Published in final edited form as:

Expert Rev Proteomics. 2016 August ; 13(8): 717–730. doi:10.1080/14789450.2016.1209414.

Top-down Proteomics: Technology Advancements and Applications to Heart Diseases

Wenxuan Cai^{a,b}, Trisha M. Tucholski^c, Zachery R. Gregorich^{a,b}, and Ying Ge^{a,c,d,*}

^aDepartment of Cell and Regenerative Biology, University of Wisconsin-Madison, 1111 Highland Ave., Madison, WI 53705

^bMolecular and Cellular Pharmacology Training Program, University of Wisconsin-Madison, 1111 Highland Ave., Madison, WI 53705

^cDepartment of Chemistry, University of Wisconsin-Madison, 1101 University Ave., Madison, WI 53706

^dHuman Proteomics Program, University of Wisconsin-Madison, 1111 Highland Ave., Madison, WI 53705

Abstract

Introduction—Diseases of the heart are a leading cause of morbidity and mortality for both men and women worldwide, and impose significant economic burdens on the healthcare systems. Despite substantial effort over the last several decades, the molecular mechanisms underlying diseases of the heart remain poorly understood.

Areas covered—Altered protein post-translational modifications (PTMs) and protein isoform switching are increasingly recognized as important disease mechanisms. Top-down high-resolution mass spectrometry (MS)-based proteomics has emerged as the most powerful method for the comprehensive analysis of PTMs and protein isoforms. Here, we will review recent technology developments in the field of top-down proteomics, as well as highlight recent studies utilizing top-down proteomics to decipher the cardiac proteome for the understanding of the molecular mechanisms underlying diseases of the heart.

Expert commentary—Top-down proteomics is a premier method for the global and comprehensive study of protein isoforms and their PTMs, enabling the identification of novel protein isoforms and PTMs, characterization of sequence variations, and quantification of disease-associated alterations. Despite significant challenges, continuous development of top-down proteomics technology will greatly aid the dissection of the molecular mechanisms underlying diseases of the hearts for the identification of novel biomarkers and therapeutic targets.

*To whom correspondence should be addressed: Dr. Ying Ge, 1300 University Ave., SMI 130, Madison, WI 53706, USA, ge2@wisc.edu.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Keywords

Top-down Mass Spectrometry; Heart Diseases; Post-translational Modifications; Quantitative Analysis; Myofilaments

1. Introduction

Diseases of the heart are a leading cause of morbidity and mortality worldwide, and are associated with significant economic costs [1–4]. In 2011, over 600,000 deaths in the United States were caused by diseases of the heart, and the annual economic cost of these diseases in the United States is estimated to be roughly \$215.6 billion [1]. Nevertheless, despite substantial effort over the past decades, the molecular mechanisms underlying diseases of the heart remain elusive. As a consequence, current therapeutic approaches for treating patients with heart disease focus primarily on symptom management and cannot alter disease progression. Thus, there is an urgent need for new methods enabling delineation of the molecular mechanisms underlying diseases of the heart for the purposes of identifying novel disease biomarkers and therapeutic targets.

The field of systems biology aims to obtain a holistic, global, and integrated understanding of the molecular constituents of biological systems, including nucleic acids, proteins, and metabolites, as well as their interactions, and, thus, holds great promise for gaining mechanistic insights into a variety of diseases, including diseases of the heart [5,6]. Following sequencing of the human genome, it became clear that genomic information alone was insufficient to understand biological functions. Therefore, in the post-genomic era, proteomics, which focuses on the analysis of proteins, the primary effectors of all cellular functions, has come to the forefront of systems biology [7]. In contrast to the genome, which is considered to be largely static, the proteome is highly complex and exhibits considerable plasticity owing to a myriad proteoforms—a term encompassing the plethora of protein species produced by a single gene via the alternative splicing of mRNA transcripts and protein post-translational modifications (PTMs) [8]—and highly homologous protein isoforms produced by multiple homologous genes (Figure 1A) [6,9]. In particular, protein PTMs play critical roles in the regulation of protein localization, activity, and degradation; and can be altered in disease [10,11]. Moreover, the differential expression of protein isoforms is increasingly recognized as an important molecular mechanism contributing to disease [12–14].

Mass spectrometry (MS)-based proteomics has quickly become the method of choice for the quantitative analysis of proteoforms and cognate isoforms produced by different genes [9]. MS-based proteomics analysis has traditionally been carried out using the “bottom-up” approach, wherein protein samples are digested into small peptides, separated using liquid chromatography (LC), and analyzed by MS (Figure 1B). While bottom-up proteomics is widely-used and enables the identification and quantification of thousands of proteins, this approach suffers from intrinsic limitations resulting from protein digestion, which makes it sub-optimal for the identification of protein isoforms (particularly those with high sequence homology), polymorphisms/mutations, and PTMs [15]. On the other hand, an emerging

“top-down” proteomics approach, in which intact proteins are analyzed rather than peptides, provides a global or “bird’s eye” view of all proteoforms and cognate isoforms produced by different genes and, thus, represents a superior approach for the comprehensive analysis of proteins and their post-translationally modified forms [9,15–19]. Following intact protein MS analysis, specific proteoforms/isoforms of interest can be isolated in the gas phase and subjected to tandem MS (MS/MS) analysis to obtain sequence information that can be used to identify proteins and localize PTMs and sequence variations. Furthermore, top-down MS is semi-quantitative, permitting the relative quantification of proteoforms present within the same spectrum [18,20,21].

Here, we will review recent advancements in the top-down proteomics methodology, including developments in sample preparation, protein separation, MS instrumentation, protein characterization, and quantification, as well as tools for top-down data analysis. Moreover, we will summarize recent applications of top-down proteomics to study the cardiac proteome towards a better understanding of the molecular mechanisms underlying diseases of the heart.

2. Top-down Proteomics Methodology

2.1 Sample preparation for top-down proteomics

Traditional sample preparation methods in biomedical research typically involve the use of high concentrations of salt (>100 mM) and/or surfactants, such as SDS or Triton X-100, for the efficient solubilization of proteins, particularly highly hydrophobic proteins [9,16,22–24]. However, most of the commonly used salts and surfactants are incompatible with protein analysis by MS as salt inhibits electrospray ionization (ESI) and surfactants can out-compete proteins for ionization and/or form multiply adducted species that are detrimental to the MS signal intensity of the proteins [9,16,24,25]. Therefore, it is critical to remove MS-incompatible salts and surfactants, or replace them with MS-compatible ones, before MS analysis.

Protein samples containing high concentrations of MS-incompatible salts can be desalted using reverse phase chromatography (RPC) [17,18,26]. Another option is to replace non-volatile MS-incompatible salts with volatile salts (e.g., ammonium acetate), which are compatible with MS analysis. The exchange of MS-incompatible for MS-compatible salts can be achieved either by substituting the compatible salt for the incompatible salts in the extraction buffer, or post-extraction using dialysis or ultrafiltration [16,27,28]. In addition to commonly employed ammonium salts, it was recently shown that ammonium tartrate is also compatible with MS. Xiu et al. demonstrated high-resolution separation of intact proteins using hydrophobic interaction chromatography (HIC) employing mobile phases containing ammonium tartrate [29].

Surfactants can be generally categorized into ionic and non-ionic; with the former being extremely incompatible with MS due to severe suppression of protein ionization. Conversely, non-ionic surfactants are compatible with MS at low concentrations [28]. The seminal work by the Robinson group identified a number of MS-compatible non-ionic surfactants that can be used for the analysis of intact membrane proteins and protein

complexes [28]. Nevertheless, protein solubilization using non-ionic surfactants is limited in comparison to that achieved with ionic surfactants and, as a result, ionic surfactants are still routinely employed during sample preparation. These MS-incompatible surfactants are then frequently removed by protein precipitation in organic solvents [30]. However, this method suffers from low recovery, and, therefore, considerable effort has been invested in the development of degradable surfactants that can be broken down prior to MS analysis. To date, the majority of the degradable surfactants that have been developed, including ProteaseMAX (Promega), RapiGest (Waters), and MaSDeS [25], are acid-labile and are degraded during LC separation due to the presence of acid in the mobile phase. Although these surfactants have shown significant efficacy in bottom-up proteomics, the successful use of degradable surfactants for intact protein analysis has not been demonstrated.

Hydrophobic proteins, such as membrane proteins, represent a particularly difficult group of proteins to analyze using top-down MS. In contrast to bottom-up proteomics, wherein some soluble peptides from membrane proteins can be recovered following protein digestion, permitting their identification and quantification, the analysis of intact membrane proteins is extremely challenging due to aggregation and precipitation of these proteins in the absence of surfactants. Despite the challenges, a number of research groups have successfully utilized a variety of systems to study intact membrane proteins and complexes *in vitro*, including detergent micelles, bicelles, amphipols, and nanodiscs [31]. As mentioned above, non-ionic surfactants were successfully used for the analysis of membrane proteins and protein complexes by the Robinson group. Membrane proteins or protein complexes can be prepared in buffer containing low concentrations of MS-compatible surfactants, which are subsequently removed in the gas phase by thermal activation. Removal of the surfactants permits MS analysis of the ionized protein or protein assembly [28]. The solubilization of membrane proteins for MS analysis can also be achieved using detergent-free systems. Bicelles and nanodiscs employ small, discoidal arrangements of phospholipids to form lipid bilayers that mimic native membrane structure to deliver membrane proteins for MS analysis [32]. Amphipols are amphipathic polymers with alternating hydrophilic and hydrophobic side chains, which can also be used for the solubilization of hydrophobic proteins. Proteins or protein assemblies can be released in the gas phase from bicelles, nanodiscs, or amphipols by collisional activation similar to the removal of non-ionic surfactants [32]. Although collisional activation aids in the removal of surfactants and lipid-like molecules, over-activation can lead to local unfolding of protein structure, dissociation of protein complexes, and even protein fragmentation or the loss of labile PTMs such as phosphorylation or glycosylation. Alternatively, investigators can exploit the chemical properties of certain proteins to maintain protein solubility in the absence of MS-incompatible surfactants or lipid-like molecules. High concentration of formic acid have proved efficacious for the solubilization of membrane proteins for top-down MS analysis [33–35]. However, prolonged storage of protein samples in formic acid can result in protein degradation and artefactual modifications (formylation), and, therefore, immediate solvent transfer to a more inert solution is needed [33]. Similarly, high concentrations of organic solvents containing MS-compatible chaotropes have also been used for the extraction of membrane proteins from purified mitochondria for top-down proteomics analysis [22]. Despite recent advancements in protein solubilization, the solubility of intact proteins, particularly membrane proteins,

remain a significant challenge that awaits further development of novel reagents and methods.

2.2 Protein purification and separation

The complexity and high dynamic range of the proteome necessitates the effective fractionation and separation of proteins prior to MS analysis [36]. Affinity purification (AP) remains a popular approach for the targeted analysis of intact proteins and protein complexes. In addition to AP for hypothesis-driven targeted proteomics, there exists a variety of gel-based and LC-based methods for large-scale protein separation for discovery-driven research.

Traditional biomedical research has heavily relied on AP for the selective purification of a single protein or protein complex of interest. AP is an effective and robust means for the targeted study of proteins with known biological relevance. The Ge group has extensively utilized AP for the purification of cardiac troponin I (cTnI) from animal and human cardiac tissue for the study of disease-associated cTnI proteoform alterations using the top-down approach [17,26,37–39].

Nevertheless, AP usually involves the use of MS-incompatible salts that need to be removed prior to MS analysis, making AP difficult to couple directly with MS. Additionally, as a targeted approach for the purification of specific proteins of interest, AP is limited by the availability and fidelity of the analyte-specific stationary phase (e.g., antibodies) [9]. Furthermore, AP is relatively low throughput and provides a very limited view of complex biological systems such that many important/disease-associated changes in the system may be overlooked. Therefore, the development of high-throughput and high-resolution protein fractionation and separation methods for the interrogation of the entire proteome or large sub-proteomes remains an active area of research.

Gel-based methods often exploit size or isoelectric point for the separation of proteins, and are widely used in biomedical research. Proteins separated in gels can be digested and analyzed by the bottom-up approach due to the relatively high recovery of small peptides from acrylamide gels. However, the recovery of intact proteins from the gel matrix remains extremely difficult. Notably, the development of Gel-Eluted Liquid Fraction Entrapment Electrophoresis (GELFrEE) has allowed for multiplexed, high-resolution size-based separation of intact proteins [30,40]. Tran et al. were able to identify over 3000 proteoforms from a HeLa cell lysate by employing solution isoelectric focusing followed by GELFrEE and online nanocapillary RPC-MS [41]. This multi-dimensional platform demonstrated the power of separation based on a number of different physiochemical properties, including isoelectric point, size, and hydrophobicity, for the analysis of intact proteins from a complex cell lysate. Using this method, Catherman et al. identified 1220 proteins, approximately 25% of which were membrane proteins, and over 5000 proteoforms from the H1299 human cell line [42], which represents the largest top-down proteomics analysis conducted to date. Nevertheless, despite the high-resolution protein separation that can be achieved using the GELFrEE system, this method requires the use of SDS, which, as mentioned above, is not compatible with MS and must be removed prior to MS analysis.

An appealing alternative to gel-based methods are LC-based methods, including RPC, HIC, ion-exchange chromatography (IEC), and size exclusion chromatography (SEC). A significant advantage of LC-based separation methods is that they can be coupled directly with MS [43]. Online LC-MS not only increases the overall throughput of sample analysis, but also reduces potential artefactual modifications associated with lengthy sample handling. RPC makes use of a hydrophobic stationary phase that proteins will interact with when low concentrations of organic solvents are included in the mobile phase. The analytes are subsequently eluted using increasing concentrations of organic solvents. Due to the high compatibility of RPC with MS, RPC remains a popular choice for intact protein separation coupled with MS. Using RPC alone with a long capillary column (80 cm) and an extended gradient, Ansong et al. were able to identify 563 proteins and 1665 proteoforms from *Salmonella typhimurium* [44]. HIC and IEC utilize buffered salt concentration gradients for protein separation based on hydrophobicity and electrostatic interactions, respectively. In HIC, a gradient with a decreasing concentration of salt is used to separate proteins based on hydrophobic interactions. Commonly used salts for achieving high-resolution separation of proteins by HIC contains sulfate, phosphate, and citrate, which are incompatible with MS [45]. Ammonium acetate is an MS-compatible salt, but HIC separation using ammonium acetate suffers from low separation power due to inadequate protein retention when used with conventional HIC materials [45]. Chen et al. recently evaluated a series of HIC columns with heightened degrees of hydrophobicity to compensate for the inadequate retention of proteins by the stationary phase when using ammonium acetate in the mobile phase. They demonstrated improved protein retention using columns with stationary phases that were more hydrophobic than conventional HIC stationary phases and a mobile phase containing ammonium acetate, which opens the door for the use of this powerful high-resolution separation method for top-down proteomics analysis [45]. Importantly, HIC and RPC are orthogonal with different selectivity for intact protein separation [46], despite the fact that both methods separate proteins based on hydrophobicity. Moreover, HIC is a non-denaturing separation method and holds great promises for the study of native protein confirmation, as well as protein-protein and protein-ligand interactions. In addition to HIC, recent work using online IEC coupled with MS also demonstrated great success in the study of native proteins and protein complexes [47]. Nevertheless, detection of high molecular weight (MW) proteins remains challenging due to an exponential decay in the signal-to-noise ratio (S/N) for proteins with increasing molecular weight [48], and, therefore, large proteins, particularly those co-eluting with low MW proteins, are often difficult to detect. As a consequence, size-based separation of proteins prior to MS analysis is necessary for the detection of high MW proteins. SEC is a highly versatile method with limited sample loss due to reduced interaction with stationary phase [49]. SEC can also utilize a variety of mobile phase solvents, both denaturing and non-denaturing for the analysis of proteins and protein interactions. Nevertheless, SEC traditionally suffers from low resolution and undesirable sample dilution due to the use of high volumes of solvent [36]. To address this, Chen et al. used ultra-high pressure SEC, and demonstrated high-resolution, high-efficiency separation of proteins in a wide mass range (6–669 kDa) in only 7 minutes [50].

In the analysis of complex protein mixtures, great separation power can be achieved by multi-dimensional LC platforms combining two or more orthogonal LC methods. Due to the

high compatibility of RPC with MS, RPC remains a popular choice for intact protein separation and is often used in combination with other orthogonal LC methods for multi-dimensional separation. Notably, the Paša-Toli lab coupled weak anion exchange (a subtype of IEC) to RPC-MS. This study used RPC with a long C5 nanoLC column (80 cm) coupled directly to a 12 Tesla Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer, and identified 750 proteins from *Shewanella oneidensis* lysates [51]. Valeja et al. employed a three-dimensional LC platform coupling IEC, HIC, and RPC with an Orbitrap mass spectrometer, and were able to identify 640 proteins in a single IEC fraction (out of 35 fractions) from a HEK 293 cell lysate [46]. There has been significant progress for intact protein separation; however, the separation power needs to further improve to address the extreme complexity and dynamic range of the human proteome. The development of separation methods that can be coupled online with MS and allow for the orthogonal and high-resolution separation of intact proteins will be essential for the advancement of high-throughput top-down proteomics.

2.3 Top-down MS instrumentation

A typical mass spectrometer consists of a sample inlet, ion source, mass analyzer, and detector. MS analysis relies on the efficient ionization of molecules of interest, and the development of “soft-ionization” methods, including matrix-assisted laser desorption ionization (MALDI) [52] and electrospray ionization (ESI) [53], which do not induce fragmentation or dissociation of intact proteins or protein complexes, has significantly advanced the MS analysis of biomacromolecules. For the analysis of intact proteins, ESI is generally more favorable due to the resulting multiply charged ions produced using this ionization method [48,54]. In contrast, MALDI produces predominantly singly or doubly charged ions. The multiply charged ions generated by ESI aid in the detection of proteins in mass analyzers with limited mass (m/z) range. In addition, multiply charged ions can be more effectively fragmented using electron-based fragmentation methods.

For the analysis of large intact proteins, high-resolution mass analyzers are desired because high-resolution instruments allow for highly accurate determination of the protein mass [55]. In addition, following MS/MS analysis, the resulting spectra are usually highly complex and contain a significant number of overlapping fragment ion peaks, which require high-resolution to differentiate. Therefore, early work on the MS analysis of intact proteins was almost exclusively conducted using FT-ICR mass spectrometers [56–58]. One difficulty of using FT-ICR mass spectrometers is that they require superconducting magnets, which can be difficult and costly to maintain, and, therefore, top-down technology was limited to only a few specialized labs. Advancements in instrumentation have now opened the field of top-down proteomics to other research groups. For example, time-of-flight (TOF) instruments have been used for the study of intact proteins and native protein complexes [28,32]. In fact, TOF instruments are typically preferred for the study of very large protein complexes due to the essentially unlimited mass (m/z) range of these instruments [31]. However, the old-generation TOF detectors, such as electron multipliers and microchannel plates, are suboptimal for the detection of high MW proteins due to the low kinetic energy of these molecules in the drift tube, which decreases the probability of generating secondary electrons and dramatically reduces the detection efficiency of high MW molecules [59].

With the nanomembrane TOF detector developed recently, the kinetic energy of the ions generates mechanical oscillations in the nanomembrane thereby significantly improving the detection of large proteins or protein assemblies [60]. In addition, the hybridization of TOF instruments with quadrupole mass analyzers (q-TOF), which can act as a mass filter or collision cell, has greatly improved the capabilities of these instruments [31]. Recent work by Mann and coworkers has demonstrated the high resolving power and mass accuracy of commercially available q-TOF instruments [61], making them an affordable and robust option for top-down protein analysis. As an alternative to FT-ICR with relatively low cost and maintenance, the newest generation of the Orbitrap mass spectrometers [62,63] can also offer good mass accuracy and high resolving power. The Kelleher group has demonstrated identification of thousands of intact proteins using Orbitrap mass spectrometers [64,65]. Nevertheless, FT-ICR mass spectrometers still remain the preferred instrument for achieving isotopic resolution of proteins that are larger than 70 kDa [56,66]. Importantly, the new dynamically harmonized FT-ICR cells developed by Nicolaev and coworkers provide significant improvements on the stability of the transient, which enhances resolution and makes it easier to achieve isotopic resolution of large proteins [67]. Significant advancements in MS instrumentation have been made in the past decades, but large-scale top-down proteomics analysis of proteins, especially high MW proteins, remains challenging. Further developments in instrumentation, including enhancements in resolution and sensitivity over a broad mass range, will significantly advance the field of top-down proteomics.

2.4 Top-down protein characterization

Following MS analysis of intact proteins, specific protein species of interest can be isolated in the mass spectrometer and fragmented using various fragmentation methods. In general, these fragmentation methods can be classified as either energetic or electron-based dissociation methods [68,69]. Energetic dissociation methods, such as collision induced dissociation (CID) and infrared multiphoton dissociation, dissociate bonds of the lowest activation energy. In the case of unmodified proteins/peptides, energetic dissociation results in cleavage of the CO-NH bonds in the protein backbone, yielding *b* and *y* fragment ions (Figure 2A). However, when labile PTMs (e.g., phosphorylation or glycosylation) are present, dissociation using energetic fragmentation methods preferentially cleaves the covalent bonds attaching these modifications to the protein, resulting in PTM loss and preventing localization of the modification (Figure 2A) [9]. On the other hand, the electron-based fragmentation methods, including electron capture dissociation (ECD) [70] and electron transfer dissociation (ETD) [71], induce local fast cleavage of N-C α bonds in the protein backbone, yielding *c* and *z*. fragment ions, and preserving labile PTMs (Figure 2B). In addition, electron-based dissociation methods are non-selective and tend to produce cleavage that is randomly distributed along the protein backbone, which can increase sequence coverage and aid protein identification and the localization of PTMs [9,57].

Although ESI is generally preferred due to the production of multiply charged ions, the use of ESI results in dispersion of the protein signal over a large number of charge states and causes detrimental reduction of the S/N, especially for high MW proteins [48]. The development of in-spray supercharging ESI following LC separation allows for preferential

formation of ions with higher charge states [72], which significantly increases S/N of the proteins, and holds great promise for the characterization of large proteins using electron-based dissociation methods. In addition to supercharging reagents for reducing charge state dispersion of protein signals, another technology, ion parking via ion/ion proton transfer reaction [73], has proved to be particularly useful in reducing the number of charge states of high MW proteins, and has recently gained significant attention [74]. Ion parking allows for the concentration of protein signal over a small number of charge states, which allows for detection of low abundant and high MW proteins, as well as gas-phase “purification” of the proteoforms of interest for comprehensive sequence characterization [74].

However, efficient fragmentation of large proteins (particularly those >50 kDa) remains challenging due to the fact that extensive intramolecular interactions promote the formation of highly compact gas phase structures that are refractory to fragmentation using conventional dissociation methods [58]. To enhance the fragmentation of large proteins, Han *et al.* utilized increased pre- and post-skimmer voltages and electrospray additives to reduce structural rearrangement of ions in the gas-phase and facilitate the fragmentation of large proteins [58]. Using their “pre-folding dissociation” strategies, the authors were able to achieve high coverage of fragmentation at the termini of large proteins up to 200 kDa [58]. More recently, the Brodbelt group has demonstrated enhanced sequence coverage of intact proteins using ultraviolet photodissociation (UVPD) in comparison to conventional fragmentation methods, including CID, high energy collision dissociation, and ETD [75]. UVPD with a 193 nm laser yields extensive fragmentation producing predominantly *a*-*x* ions and a significant amount of *b*-*y* and *c*-*z*. ions (Figure 2C), as well as other ions of side-chain cleavage [75]. Additionally, unlike ECD/ETD, which exhibit significant charge state dependence of fragmentation, fragmentation efficiency with UVPD is relatively independent on the charge state of the precursor ions. High-resolution mass spectrometers implemented with UVPD presents a powerful tool for the analysis of intact proteins and protein complexes.

Nevertheless, fragmentation of large proteins produces very complex tandem mass spectra with a mixture of ions of various molecular weight and charge states that quite often overlap. Therefore, mass spectrometers with high resolution and sensitivity are needed for the analysis of complex product ions resulted from fragmentation of high molecular weight proteins. In addition, ion parking after MS/MS analysis allows for reduction of the ion species as well as concentration of the ion signals [74], which makes it an appealing option for the characterization of high molecular weight proteins.

2.5 Top-down protein quantification

Unlike in bottom-up proteomics, where the addition of modifications to peptides can significantly influence ionization, the physiochemical properties of intact proteins are relatively unaffected by the addition of small modifications or minor sequence variations [76], which allows for relative quantification of modified and unmodified proteoforms present within the same spectrum. Relative quantification of protein PTMs and isoforms using top-down MS has been widely employed by our lab and others for uncovering disease-associated proteoform alterations and identifying biomarkers [17,18,21,44,76–78]. Using

this quantification method, in a large-scale top-down proteomics analysis of pathogenic bacterium *Salmonella typhimurium*, Ansong et al. identified and quantified 1665 bacterial proteoforms, and revealed a unique protein PTM switch from predominantly protein S-cysteinylation to S-glutathionylation under infection-like conditions [44]. Similarly, Chamot-Rooke et al. employed top-down MS to investigate the PTMs of the intact pilin of *Neisseria meningitidis* that causes cerebrospinal meningitis [78]. They discovered that enhanced glycerophosphorylation of pilin was responsible for the destabilization of type IV pili fiber bundles and enhanced bacterial detachment and migration across epithelial cells [78]. In addition to infectious diseases, top-down proteomics also revealed a decreased in the phosphorylation of cTnI in chronic heart failure [18].

Although relative quantification of proteoform abundance changes aids in dissecting the molecular mechanisms underlying various diseases, including heart diseases and infectious diseases, this method does not permit quantification of protein expression changes across samples. To quantify protein expression changes across samples, substantial efforts have been invested for developing methods to label intact proteins. Early work employing stable isotopic labeling by amino acids (SILAC) demonstrated reliable quantification of intact proteins [79,80] in a small scale. Chemical labeling such as isobaric tags for relative and absolute quantitation (iTRAQ) allows for quantification of protein expression in samples other than cell culture or mouse models, and has been widely used in bottom-up proteomics. Adapted from iTRAQ, metal-coded affinity tags (MeCAT) were developed for the quantitative analysis of intact proteins [81]. Although SILAC and MeCAT have been successfully implemented for the quantitative analysis of protein expression changes in biological samples, these methods may suffer from detrimental deterioration of protein signals due to signal spreading with partial incorporation of the isotopic labels or metal-coded tags. On the other hand, Mazur et al. utilized label-free differential MS for the quantification of intact apolipoproteins in human high density lipoproteins (HDL). By carefully monitoring instrument response and sample load, Mazur et al. showed an average 4.7 (p-value 0.017) increase of O-glycosylated apolipoprotein C-III in the low versus high plasma HDL donors [82]. This study demonstrated the feasibility of using top-down MS for reliable quantification of protein abundance in a small scale. Taking this further, the Kelleher group employed label-free MS for the quantification of intact proteins less than 30 kDa in the growing and senescent fibroblasts, and confidently quantified 751 proteoforms, among which a proteoform of the enhancer of rudimentary homologue and Isoform 1 of FAM107B were down-regulated in the senescent cells [65]. Nevertheless, this study is limited due to the biased detection of low MW proteins. High-throughput large-scale quantification of protein expression level by top-down MS remains a significant challenge, and, thus, presents significant opportunities for future top-down research.

2.6 Top-down data analysis

Greatly aided by the well-developed computational tools for data analysis, bottom-up proteomics has gained considerable popularity for protein analysis. On the contrary, data analysis tools for top-down proteomics remain relatively under-developed. The use of high-resolution mass spectrometers and the resulting complex mass spectra necessitates sophisticated tools for the deconvolution of top-down mass and tandem mass spectra.

Thorough High-Resolution Analysis of Spectra by Horn (THRASH) [83] was the first automatic computational tool developed for the deconvolution of high-resolution mass spectra, and is still widely used. THRASH compares the experimental isotopomer envelope with a theoretical envelope generated based on the average [84], computes a figure of merit value for each envelope and reports those score higher than a user-defined fit score. MS-Deconv, which adopts a combinatorial approach for the deconvolution of overlapping isotopomer envelopes [85], has also been applied for top-down data analysis. Recently, UniDec, which employs a Bayesian approach for the separation of mass and charge dimension [86], can also be used for the deconvolution of high-resolution spectra. Although these algorithms significantly facilitate data processing and interpretation, the complexity of high-resolution top-down MS and MS/MS data often results in a considerable amount of mis-assigned peaks necessitating the implementation of visualization functions for the validation of the deconvolution results. Decon2LS uses a variation of THRASH for high-resolution spectral analysis and visualization [87]. The DataAnalysis (Bruker Daltonics) utilizes SNAP2, a THRASH-based algorithm, for spectral deconvolution, and allows for direct visual comparison of the experimental isotopomer envelopes with theoretical patterns (termed SNAP pattern). Although these tools allow for the visual comparison of the experimental isotopomer envelopes with the theoretical ones for the determination of monoisotopic mass, manual adjustment of the computational output for the correction of charge state and mass error is not supported.

Spectral deconvolution of tandem mass spectra produces a mass list, which allows for protein identification. There are a number of software tools for protein identification in top-down proteomics, including ProSight PC [88], Mascot Top Down (also known as Big-Mascot) [89], MS-TopDown [90], and MS-Align+ [91]. Both ProSight PC and Mascot Top Down require specific PTMs to be defined prior to database search, which limits their utility in the identification of unexpected PTMs. MS-TopDown and MS-Align+ both employ the spectral alignment algorithm, which matches MS/MS data to proteins in the database, and allows for the identification of mass discrepancy between the experimental and theoretical fragment ions, which represents PTMs and/or sequence variations. In particular, MS-Align+ enables efficient filtering of candidate proteins when searches in a large protein database, and, therefore, allows for rapid identification of proteins and unexpected PTMs [91]. MS-Align+ also provides statistical evaluation of the identification results [91]. Recently, an upgraded version of MS-Align+, TOPPIC, was developed for faster protein identification and characterization with less computing requirements.

Recently, the Ge group has developed an integrated software tool, MASH Suite Pro [92], which incorporates both THRASH and MS-Deconv for spectral deconvolution, MS-Align+ for protein identification, as well as a number of in-house developed functionality for protein quantification and characterization. Moreover, MASH Suite Pro provides easy access to the real spectra and visualization of each deconvoluted mass, with additional features allowing for deletion of false positives and correction of the mass values or charge states when mistakes occur. The visualization features and flexibility for correction of computational outputs significantly facilitate top-down data analysis with high accuracy and reliability. Importantly, MASH Suite Pro is capable of processing both offline MS/MS and online LC-MS/MS data, which greatly simplify and speed up data analysis in top-down proteomics.

3 Application of top-down proteomics to heart diseases

Given the importance of protein isoform switching and PTM changes in cardiac contraction and relaxation, as well as in the development of heart failure [13,93], top-down proteomics has great potential to aid in the elucidation of the molecular mechanisms underlying physiological and pathophysiological processes in the heart [93]. In this section, we will summarize the applications of top-down proteomics to the advancement of our understanding of diseases of the heart. Moreover, we will highlight the potential of the technology to be translated into clinical setting for the identification of novel drug targets and the discovery and validation of biomarkers.

3.1. Identification of contractile protein isoforms and their associated proteoforms

Sarcomeres are the basic contractile units that are composed of myofilaments flanked by dense protein structure called Z-disc [21,77,94]. Myofilaments consist of the actin-based thin filaments and myosin-based thick filaments, the interaction of which mediates muscle contraction and relaxation in a Ca^{2+} dependent manner. The troponin complexes, cTnI, cardiac troponin T (cTnT), and cardiac troponin C are thin filament associated proteins, which induce conformational change of Tropomyosin (Tm) with increased intracellular Ca^{2+} , and, subsequently, release the myosin binding site on the actin filament and initiate muscle contraction [21,77].

Unlike the bottom-up approach, top-down MS analyzes intact proteins, which preserves important sequence information of the proteins present in the samples. The ability of top-down MS for maintaining full sequence coverage allows for the analysis of highly homologous protein isoforms, as well as the comprehensive characterization of their associated proteoforms. Human Tms are encoded by different genes (*Tpm1*, *Tpm2*, *Tpm3*, and *Tpm4*) and the alternative splicing event can generate a wide variety of proteoforms from a single gene. Using top-down MS, Peng et al. were able to unambiguously determine the presence and relative abundance of the Tm isoforms and their associated proteoforms in human and swine cardiac tissue, including α -Tm, β -Tm, and κ -Tm, as well as the phosphorylated α -Tm and β -Tm [95,96]. α -Tm and κ -Tm are different isoforms encoded by *Tpm1* due to alternative splicing. α -Tm is well-known to be highly expressed in cardiac tissue, while κ -Tm was only recently found to be expressed in the heart and confirmed using top-down MS [96]. Interestingly, the expression of κ -Tm is higher in the left atrium as compared to the left ventricle [96], suggesting important functional roles of κ -Tm in regional heterogeneity.

In addition to Tm, changes in the relative abundance of α -cardiac actin and α -skeletal actin in the heart have been associated with cardiac dysfunction [97]. However, precise quantification of these two isoforms is difficult because the two isoforms differ from each other by only 32 Da due to two juxtaposed amino acids [98]. Although gene expression levels are typically quantified by measuring the levels of the protein-specific mRNAs, studies have suggested that protein levels do not necessarily correlate with mRNA levels [99]. By employing high-resolution top-down MS, Chen et al. were able to distinguish between α -cardiac and α -skeletal actin in both human and swine heart tissue [98]. Moreover, top-down proteomics revealed that there is a marked upregulation of α -skeletal

actin relative to α -cardiac actin in the failing hearts as compared to the donor hearts, which can potentially serve as a biomarker for cardiac dysfunction [98].

3.2. Mapping sequence variations and PTMs in contractile proteins

Top-down MS is the premier method for the unambiguous localization of sequence variations and PTMs without prior knowledge. By taking advantage of a variety of fragmentation methods, any protein sequence variations and PTMs can be localized to a single amino acid. Using high-resolution top-down MS, Zhang et al. identified a -28 Da mass discrepancy in cTnI from swine [37]. Top-down MS/MS analysis revealed that this mass discrepancy was the result of a naturally occurring polymorphism (Val to Ala switch at position 116) in swine cTnI [37]. Similarly, Peng et al. identified two amino acid polymorphisms (Pro to Gln switch at position 38, and Pro to Leu switch at position 64) in swine α -Tm in the deep sequencing of swine Tm isoforms [96].

Protein PTMs regulate all aspects of protein biology, including protein localization, activity and degradation, and, consequently, play critical roles in cardiac physiology and pathophysiology [93]. While sequence variations/mutations may be readily detected by DNA sequencing, top-down MS is a powerful method for the comprehensive identification and localization of protein PTMs, as well as the determination of potential PTM cross-talk, which cannot be derived from DNA sequence. For examples, cTnI phosphorylation at Ser42/44 was identified *in vitro* and *in vitro* contractility assays showed that phosphorylation of these sites was detrimental for contractility [100]. However, it was not known whether these sites were phosphorylated *in vivo*. until Dong et al. employed a top-down proteomics approach and confirmed the phosphorylation of cTnI *in vivo* using rat models of spontaneously hypertensive heart failure [17]. In addition to determining the biological relevance of PTMs identified *in vitro*, top-down MS is also capable of identifying unexpected PTMs for unraveling the regulatory pathways involved in cardiac dysfunction. As demonstrated in a swine model of acute myocardial infarction (AMI), commonly known as heart attack, top-down MS revealed phosphorylation in cTnI, myosin regulatory light chain (RLC) and a previously un-identified protein (UnK) (Figure 3A–C). Subsequent MS/MS analysis identified the unknown protein as enigma homolog 2 (ENH2), a Z-disc protein that was not previously found to be a phosphoprotein (Figure 3C). Importantly, MS/MS using ECD unambiguously localized the site of ENH2 phosphorylation to Ser118 [77] (Figure 3D). As opposed to the traditional view of the Z-disc playing predominantly structural roles, this study provided definite evidence that Z-disc proteins can play critical regulatory roles post-myocardial infarction [77].

PTMs had typically been studied in isolation until, recently, emerging evidence suggested that proteins can be modified by multiple PTMs and that the presence or absence of specific PTMs can help or hinder modifications at other sites [101]. Top-down proteomics is the premier method for the analysis of co-occurring PTMs and potential PTM crosstalk, as the study of intact proteins retains the combinatorial nature of PTMs. By the top-down approach, Zhang et al. was able to determine the sequence of cTnI phosphorylation using the cardiac tissue from patients with chronic heart failure. Ser22 was the only site of phosphorylation observed in mono-phosphorylated TnI, while both Ser22/23 were present in

bis-phosphorylated cTnI, suggesting that cTnI is first phosphorylated at Ser22 followed by phosphorylation at Ser23 [18]. Cardiac myosin binding protein C (cMyBP-C) is an important component of the thick filament, and plays important roles in the regulation of actin-myosin cross-bridge and Ca^{2+} sensitivity of force generation [102]. Mutations in cMyBP-C have been shown to cause a variety of cardiac disorders, including dilated cardiomyopathy and hypertrophic cardiomyopathy [102]. A significant number of cMyBP-C mutations are caused by frame-shift or point mutations that lead to pre-mature termination of the coding sequence, resulting in truncated versions of cMyBP-C with various lengths [103]. In the study aiming to characterize the phosphorylation sites of full length and truncated cMyBP-C, Ge et al. discovered that truncation of cMyBP-C significantly altered its phosphorylation by protein kinase A (PKA) [19]. While the PKA target sites in full-length cMyBP-C locate at Ser283, Ser292, and Ser312, with phosphorylation at Ser292 occurring prior to the other sites, truncation of cMyBP-C at the C-termini leads to altered specificity of the PKA-mediated phosphorylation. Ser292, Ser312, and Ser484 were found to be phosphorylated on the truncated cMyBP-C [19]. This study suggested that truncation can significantly alter specificity of modification events and provided important insights into the potential mechanisms underlying cardiac disorder caused by cMyBP-C truncation.

3.3. Quantification of cardiac contractile proteoforms

The ability of top-down proteomics to provide a “bird’s eye” view of all proteoforms allows for robust and reliable quantification of proteoforms within a same spectrum. Top-down proteomics analysis of cTnI revealed a decline of cTnI phosphorylation with progressive cardiac hypertrophy, and a markedly abolishment of cTnI phosphorylation in chronic heart failure [18]. Subsequent MS/MS analysis of cTnI identified the sites decreased cTnI phosphorylation exclusively to Ser22/23, which are two well-established target sites of phosphorylation mediated by PKA [100]. This study represents the first application of top-down proteomics to study changes in contractile protein PTMs in clinically relevant samples, underscoring the potential of top-down proteomics for the quantification of biologically relevant PTM changes for biomarker discovery. Using a similar approach, Dong et al. quantified the phosphorylation of cTnI in a rat model of hypertensive heart failure, and uncovered an increase in cTnI phosphorylation. Interestingly, distinct from chronic heart failure in human, the sites of increases cTnI phosphorylation were localized to Ser22/23 and Ser42/44, the latter of which were established target sites of PKC *in vitro*, suggesting an important role of PKC in the development of hypertension-induced heart failure in rodents. Subsequently, Dong et al. evaluated the expression level of PKC, and found that PKC α was up-regulated in the rats with hypertensive heart failure.

By improving the strategies for contractile protein extraction and separation, Peng et al. were able to simultaneously quantify the proteoforms of multiple important contractile proteins, including cTnT, cTnI, RLC, α -Tm, β -Tm, and ENH2 in the swine model of AMI. While the phosphorylation of cTnT, α -Tm and β -Tm remained relatively unchanged following 90 min of ischemia, significant decreases of cTnI, RLC and ENH2 phosphorylation were observed (Figure 3) [77]. This represents the first study investigating the PTM changes of contractile proteins immediately following heart attack, and suggests that impaired PKA pathways are responsible for cardiac dysfunctions after heart attack.

In addition to the quantification of disease-associated proteoform changes in various animal models and human, top-down proteomics was used for the characterization of cardiac heterogeneity. Though heterogeneity of the heart is well-appreciated, transmural and regional difference of the cardiac contractile protein PTMs were not well-established, until the recent study performed by Gregorich et al. using top-down high-resolution MS [21]. Whereas minimal transmural heterogeneity in the phosphorylation of cTnT, cTnI, RLC and Tm were observed, the study revealed lower basal phosphorylation of cTnI and, conversely, higher α -Tm phosphorylation in the atria versus ventricles (Figure 4), which underscores the importance of sampling in the analysis of cardiac PTM changes [21]. Importantly, regional difference in the basal phosphorylation of cTnI and α -Tm likely contributes to the differences in the contractile function of the atrial and ventricular muscle preparations [21].

4. Expert Commentary

Top-down high-resolution MS-based proteomics is a premier method for the global and comprehensive study of protein isoforms and their associated proteoforms, enabling the identification of novel protein isoforms and PTMs, characterization of un-identified sequence variations, and quantification of disease-associated proteoform alterations. Given the importance of isoform switching and PTM changes in the regulation of cardiac contraction and relaxation, and in the development of heart failure [13,94], top-down proteomics holds great promise for uncovering the molecular mechanisms underlying diseases of the heart towards the discovery of novel therapeutic targets/strategies to combat this devastating syndrome. In addition, the capability of top-down proteomics for a global view of the essential contractile proteins allows for the discovery of novel biomarkers for various heart diseases. Furthermore, owing to the drastic heterogeneity of patients and their clinical profiles, top-down proteomics employing rapid high-throughput LC-MS profiling of the cardiac proteins can be used for the comprehensive characterization of cardiac protein isoforms and the associated proteoforms, which allows for better understanding of the patients and proper selection of treatments.

5. Five-year view

Top-down MS has proved powerful in the research setting for the discovery of novel protein isoforms (produced by multiple homologous genes) and the associated proteoforms (produced by a single gene), and the quantification of biologically relevant proteoform alterations, which holds great promise for clinical applications. However, as an emerging technology, challenges remains in the development of top-down proteomics methodology and its application to heart diseases in the research and clinical settings. Improvements in the solubilization of hydrophobic proteins, detection of large and low abundant proteins, enhancement of overall throughput and automation, and quantification of protein expression are the focus areas of ongoing research. Heart diseases remain a devastating threat to the well-being of the general population. Top-down proteomics, in combination of cardiac physiology, metabolomics and functional analysis, will further advance our understanding of the mechanisms underlying various cardiac diseases; and uncover novel targets for the development of effective therapeutic approaches.

Acknowledgments

Funding

The authors declare funding from the following U.S. Department of Health and Human Services, National Institutes of Health, National Heart, Lung, and Blood Institute NIH R01HL096971 and R01HL109810.

Abbreviation

PTMs	post-translational modifications
MS	mass spectrometry
LC	liquid chromatography
MS/MS	tandem mass spectrometry
RPC	reverse phase chromatography
AP	affinity purification
cTnI	cardiac troponin I
HIC	hydrophobic interaction chromatography
IEC	ion exchange chromatography
SEC	size exclusion chromatography
GELFrEE	Gel-Eluted Liquid Fraction Entrapment Electrophoresis
ESI	electrospray ionization
FT-ICR	Fourier transform ion cyclotron resonance
TOF	time-of-flight
CID	collision induced dissociation
ECD	electron capture dissociation
ETD	electron transfer dissociation
UVPD	ultraviolet photon dissociation
THRASH	Thorough High-Resolution Analysis of Spectra by Horn
Tm	tropomyosin
PKC	protein kinase C
AMI	acute myocardial infarction
RLC	myosin regulatory light chain
ENH2	enigma homolog isoform 2

cMyBP-C	cardiac myosin binding protein C
PKA	protein kinase A
cTnT	cardiac troponin T

References

Papers of special note have been highlighted as:

* of interest

** of considerable interest

1. Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics–2015 update: a report from the American Heart Association. *Circulation*. 2015; 131:e29–322. [PubMed: 25520374]
2. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. *Nat Rev Cardiol*. 2011; 8:30–41. [PubMed: 21060326]
3. McMurray JJ, Pfeffer MA. Heart failure. *Lancet*. 2015; 365:1877–89. [PubMed: 15924986]
4. Mudd JO, Kass DA. Tackling heart failure in the twenty-first century. *Nature*. 2008; 451:919–28. [PubMed: 18288181]
5. Drake TA, Ping P. Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Proteomics approaches to the systems biology of cardiovascular diseases. *J Lipid Res*. 2007; 48:1–8. [PubMed: 17065662]
6. Sabidó E, Selevsek N, Aebersold R. Mass spectrometry-based proteomics for systems biology. *Curr Opin Biotechnol*. 2015; 23:591–7. [PubMed: 22169889]
7. Pandey A, Mann M. Proteomics to study genes and genomes. *Nature*. 2000; 405:837–846. [PubMed: 10866210]
8. Smith LM, Kelleher NL, Proteomics CfTD. Proteoform: a single term describing protein complexity. *Nat Methods*. 2013; 10:186–7. [PubMed: 23443629]
9. Gregorich ZR, Ge Y. Top-down proteomics in health and disease: challenges and opportunities. *Proteomics*. 2014; 14:1195–210. [PubMed: 24723472]
10. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol*. 2003; 21:255–61. [PubMed: 12610572]
11. Karve TM, Cheema AK. Small changes huge impact: the role of protein posttranslational modifications in cellular homeostasis and disease. *J Amino Acids*. 2011; 207691
12. Geraldès P, King GL. Activation of protein kinase C isoforms and its impact on diabetic complications. *Circ Res*. 2010; 106:1319–31. [PubMed: 20431074]
13. Marston SB, Redwood CS. Modulation of thin filament activation by breakdown or isoform switching of thin filament proteins: physiological and pathological implications. *Circ Res*. 2003; 93:1170–8. [PubMed: 14670832]
14. Risco A, Cuenda A. New Insights into the p38 γ and p38 δ MAPK Pathways. *J Signal Transduct*. 2012; 520289
15. Chait BT. Chemistry. Mass spectrometry: bottom-up or top-down? *Science*. 2006; 314:65–6. [PubMed: 17023639]
16. Zhang H, Ge Y. Comprehensive analysis of protein modifications by top-down mass spectrometry. *Circ Cardiovasc Genet*. 2011; 4:711. [PubMed: 22187450]
17. Dong X, Sumandea CA, Chen YC, et al. Augmented phosphorylation of cardiac troponin I in hypertensive heart failure. *J Biol Chem*. 2012; 287:848–57. [PubMed: 22052912]
18. Zhang J, Guy MJ, Norman HS, et al. Top-down quantitative proteomics identified phosphorylation of cardiac troponin I as a candidate biomarker for chronic heart failure. *J Proteome Res*. 2011; 10:4054–65. [PubMed: 21751783]
- 19*. Ge Y, Rybakova IN, Xu Q, et al. Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state. *Proc Natl Acad*

- Sci U S A. 2009; 106:12658–63. Demonstrated alteration of protein phosphorylation caused by truncation using top-down MS. [PubMed: 19541641]
20. Peng Y, Ayaz-Guner S, Yu D, et al. Top-down mass spectrometry of cardiac myofilament proteins in health and disease. *Proteomics Clin Appl*. 2014; 8:554–68. [PubMed: 24945106]
 - 21*. Gregorich ZR, Peng Y, Lane NM, et al. Comprehensive assessment of chamber-specific and transmural heterogeneity in myofilament protein phosphorylation by top-down mass spectrometry. *J Mol Cell Cardiol*. 2015; 87:102–12. First comprehensive assessment of regional and transmural heterogeneity of contractile proteoforms using top-down proteomics. [PubMed: 26268593]
 22. Carroll J, Fearnley IM, Walker JE. Definition of the mitochondrial proteome by measurement of molecular masses of membrane proteins. *Proc Natl Acad Sci U S A*. 2006; 103:16170–5. [PubMed: 17060615]
 23. Catherman AD, Li M, Tran JC, et al. Top down proteomics of human membrane proteins from enriched mitochondrial fractions. *Anal Chem*. 2013; 85:1880–8. [PubMed: 23305238]
 24. Gregorich ZR, Chang YH, Ge Y. Proteomics in heart failure: top-down or bottom-up? *Pflugers Arch*. 2014; 466:1199–209. [PubMed: 24619480]
 25. Chang YH, Gregorich ZR, Chen AJ, et al. New mass-spectrometry-compatible degradable surfactant for tissue proteomics. *J Proteome Res*. 2015; 14:1587–1599. [PubMed: 25589168]
 26. Ayaz-Guner S, Zhang J, Li L, et al. In vivo phosphorylation site mapping in mouse cardiac troponin I by high resolution top-down electron capture dissociation mass spectrometry: Ser22/23 are the only sites basally phosphorylated. *Biochemistry*. 2009; 48:8161–70. [PubMed: 19637843]
 27. Sterling HJ, Batchelor JD, Wemmer DE, et al. Effects of buffer loading for electrospray ionization mass spectrometry of a noncovalent protein complex that requires high concentrations of essential salts. *J Am Soc Mass Spectrom*. 2010; 21:1045–9. [PubMed: 20226685]
 - 28*. Laganowsky A, Reading E, Hopper JT, et al. Mass spectrometry of intact membrane protein complexes. *Nat Protoc*. 2013; 8:639–51. Seminal work identified MS-compatible surfactants for the analysis of intact membrane proteins and protein complexes. [PubMed: 23471109]
 29. Xiu L, Valeja SG, Alpert AJ, et al. Effective protein separation by coupling hydrophobic interaction and reverse phase chromatography for top-down proteomics. *Anal Chem*. 2014; 86:7899–906. [PubMed: 24968279]
 - 30*. Tran JC, Doucette AA. Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. *Anal Chem*. 2008; 80:1568–73. Development of high-resolution gel-based separation and fractionation of proteins. [PubMed: 18229945]
 31. Mehmood S, Allison TM, Robinson CV. Mass spectrometry of protein complexes: from origins to applications. *Annu Rev Phys Chem*. 2015; 66:453–74. [PubMed: 25594852]
 - 32*. Hopper JT, Yu YT, Li D, et al. Detergent-free mass spectrometry of membrane protein complexes. *Nat Methods*. 2013; 10:1206–8. Development of detergent-free systems for the analysis of intact membrane proteins and protein complexes. [PubMed: 24122040]
 33. Whitelegge JP. HPLC and mass spectrometry of intrinsic membrane proteins. *Methods Mol Biol*. 2004; 251:323–40. [PubMed: 14704455]
 34. Whitelegge J. Intact protein mass spectrometry and top-down proteomics. *Expert Rev Proteomics*. 2013; 10:127–9. [PubMed: 23573779]
 35. Souda P, Ryan CM, Cramer WA, et al. Profiling of integral membrane proteins and their post translational modifications using high-resolution mass spectrometry. *Methods*. 2011; 55:330–6. [PubMed: 21982782]
 36. Doucette AA, Tran JC, Wall MJ, et al. Intact proteome fractionation strategies compatible with mass spectrometry. *Expert Rev Proteomics*. 2011; 8:787–800. [PubMed: 22087661]
 37. Zhang J, Dong X, Hacker TA, et al. Deciphering modifications in swine cardiac troponin I by top-down high-resolution tandem mass spectrometry. *J Am Soc Mass Spectrom*. 2010; 21:940–8. [PubMed: 20223681]
 38. Wang Y, Pinto JR, Solis RS, et al. Generation and functional characterization of knock-in mice harboring the cardiac troponin I-R21C mutation associated with hypertrophic cardiomyopathy. *J Biol Chem*. 2012; 287:2156–67. [PubMed: 22086914]

39. Sancho Solis R, Ge Y, Walker JW. Single amino acid sequence polymorphisms in rat cardiac troponin revealed by top-down tandem mass spectrometry. *J Muscle Res Cell Motil.* 2008; 29:203–12. [PubMed: 19165611]
- 40*. Tran JC, Doucette AA. Multiplexed size separation of intact proteins in solution phase for mass spectrometry. *Anal Chem.* 2009; 81:6201–9. Novel gel-based, multiplexed separation method for application in top-down proteomics. [PubMed: 19572727]
- 41*. Tran JC, Zamdborg L, Ahlf DR, et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature.* 2011; 480:254–8. First study utilizing multi-dimensional separation coupling GELFrEE, isoelectric focusing and RPC to top-down MS, and allows for identification of over 3000 proteoforms. [PubMed: 22037311]
42. Catherman AD, Durbin KR, Ahlf DR, et al. Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Mol Cell Proteomics.* 2013; 12:3465–73. [PubMed: 24023390]
43. Neverova I, Van Eyk JE. Role of chromatographic techniques in proteomic analysis. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2005; 815:51–63.
44. Ansong C, Wu S, Meng D, et al. Top-down proteomics reveals a unique protein S-thiolation switch in *Salmonella Typhimurium* in response to infection-like conditions. *Proc Natl Acad Sci U S A.* 2013; 110:10153–8. [PubMed: 23720318]
- 45*. Chen B, Peng Y, Valeja SG, et al. Online Hydrophobic Interaction Chromatography-Mass Spectrometry for Top-Down Proteomics. *Anal Chem.* 2016; 88:1885–91. Development of novel column materials for HIC that allows for direct coupling to top-down MS. [PubMed: 26729044]
- 46*. Valeja SG, Xiu L, Gregorich ZR, et al. Three dimensional liquid chromatography coupling ion exchange chromatography/hydrophobic interaction chromatography/reverse phase chromatography for effective protein separation in top-down proteomics. *Anal Chem.* 2015; 87:5363–71. Demonstrated unprecedented separation power of multi-dimensional liquid chromatography platform for the analysis of intact proteins by top-down MS. [PubMed: 25867201]
47. Muneeruddin K, Nazzaro M, Kaltashov IA. Characterization of intact protein conjugates and biopharmaceuticals using ion-exchange chromatography with online detection by native electrospray ionization mass spectrometry and top-down tandem mass spectrometry. *Anal Chem.* 2015; 87:10138–45. [PubMed: 26360183]
48. Compton PD, Zamdborg L, Thomas PM, et al. On the scalability and requirements of whole protein mass spectrometry. *Anal Chem.* 2011; 83:6868–74. [PubMed: 21744800]
49. Alpert, AJ. Column Handbook for Size Exclusion Chromatography. Academic Press; San Diego: 1999. Size Exclusion High-Performance Liquid Chromatography of Small Solutes; p. 49-66.
50. Chen X, Ge Y. Ultrahigh pressure fast size exclusion chromatography for top-down proteomics. *Proteomics.* 2013; 13:2563–6. [PubMed: 23794208]
51. Sharma S, Simpson DC, Toli N, et al. Proteomic profiling of intact proteins using WAX-RPLC 2-D separations and FTICR mass spectrometry. *J Proteome Res.* 2007; 6:602–10. [PubMed: 17269717]
52. Karas M, Krüger R. Ion formation in MALDI: the cluster ionization mechanism. *Chem Rev.* 2003; 103:427–40. [PubMed: 12580637]
53. Fenn JB, Mann M, Meng CK, et al. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* 1989; 246:64–71. [PubMed: 2675315]
54. Scherperel G, Reid GE. Emerging methods in proteomics: top-down protein characterization by multistage tandem mass spectrometry. *Analyst.* 2007; 132:500–6. [PubMed: 17525804]
55. Mann M, Kelleher NL. Precision proteomics: the case for high resolution and high mass accuracy. *Proc Natl Acad Sci U S A.* 2008; 105:18132–8. [PubMed: 18818311]
56. Kelleher NL, Senko MW, Siegel MM, et al. Unit resolution mass spectra of 112 kDa molecules with 3 Da accuracy. *Journal of the American Society for Mass Spectrometry.* 1997; 8:380–3.
57. Ge Y, Lawhorn BG, ElNaggar M, et al. Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J Am Chem Soc.* 2002; 124:672–8. [PubMed: 11804498]

- 58*. Han X, Jin M, Breuker K, et al. Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science*. 2006; 314:109–12. Novel pre- and post-skimmer dissociation, and capillary additives for the characterization of proteins up to 200 kDa. [PubMed: 17023655]
59. Chen X, Westphall MS, Smith LM. Mass spectrometric analysis of DNA mixtures: instrumental effects responsible for decreased sensitivity with increasing mass. *Anal Chem*. 2003; 75:5944–52. [PubMed: 14588036]
60. Park J, Qin H, Scalf M, et al. A mechanical nanomembrane detector for time-of-flight mass spectrometry. *Nano Lett*. 2011; 11:3681–4. [PubMed: 21806063]
61. Beck S, Michalski A, Raether O, et al. The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics. *Mol Cell Proteomics*. 2015; 14:2014–29. [PubMed: 25991688]
62. Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem*. 2000; 72:1156–62. [PubMed: 10740853]
63. Perry RH, Cooks RG, Noll RJ. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom Rev*. 2008; 27:661–99. [PubMed: 18683895]
64. Tran JC, Zamdborg L, Ahlf DR, et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*. 2011; 480:254–8. [PubMed: 22037311]
- 65*. Durbin KR, Fornelli L, Fellers RT, et al. Quantitation and Identification of Thousands of Human Proteoforms below 30 kDa. *J Proteome Res*. 2016; 15:976–82. Large-scale quantification of protein expression level using label-free top-down MS. [PubMed: 26795204]
66. Valeja SG, Kaiser NK, Xian F, et al. Unit mass baseline resolution for an intact 148 kDa therapeutic monoclonal antibody by Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem*. 2011; 83:8391–5. [PubMed: 22011246]
67. Kostyukovich YI, Vladimirov GN, Nikolaev EN. Dynamically harmonized FT-ICR cell with specially shaped electrodes for compensation of inhomogeneity of the magnetic field. Computer simulations of the electric field and ion motion dynamics. *J Am Soc Mass Spectrom*. 2012; 23:2198–207. [PubMed: 22993044]
68. McLafferty FW, Breuker K, Jin M, et al. Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J*. 2007; 274:6256–68. [PubMed: 18021240]
69. Siuti N, Kelleher NL. Decoding protein modifications using top-down mass spectrometry. *Nat Methods*. 2007; 4:817–821. [PubMed: 17901871]
70. Zubarev RA, Horn DM, Fridriksson EK, et al. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem*. 2000; 72:563–73. [PubMed: 10695143]
71. Syka JE, Coon JJ, Schroeder MJ, et al. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A*. 2004; 101:9528–33. [PubMed: 15210983]
72. Miladinovi SM, Fornelli L, Lu Y, et al. In-spray supercharging of peptides and proteins in electrospray ionization mass spectrometry. *Anal Chem*. 2012; 84:4647–51. [PubMed: 22571167]
73. McLuckey SA, Reid GE, Wells JM. Ion parking during ion/ion reactions in electrodynamic ion traps. *Anal Chem*. 2002; 74:336–46. [PubMed: 11811406]
74. Huang TY, McLuckey SA. Top-down protein characterization facilitated by ion/ion reactions on a quadrupole/time of flight platform. *Proteomics*. 2010; 10:3577–88. [PubMed: 20848674]
- 75*. Shaw JB, Li W, Holden DD, et al. Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J Am Chem Soc*. 2013; 135:12646–51. Demonstrated extensive fragmentation of UVPD for intact protein characterization. [PubMed: 23697802]
76. Pesavento JJ, Mizzen CA, Kelleher NL. Quantitative analysis of modified proteins and their positional isomers by tandem mass spectrometry: human histone H4. *Anal Chem*. 2006; 78:4271–80. [PubMed: 16808433]
- 77*. Peng Y, Gregorich ZR, Valeja SG, et al. Top-down Proteomics Reveals Concerted Reductions in Myofilament and Z-disc Protein Phosphorylation after Acute Myocardial Infarction. *Molecular &*

- Cellular Proteomics. 2014; 13:2752–64. Demonstrated concerted reduction of phosphorylation in myofilament and Z-disc proteins in a novel swine model of AMI. [PubMed: 24969035]
78. Chamot-Rooke J, Mikaty G, Malosse C, et al. Posttranslational modification of pili upon cell contact triggers *N. meningitidis* dissemination. *Science*. 2011; 331:778–82. [PubMed: 21311024]
 79. Du Y, Parks BA, Sohn S, et al. Top-down approaches for measuring expression ratios of intact yeast proteins using Fourier transform mass spectrometry. *Anal Chem*. 2006; 78:686–94. [PubMed: 16448040]
 80. Waanders LF, Hanke S, Mann M. Top-down quantitation and characterization of SILAC-labeled proteins. *J Am Soc Mass Spectrom*. 2007; 18:2058–64. [PubMed: 17920290]
 81. Bergmann U, Ahrends R, Neumann B, et al. Application of metal-coded affinity tags (MeCAT): absolute protein quantification with top-down and bottom-up workflows by metal-coded tagging. *Anal Chem*. 2012; 84:5268–75. [PubMed: 22659083]
 82. Mazur MT, Cardasis HL, Spellman DS, et al. Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry. *Proc Natl Acad Sci U S A*. 2010; 107:7728–33. [PubMed: 20388904]
 83. Horn DM, Zubarev RA, McLafferty FW. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J Am Soc Mass Spectrom*. 2000; 11:320–32. [PubMed: 10757168]
 84. Senko MW, Beu SC, McLafferty FW. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J Am Soc Mass Spectrom*. 1995; 6:229–33. [PubMed: 24214167]
 85. Liu X, Inbar Y, Dorrestein PC, et al. Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Mol Cell Proteomics*. 2010; 9:2772–82. [PubMed: 20855543]
 86. Marty MT, Baldwin AJ, Marklund EG, et al. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal Chem*. 2015; 87:4370–6. [PubMed: 25799115]
 87. Jaitly N, Mayampurath A, Littlefield K, et al. Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics*. 2009; 10:87. [PubMed: 19292916]
 88. Zamdborg L, LeDuc RD, Glowacz KJ, et al. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res*. 2007; 35:W701–6. [PubMed: 17586823]
 89. Karabacak NM, Li L, Tiwari A, et al. Sensitive and specific identification of wild type and variant proteins from 8 to 669 kDa using top-down mass spectrometry. *Mol Cell Proteomics*. 2009; 8:846–56. [PubMed: 19074999]
 90. Frank AM, Pesavento JJ, Mizzen CA, et al. Interpreting top-down mass spectra using spectral alignment. *Anal Chem*. 2008; 80:2499–505. [PubMed: 18302345]
 91. Liu X, Sirotkin Y, Shen Y, et al. Protein identification using top-down. *Mol Cell Proteomics*. 2012; 11:M111.008524.
 - 92*. Cai W, Guner H, Gregorich ZR, et al. MASH Suite Pro: A Comprehensive Software Tool for Top-down Proteomics. 2015 First software interface integrating spectral deconvolution, protein identification, quantification and characterization in a suite with visual components for data validation.
 93. de Tombe PP, Solaro RJ. Integration of cardiac myofilament activity and regulation with pathways signaling hypertrophy and failure. *Ann Biomed Eng*. 2000; 28:991–1001. [PubMed: 11144684]
 94. Yin Z, Ren J, Guo W. Sarcomeric protein isoform transitions in cardiac muscle: a journey to heart failure. *Biochim Biophys Acta*. 2015; 1852:47–52. [PubMed: 25446994]
 95. Peng Y, Yu D, Gregorich Z, et al. In-depth proteomic analysis of human tropomyosin by top-down mass spectrometry. *J Muscle Res Cell Motil*. 2013; 34:199–210. [PubMed: 23881156]
 96. Peng Y, Chen X, Zhang H, et al. Top-down targeted proteomics for deep sequencing of tropomyosin isoforms. *J Proteome Res*. 2013; 12:187–98. [PubMed: 23256820]

97. Copeland O, Nowak KJ, Laing NG, et al. Investigation of changes in skeletal muscle alpha-actin expression in normal and pathological human and mouse hearts. *J Muscle Res Cell Motil.* 2010; 31:207–14. [PubMed: 20706863]
98. Chen YC, Ayaz-Guner S, Peng Y, et al. Effective Top-Down LC/MS+ Method for Assessing Actin Isoforms as a Potential Cardiac Disease Marker. *Anal Chem.* 2015; 87:8399–406. [PubMed: 26189812]
99. Gry M, Rimini R, Strömberg S, et al. Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genomics.* 2009; 10:365. [PubMed: 19660143]
100. Solaro RJ, van der Velden J. Why does troponin I have so many phosphorylation sites? Fact and fancy. *J Mol Cell Cardiol.* 2010; 48:810–6. [PubMed: 20188739]
101. Hunter T. The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol Cell.* 2007; 28:730–8. [PubMed: 18082598]
102. de Tombe PP. Myosin binding protein C in the heart. *Circ Res.* 2006; 98:1234–6. [PubMed: 16728667]
103. Flashman E, Redwood C, Moolman-Smook J, et al. Cardiac myosin binding protein C: its role in physiology and disease. *Circ Res.* 2004; 94:1279–1289. [PubMed: 15166115]

Key issues

- Top-down proteomics analyzes intact proteins and provides a global or “bird’s eye” view of protein isoforms and their associated proteoforms, and, thus, represents a superior approach for the comprehensive analysis of protein isoforms and PTMs for the elucidation of disease mechanisms and identification of disease markers.
- Notable advancements have been made in the top-down proteomics work flow to improve intact protein solubility, high-throughput protein separation, MS and MS/MS analysis, and protein quantification.
- The development of MS-compatible reagents for protein solubilization and multi-dimensional orthogonal platforms for protein separation have allowed for high-throughput and large-scale interrogation of human proteome. High-resolution and high-sensitivity instruments and a variety of fragmentation methods have enabled detection and characterization of large proteins. Significant efforts have also been invested in the development of robust quantification methods.
- Nevertheless, significant challenges remain for the solubilization of highly hydrophobic membrane proteins, efficient and high-resolution separation of the complex proteome, as well as the detection and fragmentation of very large proteins. High-throughput platforms for simultaneous quantification of relative proteoform abundance within a sample and protein expression changes across samples are still lacking. These challenges need to be overcome before the technology can realize its full potential for biomedical research and clinical application.
- The top-down approach has enabled differentiation of cardiac contractile protein isoforms and localization of sequence variations and PTMs, leading to the discoveries of previously unknown proteoforms and PTMs that play important roles in the regulation of cardiac physiology and pathophysiology.
- Top-down proteomics allows for the relative quantification of the contractile proteoforms in health and disease, which aids in the deciphering of the disease mechanisms and identification of potential biomarkers.
- Future research combining top-down proteomics, cardiac physiology and functional studies hold great promise for providing novel mechanistic insights towards the understanding of diseases of the heart, and, more importantly, identifying novel therapeutic targets and diagnostic tools for combating this devastating syndrome.

Figure 1.

A. Exponential increase of proteome complexity owing to alternative splicing of mRNAs and post-translational modifications of proteins. B. Schematic depicting the basic workflow for top-down proteomics as compared with that of bottom-up proteomics. Adapted from [24] originally published in *Pflügers Archiv European Journal of Physiology*, with permission from Springer and authors. © Springer

Figure 2. Schematics showing the fragment ions produced by CID (A), ECD/ETD (B), and UVPD (C). Noted that the dissociation pathways delineated are not exclusive. CID usually resulted in the loss of labile PTMs (A), whereas ECD/ETD typically preserves them (B).

Figure 3.

Top-down proteomics revealed a concerted reduction of protein phosphorylation in cTnI (A), RLC (B), and an unknown protein (UnK) (C), which was subsequently identified as ENH2. D. Top-down MS/MS analysis unambiguously localized the site of ENH2 phosphorylation to Ser118, and detected a previously unknown amino acid polymorphism at Cys72. This research was originally published in *Mol Cell Proteomics* [77]. Adapted from [77] with permission. © The American Society for Biochemistry and Molecular Biology.

Figure 4.

Top-down proteomics revealed regional heterogeneity of cTnI and α-Tm proteoforms. Basal phosphorylation of cTnI is higher in left and right ventricles versus atria (A-B), and basal phosphorylation of α-Tm is higher in the atria versus ventricles (D-E). This research was originally published in J Mol Cell Cardiol [21]. Adapted from [21] with permission from Elsevier and authors. © Elsevier