Anderson LC; Blakney GT; Shabanowitz J; Hendrickson CL; Marshall AG; Hunt DF J. Am. Soc. Mass. Spectrom 2017, doi: 10.1007/s13361-13017-11702-13363.

(16). Ge Y; Rybakova IN; Xu QG; Moss RL Proc. Natl. Acad. Sci. U. S. A 2009, 106, 12658–12663. [PubMed: 19541641]

(17). Chamot-Rooke J; Mikaty G; Malosse C; Soyer M; Dumont A; Gault J; Imhaus AF; Martin P; Trellet M; Clary G; Chafey P; Camoin L; Nilges M; Nassif X; Dumenil G Science 2011, 331, 778–782. [PubMed: 21311024]

(18). Zhou H; Ning Z; Starr AE; Abu-Farha M; Figeys D Anal. Chem 2012, 84, 720–734. [PubMed: 22047528]

(19). Savaryn JP; Catherman AD; Thomas PM; Abecassis MM; Kelleher NL Genome Med 2013, 5, 53. [PubMed: 23806018]

(20). Durbin KR; Fellers RT; Ntai I; Kelleher NL; Compton PD Anal. Chem 2014, 86, 1485–1492. [PubMed: 24400813]

(21). Cai W; Tucholski T; Chen B; Alpert AJ; McIlwain S; Kohmoto T; Jin S; Ge Y Anal. Chem 2017, 89, 5467–5475. [PubMed: 28406609]

(22). Pan P; McLuckey SA Anal. Chem 2003, 75, 5468–5474. [PubMed: 14710826]

(23). Metwally H; McAllister RG; Konermann L Anal. Chem 2015, 87, 2434–2442. [PubMed: 25594702]

(24). Laganowsky A; Reading E; Hopper JTS; Robinson CV Nat. Protocols 2013, 8, 639–651. [PubMed: 23471109]

(25). Pohl T; Kamp RM Anal. Biochem 1987, 160, 388–391. [PubMed: 3034092]

(26). Zhang H; Ge Y Circ. Cardiovasc. Genet 2011, 4, 711–711. [PubMed: 22187450]

(27). Susa AC; Xia Z; Williams ER Anal. Chem 2017, 89, 3116–3122. [PubMed: 28192954]

(28). Speers AE; Wu CC Chem. Rev 2007, 107, 3687–3714. [PubMed: 17683161]

(29). Catherman AD; Li M; Tran JC; Durbin KR; Compton PD; Early BP; Thomas PM; Kelleher NL Anal. Chem 2013, 85, 1880–1888. [PubMed: 23305238]

(30). Skinner OS; Catherman AD; Early BP; Thomas PM; Compton PD; Kelleher NL Anal. Chem 2014, 86, 4627–4634. [PubMed: 24689519]

(31). Loo RR; Dales N; Andrews PC Protein Sci 1994, 3, 1975–1983. [PubMed: 7703844]

(32). Wessel D; Flugge UI Anal. Biochem 1984, 138, 141–143. [PubMed: 6731838]

(33). Doucette AA; Vieira DB; Orton DJ; Wall MJ J. Proteome Res 2014, 13, 6001–6012. [PubMed: 25384094]

(34). Moore SM; Hess SM; Jorgenson JW J. Proteome Res 2016, 15, 1243–1252. [PubMed: 26979493]

(35). Whitelegge JP; Gundersen CB; Faull KF Protein Sci 1998, 7, 1423–1430. [PubMed: 9655347]

(36). Hengel SM; Floyd E; Baker ES; Zhao R; Wu S; Paša-Tolić L Proteomics 2012, 12, 3138–3142. [PubMed: 22936678]

(37). Crowell AM; MacLellan DL; Doucette AA J. Proteomics 2015, 118, 140–150. [PubMed: 25316050]

(38). Kachuk C; Faulkner M; Liu F; Doucette AA J. Proteome Res 2016, 15, 2634–2642. [PubMed: 27376408]

(39). Kim KH; Compton PD; Tran JC; Kelleher NL J. Proteome Res 2015, 14, 2199–2206. [PubMed: 25836738]

(40). Chang YH; Gregorich ZR; Chen AJ; Hwang L; Guner H; Yu D; Zhang J; Ge YJ Proteome Res 2015, 14, 1587–1599.

(41). Barrera NP; Robinson CV Annu. Rev. Biochem 2011, 80, 247–271. [PubMed: 21548785]

(42). Yen HY; Hopper JTS; Liko I; Allison TM; Zhu Y; Wang DJ; Stegmann M; Mohammed S; Wu BL; Robinson CV Sci. Adv 2017, 3.

(43). Heck AJR Nat. Methods 2008, 5, 927–933. [PubMed: 18974734]

(44). Carroll J; Fearnley IM; Walker JE Proc. Natl. Acad. Sci. U. S. A 2006, 103, 16170–16175. [PubMed: 17060615]

(45). Waas M; Bhattacharya S; Chuppa S; Wu X; Jensen DR; Omasits U; Wollscheid B; Volkman BF; Noon KR; Gundry RL Anal. Chem 2014, 86, 1551–1559. [PubMed: 24392666]

(46). Cox B; Emili A Nat. Protoc 2006, 1, 1872–1878. [PubMed: 17487171]

(47). Catherman AD; Durbin KR; Ahlf DR; Early BP; Fellers RT; Tran JC; Thomas PM; Kelleher NL Mol. Cell. Proteomics 2013, 12, 3465–3473. [PubMed: 24023390]

(48).) Zhou MW; Wu S; Stenoien DL; Zhang ZR; Connolly L; Freitag M; Pasa-Tolic L In Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation, Wajapeyee N; Gupta R, Eds., 2017, pp 153–168.

(49). Fei R; Zhang T; Huang Y; Hu Y Anal. Chim. Acta 2017, 986, 161–170. [PubMed: 28870322]

(50). He YT; Liu W; Chen L; Lin G; Xiao Q; Gao CL; Wu JL; Lin ZJ Sep. Sci 2017, 40, 1516–1523.

(51). Hwang L; Ayaz-Guner S; Gregorich ZR; Cai W; Valeja SG; Jin S; Ge YJ Am. Chem. Soc 2015, 137, 2432–2435.

(52). Chen B; Hwang L; Ochowicz W; Lin Z; Guardado-Alvarez TM; Cai W; Xiu L; Dani K; Colah C; Jin S; Ge Y Chem. Sci 2017, 8, 4306–4311. [PubMed: 28660060]

(53). Zhang J; Guy MJ; Norman HS; Chen YC; Xu QG; Dong XT; Guner H; Wang SJ; Kohmoto T; Young KH; Moss RL; Ge YJ Proteome Res 2011, 10, 4054–4065.

(54). Xu FM; Xu QG; Dong XT; Guy M; Guner H; Hacker TA; Ge Y Int. J. Mass spectrom 2011, 305, 95–102.

(55). Dong XT; Sumandea CA; Chen YC; Garcia-Cazarin ML; Zhang J; Balke CW; Sumandea MP; Ge YJ Biol. Chem 2012, 287, 848–857.

(56). Peng Y; Chen X; Sato T; Rankin SA; Tsuji RF; Ge Y Anal. Chem 2012, 84, 3339–3346. [PubMed: 22390166]

(57). Burnaevskiy N; Fox TG; Plymire DA; Ertelt JM; Weigele BA; Selyunin AS; Way SS; Patrie SM; Alto NM Nature 2013, 496, 106–+. [PubMed: 23535599]

(58). Savaryn JP; Skinner OS; Fornelli L; Fellers RT; Compton PD; Terhune SS; Abecassis MM; Kelleher NL J. Proteomics 2016, 134, 76–84. [PubMed: 25952688]

(59). Doucette AA; Tran JC; Wall MJ; Fitzsimmons S Expert Rev. Proteomics 2011, 8, 787–800. [PubMed: 22087661]

(60). Capriotti AL; Cavaliere C; Foglia P; Samperi R; Lagana AJ Chromatogr. A 2011, 1218, 8760–8776.

(61). Tran JC; Doucette AA Anal. Chem 2008, 80, 1568–1573. [PubMed: 18229945]

(62). Vellaichamy A; Tran JC; Catherman AD; Lee JE; Kellie JF; Sweet SM; Zamdborg L; Thomas PM; Ahlf DR; Durbin KR; Valaskovic GA; Kelleher NL Anal. Chem 2010, 82, 1234–1244. [PubMed: 20073486]

(63). Sun L; Knierman MD; Zhu G; Dovichi NJ Anal. Chem 2013, 85, 5989–5995. [PubMed: 23692435]

(64). Xiu LC; Valeja SG; Alpert AJ; Jin S; Ge Y Anal. Chem 2014, 86, 7899–7906. [PubMed: 24968279]

(65). Rogers BA; Wu Z; Wei B; Zhang X; Cao X; Alabi O; Wirth MJ Anal. Chem 2015, 87, 2520–2526. [PubMed: 25646567]

(66). Valeja SG; Xiu LC; Gregorich ZR; Guner H; Jin S; Ge Y Anal. Chem 2015, 87, 5363–5371. [PubMed: 25867201]

(67). Chen BF; Peng Y; Valeja SG; Xiu LC; Alpert AJ; Ge Y Anal. Chem 2016, 88, 1885–1891. [PubMed: 26729044]

(68). Shen Y; Tolic N; Piehowski PD; Shukla AK; Kim S; Zhao R; Qu Y; Robinson E; Smith RD; Pasa-Tolic LJ Chromatogr. A 2017, 1498, 99–110.

(69). Cai W; Tucholski T; Chen B; Alpert AJ; McIlwain S; Kohmoto T; Jin S; Ge Y Anal. Chem 2017, 89, 5467–5475. [PubMed: 28406609]

(70). Zhen W; Bingchuan W; Ximo Z; Wirth MJ Anal. Chem 2014, 86, 1592–1598. [PubMed: 24383398]

(71). Eeltink S; Wouters B; Desmet G; Ursem M; Blinco D; Kemp GD; Treumann AJ Chromatogr. A 2011, 1218, 5504–5511.

(72). Guiochon GJ Chromatogr. A 2007, 1168, 101–168.

(73). Simone P; Pierri G; Foglia P; Gasparrini F; Mazzoccanti G; Capriotti AL; Ursini O; Ciogli A; Lagana AJ Sep. Sci 2016, 39, 264–271.

(74). Muneeruddin K; Thomas JJ; Salinas PA; Kaltashov IA Anal. Chem 2014, 86, 10692–10699. [PubMed: 25310183]

(75). Muneeruddin K; Nazzaro M; Kaltashov IA Anal. Chem 2015, 87, 10138–10145. [PubMed: 26360183]

(76). Muneeruddin K; Bobst CE; Frenkel R; Houde D; Turyan I; Sosic Z; Kaltashov IA Analyst 2017, 142, 336–344. [PubMed: 27965993]

(77). Valaskovic GA; Kelleher NL; McLafferty FW Science 1996, 273, 1199–1202. [PubMed: 8703047]

(78). Zhou F; Johnston MV Anal. Chem 2004, 76, 2734–2740. [PubMed: 15144182]

(79). Haselberg R; de Jong GJ; Somsen GW Anal. Chem 2013, 85, 2289–2296. [PubMed: 23323765]

(80). Zhao Y; Sun L; Champion MM; Knierman MD; Dovichi NJ Anal. Chem 2014, 86, 4873–4878. [PubMed: 24725189]

(81). Han XM; Wang YJ; Aslanian A; Fonslow B; Graczyk B; Davis TN; Yates JR J. Proteome Res 2014, 13, 6078–6086. [PubMed: 25382489]

(82). Zhao Y; Riley NM; Sun L; Hebert AS; Yan X; Westphall MS; Rush MJ; Zhu G; Champion MM; Mba Medie F; Champion PA; Coon JJ; Dovichi NJ Anal. Chem 2015, 87, 5422–5429. [PubMed: 25893372]

(83). Bush DR; Zang L; Belov AM; Ivanov AR; Karger BL Anal. Chem 2016, 88, 1138–1146. [PubMed: 26641950]

(84). Stepanova S; Kasicka V Anal. Chim. Acta 2016, 933, 23–42. [PubMed: 27496994]

(85). Zhao Y; Sun L; Zhu G; Dovichi NJ J. Proteome Res 2016, 15, 3679–3685. [PubMed: 27490796]

(86). Wang Z; Ma H; Smith K; Wu S Int. J. Mass spectrom 2017, doi: 10.1016/j.ijms.2017.1009.1001.

(87). Compton PD; Zamdborg L; Thomas PM; Kelleher NL Anal. Chem 2011, 83, 6868–6874. [PubMed: 21744800]

(88). Tran JC; Zamdborg L; Ahlf DR; Lee JE; Catherman AD; Durbin KR; Tipton JD; Vellaichamy A; Kellie JF; Li MX; Wu C; Sweet SMM; Early BP; Siuti N; LeDuc RD; Compton PD; Thomas PM; Kelleher NL Nature 2011, 480, 254–U141. [PubMed: 22037311]

(89). Catherman AD; Durbin KR; Ahlf DR; Early BP; Fellers RT; Tran JC; Thomas PM; Kelleher NL Mol. Cell. Proteomics 2013, 12, 3465–3473. [PubMed: 24023390]

(90). Morrison LJ; Brodbelt JS J. Am. Chem. Soc 2016, 138, 10849–10859. [PubMed: 27480400]

(91). Tian ZX; Tolic N; Zhao R; Moore RJ; Hengel SM; Robinson EW; Stenoien DL; Wu S; Smith RD; Pasa-Tolic L Genome Biol 2012, 13.

(92). Olsen JV; Macek B; Lange O; Makarov A; Horning S; Mann M Nat. Methods 2007, 4, 709–712. [PubMed: 17721543]

(93). Durbin KR; Fornelli L; Fellers RT; Doubleday PF; Narita M; Kelleher NL J. Proteome Res 2016, 15, 976–982. [PubMed: 26795204]

(94). Siuti N; Kelleher NL Nat. Methods 2007, 4, 817–821. [PubMed: 17901871]

(95). Zubarev RA; Horn DM; Fridriksson EK; Kelleher NL; Kruger NA; Lewis MA; Carpenter BK; McLafferty FW Anal. Chem 2000, 72, 563–573. [PubMed: 10695143]

(96). Syka JEP; Coon JJ; Schroeder MJ; Shabanowitz J; Hunt DF Proc. Natl. Acad. Sci. U. S. A 2004, 101, 9528–9533. [PubMed: 15210983]

(97). Peng Y; Gregorich ZR; Valeja SG; Zhang H; Cai WX; Chen YC; Guner H; Chen AJ; Schwahn DJ; Hacker TA; Liu XW; Ge Y Mol. Cell. Proteomics 2014, 13, 2752–2764. [PubMed: 24969035]

(98). Gregorich ZR; Peng Y; Lane NM; Wolff JJ; Wang SJ; Guo W; Guner H; Doop J; Hacker TA; Ge YJ Mol. Cell. Cardiol 2015, 87, 102–112.

(99). Jin YT; Peng Y; Lin ZQ; Chen YC; Wei LM; Hacker TA; Larsson L; Ge YJ Muscle Res. Cell Motil 2016, 37, 41–52.

(100). Yu DY; Peng Y; Ayaz-Guner S; Gregorich ZR; Ge YJ Am. Soc. Mass. Spectrom 2016, 27, 220–232.

(101). Pan JX; Zhang SP; Borchers CH J. Proteomics 2016, 134, 138–143. [PubMed: 26675311]

(102). Fornelli L; Damoc E; Thomas PM; Kelleher NL; Aizikov K; Denisov E; Makarov A; Tsybin YO Mol. Cell. Proteomics 2012, 11, 1758–1767. [PubMed: 22964222]

(103). Mao Y; Valeja SG; Rouse JC; Hendrickson CL; Marshall AG Anal. Chem 2013, 85, 4239–4246. [PubMed: 23551206]

(104). Tran BQ; Barton C; Feng JH; Sandjong A; Yoon SH; Awasthi S; Liang T; Khan MM; Kilgour DPA; Goodlett DR; Goo YA J. Proteomics 2016, 134, 93–101. [PubMed: 26485299]

(105). He LD; Anderson LC; Barnidge DR; Murray DL; Hendrickson CL; Marshall AG J. Am. Soc. Mass. Spectrom 2017, 28, 827–838. [PubMed: 28247297]

(106). Fornelli L; Ayoub D; Aizikov K; Liu XW; Damoc E; Pevzner PA; Makarov A; Beck A; Tsybin YO J. Proteomics 2017, 159, 67–76. [PubMed: 28242452]

(107). Brunner AM; Lossl P; Liu F; Huguet R; Mullen C; Yamashita M; Zabrouskov V; Makarov A; Altelaar AFM; Heck AJ R. Anal. Chem 2015, 87, 4152–4158.

(108). Riley NM; Westphall MS; Coon JJ Anal. Chem 2015, 87, 7109–7116. [PubMed: 26067513]

(109). Riley NM; Westphall MS; Coon JJ J. Proteome Res 2017, 16, 2653–2659. [PubMed: 28608681]

(110). Horn DM; Ge Y; McLafferty FW Anal. Chem 2000, 72, 4778–4784. [PubMed: 11055690]

(111). Shaw JB; Li WZ; Holden DD; Zhang Y; Griep-Raming J; Fellers RT; Early BP; Thomas PM; Kelleher NL; Brodbelt JS J. Am. Chem. Soc 2013, 135, 12646–12651. [PubMed: 23697802]

(112). Thyer R; Robotham SA; Brodbelt JS; Ellington AD J. Am. Chem. Soc 2015, 137, 46–49. [PubMed: 25521771]

(113). Cammarata MB; Thyer R; Rosenberg J; Ellington A; Brodbelt JS J. Am. Chem. Soc 2015, 137, 9128–9135. [PubMed: 26125523]

(114). Holden DD; McGee WM; Brodbelt JS Anal. Chem 2016, 88, 1008–1016. [PubMed: 26633754]

(115). Mayfield JE; Robinson MR; Cotham VC; Irani S; Matthews WL; Ram A; Gilmour DS; Cannon JR; Zhang YJ; Brodbelt JS ACS Chem. Biol 2017, 12, 153–162. [PubMed: 28103682]

(116). Cleland TP; DeHart CJ; Fellers RT; VanNispen AJ; Greer JB; Leduc RD; Parker WR; Thomas PM; Kelleher NL; Brodbelt JS J. Proteome Res 2017, 16, 2072–2079. [PubMed: 28412815]

(117). McLuckey SA; Reid GE; Wells JM Anal. Chem 2002, 74, 336–346. [PubMed: 11811406]

(118). Earley L; Anderson LC; Bai DL; Mullen C; Syka JE; English AM; Dunyach JJ; Stafford GC, Jr.; Shabanowitz J; Hunt DF; Compton PD Anal. Chem 2013, 85, 8385–8390. [PubMed: 23909443]

(119). Anderson LC; Karch KR; Ugrin SA; Coradin M; English AM; Sidoli S; Shabanowitz J; Garcia BA; Hunt DF Mol. Cell. Proteomics 2016, 15, 975–988. [PubMed: 26785730]

(120). Riley NM; Mullen C; Weisbrod CR; Sharma S; Senko MW; Zabrouskov V; Westphall MS; Syka JEP; Coon JJ J. Am. Soc. Mass. Spectrom 2016, 27, 520–531. [PubMed: 26589699]

(121). Holden DD; Brodbelt JS Anal. Chem 2017, 89, 837–846. [PubMed: 28105830]

(122). Hofstadler SA; Bruce JE; Rockwood AL; Anderson GA; Winger BE; Smith RD Int. J. Mass Spectrom. Ion Processes 1994, 132, 109–127.

(123). Anderson LC; DeHart CJ; Kaiser NK; Fellers RT; Smith DF; Greer JB; LeDuc RD; Blakney GT; Thomas PM; Kelleher NL; Hendrickson CL J. Proteome Res 2017, 16, 1087–1096. [PubMed: 27936753]

(124). Kaiser NK; Savory JJ; Hendrickson CL J. Am. Soc. Mass. Spectrom 2014, 25, 943–949. [PubMed: 24692045]

(125). Liu X; Inbar Y; Dorrestein PC; Wynne C; Edwards N; Souda P; Whitelegge JP; Bafna V; Pevzner PA Mol. Cell. Proteomics 2010, 9, 2772–2782. [PubMed: 20855543]

(126). Kou Q; Wu S; Liu X BMC Genomics 2014, 15, 1140. [PubMed: 25523396]

(127). Fellers RT; Greer JB; Early BP; Yu X; LeDuc RD; Kelleher NL; Thomas PM Proteomics 2015, 15, 1235–1238. [PubMed: 25828799]

(128). Cai WX; Guner H; Gregorich ZR; Chen AJ; Ayaz-Guner S; Peng Y; Valeja SG; Liu XW; Ge Y Mol. Cell. Proteomics 2016, 15, 703–714. [PubMed: 26598644]

(129). Xiao KJ; Yu F; Tian ZX J. Proteomics 2017, 152, 41–47. [PubMed: 27989944]

(130). Sun RX; Luo L; Wu L; Wang RM; Zeng WF; Chi H; Liu C; He SM Anal. Chem 2016, 88, 3082–3090. [PubMed: 26844380]

(131). Kou Q; Xun LK; Liu XW Bioinformatics 2016, 32, 3495–3497. [PubMed: 27423895]

(132). Petrotchenko EV; Borchers CH J. Am. Soc. Mass. Spectrom 2015, 26, 1895–1898. [PubMed: 26162650]

(133). Vyatkina K; Wu S; Dekker LJM; VanDuijn MM; Liu XW; Tolic N; Luider TM; Pasa-Tolic L; Pevzner PA Bioinformatics 2016, 32, 2753–2759. [PubMed: 27187201]

(134). Horn DM; Zubarev RA; McLafferty FW J. Am. Soc. Mass. Spectrom 2000, 11, 320–332. [PubMed: 10757168]

(135). Jaitly N; Mayampurath A; Littlefield K; Adkins JN; Anderson GA; Smith RD BMC Bioinformatics 2009, 10, 87. [PubMed: 19292916]

(136). Mayampurath AM; Jaitly N; Purvine SO; Monroe ME; Auberry KJ; Adkins JN; Smith RD Bioinformatics 2008, 24, 1021–1023. [PubMed: 18304935]

(137). Zamdborg L; LeDuc RD; Glowacz KJ; Kim YB; Viswanathan V; Spaulding IT; Early BP; Bluhm EJ; Babai S; Kelleher NL Nucleic Acids Res 2007, 35, W701–W706. [PubMed: 17586823]

(138). Guner H; Close PL; Cai WX; Zhang H; Peng Y; Gregorich ZR; Ge YJ Am. Soc. Mass. Spectrom 2014, 25, 464–470.

(139). Marty MT; Baldwin AJ; Marklund EG; Hochberg GK; Benesch JL; Robinson CV Anal. Chem 2015, 87, 4370–4376. [PubMed: 25799115]

(140). Li L; Tian ZX Rapid Commun. Mass Spectrom 2013, 27, 1267–1277. [PubMed: 23650040]

(141). Liu XW; Sirotkin Y; Shen YF; Anderson G; Tsai YS; Ting YS; Goodlett DR; Smith RD; Bafna V; Pevzner PA Mol. Cell. Proteomics 2012, 11.

(142). Tsai YS; Scherl A; Shaw JL; MacKay CL; Shaffer SA; Langridge-Smith PRR; Goodlett DR J. Am. Soc. Mass. Spectrom 2009, 20, 2154–2166. [PubMed: 19773183]

(143). Liu XW; Hengel S; Wu S; Tolic N; Pasa-Tolic L; Pevzner PA J. Proteome Res 2013, 12, 5830–5838. [PubMed: 24188097]

(144). Kou Q; Zhu BH; Wu S; Ansong C; Tolic N; Pasa-Tolic L; Liu XW J. Proteome Res 2016, 15, 2422–2432. [PubMed: 27291504]

(145). Kou Q; Wu S; Tolic N; Pasa-Tolic L; Liu YL; Liu XW Bioinformatics 2017, 33, 1309–1316. [PubMed: 28453668]

(146). LeDuc RD; Fellers RT; Early BP; Greer JB; Thomas PM; Kelleher NL J. Proteome Res 2014, 13, 3231–3240. [PubMed: 24922115]

(147). Durbin KR; Skinner OS; Fellers RT; Kelleher NL J. Am. Soc. Mass. Spectrom 2015, 26, 782–787. [PubMed: 25716753]

(148). Skinner OS; Do Vale LHF; Catherman AD; Havugimana PC; de Sousa MV; Compton PD; Kelleher NL Anal. Chem 2015, 87, 3032–3038. [PubMed: 25664979]

(149). Belov AM; Viner R; Santos MR; Horn DM; Bern M; Karger BL; Ivanov AR J. Am. Soc. Mass. Spectrom 2017.

(150). Belov ME; Damoc E; Denisov E; Compton PD; Horning S; Makarov AA; Kelleher NL Anal. Chem 2013, 85, 11163–11173. [PubMed: 24237199]

(151). Skinner OS; Havugimana PC; Haverland NA; Fornelli L; Early BP; Greer JB; Fellers RT; Durbin KR; Do Vale LHF; Melani RD; Seckler HS; Nelp MT; Belov ME; Horning SR; Makarov AA; LeDuc RD; Bandarian V; Compton PD; Kelleher NL Nat. Methods 2016, 13, 237–+. [PubMed: 26780093]

(152). Melani RD; Skinner OS; Fornelli L; Domont GB; Compton PD; Kelleher NL Mol. Cell. Proteomics 2016, 15, 2423–2434. [PubMed: 27178327]

(153). Haverland NA; Skinner OS; Fellers RT; Tariq AA; Early BP; Leduc RD; Fornelli L; Compton PD; Kelleher NL J. Am. Soc. Mass. Spectrom 2017, 28, 1203–1215. [PubMed: 28374312]

(154). Li HL; Wolff JJ; Van Orden SL; Loo JA Anal. Chem 2014, 86, 317–320. [PubMed: 24313806]

(155). Schneeberger EM; Breuker K Angew. Chem. Int. Ed 2017, 56, 1254–1258.

(156). Yan J; Zhou M; Gilbert JD; Wolff JJ; Somogyi A; Pedder RE; Quintyn RS; Morrison LJ; Easterling ML; Pasa-Tolic L; Wysocki VH Anal. Chem 2017, 89, 895–901. [PubMed: 27977147]

(157). Skinner OS; McAnally MO; Van Duyne RP; Schatz GC; Breuker K; Compton PD; Kelleher NL Anal. Chem 2017, 89, 10711–10716. [PubMed: 28938074]

(158). Pesavento JJ; Mizzen CA; Kelleher NL Anal. Chem 2006, 78, 4271–4280. [PubMed: 16808433]

(159). Ansong C; Wu S; Meng D; Liu XW; Brewer HM; Kaiser BLD; Nakayasu ES; Cort JR; Pevzner P; Smith RD; Heffron F; Adkins JN; Pasa-Tolic L Proc. Natl. Acad. Sci. U. S. A 2013, 110, 10153–10158. [PubMed: 23720318]

(160). Dang XB; Singh A; Spetman BD; Nolan KD; Isaacs JS; Dennis JH; Dalton S; Marshall AG; Young NL J. Proteome Res 2016, 15, 3196–3203. [PubMed: 27431976]

(161). Rea M; Jiang TT; Eleazer R; Eckstein M; Marshall AG; Fondufe-Mittendorf YN Mol. Cell. Proteomics 2016, 15, 2411–2422. [PubMed: 27169413]

(162). Zheng YP; Fornelli L; Compton PD; Sharma S; Canterbury J; Mullen C; Zabrouskov V; Fellers RT; Thomas PM; Licht JD; Senko MW; Kelleher NL Mol. Cell. Proteomics 2016, 15, 776–790. [PubMed: 26272979]

(163). Stefanowicz P; Kijewska M; Szewczuk Z Anal. Chem 2014, 86, 7247–7251. [PubMed: 25029396]

(164). Chen BF; Guo X; Tucholski T; Lin ZQ; McIlwain S; Ge YJ Am. Soc. Mass. Spectrom 2017, 28, 1805–1814.

(165). Collier TS; Sarkar P; Rao B; Muddiman DC J. Am. Soc. Mass. Spectrom 2010, 21, 879–889. [PubMed: 20199872]

(166). Wiese S; Reidegeld KA; Meyer HE; Warscheid B Proteomics 2007, 7, 340–350. [PubMed: 17177251]

(167). Hung CW; Tholey A Anal. Chem 2012, 84, 161–170. [PubMed: 22103715]

(168). Fang HQ; Xiao KJ; Li YH; Yu F; Liu Y; Xue BB; Tian ZX Anal. Chem 2016, 88, 7198–7205. [PubMed: 27359340]

(169). Rhoads TW; Rose CM; Bailey DJ; Riley NM; Molden RC; Nestler AJ; Merrill AE; Smith LM; Hebert AS; Westphall MS; Pagliarini DJ; Garcia BA; Coon JJ Anal. Chem 2014, 86, 2314–2319. [PubMed: 24475910]

(170). Shortreed MR; Frey BL; Scalf M; Knoener RA; Cesnik AJ; Smith LM J. Proteome Res 2016, 15, 1213–1221. [PubMed: 26941048]

(171). Dai Y; Shortreed MR; Scalf M; Frey BL; Cesnik AJ; Solntsev S; Schaffer LV; Smith LM J. Proteome Res 2017, 16, 4156–4165. [PubMed: 28968100]

(172). Waanders LF; Hanke S; Mann MJ Am. Soc. Mass. Spectrom 2007, 18, 2058–2064.

(173). Geis-Asteggiante L; Ostrand-Rosenberg S; Fenselau C; Edwards NJ Anal. Chem 2016, 88, 10900–10907. [PubMed: 27748581]

(174). Quijada JV; Schmitt ND; Salisbury JP; Auclair JR; Agar JN Anal. Chem 2016, 88, 11139–11146. [PubMed: 27744677]

(175). Mazur MT; Cardasis HL; Spellman DS; Liaw A; Yates NA; Hendrickson RC Proc. Natl. Acad. Sci. U. S. A 2010, 107, 7728–7733. [PubMed: 20388904]

(176). Wu S; Brown JN; Tolic N; Meng D; Liu X; Zhang H; Zhao R; Moore RJ; Pevzner P; Smith RD; Pasa-Tolic L Proteomics 2014, 14, 1211–1222. [PubMed: 24591407]

(177). Ntai I; Kim K; Fellers RT; Skinner OS; Smith A. D. t.; Early BP; Savaryn JP; LeDuc RD; Thomas PM; Kelleher NL Anal. Chem 2014, 86, 4961–4968. [PubMed: 24807621]

(178). Ntai I; LeDuc RD; Fellers RT; Erdmann-Gilmore P; Davies SR; Rumsey J; Early BP; Thomas PM; Li S; Compton PD; Ellis MJ; Ruggles KV; Fenyo D; Boja ES; Rodriguez H; Townsend RR; Kelleher NL Mol. Cell. Proteomics 2016, 15, 45–56. [PubMed: 26503891]

(179). Schmit PO; Vialaret J; Wessels HJCT; van Gool AJ; Lehmann S; Gabelle A; Wood J; Bern M; Paape R; Suckau D; Kruppa G; Hirtz CJ Proteomics 2017, doi: 10.1016/j.jprot.2017.1008.1003.

(180). Wang EH; Combe PC; Schug KA J. Am. Soc. Mass. Spectrom 2016, 27, 886–896. [PubMed: 26956437]

(181). Wang EH; Appulage DK; McAllister EA; Schug KA J. Am. Soc. Mass. Spectrom 2017.

(182). Kilpatrick LE; Kilpatrick EL J. Proteome Res 2017, 16, 3255–3265. [PubMed: 28738681]

(183). Petras D; Heiss P; Sussmuth RD; Calvete JJ J. Proteome Res 2015, 14, 2539–2556. [PubMed: 25896403]

(184). Petras D; Heiss P; Harrison RA; Sussmuth RD; Calvete JJ J. Proteomics 2016, 146, 148–164. [PubMed: 27318176]

(185). Mao P; Wang DJ J. Proteome Res 2014, 13, 1560–1569. [PubMed: 24533899]

(186). Cheon DH; Nam EJ; Park KH; Woo SJ; Lee HJ; Kim HC; Yang EG; Lee C; Lee JE J. Proteome Res 2016, 15, 229–244. [PubMed: 26576621]

(187). Savaryn JP; Toby TK; Catherman AD; Fellers RT; LeDuc RD; Thomas PM; Friedewald JJ; Salomon DR; Abecassis MM; Kelleher NL Proteomics 2016, 16, 2048–2058. [PubMed: 27120713]

(188). Toby TK; Abecassis M; Kim K; Thomas PM; Fellers RT; LeDuc RD; Kelleher NL; Demetris J; Levitsky J Am. J. Transplantation 2017, 17, 2458–2467.

(189). Desiderio C; D’Angelo L; Rossetti DV; Iavarone F; Giardina B; Castagnola M; Massimi L; Tamburrini G; Di Rocco C Proteomics 2012, 12, 2158–2166. [PubMed: 22623401]

(190). Fania C; Arosio B; Capitanio D; Torretta E; Gussago C; Ferri E; Mari D; Gelfi C PLoS One 2017, 12, e0179280. [PubMed: 28628634]

(191). Cabras T; Pisano E; Montaldo C; Giuca MR; Iavarone F; Zampino G; Castagnola M; Messana I Mol. Cell. Proteomics 2013, 12, 1844–1852. [PubMed: 23533003]

(192). Iauarone F; Melis M; Platania G; Cabras T; Manconi B; Petruzzelli R; Cordaro M; Siracusano A; Faa G; Messana I; Zanasi M; Castagnola MJ Proteomics 2014, 103, 15–22. [PubMed: 24690516]

(193). Labas V; Spina L; Belleannee C; Teixeira-Gomes AP; Gargaros A; Dacheux F; Dacheux JL J. Proteomics 2015, 113, 226–243. [PubMed: 25452132]

(194). Bystrom C; Sheng SJ; Zhang K; Caulfield M; Clarke NJ; Reitz R PLoS One 2012, 7, e43457. [PubMed: 22984427]

(195). Kellie JF; Higgs RE; Ryder JW; Major A; Beach TG; Adler CH; Merchant K; Knierman MD Sci. Rep 2014, 4.

(196). Steffen P; Kwiatkowski M; Robertson WD; Zarrine-Afsar M; Deterra D; Richter V; Schluter HJ Proteomics 2016, 134, 5–18. [PubMed: 26721442]

(197). Gafvels M; Bengtson P Clin. Chim. Acta 2015, 442, 87–95. [PubMed: 25603406]

(198). Carel C; Marcoux J; Reat V; Parra J; Latge G; Laval F; Demange P; Burlet-Schiltz O; Milon A; Daffe M; Tropis MG; Renault MA M. Proc. Natl. Acad. Sci. U. S. A 2017, 114, 4231–4236.

(199). Parra J; Marcoux J; Poncin I; Canaan S; Herrmann JL; Nigou J; Burlet-Schiltz O; Riviere M Sci. Rep 2017, 7, 43682. [PubMed: 28272507]

(200). Gregorich ZR; Peng Y; Cai WX; Jin YT; Wei LM; Chen AJ; McKiernan SH; Aiken JM; Moss RL; Diffee GM; Ge YJ Proteome Res 2016, 15, 2706–2716.

(201). Wei L; Gregorich ZR; Lin Z; Cai W; Jin Y; McKiernan SH; McIlwain S; Aiken JM; Moss RL; Diffee GM; Ge Y Mol. Cell. Proteomics 2017, doi: 10.1074/mcp.RA1117.000124.

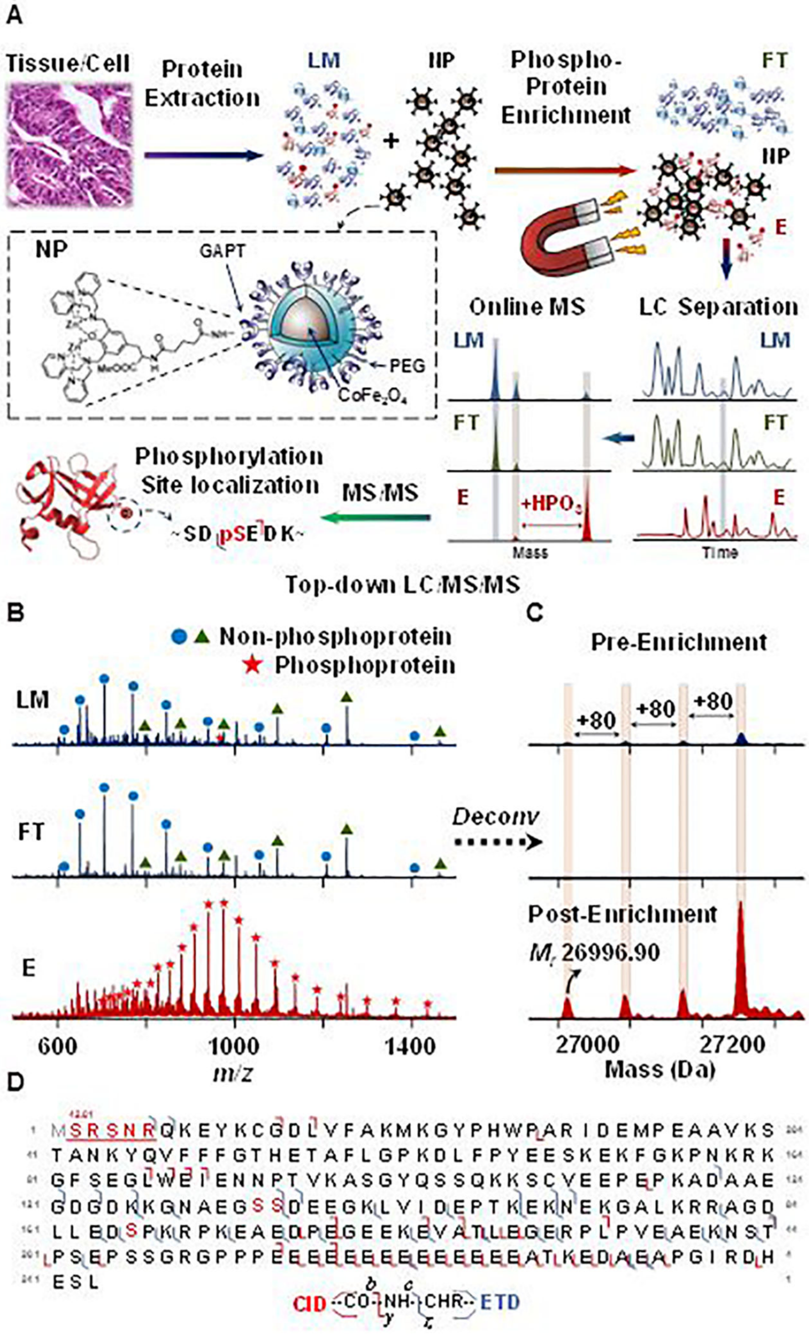

A
Tissue/Cell
Protein Extraction
LM
NP
+
Phospho-Protein Enrichment
FT
NP
E
NP
GAPT
PEG
CoFe2O4
Online MS
LC Separation
LM
LM
FT
FT
E
+HPO3
E
Mass
Time
Phosphorylation Site localization
MS/MS
~SD pSE D K~
Top-down LC/MS/MS
B
● ▲ Non-phosphoprotein
★ Phosphoprotein
LM
FT
E
600
1000
1400
m/z
C
Pre-Enrichment
+80
+80
+80
Deconv
Post-Enrichment
Mr 26996.90
27000
27200
Mass (Da)
D
42.01
1 MSRSNRQKEYKCGDLVFAKMKGYPHWPARIDEMPEAAVKS 204
41 TANKYQVFFFGTHETAFLGPKDLFPYEESKEKFGKPNKRK 164
81 GFSEGLWEIENNPTVKASGYQSSQKKSCVEEPEPKADAAE 124
121 GDGDKKGNAEGSSDEEGKLVIDEPTKEKNEKGALKRRAGD 84
161 LLEDSPKRPKEAEDPEGEEKEVATLEGERPLPVEAEKNST 44
201 PSEPSSGRGPPPEEEEEEEEEEEEEEATKEDAEAPGIRDH 4
241 ESL 1
CID
CO
NH
CHR
ETD
b
c
y
z

Figure 1.

(A) Schematic illustration of intact phosphoprotein enrichment using functionalized magnetic nanoparticles (NPs) coupled with online LC-MS/MS. (B-D) Representative LC-MS/MS analysis of low-abundance phosphoprotein enabled by effective $CoFe_2O_4$ NP-based enrichment from a complex swine heart tissue extract. (B) MS spectra of loading mixture (LM, dark blue), flow through (FT, light blue), and elution (E, red) from 32.3 min to 32.7 min; (C) the corresponding deconvoluted spectra. Ion intensities were normalized in both A and B. (D) Fragment ion map of online LC-MS/MS analysis with CID and ETD from triply phosphorylated precursor ion (29+). Grey “M” indicates methionine excision and red “S” indicates phosphorylation sites. Red numbers above the underlined red sequence reveal modifications with their additional mass. Adapted and reproduced from Chen, B.; Hwang, L.; Ochowicz, W.; Lin, Z.; Guardado-Alvarez, T. M.; Cai, W.; Xiu, L.; Dani, K.; Colah, C.; Jin, S.; Ge, Y. Chem. Sci. 2017, 8, 4306–4311 (ref 52), with permission of The Royal Society of Chemistry.

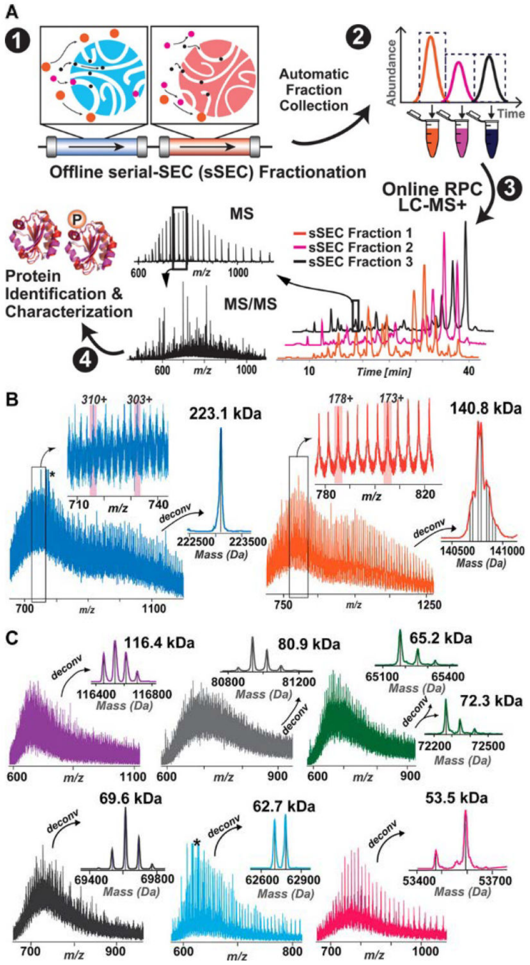


Figure 2.

(A) Schematic illustration of the serial size exclusion chromatography strategy enabling high-resolution size based separation to facilitate the detection of large molecular-weight intact proteins. (B) Representative mass spectra for 223.1 and 140.8 kDa with zoom-in views of the charge states and the corresponding deconvoluted spectra. The deconvoluted spectrum of the 140.8 kDa protein shows multiple proteoforms. (C) Representative mass spectra and the deconvoluted spectra of proteins with MW 116.4, 80.9, 65.2, 72.3, 69.6, 62.7, and 53.5 kDa. Adapted and reproduced from Cai, W.; Tucholski, T.; Chen, B.; Alpert, A. J.; McIlwain, S.; Kohmoto, T.; Jin, S.; Ge, Y. Anal. Chem. 2017, 89, 5467–5475 (ref 69). Copyright 2017 American Chemical Society.

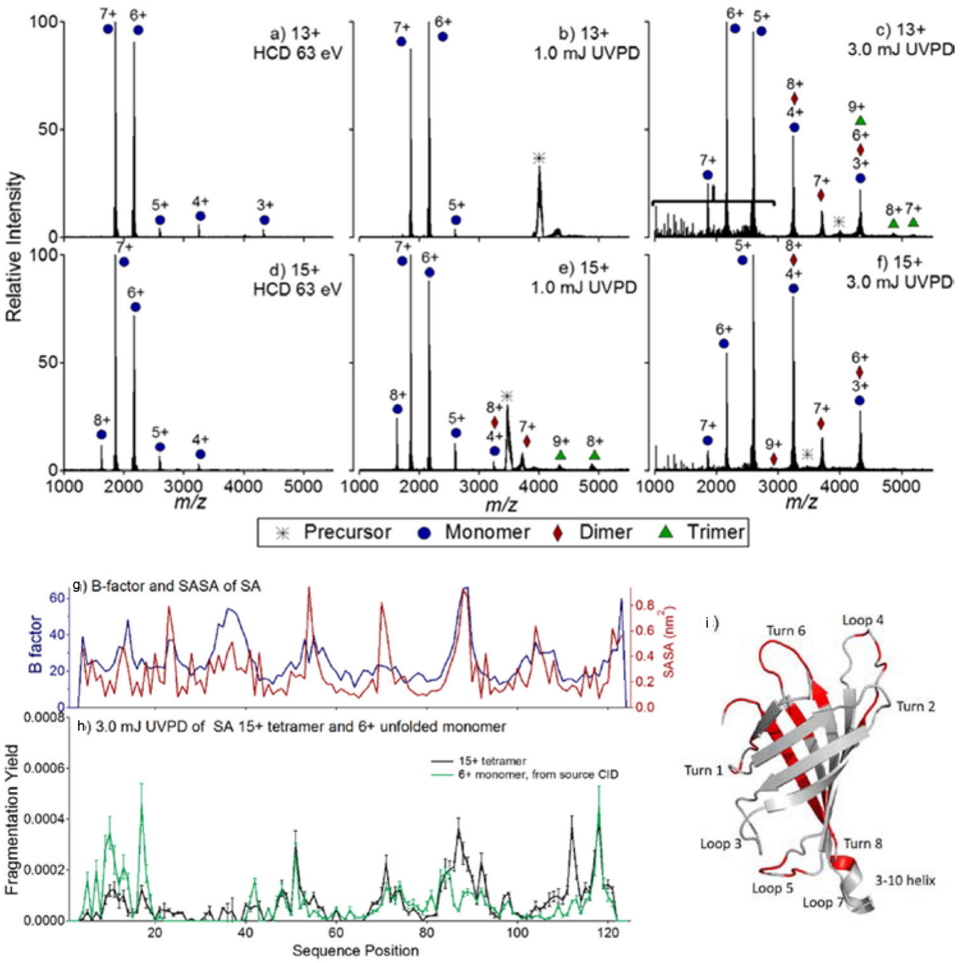


Figure 3.

HCD and UVPD of 13+ and 15+ tetrameric streptavidin. In (a) and (d) HCD of the 13+ and 15+ charge states is shown, respectively. In (b) and (e) 1.0 mJ UVPD of the 13+ and 15+ charge states is shown, and in (c) and (f) 3.0 mJ UVPD of the 13+ and 15+ charge states is shown. The bracket in (c) denotes the region populated by fragments originating from cleavages of the protein backbone (i.e., sequence-type ions). For streptavidin: (g) B-factor values and SASA, (h) UVPD fragmentation yields (15+), and (i) the crystal structure of SA (pdb 1SWB) is highlighted such that regions featuring enhanced UVPD fragmentation are shown in red. In (b), the UVPD fragmentation yield of the 6+ monomer of SA, generated from source CID, is shown in green. Adapted and reproduced from Morrison, L. J.; Brodbelt, J. S. J. Am. Chem. Soc. 2016, 138, 10849–10859 (Ref 90). Copyright 2016 American Chemical Society.

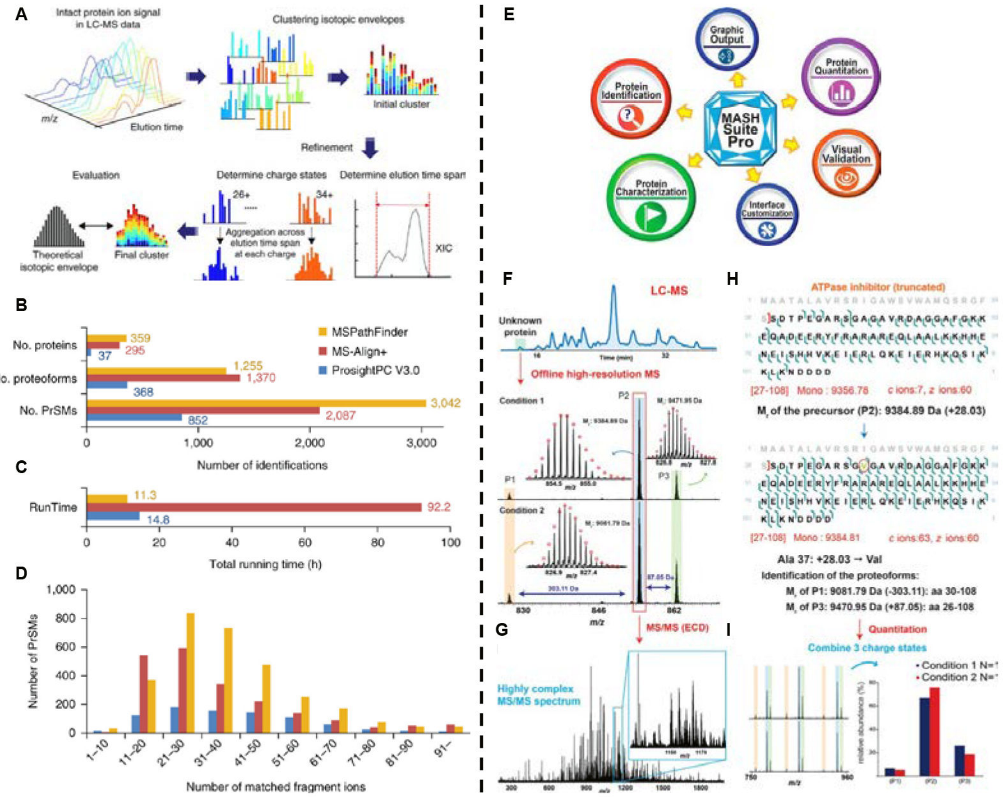


Figure 4.

Left: (A) LC-MS feature finding in ProMex. An LC-MS feature refers to a group of isotopomer envelopes corresponding to the same proteoform species across all charge states and LC elution times. The ProMex algorithm begins with clustering isotopomer envelopes across adjacent time and charge state. The initial cluster is refined to accurately determine its elution time span and range of charge states. After refinement, ProMex calculates the likelihood that the final cluster is a true LC-MS feature. (B-D) Protein identification and characterization results for a human ovarian tumor. (B) The number of proteins, proteoforms, and protein- spectrum matches (PrSMs) identified by ProsightPC V3.0 (E-value ≤ 10−4), MS-Align+ (1% FDR), and MSPathFinder (1% FDR). (C)Total running time for deconvolution and database search. (D) Histogram of the number of matched fragment ions. Adapted and reprinted by permission from Macmillan Publishers Ltd: NATURE METHODS, Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; Moore, R. J.; Liu, T.; Petyuk, V. A.; Tolic, N.; Pasa-Tolic, L.; Smith, R. D.; Payne, S. H.; Kim, S. Nat. Methods 2017, *14*, 909–914 (ref #11).Copyright 2017. Right: (E) Schematic summarizing the various functions and features of MASH Suite Pro. The major functions of MASH Suite Pro include protein identification, quantitation, and characterization of protein PTMs from top-down MS and MS/MS experiments. The program is equipped with various visualization components for the validation of the deconvolution results, identification results, and fragment ion assignments. Additional features include direct output of the graphics and customization of the program interface. (F-J) Identification and characterization of an unknown protein followed by quantitation of the proteoforms using MASH Suite Pro. (F) An unknown protein was detected in LC-MS and the fraction containing the protein was collected and analyzed by high-resolution MS. Top-down MS revealed three proteoforms (P1, P2 and P3). The relative abundance of each proteoform varied in two different experimental conditions. (G) The proteoform (P2) was selected for fragmentation using electron capture dissociation, resulting in a highly complex MS/MS spectrum. (H) MASH Suite Pro was used for spectral deconvolution and protein identification. The proteoform P2 was identified to be a truncated form of the ATPase inhibitor containing amino acids (aa) 27–108. Characterization of the protein sequence using MASH Suite Pro identified a sequence variation (Ala37Val). Based on the protein sequence, the proteoforms (P1) and (P3) were deduced to be aa 30–108 and aa 26–108 of ATPase inhibitor. (I) MASH Suite Pro provided rapid quantitation to determine the relative abundances of the different proteoforms in different experimental conditions. Reproduced from Cai, W. X.; Guner, H.; Gregorich, Z. R.; Chen, A. J.; Ayaz-Guner, S.; Peng, Y.; Valeja, S. G.; Liu, X. W.; Ge, Y. Mol. Cell. Proteomics 2016, 15, 703–714 (ref xxx). Copyright 2016 American Society for Biochemistry and Molecular Biology.

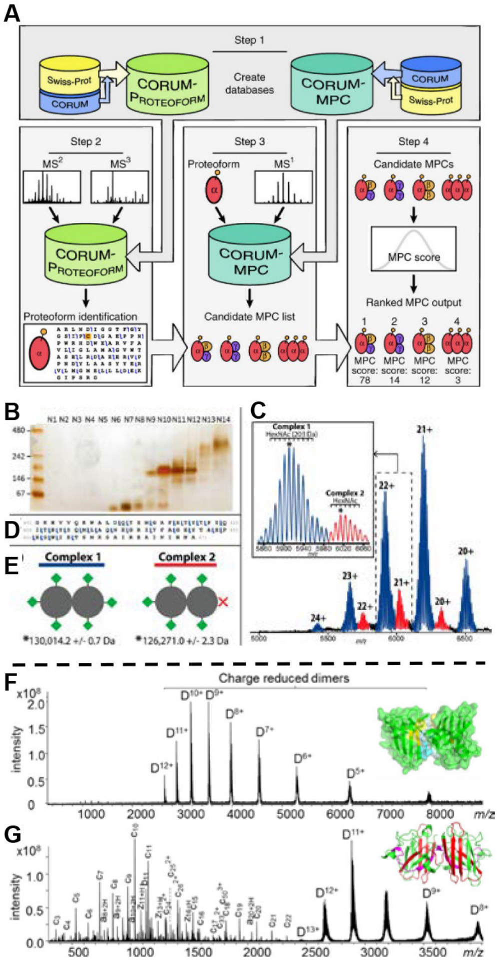


Figure 5.

Top: (A) Computational platform and workflow for the characterization of human multi-proteoform complexes (MPCs). Adapted and reprinted by permission from Macmillan Publishers Ltd: NATURE METHODS, Skinner, O. S.; Havugimana, P. C.; Haverland, N. A.; Fornelli, L.; Early, B. P.; Greer, J. B.; Fellers, R. T.; Durbin, K. R.; Do Vale, L. H. F.; Melani, R. D.; Seckler, H. S.; Nelp, M. T.; Belov, M. E.; Horning, S. R.; Makarov, A. A.; LeDuc, R. D.; Bandarian, V.; Compton, P. D.; Kelleher, N. L. Nat. Methods 2016, 13, 237-(ref #xxx). (B-E) Identification and purification of L-amino acid oxidase multiproteoform complexes. A native GELFrEE separation of whole O. hannah venom visualized using a native, silver-stained slab gel (B). An intact mass spectrum of the homodimeric L-amino acid oxidase - LAAO (P81383) multiproteoform complexes (MPCs) is shown in (C). In detail, the observed microheterogenity in the two MPCs corresponds to HexNAc mass differences (increments of 203 Da, see scale bar in the inset of Panel B). The partial fragment map (D), showing selected fragment ions from the C-terminal region of LAAO, which enabled its unambiguous identification by database retrieval. Blue and red MPCs in B are consistent with the presence of 6 (Complex 1) and 5 occupied N-glycosites (Complex 2), respectively; their graphical representation is shown at lower left (E, green diamond corresponds to a glycosylation moiety with average mass of 3,743 Da). Mass values for

Complexes 1 and 2 are given in (E) and correspond to the peaks marked with the asterisks in the inset of Panel (C). Reproduced from Melani, R. D.; Skinner, O. S.; Fornelli, L.; Domont, G. B.; Compton, P. D.; Kelleher, N. L. Mol. Cell. Proteomics 2016, 15, 2423–2434 (ref xxx). Copyright 2016 American Society for Biochemistry and Molecular Biology. Bottom: Native top-down (F) ECD and (G) EID MS of the apo-SOD1 dimer (10+ ~ 12+). EID fragment ions from the N-terminal regions are color-coded in red and products from the C-terminal regions are in magenta. (Representative spectra are shown; each spectrum was acquired from 200 scans.) Adapted and reproduced from Li, H. L.; Sheng, Y. W.; McGee, W.; Cammarata, M.; Holden, D.; Loo, J. A. Anal. Chem. 2017, 89, 2731–2738 (ref xxx). Copyright 2017 American Chemical Society.

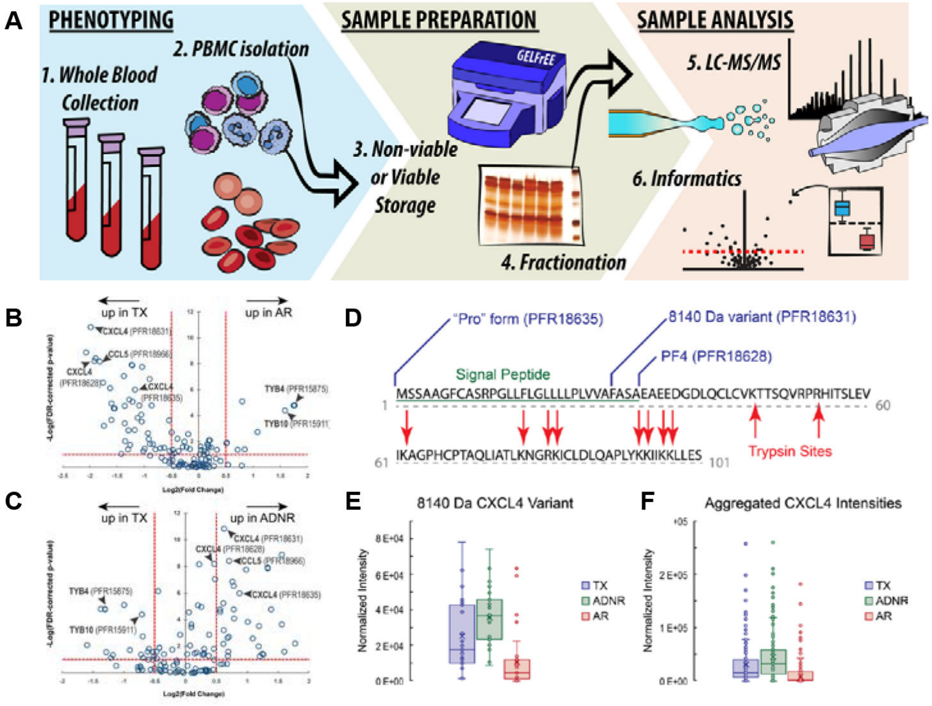


Figure 6.

(A) Schematic workflow for discovery-mode, translational, top-down proteomics applied to liver transplant patient groups (acute rejection [AR]; transplant excellent [TX]; acute dysfunction, no rejection [ADNR]) from blood collection to informatics analysis. Representative label-free, top-down, quantitative analysis describing differentially expressed proteoforms between TX and AR (B) and between TX and ADRN (C) patient group, respectively. For all proteoforms detected in the majority of data files across the data set (open circles), ANOVA was used to assign variation in signal intensity to phenotype-specific effects after accounting for patient-to-patient and technical variation. The *x*-axis represents the effect size as measured by fold-change ($\log_2$transformed) between patient groups. The *y*-axis (FDR-corrected p-value) is a measure of the statistical confidence that signal variation is associated with phenotype. The dashed lines represent our arbitrary thresholds for delineating significant hits: The horizontal dashed line corresponds to a 5% FDR, and vertical dashed lines represent effect sizes 1.4-fold above and below no change. (D-F) Proteoform-resolved analysis of PF4/CXCL4 differentially expressed proteoforms characterized by top-down proteomics and an *in silico* comparison to tryptic peptide-based approaches. (D) The full-length canonical sequence of PF4/CXCL4 (accession no. P02776). Blue flags depict the cleavage sites of the three differentially abundant PF4/CXCL4 proteoforms and are labeled with their identity. The sequence underlined in green is the signal peptide, and red arrows delineate trypsin cleavage sites generated *in silico*. Notably, tryptic peptides do not span the region of sequence variability describing the three proteoforms of interest in this study. (E) Box-and-whisker plot comparison of the 8140-Da PF4/CXCL4 proteoform intensities across all patients and injections, which were found to be significantly decreased in AR patients. (F) Box-and-whisker plots made from aggregating all PF4/CXCL4 proteoform intensities per patient group to emulate a quantitative comparison using intensities of tryptic peptides, which cannot distinguish the proteoforms. Notably, the effect size is lost to noise in this *in silico* experiment, and the analysis would return a false negative by bottom-up proteomics. For the box-and-whisker plots, data points represent the normalized intensities of the proteoform of interest yielded from every technical replicate (data file) per patient in which the proteoform was detected (TX: n = 8 patients, 31 data files; ADNR n = 9 patients, 31 data files; AR: n = 9 patients, 33 data files). Adapted and reproduced from Proteoforms in Peripheral Blood Mononuclear Cells as Novel Rejection Biomarkers in Liver Transplant Recipients, Toby, T. K.; Abecassis, M.; Kim, K.; Thomas, P. M.; Fellers, R. T.; LeDuc, R. D.; Kelleher, N. L.; Demetris, J.; Levitsky. Am. J. Tranplant., Vol. 17, Issue 9 (ref XXX). Copyright 2017 Wiley

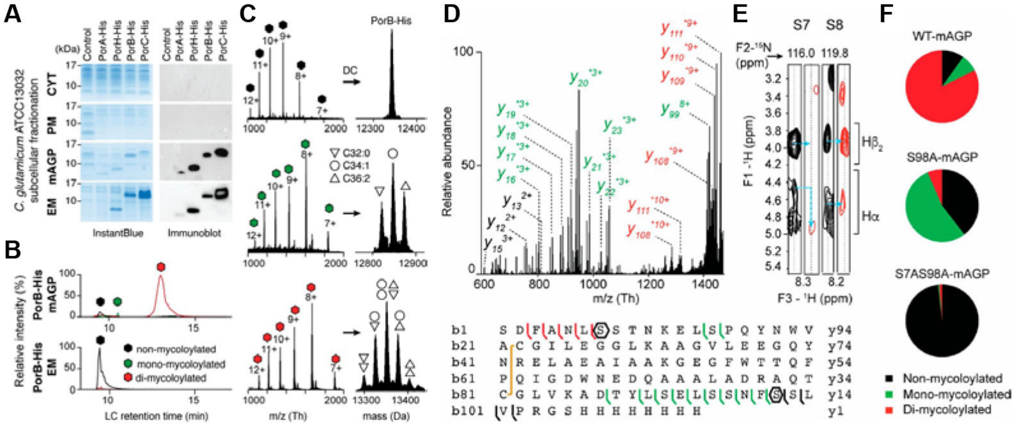


Figure 7.

Identification and characterization of distinct proteoforms for C. glutamicum OMPs associated with the mAGP complex and secreted in the extracellular medium. (A) C. glutamicum ATCC13032 cells expressing recombinant PorA-His, PorHHis, PorB-His, and PorC-His were cultured under identical conditions. Subcellular fractions corresponding to the CYT, the PM, the mAGP complex, and the extracellular medium were analyzed by SDS/PAGE after staining with InstantBlue (Left) and Western blotting (Right) with antibodies against the protein His tag. Fractions isolated from WT, untransformed cells were coanalyzed as control. Molecular mass markers (in kilodaltons) are indicated next to the gel. (B) Extracted ion chromatograms of PorB-His purified from mAGP (Upper) and extracellular medium (Lower) fractions containing nonmycoloylated (black), monomycoloylated (green), and dimycoloylated (red) proteoforms. (C) Representation of the multicharged MS spectra (Left) and deconvoluted (DC) spectra obtained for PorB-His proteoforms with isotopic resolution (Right). The mycolic acid compositions of each proteoform are indicated by triangle and circle symbols. EM, extracellular medium. Example of PorB. (D) Top-down CID of dimycoloylated PorB-His-10+ charge state (m/z 1,336.10 Th) with nonmycoloylated, monomycoloylated, and dimycoloylated y and b fragments colored black, green, and red, respectively. The sequence coverage was obtained by fragmenting the 8+, 9+, and 10+ charge states of PorB-His, identifying S98 and S7/ S8 residues (polygons) as putative mycoloylation sites and a disulfide bond between C22 and C81 (yellow line). (E) Solution NMR analysis of nonmycoloylated (black) and mycoloylated (red) PorB-His. Selected strips extracted from 3D 1 H, 15N, 1 H heteronuclear single quantum coherence–total correlation spectroscopy (HSQC-TOCSY) spectra obtained on (U-15N)-labeled PorB-His showing Hα and Hβ2 chemical shifts of S7 and S8 residues for the two proteoforms. Although 1 H resonances from S8 were not affected, significant spectral changes were observed for Hα of residue S7 (blue arrows), thus identifying the O-acylation of the S7 hydroxyl of PorB-His. (F) The positions of PTM within the protein sequence were validated by site-directed mutagenesis of S7 and S98 residues and subsequent MS analysis of PorB-His WT (WT-mAGP; Top) and its mutant derivatives PorB-S98A (S98A-mAGP; Middle) and PorB-S7AS98A (S7AS98A-mAGP; Bottom). Nonmycoloylated (black), monomycoloylated (green), and dimycoloylated (red) proteoforms were semiquantified from extracted ion chromatograms of the corresponding 9+ charge states. Adapted and reproduced with permission from *Proceedings of the National Academy of*

Sciences USA Carel, C.; Marcoux, J.; Reat, V.; Parra, J.; Latge, G.; Laval, F.; Demange, P.; Burlet-Schiltz, O.; Milon, A.; Daffe, M.; Tropis, M. G.; Renault, M. A. M. Proc. Natl. Acad. Sci. U.S.A. 2017, 114, 4231–4236 (ref # XXX)

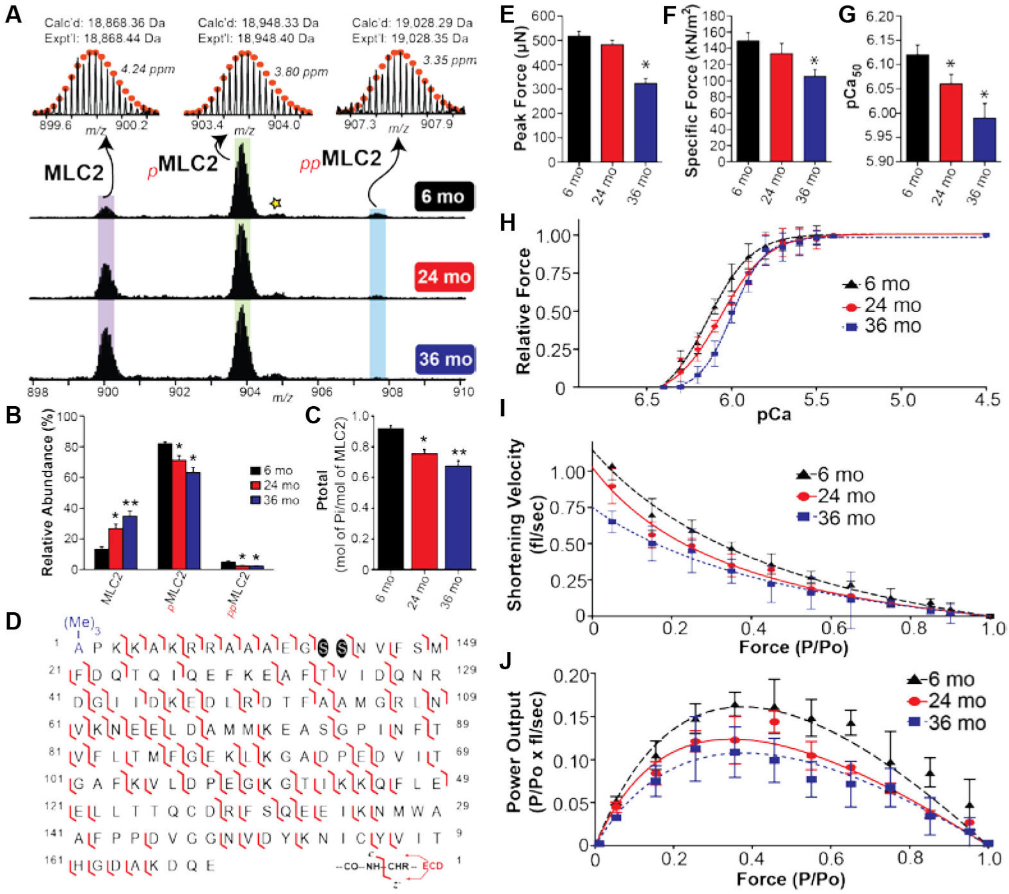


Figure 8.

Progressive decrease in the phosphorylation of fast skeletal RLC is observed with advancing age. (A) Representative mass spectra of RLC from 6, 24, and 36 month old rats. Circles represent the theoretical isotopic abundance distribution of the isotopomer peaks corresponding to the assigned monoisotopic mass. Star represents oxidized pRLC. *m/z*, mass-to-charge ratio. Calc'd, calculated monoisotopic molecular mass based on protein sequence. Expt'l, experimentally determined molecular mass. (B) Graph showing the relative abundances of RLC, pRLC, and ppRLC proteoforms in the gastrocnemius muscle of rats from different age groups. (C) Graph showing decrease in total RLC phosphorylation (expressed as mol Pi / mol of RLC) with advancing age. n = 5 for 6 and 24 month old groups, and n = 6 for 36 month old group. mo, month. All values represent mean ± SEM *p < 0.05 versus 6 mo, **p < 0.001 versus 6 mo. (D) Representative fragmentation map for ppRLC proteoforms. Phosphorylation sites are highlighted by circles. (Me)$_3$- represents Nα - trimethylation. Reproduced from Gregorich, Z. R.; Peng, Y.; Cai, W. X.; Jin, Y. T.; Wei, L. M.; Chen, A. J.; McKiernan, S. H.; Aiken, J. M.; Moss, R. L.; Diffee, G. M.; Ge, Y. J. Proteome Res. 2016, 15, 2706−2716 (ref xxx). Copyright 2016 American Chemical Society.