Proteomics

Comprehensive Analysis of Protein Modifications by Top-Down Mass Spectrometry

Han Zhang, PhD; Ying Ge, PhD

Mass spectrometry (MS)-based proteomics is playing an increasingly important role in cardiovascular research. Proteomics includes identification and quantification of proteins and the characterization of protein modifications, such as posttranslational modifications and sequence variants. The conventional bottom-up approach, involving proteolytic digestion of proteins into small peptides before MS analysis, is routinely used for protein identification and quantification with high throughput and automation. Nevertheless, it has limitations in the analysis of protein modifications, mainly because of the partial sequence coverage and loss of connections among modifications on disparate portions of a protein. An alternative approach, top-down MS, has emerged as a powerful tool for the analysis of protein modifications. The top-down approach analyzes whole proteins directly, providing a “bird’s-eye” view of all existing modifications. Subsequently, each modified protein form can be isolated and fragmented in the mass spectrometer to locate the modification site. The incorporation of the nonergodic dissociation methods, such as electron-capture dissociation (ECD), greatly enhances the top-down capabilities. ECD is especially useful for mapping labile posttranslational modifications that are well preserved during the ECD fragmentation process. Top-down MS with ECD has been successfully applied to cardiovascular research, with the unique advantages in unraveling the molecular complexity, quantifying modified protein forms, complete mapping of modifications with full-sequence coverage, discovering unexpected modifications, identifying and quantifying positional isomers, and determining the order of multiple modifications. Nevertheless, top-down MS still needs to overcome some technical challenges to realize its full potential. Herein, we reviewed the advantages and challenges of the top-down method, with a focus on its application in cardiovascular research.

KEY WORDS: cardiovascular diseases ■ proteomics ■ electron capture dissociation ■ posttranslational modification ■ top-down mass spectrometry
Comprehensive Analysis of Protein Modifications by Top-Down Mass Spectrometry
Han Zhang and Ying Ge

Circ Cardiovasc Genet. 2011;4:711
doi: 10.1161/CIRCGENETICS.110.957829
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2011 American Heart Association, Inc. All rights reserved.
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/4/6/711

Data Supplement (unedited) at:
http://circgenetics.ahajournals.org/content/suppl/2011/12/23/4.6.711.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Genetics can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Genetics is online at:
http://circgenetics.ahajournals.org//subscriptions/
Proteomics

Comprehensive Analysis of Protein Modifications by Top-Down Mass Spectrometry

Han Zhang, PhD; Ying Ge, PhD

Abstract—Mass spectrometry (MS)–based proteomics is playing an increasingly important role in cardiovascular research. Proteomics includes identification and quantification of proteins and the characterization of protein modifications, such as posttranslational modifications and sequence variants. The conventional bottom-up approach, involving proteolytic digestion of proteins into small peptides before MS analysis, is routinely used for protein identification and quantification with high throughput and automation. Nevertheless, it has limitations in the analysis of protein modifications, mainly because of the partial sequence coverage and loss of connections among modifications on disparate portions of a protein. An alternative approach, top-down MS, has emerged as a powerful tool for the analysis of protein modifications. The top-down approach analyzes whole proteins directly, providing a “bird’s-eye” view of all existing modifications. Subsequently, each modified protein form can be isolated and fragmented in the mass spectrometer to locate the modification site. The incorporation of the nonergodic dissociation methods, such as electron-capture dissociation (ECD), greatly enhances the top-down capabilities. ECD is especially useful for mapping labile posttranslational modifications that are well preserved during the ECD fragmentation process. Top-down MS with ECD has been successfully applied to cardiovascular research, with the unique advantages in unraveling the molecular complexity, quantifying modified protein forms, complete mapping of modifications with full-sequence coverage, discovering unexpected modifications, identifying and quantifying positional isomers, and determining the order of multiple modifications. Nevertheless, top-down MS still needs to overcome some technical challenges to realize its full potential. Herein, we reviewed the advantages and challenges of the top-down method, with a focus on its application in cardiovascular research. (Circ Cardiovasc Genet. 2011;4:00-00.)

Key Words: cardiovascular diseases \[\text{top-down mass spectrometry}\] proteomics electron capture dissociation posttranslational modification

Proteomics is playing an increasingly important role in cardiovascular research.1–8 Proteomics includes the separation, identification, and quantification of proteins, as well as the characterization of protein modifications, such as posttranslational modifications (PTMs) and sequence variants (eg, mutants, alternatively spliced isoforms, and amino acid polymorphisms).9,10 The PTMs (eg, phosphorylation, glycosylation, acetylation, and proteolysis) are covalent modifications of a protein after its translation. The PTMs can modulate the activity, stability, and function of a protein.11 Disease-induced PTMs can occur in concert with altered gene expression, which will substantially affect protein function and its interaction with other proteins in the signaling network.1 Aberrant protein PTMs, together with mutations and alternatively spliced isoforms, are increasingly recognized as important underlying mechanisms for cardiovascular diseases.12–14 Hence, a comprehensive analysis of protein modifications is important for understanding the disease mechanisms.

Traditional strategies for the analysis of protein modifications, such as radioactive labeling and Western blotting, can be specific and relatively quantitative, but they require prior knowledge of the modification type and are limited by antibody availability and specificity. The recently developed Pro-Q diamond staining can globally reveal the level of protein phosphorylation, but it cannot provide the identification of proteins or their modification sites, which are essential for understanding the disease mechanisms.15,16 Biological mass spectrometry (MS) is the preferred method for the analysis of protein modifications because it is capable of providing universal information about protein modifications without a priori knowledge and locating the sites of modification.11

Currently, there are 2 complementary approaches in MS-based proteomics: “bottom-up” and “top-down.”17–20 The conventional peptide-based bottom-up shotgun proteomics involves in-gel or in-solution proteolytic digestion of proteins with enzymes, usually trypsin, into many pieces of small peptides (1–3 kDa) before MS analysis (Figure 1A). This approach is well suited for protein identification, which only requires a small portion of sequence coverage (>10–20 amino acid residues) to identify the protein from the database.21 With the tremendous...
Top-down MS is becoming a powerful technology for comprehensive analysis of protein modifications. In contrast to bottom-up MS, top-down MS analyzes intact proteins without proteolytic digestion (Figure 1B). This strategy preserves the labile structural characteristics that are mostly destroyed in bottom-up MS. It can universally detect all the existing modifications, including PTMs (ie, phosphorylation, proteolysis, and acetylation) and sequence variants (ie, mutants, alternatively spliced isoforms, and amino acid polymorphisms) simultaneously in 1 spectrum (a “bird’s-eye” view) without a priori knowledge. Top-down MS first measures the molecular weight (MW) of an intact protein and compares it with the calculated value based on the DNA-predicted protein sequence, which can easily reveal any changes/modifications in the protein sequence globally (the “top” part). Then, a specific modified form of interest can be directly isolated in the mass spectrometer (“a gas-phase purification”) and subsequently fragmented in the mass spectrometer by tandem MS (MS/MS), such as collision-induced dissociation (CID) and electron-capture dissociation (ECD), for highly reliable mapping of the modification sites (the “down” part). The incorporation of the novel MS/MS technique, ECD, has greatly enhanced the capability of top-down MS in structural analysis of biomolecules. As a nonergodic fragmentation method, ECD preserves labile PTMs during the fragmentation process; thus, it is particularly suitable for the localization of labile PTMs. Top-down MS with ECD has been successfully applied to cardiovascular research with the unique advantages in unraveling the molecular complexity, quantifying multiple modified protein forms, complete mapping of modifications with full-sequence coverage, discovering unexpected PTMs and amino acid polymorphisms, identifying and quantifying positional isomers, and determining the order of multiple modifications.

In contrast to the well-established bottom-up proteomics, the top-down proteomics is still in its early developmental stage and has yet to fully overcome its technical challenges in sample preparation, instrument sensitivity/detection limit, and throughput/automation. New technological developments are needed to advance top-down proteomics for the analysis of complex samples of cell/tissue lysate and biological fluid. Herein, we provide an overview on the top-down MS method, the basic information needed for top-down MS analysis, and its advantages and technical challenges, with a focus on its application in cardiovascular disease.

**Top-Down MS-Based Proteomics Method**

**Sample Preparation for Top-Down MS**

Typically, intact proteins need to be extracted from cell/tissue lysate, solubilized, and separated/purified before MS analysis. Protein samples then need to be introduced to a mass spectrometer in buffer conditions compatible with MS analysis. Typical buffers used to extract/solubilize proteins involve a high salt concentration, usually with the addition of detergents, such as SDS, to increase the solubility of the protein. These salts and detergents interfere with MS detec-
tion of proteins because they are present in huge excess relative to proteins and have much higher ionization efficiency, which will, therefore, suppress protein signals.

The high salt content can be easily removed by a desalting step using a reverse-phase (RP) column/trap either on-line or off-line before MS analysis. Alternatively, MS-compatible volatile salt buffers, such as ammonium bicarbonate and ammonium acetate, can also be used to replace the common salt present in biological samples (ie, NaCl, KCl, and CaCl₂) via a buffer exchange step such as dialysis or ultrafiltration. Typical procedures for detergent removal involve precipitation and resolubilization of proteins in detergent-free buffers, which may result in sample loss because some portion of protein may become insoluble in detergent-free buffers. Recently, there are efforts allocated in designing MS-compatible acid labile detergents with the hope of replacing these traditional detergents. Alternatively, proteins can also be selectively solubilized based on their inherent chemical properties, such as biospecificity, hydrophobicity, and charge without the use of a detergent. This can also be used to fractionate a specific subproteome before chromatographic separation.

Separation/Purification of Intact Proteins
A complex protein mixture can be separated by SDS-PAGE, liquid chromatography (LC), and high-resolution MS. The methods of 1- and 2-dimensional SDS-PAGE are classic for protein separation and visualization. The gel-based separation is widely used in bottom-up proteomics because trypsin-digested peptides can be effectively retrieved from gels. However, it is technically challenging to extract the intact proteins from gel matrices with a high recovery rate. Thus, gel-based separation is not applicable in top-down MS. Recently, solution-based isoelectric focusing, coupled with a multiplex tube gel electrophoresis separation device, referred to as gel-eluted liquid fraction entrapment electrophoresis, has been developed for intact protein separation based on their MWs and applied to proteins (10–250 kDa) with a high resolution and a high recovery rate. Nevertheless, the surfactant SDS is still present in the sample so the proteins need to be precipitated in organic solvent and resolubilized in MS-compatible buffers.

LC is ideally suited for proteomics because it can be conveniently interfaced with MS. The major LC techniques typically used for intact protein separation include affinity, ion-exchange chromatography, size-exclusion chromatography, and RP chromatography. Affinity chromatography is by far the most effective and specific protein purification method. For example, immunoaffinity methods have been used to effectively purify cardiac troponin I (cTnI) from animal and human myocardium. Nonetheless, most of the affinity methods have been performed off-line, requiring an additional separation/desalting procedure using RPLC. Ion-exchange chromatography and size-exclusion chromatography have also been used for intact protein separation. Generally, they are used to perform the first-dimension separation, followed by RP chromatography in the second dimension. RP chromatography enhances the separation from the previous step and performs desalting as the last sample preparation step before MS analysis. The 2-dimensional LC approach has the advantage of preconcentrating and desalting the species of interest simultaneously, yielding a higher peak capacity and better separation and, if connected on-line, minimizing sample loss. Chromatographic focusing, as a promising alternative to salt-gradient ion-exchange chromatography, is also capable of separating intact proteins. Sheng and coworkers reported the separation of intact proteins from human serum proteome with the use of chromatographic focusing and RPLC in a commercial ProteomeLab PF2D system.

With high-resolution mass spectrometers (eg, Fourier-transform ion cyclotron resonance [FTICR] MS), protein mixtures with high complexity may be resolved during MS analysis in addition to other traditional separation techniques. Also, >100 protein components from a crude cell lysate were successfully separated in 1 FTICR MS spectrum using the high-resolving power of FTICR.

Top-Down MS Instrumentation
A typical mass spectrometer consists of 4 components: (1) a sample inlet; (2) an ion source, which converts sample molecules to ions, such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI); (3) a mass analyzer; and (4) a detector. Top-down MS is usually performed in high-resolution instruments because of the need to resolve the high MW of intact proteins. The early top-down MS studies were performed in home-built FTICR mass spectrometers with high resolution and accuracy. Recently, the development of several commercial high-resolution hybrid mass spectrometers, such as linear trap quadrupole (LTQ)/FT, LTQ-orbitrap (a hybrid between ion trap and FT), and quadrupole time-of-flight, substantially boosted the growth of top-down MS applications. ESI is preferred over MALDI in top-down proteomics because it produces multiply charged precursor ions for more efficient dissociation of large protein ions and provides more MS/MS options than MALDI, which mainly produces singly charged species. The inability to resolve the large MW ions substantially limits the practice of top-down MS in low-resolution instruments.

Fragmentation Methods for Protein Molecular Ions
Generally, there are 2 categories of fragmentation methods: energetic versus nonenergetic (Figure 1C). The well-developed energetic dissociation methods, such as CID, infrared multiphoton dissociation, and postsource decay, cleave CO-NH backbone bonds to produce b and y fragment ions at high efficiency. All these energetic dissociation methods preferably dissociate bonds of the lowest activation energy. Because energy levels of the labile PTMs, such as phosphorylation bonds, are weaker than the backbone amide bonds, they are usually cleaved first in the energetic dissociation methods, resulting in the neutral loss of phosphoric acid (H₃PO₄, 98 Da) or metaphosphoric acid (HPO₃, 80 Da), which makes it difficult to map phosphorylation sites. The nonenergetic electron-based MS/MS techniques, ECD and electron transfer dissociation, are particularly suitable for the localization of labile PTMs, such as phosphorylation. In ECD, low-energy thermal electrons are captured
by a protonated peptide or a protein causing local fast (<10^{-12} second) cleavages of backbone covalent bonds, NH-CHR, producing mainly \( c \) and \( z \) ions (Figure 1C).34 Hence, the labile PTMs are well preserved in the ECD fragmentation process, and no loss of labile phosphate has been observed in ECD spectra.25,26,30 ECD is nonselective, so that it often provides far more cleavages than CID,21 which greatly enhances the capability of top-down MS in identifying PTMs and sequence variants.23,25,27,29,30,59 Nevertheless, ECD is only available at high-end FTICR instruments, which substantially limits its wide use in proteomic laboratories. More recently, electron transfer dissociation,58 a sister version of ECD, in which the electrons are delivered by an anion that transfers its electron to a multiply charged peptide or protein, has been implemented in hybrid LTQ/orbitrap instruments,60 which begins to play a role in top-down proteomics.61

**Top-Down MS Data Analysis Tools**

Because ESI, the commonly used ionization method in top-down MS, produces multiply charged ions, Horn et al62 first developed a fully automated computer algorithm, THRASH, to determine the charge states and subsequently the MWs of precursor and product ions from complex high-resolution mass spectra. Unfortunately, this software is not compatible with the mainstream computers. Smith’s laboratory developed free downloadable PC- and Mac-compatible software (http://omic-s.s.npl.gov/software), DeconMS and Decon2LS, based on the THRASH algorithm. Thermo has also developed the ManualXtract algorithm to process high-resolution data of multiply charged ions from ESI.

Furthermore, bioinformatics tools for identification of proteins using top-down MS are also available. There are 3 major search engines: ProSight, PIITA, and MascotTD (also known as “big Mascot”).37,38 ProSight, developed by Meng et al.63 is the first and most flexible search engine for top-down proteomics. It allows the identification of proteins from complex mixtures by searching both the precursor ion MS and product ion MS/MS data (processed via the THRASH algorithm) in the databases. ProSight can be used to map both known and unknown PTMs.37 PIITA is based on a precursor ion-independent top-down algorithm using fragmentation data without a priori expectation of PTMs or sequence errors.64 A recently released big Mascot extended the classic Mascot to large proteins (>16 kDa). The big Mascot search requires both precursor and fragment ions and can identify protein PTMs and sequence variants.65

**Application of Top-Down MS to Cardiovascular Research**

**Unraveling the Molecular Complexity of Intact Proteins by High-Resolution MS**

The major advantage of top-down MS is that it can easily reveal the full extent of molecular complexity of a protein because it analyzes whole proteins instead of small peptides.32 The presence of PTMs can be detected by the mass difference (\( \Delta m \)) between unmodified and modified forms.11 Moreover, most of the common amino acids have a distinct mass (except Leu and Ile) so that the top-down approach can also be used to reveal sequence variations, resulting from proteolytic degradation, mutation, amino acid polymorphism, and alternatively spliced isoform.

We recently used the high resolving power of FTICR MS to unravel the complexity of PTMs for a commercially available cTnI sample purified from healthy human heart tissues (Figure 2).28 cTnI is the inhibitory subunit of the thin filament troponin-tropomyosin regulatory complex, playing a critical role in \( Ca^{2+} \)-mediated cardiac muscle contractility.66,67 The ESI/FTMS spectrum of cTnI revealed a total of 36 modified molecular ions (Figure 2A), despite the protein running as a single sharp band on SDS-PAGE (Figure 2B). The accurate mass measurements of all the well-resolved cTnI forms suggest several mass differences (\( \Delta m \)) (Figure 2A). For example, \( \Delta m \) of 79.971 Da matches with phosphorylation (calculated as 79.966 Da), and \( \Delta m \) of 15.995 Da is commonly assigned as oxidation (calculated as 15.995 Da), of the proteins. Moreover, \( \Delta m \) values of 128.130, 147.069, and 216.085 Da correspond to amino acids Gln/Lys (calculated as follows: Gln, 128.059 Da; Lys, 128.095 Da), Phe (calculated as 147.068 Da), and Glu-Ser (calculated as 216.075 Da), respectively, which resulted from truncations of Glu-Ser, Phe-Glu-Ser, and Lys-Phe-Glu-Ser residues, respectively, from the C-terminus of a full-length cTnI. The high mass accuracy afforded by high-resolution MS data greatly increased the confidence in assigning the protein modification. All forms detected herein matched with the predicted sequence (TNNI3_Human) obtained from the Swissprot database, with the removal of N-terminal Met and the addition of acetylation at the new N-terminus. Thus, we find acetylation, N-terminal Met removal, monophosphorylation and bisphosphorylation, oxidation, and proteolytic degradation for this commercially available cTnI.28 The major modified forms of cTnI, resulting from the combinations of these multiple modifications, are illustrated in Figure 2C. Such a high level of molecular complexity cannot be accounted for by any other technique.

**Complete PTM Mapping by MS/MS**

Top-down MS can reveal the molecular complexity and precisely map sites of modifications by isolating and fragmenting the modified protein ions in the mass spectrometer.10,68 MS/MS produces \( b/y \) or \( c/z \) fragment ions with \( b/c \) ions counting from the N-terminus and \( y/z \) ions from the C-terminus (Figure 1C). In a top-down MS/MS experiment, full-sequence coverage can be easily achieved for a protein <60 kDa.25,27,29,30,68 We have used top-down MS with ECD for the comprehensive mapping of all present modifications in mouse, rat, pig, and primate cTnI.25,27,29,30 In the case of mouse cTnI, high-resolution ESI/MS revealed the coexistence of unphosphorylated, monophosphorylated, and bisphosphorylated forms, together with minor N-terminal proteolytic fragments.25 The subsequent MS/MS of individually isolated unphosphorylated, monophosphorylated, and bisphosphorylated forms generated extensive fragmentation ions, including complementary pairs that covered the entire sequence. For a 210–amino acid mouse cTnI, \( e_{134}/z_{166} \) is a complementary pair, resulting from the cleavage of the same bond between amino acid 134 and 135. \( e_{134} \) covers the first 134 amino acids from the N-terminus, and \( z_{166} \) covers the last 166...
amino acids from the C-terminus. All the MS/MS data unambiguously identified Ser22/23, the *bona fide* sites for protein kinase A (PKA), as the only phosphorylation sites in cTnI immunoaffinity purified from healthy wild-type mouse hearts, consistent with our other studies\(^{27,29,30}\) of rat, pig, and primate cTnI. Furthermore, high-resolution MS and Pro-Q diamond gel analysis consistently showed the lack of phosphorylation in transgenic mouse cTnI (cTnI-\(\text{Ala}^2\)), where Ser22/23 in cTnI had been rendered nonphosphorylatable by mutations to Ala.\(^{25}\) These data confirmed that top-down MS mapped all phosphorylation sites in mouse cTnI. However, cTnI is also well-known to be phosphorylated by PKC at Ser43/45 and Thr144 (mouse sequence counting the N-terminal Met). Thus, the question is whether these PKC sites, mainly identified via in vitro phosphorylation assays, are fact or fancy.\(^{69}\) It is possible that, in the healthy mouse hearts, the phosphorylation occupancy (\(<1\%\)) of these PKC sites is lower than the limit of detection at the current stage of top-down method development.\(^{25}\) More importantly, because we have only analyzed the basal phosphorylation state in healthy animals, the phosphorylation of the PKC sites in cTnI might be related to cardiac dysfunction.\(^{67}\) The transgenic mouse lines that overexpressed PKC in the myocardium exhibit a steady progression to heart failure,\(^{70,71}\) and partial replacement of cTnI with a nonphosphorylatable mutant at Ser43/45 attenuates the contractile dysfunction, suggesting that PKC sites Ser43/45 may contribute to the progression of failure.\(^{72}\) Indeed, our recent study has precisely identified PKC sites of Ser43/45 in cTnI affinity purified from spontaneously hypertensive heart failure rat (X. Dong et al, unpublished data).

**Discovery of the Unexpected Modifications**

Top-down MS is also able to discover unexpected PTMs or amino acid polymorphisms (ie, protein-sequence polymorphisms).\(^{27,29,68,73}\) Typically, the first step in the top-down MS approach is to compare the experimental MW with that predicted from the DNA sequence. The “incorrect” MW could immediately caution that there might be sequence discrepancy (“error”) present in the protein sequence. Then, the precursor ion containing the error will be isolated and fragmented to locate the precise error position. For example, Zhang et al\(^{29}\) effectively localized an unexpected single amino acid polymorphism, V116A, in addition to phosphorylation, acetylation, and removal of N-terminal Met for swine cTnI (Figure 3), which was extracted, immunoaffinity enriched, from the left ventricle of domestic pig myocardium and separated and desalted using off-line RPLC before MS analysis (Figure 3A). The high-resolution MS analysis revealed 3 major MW forms, corresponding to unphosphorylated, monophosphorylated, and bisphosphorylated swine cTnI. Surprisingly, all 3 cTnI forms contained an error of \(\Delta m = -28\) Da from the predicted value (Figure 3B). The bisphosphorylated cTnI was individually isolated and frag-
mented by CID and ECD. All fragment ions containing the first 115 amino acids from the N-terminus, and the fragment ions containing the first 88 amino acids from the C-terminus, were error free, which essentially narrows the error down to the 7–amino acid region after the first 115 and before the last 88 amino acids, V116-I122 (VKVTKNI) (Figure 3C). The only possibility to account for the -28 Da mass discrepancy is the replacement of Val for Ala and most likely to be V116A based on the sequence homology alignment (Figure 3D). We assigned it as an amino acid polymorphism instead of a mutation because the latter typically is associated with pathological characteristics; healthy pig heart tissue was used in this study. An amino acid polymorphism of Ala/Ser at residue 7 and the addition of Gln (Q) at residue 192 were also identified for rat cTnI and cardiac troponin T (cTnT).27

Quantification of Modified Protein Levels
Protein quantification is becoming an increasingly important subject in MS-based proteomics.75,76 The top-down MS approach is especially attractive for quantification of the relative abundance of modified protein species, because the effect of the modifying groups on the physicochemical properties of the intact proteins in the top-down approach is much less compared with that of the peptides in the bottom-up approach.24,28,77 Gordon and Muddiman78 demonstrated a quantitative relationship between the concentration ratios of 2 proteins that are 97% identical and their ESI/FTMS peak ratios, as well as a linear relationship between the protein concentration and the FTMS peak intensity.77 The Kelleher group developed a method called protein ion relative ratio, in which the relative ratio of MS signal intensity values was used to calculate the relative amount of modified protein forms, and applied it to the global assessment of the combinatorial PTMs of core histones79 and histone H4.80 Our group has modified such a protein ion relative ratio method and quantitatively deter-

Figure 3. Top-down mass spectrometry (MS) for the discovery of unexpected modifications. A, The workflow of extraction and purification of cardiac troponin (cTn) from domestic swine hearts for MS analysis. B, FTMS spectrum of swine cTnI for the charge state 28+ precursor ions, showing its distribution in unphosphorylated, monophosphorylated, and bisphosphorylated forms. Insets: Isotopically resolved molecular ions of unphosphorylated, monophosphorylated, and bisphosphorylated cTnI (M+28+) with high-accuracy molecular weight measurements. cTnI and ppTcTnI represent monophosphorylated (79.95 Da) and bisphosphorylated (159.92 Da) cTnI, respectively. Circles represent the theoretical isotopic abundance distribution of the isotopomer peaks corresponding to the assigned mass. Calcd, calculated most abundant mass; Exptl, experimental most abundant mass. C, MS/MS fragmentation and product ion map from electron-capture dissociation (ECD) and collision-induced dissociation (CID) spectra for bisphosphorylated swine cTnI (ppcTnI). Bisphosphorylated cTnI was isolated and fragmented by ECD and CID. Fragment assignments were made to the swine cTnI (UniProtKB/Swiss-Prot A5X5T5, TNNI3_pig) with the removal of N-terminal Met and acetylation at the new terminus. Bisphosphorylation sites of Ser22/23 were highlighted by circles. The fragmentation ions carrying the mass discrepancy (-28 Da) were indicated in dots. The potential amino acids containing the mass discrepancy (-28 Da) were highlighted in shades. D, Schematic representation of all identified modifications for bisphosphorylated swine cTnI. Modified based on Zhang et al.29 with permission.
mined the phosphorylation levels in human, mouse, rat, swine, and primate cTnI25,28–30 and recombinant mouse cardiac myosin-binding protein C (cMyBP-C).26

Wanders et al81 extended the stable isotope labeling by amino acids in cell culture method to intact proteins for the quantification of their relative expression levels and the degrees of modification between different samples. More recently, Collier et al82 reported the quantitative top-down proteomics of stable isotope labeling by amino acids in cell culture–labeled human embryonic stem cells.

Identification and Quantification of Modified Positional Isomers

Top-down MS also appears to be uniquely advantageous for identification and quantification of modified positional isomers that are results from the occurrence of the same modification group at a different location (ie, position A versus B). How top-down MS can be used to identify positional isomers is illustrated in Figure 4A. In a hypothetical example of positional isomers, all 3 cases (case 1 of proteins carrying the modification entirely at position A, case 2 of proteins carrying the same modification entirely at position B, and case 3 of a mixture of proteins with modification at either position A or position B) showed the identical molecular ion spectra (Figure 4A). The subsequent MS/MS can generate different product ion spectra for fragments with bond cleavages between position A and B (eg, hypothetical fragments Cx', with modification, and Cx, without modification), which distinguishes the positional isomers in all 3 cases (Figure 4A). We have identified positional isomers for monophosphorylated human cTnI.28 The ECD data suggested 2 possible sites for the monophosphorylation: the well-known PKA site Ser22 and a novel site at Ser76/Thr77 (Figure 4B).

Figure 4. Top-down mass spectrometry (MS) for identification and quantification of positional isomers. A, Schematic representation of positional isomers determined by top-down MS. The MS spectra of hypothetical positional isomers: (1) protein modified entirely at position A, (2) protein modified entirely at position B, and (3) a mixture of proteins modified at positions A and B yields identical mass. The subsequent MS/MS spectra of these positional isomers with bond cleavages between position A and B (eg, hypothetical fragments Cx’, with modification, and Cx, without modification) yield different masses, which distinguishes the positional isomers in 1 through 3 and quantifies the percentages of modification on the position A vs B accordingly. B, Quantification of phosphorylated positional isomers in human cTnI. The normalized absolute abundance ratios and phosphorylation occupancy (PPO%) of unphosphorylated c fragment ions from electron-capture dissociation (ECD) spectra of both unphosphorylated and monophosphorylated molecular ions are plotted vs the amino acid sequence (only 114 N-terminal residues shown). Positional isomers were identified and quantified as Ser22 (53 ± 4%) and Ser76/77 (36 ± 3%). Modified based on Zabrous- kov et al,28 with permission.
phosphorylation.\textsuperscript{28} Our method used only the abundance of unphosphorylated fragment ions via \textit{intraspectrum} normalization versus “internal standard” fragments, followed by a comparison between ECD spectra of unphosphorylated and phosphorylated precursor ions. This allowed for an accurate quantitative determination of partial phosphorylation occupancies of the corresponding sites because the abundance ratios revealed the difference attributed to phosphorylation events. By using this method, we quantified the phosphorylated positional isomers in human cTnI as follows: Ser22, 53±4%; Ser76/Thr77, 36±3% (Figure 4B).\textsuperscript{28} On the other hand, it is difficult to use the bottom-up approach to determine the positional isomers because of the partial sequence coverage and loss of connectivity among different modifications.

**Determination of the Order of Multiple Phosphorylation Sites**

If a protein harbors multiple phosphorylation sites, 1 question is whether all modifications occurred simultaneously or sequentially. Top-down MS is well situated to answer this question because it can isolate each specific phosphorylation form and locate the sites. Recently, we have used top-down MS to characterize the phosphorylation order in cTnI purified from nonhuman primate myocardial tissues.\textsuperscript{30} cTnI is present at unphosphorylated, monophosphorylated, and bisphosphorylated forms, with bisphosphorylation as the predominant form. We have individually “gas phase” isolated unphosphorylated, monophosphorylated, and bisphosphorylated forms in the mass spectrometer and subsequently fragmented each form with ECD for localization of the modification sites. Ser23 was the only site phosphorylated in the monophosphorylated cTnI form, but Ser22 and Ser23 were both phosphorylated in the bisphosphorylated cTnI form, consistent with the phosphorylation that occurred at Ser23 before Ser22.

We have also unveiled the phosphorylation order in both truncated and full-length recombinant mouse cMyBP-C,\textsuperscript{26} an important regulator of cardiac contractility located in the sarcomere’s thick filament.\textsuperscript{33–36} We have identified the phosphorylation sites in full-length recombinant cMyBP-C as Ser283, Ser292, and Ser312 (corresponding to Ser273/282/302 in endogenous mouse cMyBP-C), with a sequential phosphorylation among them: the phosphorylation of Ser292 occurs before the phosphorylation of Ser312 and Ser283. Gautel et al\textsuperscript{37} previously identified these 3 sites as substrates for PKA. Ser292 is located on a cardiac-specific loop (LAGARRTS), and the phosphorylation of Ser292 could induce a conformational change that makes the other sites accessible to PKs.\textsuperscript{87} Surprisingly, the identified phosphorylation sites in truncated cMyBP-C protein are different from that of full-length cMyBP-C. We have localized the phosphorylation sites in C0-C4 (containing the intact N-terminal 4 domains and the cMyBP-C motif) to Ser292, Ser312, and Ser484, with no apparent phosphorylation order because both Ser292 and Ser312 were concurrently phosphorylated. This suggested that sequence truncations can alter the protein PTM state, which can potentially lead to variations in protein structure and function.\textsuperscript{26}

**Technical Challenges of Top-Down MS**

As previously illustrated, top-down MS has many advantages for proteomics and is superior to the bottom-up approach for protein modification analysis. However, there are still technical challenges yet to be resolved to realize its full potential.

**Protein Solubility**

Proteins are generally much more difficult to handle than small peptides, mainly because of their poor solubility.\textsuperscript{21} Unlike the tryptic peptides, which are highly soluble under the general LC/MS condition, proteins may not be all soluble under the same condition. In addition, some large proteins (MW >50 kDa) and almost all membrane proteins are difficult to solubilize without classic detergents (ie, SDS and Triton X-100). The available MS-compatible detergents are not as effective as the classic detergents and degrade rapidly in a typical acidic LC/MS buffer (0.1% formic acid). The “magic” top-down MS-compatible detergent, which can solubilize all types of proteins, is yet to be developed.

**Sensitivity and Detection Limit**

The sensitivity and detection limit of the mass spectrometer for proteins are much poorer than for peptides.\textsuperscript{21} Top-down MS could require much more concentrated and usually 10 or 100 folds more material than required for the bottom-up approach.\textsuperscript{32} In any type of MS instrument, the sensitivity decreases drastically with the increase of MW. At the current stage, it is still relatively difficult to analyze intact proteins >70 kDa. Moreover, as the MW of the protein increases, the tertiary structure of proteins becomes more difficult to disrupt, thereby limiting the MS/MS fragmentation efficiency of intact proteins. Thus, most of the top-down applications focused on proteins <50 kDa, and there are few applications on larger proteins (>100 kDa).\textsuperscript{22–32} Our laboratory has recently applied top-down ECD MS/MS to a 142-kDa recombinant cMyBP-C and isotope resolved a 115-kDa truncated form of cMyBP-C (the largest protein resolved isotopically to date).\textsuperscript{26} Han and coworkers\textsuperscript{31} extended the top-down MS to proteins >200 kDa by prefolding dissociation, which uses variable thermal and collisional activation immediately after ESI. Nevertheless, it required high protein purity (>80% homogeneity) and high concentration (0.5–1 μg/μL) for top-down analysis of the proteins >100 kDa. Hence, it is essential to develop MS instruments with increased sensitivity and a lower detection limit for large proteins to generalize the use of top-down MS because most of the proteins in the proteome are >50 kDa.

**Requirement of High-End Instruments**

Another reason for the popularity of the bottom-up approach is that small peptides (500–2000 Da) can be easily analyzed in low-end instruments, such as an ion trap mass spectrometer, with low resolution (and at a low cost).\textsuperscript{21,32} In contrast, top-down generally requires high-end instruments with a high resolution (and high cost) to resolve the protein precursors and the large fragmentation ions. Although top-down MS has been extended to the ion trap mass spectrometer, especially with the help of an ion-ion reaction, such an approach is still limited to a few academic laboratories and is not generally available.\textsuperscript{88}
Throughput
Historically, top-down MS has been used for elegant and comprehensive characterization of a purified single protein or a protein mixture of relatively low to moderate complexity, with relatively low throughput. In the past 10 years, Kelleher and other laboratories have been working on a high-throughput version of top-down proteomics, integrating on-line separation of intact proteins, automatic MS, and MS/MS data acquisition and bioinformatics tool for data processing.37 For example, Roth et al.89 detected >600 unique intact masses (≤63 kDa) in the human primary leukocytes using a multidimensional protein characterization by automated top-down approach. A subsequent MS/MS experiment identified 133 proteins from 67 unique genes, with 32 of the identified proteins harboring coding polymorphisms and PTMs, suggesting the diversity of the human proteome. Admittedly, the throughput of top-down proteomics, although increasing at a good rate, is still not comparable to bottom-up shotgun proteomics at its present stage. The question will remain whether 1 day in the future top-down proteomics will be able to bottom-up shotgun proteomics at its present stage. The throughput of top-down MS is still in its early developmental stage and in the process of overcoming several technical challenges, such as solubility, sensitivity, and throughput issues. Perceptibly, bottom-up proteomics as a mature technology will continue to serve as a workhorse in modern proteomics at present and in the near future. In some sense, the bottom-up and top-down approaches are really complementary in proteomics. Bottom-up MS is sufficient for the identification of protein from the database and quantification of protein expression level. Meanwhile, top-down MS is ideally suited for the comprehensive analysis of protein PTMs and sequence variations. The synergy between bottom-up and top-down approaches will yield best proteomic results and broaden the application of proteomics in biomedical research. A combined bottom-up/top-down hybrid approach50 and a “middle-down” proteomics approach (MS on large peptides at ~3–20 kDa from limited digestion of proteins) will play potentially important roles during this high-mass “void” until the point that the top-down approach overcomes the technical challenges to realize its full potential.38 Evidently, top-down MS has achieved substantial progress in the past few years.37 Hence, with continuous developments at such a speed on the new methods, including the front end separation/protein solubility, MS instrument sensitivity and detection limit, and user-friendly data processing and automation software, it is hopeful that top-down MS soon can become widely used in the near future.

Conclusions and Perspectives
As reviewed herein, top-down MS emerges as a powerful technology in proteomics, particularly for the analysis of protein PTMs and sequence variants. It has offered unique opportunities to cardiovascular proteomics, including unraveling molecular complexity, complete PTM mapping, quantification of modified protein forms, identification and quantification of positional isomers, and determination of the order of multiple phosphorylations, which can be difficult to achieve by the bottom-up MS approach. However, top-down MS is still in its early developmental stage and in the process of overcoming several technical challenges, such as solubility, sensitivity, and throughput issues. Perceptibly, bottom-up proteomics as a mature technology will continue to serve as a workhorse in modern proteomics at present and in the near future. In some sense, the bottom-up and top-down approaches are really complementary in proteomics. Bottom-up MS is sufficient for the identification of protein from the database and quantification of protein expression level. Meanwhile, top-down MS is ideally suited for the comprehensive analysis of protein PTMs and sequence variations. The synergy between bottom-up and top-down approaches will yield best proteomic results and broaden the application of proteomics in biomedical research. A combined bottom-up/top-down hybrid approach50 and a “middle-down” proteomics approach (MS on large peptides at ~3–20 kDa from limited digestion of proteins) will play potentially important roles during this high-mass “void” until the point that the top-down approach overcomes the technical challenges to realize its full potential.38 Evidently, top-down MS has achieved substantial progress in the past few years.37 Hence, with continuous developments at such a speed on the new methods, including the front end separation/protein solubility, MS instrument sensitivity and detection limit, and user-friendly data processing and automation software, it is hopeful that top-down MS soon can become widely used in the near future.

The ultimate goal of developing proteomic technologies is to understand molecular mechanism of diseases and to diagnose them at the early stage.1,5–7 There are 2 major issues we must address to advance the application of top-down MS in cardiovascular research. The first issue is to comprehensively characterize the protein modification states in disease models, including universal and unbiased detection of all PTMs and sequence variants, precise localization of the modification sites, and identification and quantification of the changes in the distribution of PTMs (ie, phosphorylation) among multiple targeted sites during disease progression. This goal can be readily achieved. In fact, our laboratory has already applied top-down MS to characterize cTnI from diseased cardiac tissues of humans and animal models (unpublished data). The second issue is to detect specific modifications of intact protein biomarkers (eg, cTnI) in the general circulation system and to subsequently characterize their modifications. Blood is certainly a much more practical source for routine clinical diagnosis than tissues (especially cardiac tissue). Therefore, the ability to detect protein modifications in blood will significantly increase the impact of top-down MS. Obviously, this is much more challenging because of the high complexity and dynamic range in the blood proteome, which will require the development of a high-resolution mass spectrometer with an ultrahigh sensitivity. In the near term, we can discover the specific modified forms in tissues as potential candidate biomarkers and then validate them in serum/plasma using either an antibody approach (by developing antibodies recognizing the protein and its specific modifications) or a targeted-MS approach (by developing multiple reaction monitoring with stable isotope dilution methods).91 Overall, we believe that with the continuous development in the technology front and with a collaborative effort among the MS/proteomics researchers, instrument engineers, biologists, clinicians, and bioinformatics specialists, top-down MS-based proteomics has great potential to transform the approach to cardiovascular research.

Acknowledgments
We thank Drs. Raquel Sancho Solis, Beth Altschafl, and Wei Xu for critical reading of this review.

Sources of Funding
This study was supported by the Wisconsin Partnership Fund for a Healthy Future American Heart Association Scientist Development Grant and NIH R01HL096971 (to Dr Ge).

Disclosures
None.

References


Zhang and Ge

Top-Down MS for Intact Protein Analysis


