



Published in final edited form as:

*Circ Cardiovasc Genet.* 2011 December ; 4(6): 711. doi:10.1161/CIRCGENETICS.110.957829.

## Comprehensive Analysis of Protein Modifications by Top-down Mass Spectrometry

**Han Zhang, PhD and Ying Ge, PhD**

Departments of Physiology (Y.G.) and Human Proteomics Program (H.Z., Y.G.), School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, 53706

### Abstract

Mass spectrometry (MS)-based proteomics is playing an increasingly important role in cardiovascular research. Proteomics includes not only identification and quantification of proteins, but also the characterization of protein modifications such as post-translational modifications and sequence variants. The conventional bottom-up approach, involving proteolytic digestion of proteins into small peptides prior to 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 due to 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 non-ergodic dissociation methods such as electron capture dissociation (ECD) greatly enhances the top-down capabilities. ECD is especially useful for mapping labile post-translational modifications which 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, and 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 top-down methodology with a focus on its application in cardiovascular research.

### Keywords

Cardiovascular diseases; Proteomics; Electron Capture dissociation; Post-translational modification; Top-Down Mass Spectrometry

### Introduction

Proteomics is playing an increasingly important role in cardiovascular research (see reviews<sup>1-8</sup>). Proteomics includes not only the separation, identification and quantification of proteins, but also the characterization of protein modifications such as post-translational

---

Corresponding author: Ying Ge, PhD. 1300 University Ave., SMI 130, Madison, WI 53706, Tel: 608-263-9212, Fax: 608-265-5512. yge@physiology.wisc.edu.

#### Disclosures

Dr. Han Zhang and Dr. Ying Ge have no potential conflicts of interest.

modification (PTMs) and sequence variants (e.g. mutants, alternatively spliced isoforms, and amino acid polymorphism).<sup>9–10</sup> PTMs (e.g. phosphorylation, glycosylation, acetylation, proteolysis) are covalent modifications of a protein after its translation. PTMs can modulate the activity, stability, and function of a protein.<sup>11</sup> 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.<sup>1</sup> Aberrant protein PTMs together with mutations and alternatively spliced isoforms are increasingly recognized as important underlying mechanisms for cardiovascular diseases.<sup>12–14</sup> Hence, a comprehensive analysis of protein modifications is of high importance for understanding the disease mechanisms.

Traditional strategies for 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 cannot provide the identification of proteins or their modification sites which are essential for understanding the disease mechanisms.<sup>15–16</sup> Biological mass spectrometry (MS) is the preferred method for the analysis of protein modifications since it is capable of not only providing universal information about protein modifications without *a priori* knowledge but also locating the sites of modification.<sup>11</sup>

Currently, there are two complementary approaches in MS-based proteomics: “bottom-up” and “top-down”.<sup>17–20</sup> 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) prior to MS analysis (Fig. 1A). This approach is well-suited for protein identification which only requires a very small portion of sequence coverage (~10–20 amino acid residues) to identify the protein from the database.<sup>21</sup> With the tremendous efforts dedicated to the development of hardware and software in the past decade, bottom-up shotgun proteomics currently serves as a workhorse in modern proteomics with high throughput and automation. Nevertheless, the bottom-up approach has intrinsic limitations in characterizing protein modifications since only a small and variable fraction of peptides is recovered from digestion resulting in low percentage coverage of the protein sequence. In other words, even if one identifies all modifications present in the recovered peptides, the modification status of the unrecovered sequence portion remains unknown. In addition, the connection between modifications on disparate portions of a protein can be lost since the typical peptides from tryptic digestion contain only ~5–20 amino acids.<sup>17</sup> Furthermore, since each protein is digested into many small peptide components, the overall complexity of the sample is increased.

Top-down MS is becoming a powerful technology for comprehensive analysis of protein modifications.<sup>22–33</sup> In contrast to bottom-up MS, top-down MS analyzes intact proteins without proteolytic digestion (Fig. 1B). This strategy preserves the labile structural characteristics which are mostly destroyed in bottom-up MS.<sup>32</sup> It can universally detect all the existing modifications including PTMs (i.e. phosphorylation, proteolysis, acetylation) and sequence variants (i.e. mutants, alternatively spliced isoforms, amino acid polymorphisms) simultaneously in one spectrum (a “bird’s eye” view) without *a priori* knowledge.<sup>32</sup> 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).<sup>10, 32</sup> The incorporation of the novel MS/MS technique, ECD,<sup>34</sup> has

greatly enhanced the capability of top-down MS in structural analysis of biomolecules (see review<sup>35</sup>). As a non-ergodic fragmentation method,<sup>34</sup> ECD preserves labile PTMs during the fragmentation process thus it is particularly suitable for the localization of labile PTMs.<sup>25–26</sup> 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, and identifying and quantifying positional isomers and determining the order of multiple modifications.<sup>25–30</sup>

In contrast to the well-established bottom-up proteomics, the top-down proteomics is still in its early developmental stage and yet to fully overcome its technical challenges in sample preparation, instrument sensitivity/detection limit, and throughput/automation.<sup>32, 36–39</sup> 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 methodology, 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 Methodology

### Sample preparation for top-down MS

Typically, intact proteins need to be extracted from cell/tissue lysate, solubilized and separated/purified prior to 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 high salt concentration usually with the addition of detergents such as sodium dodecyl sulfate (SDS) to increase the solubility of the protein. These salts and detergents interfere with MS detection 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 column/trap either on-line or off-line prior to MS analysis. Alternatively, MS-compatible volatile salt buffers such as ammonia bicarbonate and ammonia acetate can also be used to replace the common salt present in biological samples (i.e. NaCl, KCl, CaCl<sub>2</sub>) via a buffer exchange step such as dialysis or ultra-filtration. Typical procedures for detergent removal involve precipitation and resolubilization of proteins in detergent-free buffers, which may result in sample loss since 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.<sup>40,41</sup> Alternatively, proteins can also be selectively solubilized based on their inherent chemical properties such as biospecificity, hydrophobicity and charge without the use of detergent.<sup>1</sup> This can also be used to fractionate specific subproteome prior to chromatographic separation.

### Separation/purification of intact proteins

A complex protein mixture can be separated by SDS-polyacrylamide gel electrophoresis (PAGE), liquid chromatography (LC) as well as the high resolution mass spectrometry. One dimensional (1D) and 2D SDS-PAGE are the classical methods for protein separation and visualization.<sup>1, 42–44</sup> The gel-based separation is widely used in bottom-up proteomics since trypsin digested peptides can be effectively retrieved from gels.<sup>37, 39</sup> However it is technically challenging to extract the intact proteins from gel matrixes with a high recovery rate.<sup>45</sup> Thus gel-based separation is not applicable in top-down MS. Recently, a solution-based isoelectric focusing coupled with a multiplex tube gel electrophoresis separation

device referred to as gel-eluted liquid fraction entrapment electrophoresis (GELFrEE)<sup>46</sup> has been developed for intact protein separation based on their MWs and applied to proteins (10–250 kDa) with high resolution and high recovery rate.<sup>24, 37, 47</sup> 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 since it can be conveniently interfaced with MS (see reviews<sup>20, 48–49</sup>). The major LC techniques typically used for intact protein separation include affinity, ion exchange chromatography (IEC), size exclusion chromatography (SEC), and reverse-phase (RP) chromatography.<sup>48</sup> Affinity chromatography is by far the most effective and specific protein purification method.<sup>50</sup> For example, immunoaffinity methods have been utilized to effectively purify cardiac troponin I (cTnI) from animal and human myocardium.<sup>15, 25, 27, 30, 42</sup> Nonetheless, most of the affinity methods have been performed offline, requiring an additional separation/desalting procedure using RPLC. IEC and SEC have also been employed for intact protein separation.<sup>48, 51–53</sup> Generally, they are used to carry out the first dimension separation followed by RP chromatography in the second dimension. RP chromatography not only enhances the separation from the previous step but also performs desalting as the last sample preparation step before MS analysis.<sup>52–53</sup> The 2DLC approach has the advantage of concentrating and desalting the species of interest simultaneously, yielding higher peak capacity and better separation, and, if connected on-line, minimizing sample loss. Chromatographic focusing, as a promising alternative to salt gradient IEC, is also capable of separating intact proteins.<sup>37</sup> Van Eyk and co-workers reported the separation of intact proteins from human serum proteome with the use of chromatographic focusing and RPLC in a commercial ProteomeLab PF2D system.<sup>54</sup>

With high resolution mass spectrometers e.g. Fourier transform ion cyclotron resonance (FTICR) MS (see review<sup>55</sup>), protein mixtures with high complexity may be resolved during MS analysis in addition to other traditional separation techniques. It has been demonstrated that over 100 protein components from a crude cell lysate were successfully separated in one FTICR MS spectrum using the high resolving power of FTICR.<sup>23</sup>

### Top-down MS instrumentation

A typical mass spectrometer consists of four components: 1) sample inlet, 2) ion source which converts sample molecules to ions such as electrospray ionization (ESI) and matrix assisted laser desorption/ionization (MALDI), 3) mass analyzer, and 4) detector (see reviews<sup>20, 37, 56</sup>). Top-down MS is usually performed in high resolution instruments due to the need to resolve the high MW of intact proteins.<sup>32, 55</sup> The early top-down MS studies were done in home-built FTICR mass spectrometers with high resolution and accuracy.<sup>22–23</sup> Recently, the development of a number of commercial high resolution hybrid mass spectrometers such as LTQ/FT, LTQ-orbitrap (a hybrid between ion trap and FT), and quadrupole time-of-flight (Q-TOF) substantially boosted the growth of top-down MS applications.<sup>39</sup> ESI is preferred over MALDI in top-down proteomics since 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.<sup>57</sup> 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 two categories of fragmentation methods: energetic vs. non-ergodic (Fig. 1C).<sup>10, 32, 39</sup> The well-developed energetic dissociation methods such as CID, infrared multiphoton dissociation (IRMPD) and post-source decay (PSD), cleave CO-NH backbone bonds to produce *b* and *y* fragment ions at very high efficiency. All these energetic

dissociation methods preferably dissociate bonds of the lowest activation energy. Since 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 ( $\text{H}_3\text{PO}_4$ , 98 Da) or metaphosphoric acid ( $\text{HPO}_3$ , 80 Da), which makes it difficult to map phosphorylation sites.

The non-ergodic electron-based MS/MS techniques, ECD<sup>34</sup> and electron transfer dissociation (ETD),<sup>58</sup> are particularly suitable for the localization of labile PTMs like phosphorylation.<sup>25–26, 30, 35</sup> In ECD, low energy thermal electrons are captured by a protonated peptide or protein causing local fast ( $<10^{-12}$  s) cleavages of backbone covalent bonds, NH-CHR, producing mainly  $c$  and  $z^*$  ions (Fig. 1C).<sup>34</sup> Hence the labile PTMs are well-preserved in the ECD fragmentation process and no loss of labile phosphate has been observed in ECD spectra.<sup>25–26, 30</sup> ECD is non-selective so that it often provides far more cleavages than CID,<sup>23</sup> which greatly enhances the capability of top-down MS in identifying PTMs and sequence variants.<sup>23, 25, 27, 29–30, 59</sup> Nevertheless, ECD is only available at high-end FTICR instruments, which substantially limits its wide use in proteomic laboratories. More recently, ETD,<sup>58</sup> 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,<sup>60</sup> which begins to play a role in top-down proteomics.<sup>61</sup>

### Top-down MS data analysis tools

Since ESI, the commonly used ionization method in top-down MS, produces multiply-charged ions, Horn *et al.* 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.<sup>62</sup> Unfortunately, this software is not compatible with the mainstream computers. Smith's lab developed free downloadable PC and Mac compatible software (<http://omics.pnl.gov/software>), DeconMSn and Decon2LS, based on 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 currently available. There are three major search engines: ProSight, PIITA, and MascotTD (also known as "big Mascot") (for reviews, see<sup>37–38</sup>). ProSight, developed by Kelleher group, is the first and most flexible search engine for top-down proteomics.<sup>63</sup> It allows the identification of proteins from complex mixtures by searching both the precursor ion MS and product ion MS/MS data (processed via THRASH algorithm) in the databases. ProSight can be used to map both known and unknown PTMs.<sup>37</sup> PIITA is based on precursor ion independent top-down algorithm using fragmentation data without *a priori* expectation of PTMs or sequence errors.<sup>64</sup> A recently released big Mascot extended the classical Mascot to large proteins (>16 kDa). Big Mascot search requires both precursor and fragment ions and can identify protein PTMs and sequence variants.<sup>65</sup>

## Application of Top-down MS to Cardiovascular Research

### Unraveling 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 since it analyzes whole proteins instead of small peptides.<sup>32</sup> The presence of PTMs can be detected by the mass difference ( $\Delta m$ ) between unmodified and modified forms (See review<sup>11</sup>). Moreover, most of the common amino acids have distinct mass (except Leu, 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 have recently employed the high resolving power of FTICR MS to unravel the complexity of PTMs for a commercially available cTnI sample said purified from healthy human heart tissues (Fig. 2).<sup>28</sup> cTnI is the inhibitory subunit of the thin filament troponin-tropomyosin regulatory complex, playing a critical role in Ca<sup>2+</sup>-mediated cardiac muscle contractility.<sup>66–67</sup> ESI/FTMS spectrum of cTnI revealed a total of thirty-six modified molecular ions (Fig. 2A) despite the protein running as a single sharp band on SDS-PAGE (Fig. 2B). The accurate mass measurements of all the well resolved cTnI forms suggest several mass differences ( $\Delta m$ ) (Fig. 2A). For example,  $\Delta m$  of +79.971 Da matches with phosphorylation (Calc'd: 79.966 Da) and  $\Delta m$  of +15.996 Da is commonly assigned as oxidations (Calc'd: 15.995 Da) of the proteins. Moreover,  $\Delta m$  values of 128.130, 147.069, and 216.085 Da correspond to amino acids Q/K (Calc'd: Q, 128.059 Da; K, 128.095 Da), F (Calc'd, 147.068 Da), and ES (Calc'd, 216.075 Da) respectively, which were resulted from truncations of *Glu-Ser*, *Phe-Glu-Ser* and *Lys-Phe-Glu-Ser* residues 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 here matched with the predicted sequence (TNNI3\_human) obtained from 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, mono- and bis-phosphorylation, oxidation, and proteolytic degradation for this commercially available cTnI.<sup>28</sup> The major modified forms of cTnI, resulting from the combinations of these multiple modifications, were illustrated in Fig. 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 not only can reveal the molecular complexity but also precisely map sites of modifications by isolating and fragmenting the modified protein ions in the mass spectrometer.<sup>10,68</sup> MS/MS produces *b/y* or *c/z*<sup>+</sup> fragment ions with *b/c* ions counting from the N-terminus and *y/z*<sup>+</sup> ions from the C-terminus (Fig. 1C). In a top-down MS/MS experiment, full sequence coverage can be easily achieved for a protein less than 60 kDa.<sup>25, 27, 29–30, 68</sup> We have utilized top-down MS with ECD for the comprehensive mapping of all present modifications in mouse, rat, pig, and primate cTnI.<sup>25, 27, 29–30</sup> In the case of mouse cTnI, high resolution ESI/MS revealed the coexistence of un-, mono-, and bis-phosphorylated forms, together with minor N-terminal proteolytic fragments.<sup>25</sup> The subsequent MS/MS of individually isolated un-, mono-, and bis-phosphorylation forms generated extensive fragmentation ions including complementary pairs which covered the entire sequence (e.g. for a 210 amino acid mouse cTnI, *c*<sub>134</sub>/*z*<sup>+</sup><sub>166</sub> is a complementary pair resulting from the cleavage of the same bond between amino acid 134 and 135. *c*<sub>134</sub> covers the first 134 amino acids from the N-terminus and *z*<sup>+</sup><sub>166</sub> covers the last 166 amino acids for 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 of rat, pig and primate cTnI.<sup>27, 29–30</sup> Furthermore, high resolution MS and Pro-Q diamond gel analysis consistently showed the lack of phosphorylation in transgenic mouse cTnI (cTnI-*Ala2*) where Ser22/23 in cTnI had been rendered non-phosphorylatable by mutations to Alanine.<sup>25</sup> This data confirmed that top-down MS mapped all phosphorylation sites in mouse cTnI. However, cTnI is also well-known to be phosphorylated by protein kinase C (PKC) at Ser43/45 and Thr144 (mouse sequence counting the N-terminal Met). So the question is whether these PKC sites, mainly identified via *in vitro* phosphorylation assays, are fact or fancy.<sup>69</sup> It is possible that in the healthy mouse hearts the phosphorylation occupancy (<1%) of these PKC sites are below the limit of detection at current stage of the top-down methodology development.<sup>25</sup> More importantly, since 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.<sup>67</sup> The transgenic mouse lines that over-expressed PKC in the myocardium exhibit a steady progression to heart failure<sup>70–71</sup> and partial replacement of cTnI with a non-phosphorylatable mutant at Ser43/45 attenuates the contractile dysfunction suggesting that PKC sites Ser43/45 may contribute to the progression of failure.<sup>72</sup> Indeed our recent study has precisely identified PKC sites of Ser43/45 in cTnI affinity purified from spontaneously hypertensive heart failure rat (Dong X., *et al.* unpublished data).

### Discovery of the unexpected modifications

Top-down MS is also able to discover unexpected PTMs or amino acid polymorphisms (a.k.a. protein-sequence polymorphisms).<sup>27, 29, 68, 73</sup> Typically, the first step in top-down MS approach is to compare the experimental MW with that predicted from 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.* effectively localized an unexpected single amino acid polymorphism, V116A in addition to phosphorylation, acetylation, and removal of N-terminal Met for swine cTnI (Fig. 3)<sup>29</sup> which was extracted, immunoaffinity enriched from the left ventricle of domestic pig myocardium, and separated and desalted using off-line RPLC prior to MS analysis (Fig. 3A). The high resolution MS analysis revealed three major MW forms corresponding to un-, mono-, and bis-phosphorylated swine cTnI. Surprisingly, all three cTnI forms contained an “error” of  $\Delta m = -28$  Da from the predicted value (Fig. 3B). The bis-phosphorylated cTnI was individually isolated and fragmented by CID and ECD. It was found that all fragment ions containing the first 115 amino acids from the N-terminus, as well as the fragment ions containing the first 88 amino acids from the C-terminus, were “error”-free, which essentially narrow the “error” down to the 7 amino acid region after the first 115 and before the last 88 amino acids, V116-I122 (VKVTKNI) (Fig. 3C). The only possibility to account for the  $-28$  Da mass discrepancy is the replacement of valine for alanine (V->A) and most likely to be V116A based on the sequence homology alignment (Fig. 3D). We assigned it as amino acid polymorphism instead of mutation since the latter typically is associated with pathology while healthy pig heart tissue was used in this study. 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).<sup>27</sup>

There are occasions even if the measured MW matches exactly with the predicted value; errors still exist in the sequence. A remarkable example is reported by Sze *et al.* where the comprehensive top-down MS sequencing of a 29 kDa protein, carbonic anhydrase was accomplished.<sup>74</sup> A total of 250 of the 258 interresidue bonds were extensively cleaved which essentially can define any modifications to within one residue. It was suggested that there are two “errors” in the first 10–31 amino acid region: Asp10 should be Asn10 and Asn31 should be Asp31.<sup>74</sup> Top-down MS has also been employed to discover unexpected PTMs such as an unexpected automethylation event.<sup>68</sup>

### Quantification of modified protein levels

Protein quantification is becoming an increasingly important subject in MS-based proteomics.<sup>75–76</sup> The top-down MS approach is especially attractive for quantification of the relative abundance of modified protein species, since the effect of the modifying groups on the physico-chemical properties of the intact proteins in the top-down approach is much less compared with that of the peptides in the bottom-up approach.<sup>24, 28, 77</sup> Muddiman and co-workers demonstrated a quantitative relationship between the concentration ratios of two proteins which are 97% identical and their ESI/FTMS peak ratios as well as a linear relationship between the protein concentration and the FTMS peak intensity over  $\sim 1.5$  orders of magnitude.<sup>78</sup> Kelleher group has developed a method called protein ion relative

ratio (PIRR) where the relative ratio of MS signal intensity values were used to calculate the relative amount of modified protein forms<sup>77</sup> and applied it to the global assessment of the combinatorial PTMs of core histones<sup>79</sup> and histone H4.<sup>80</sup> Our group has modified such a PIRR method and quantitatively determined the phosphorylation levels in human, mouse, rat, swine, and primate cTnI<sup>25, 28–30</sup> as well as recombinant mouse cardiac myosin binding protein C (cMyBP-C).<sup>26</sup>

Wannders *et al.* extended the stable isotope labeling by amino acids in cell culture (SILAC) methodology to intact proteins for the quantification of their relative expression levels and the degrees of modification between different samples.<sup>81</sup> More recently, Collier *et al.* reported the quantitative top-down proteomics of SILAC labeled human embryonic stem cells.<sup>82</sup>

### Identification and quantification of modified positional isomers

Top-down MS also appears to be uniquely advantageous for identification and quantification of modified positional isomers which are results from the occurrence of the same modification group at a different location (i.e. position A vs. B). How top-down MS can be used to identify positional isomers is illustrated in Fig. 4A. In a hypothetical example of positional isomers, all three 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 B) showed the identical molecular ion spectra (Fig. 4A). The subsequent MS/MS can generate different product ion spectra for fragments with bond cleavages between position A and B (e.g. hypothetical fragments  $C_x'$  with modification and  $C_x$  without modification), which distinguishes the positional isomers in all three cases (Fig. 4A). We have identified positional isomers for mono-phosphorylated human cTnI.<sup>28</sup> ECD data suggested two possible sites for the mono-phosphorylation: the well known PKA site Ser22 and a novel site at Ser76/Thr77 (Fig. 4B).

After the identification of positional isomer sites, the next step is to quantify the occupancy of the modification on each site. In other words, how much (e.g. percent) of the total protein population is modified on position A vs. B? The top-down MS strategy with ECD has been successfully employed for quantitative characterization of positional isomers since ECD generates fragment-ion-rich data.<sup>24, 28, 77</sup> To quantify positional isomers carrying the stable modifications such as deamidation or acetylation, the ratio of unmodified to modified fragments generated in ECD spectra was measured in a single spectrum.<sup>28, 77</sup> However, the accuracy in quantification of labile phosphorylated isomers would be compromised in a direct measurement of the abundance ratios of phosphorylated to unphosphorylated fragment ions ( $c/z^*$ ) within one spectrum. This is due to the possible effects from the presence of the phosphate moiety on the ECD cleavage efficiency. So the relative abundance of phosphorylated  $c$  ions derived from ECD are *not* necessarily proportional to the abundance of the corresponding phosphorylated isomeric precursors. Hence, we have developed a novel method to quantify positional isomers with labile modification such as phosphorylation.<sup>28</sup> Our method utilized only the abundance of unphosphorylated fragment ions via *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. Using this methodology, we have quantified the phosphorylated positional isomers in human cTnI as Ser22: 53±4%; Ser76/Thr77: 36±3% (Fig. 4B).<sup>28</sup> On the other hand, it is very difficult to use the bottom-up approach to determine the positional isomers due to 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, one question is whether all modifications occurred simultaneously or sequentially. Top-down MS is well situated to answer this question since it can isolate each specific phosphorylation form and locate the sites. Recently, we have employed top-down MS to characterize the phosphorylation order in cTnI purified from non-human primate myocardial tissues.<sup>30</sup> cTnI is present at un-, mono- and bis-phosphorylated forms with bis-phosphorylation as the predominant form. We have individually “gas-phase” isolated un-, mono-, bis-phosphorylated 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 monophosphorylated cTnI form but Ser22 and Ser23 were both phosphorylated in bis-phosphorylated cTnI form, consistent with the phosphorylation occurred at Ser23 prior to Ser22.

We have also unveiled the phosphorylation order in both truncated and full-length recombinant mouse cMyBP-C,<sup>26</sup> an important regulator of cardiac contractility located in the sarcomere’s thick filament.<sup>83–86</sup> 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 phosphorylation of Ser312 and Ser283. Gautel *et al.* has previously identified these three sites as substrates for PKA.<sup>87</sup> Ser292 is located on a cardiac-specific loop LAGARRTS and the phosphorylation of Ser292 could induce a conformational change which makes the other sites accessible to protein kinases.<sup>87</sup> Surprisingly, the identified phosphorylation sites in truncated cMyBP-C protein are quite 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 as both Ser292 and Ser312 were phosphorylated concurrently. This suggested that sequence truncations can alter the protein PTM state, which can potentially lead to variations in protein structure and function.<sup>26</sup>

## Technical Challenges of Top-down MS

As illustrated above, 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 due to their poor solubility.<sup>21</sup> 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 (i.e. SDS and Triton<sup>R</sup>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 is much lower than that for peptides.<sup>21</sup> Top-down MS could require much more concentrated and usually 10 or 100 folds more materials than that required for the bottom-up approach.<sup>32</sup> In any type of MS instrument, the sensitivity decreases drastically with the increase of MW. At current stage it is still relatively difficult to analyze intact proteins larger than 70 kDa. Moreover, as the MW of the protein increases, the tertiary structure of proteins becomes more difficult to

disrupt, which thereby limits the MS/MS fragmentation efficiency of intact proteins. Thus most of the top-down applications focused on proteins less than 50 kDa and there are very few applications to date on larger proteins (>100 kDa).<sup>22–32</sup> Our lab has recently applied the top-down ECD MS/MS to a 142 kDa recombinant cMyBP-C and isotopically resolved a 115 kDa truncated form of cMyBP-C – the largest protein resolved isotopically to date.<sup>26</sup> McLafferty and co-workers extended the top-down MS to proteins larger than 200 kDa by prefolding dissociation which utilizes variable thermal and collisional activation immediately after ESI.<sup>31</sup> Nevertheless, it required high protein purity (>80% homogeneity) and high concentration (0.5–1 µg/µl) for top-down analysis of the proteins larger than 100 kDa. Hence, it is essential to develop MS instruments with increased sensitivity and lower detection limit for large proteins in order to generalize the use of top-down MS since the majority of the proteins in the proteome are larger than 50 kDa.

### Requirement of high-end instruments

Another reason for the popularity of bottom-up approach is that small peptides (500–2000 Da) can be easily analyzed in low-end instruments such as ion trap mass spectrometer with low resolution (also at a low cost).<sup>21,32</sup> In contrast, top-down generally requires high-end instruments with high-resolution (and high cost) to resolve the protein precursors as well as the large fragmentation ions. Although top-down MS has been extended to ion trap mass spectrometer especially with the help of ion-ion reaction, such an approach is still limited to a small number of academic labs and not generally available.<sup>88</sup>

### Throughput

Historically, top-down MS has been employed for elegant and comprehensive characterization of purified single protein or protein mixture of relatively low to moderate complexity but with relatively low throughput. In the past ten years, Kelleher and other labs have been working on a high-throughput version of top-down proteomics integrating online separation of intact proteins, automatic MS and MS/MS data acquisition and bioinformatics tool for data processing (see review<sup>37</sup>). For example, Roth *et al.* detected > 600 unique intact masses (up to 63 kDa) in the human primary leukocytes using a multidimensional protein characterization by automated top-down approach.<sup>89</sup> 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 one day in the future top-down proteomics can become a truly high-throughput complement to bottom-up proteomics.<sup>38</sup>

### Conclusions and Perspectives

As reviewed here, 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 very 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 are really complementary approaches in proteomics. Bottom-up MS is sufficient for 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 approach<sup>90</sup> and a “middle-down” proteomics (MS on large peptides at about 3–20 kDa from limited digestion of proteins) will play potentially important roles during this high mass “void” till the point that top-down overcomes the technical challenges to realize its full potential.<sup>38</sup> Evidently, the top-down MS has achieved substantial progress in the last few years.<sup>37</sup> Hence, with continuous developments at such a speed on the new methodologies 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.

Needless to say, the ultimate goal of developing proteomic technologies is to understand molecular mechanism of diseases and to diagnose them at the early stage.<sup>1, 5–7</sup> There are two major issues we must address to advance the application of top-down MS in cardiovascular research. First 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, identification and quantification of the changes in the distribution of PTMs (i.e. 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). Second is to detect specific modifications of intact protein biomarkers (e.g. cTnI) in the general circulation system and 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 due to the high complexity and dynamic range in the blood proteome, which will require the development of a high resolution mass spectrometer with ultra-high sensitivity. In the near term, we can discover the specific modified form 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 targeted-MS approach (by developing multiple reaction monitoring with stable isotope dilution methods).<sup>91</sup> 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 Altschaf, and Wei Xu for critical reading of this review.

### Funding Sources:

This work was supported by the Wisconsin Partnership Fund for a Healthy Future and American Heart Association Scientist Development Grant 0735443Z and National Institutes of Health Grant R01HL096971 (to Y.G.).

## Abbreviations and Acronyms

PTMs	post-translational modifications
MS	mass spectrometry

<b>MS/MS</b>	tandem mass spectrometry
<b>MW</b>	molecular weight
<b>CID</b>	collision-induced dissociation
<b>ECD</b>	electron capture dissociation
<b>SDS</b>	sodium dodecyl sulfate
<b>PAGE</b>	polyacrylamide gel electrophoresis
<b>LC</b>	liquid chromatography
<b>GELFrEE</b>	gel-eluted liquid fraction entrapment electrophoresis
<b>IEC</b>	ion exchange chromatography
<b>SEC</b>	size exclusion chromatography
<b>RP</b>	reverse-phase
<b>cTnI</b>	cardiac troponin I
<b>FTICR</b>	Fourier transform ion cyclotron resonance
<b>CIEF</b>	capillary isoelectrofocusing
<b>ESI</b>	electrospray ionization
<b>MALDI</b>	matrix assisted laser desorption/ionization
<b>Q-TOF</b>	quadrupole time-of-flight
<b>IRMPD</b>	infrared multiphoton dissociation
<b>PSD</b>	post-source decay
<b>ETD</b>	electron transfer dissociation
<b>SILAC</b>	stable isotope labeling by amino acids in cell culture
<b>PKA</b>	protein kinase A
<b>PKC</b>	protein kinase C
<b>cTnT</b>	cardiac troponin T
<b>cMyBP-C</b>	cardiac myosin binding protein C
<b>Calc'd</b>	calculated mass
<b>Expt'l</b>	experimental mass

## References

1. Arrell DK, Neverova I, Van Eyk JE. Cardiovascular proteomics - evolution and potential. *Circ Res.* 2001; 88:763–773. [PubMed: 11325867]
2. Arab S, Gramolini AO, Ping PP, Kislinger T, Stanley B, van Eyk J, Ouzounian M, MacLennan DH, Emili A, Liu PP. Cardiovascular proteomics - tools to develop novel biomarkers and potential applications. *J Am Coll Cardiol.* 2006; 48:1733–1741. [PubMed: 17084242]
3. White MY, Van Eyk JE. Cardiovascular proteomics - past, present, and future. *Mol Diag Ther.* 2007; 11:83–95.
4. Edwards AVG, White MY, Cordwell SJ. The role of proteomics in clinical cardiovascular biomarker discovery. *Mol Cell Proteomics.* 2008; 7:1824–1837. [PubMed: 18667414]

5. Ping PP, Chan DW, Srinivas P. Advancing cardiovascular biology and medicine via proteomics opportunities and present challenges of cardiovascular proteomics. *Circulation*. 2010; 121:2326–2328. [PubMed: 20516387]
6. Kullo IJ, Cooper LT. Early identification of cardiovascular risk using genomics and proteomics. *Nat Rev Cardiol*. 2010; 7:309–317. [PubMed: 20440292]
7. Vasan RS. Biomarkers of cardiovascular disease - molecular basis and practical considerations. *Circulation*. 2006; 113:2335–2362. [PubMed: 16702488]
8. Mayr M, Zhang J, Greene AS, Gutterman D, Perloff J, Ping PP. Proteomics-based development of biomarkers in cardiovascular disease - mechanistic, clinical, and therapeutic insights. *Mol Cell Proteomics*. 2006; 5:1853–1864. [PubMed: 16733263]
9. Fields S. Proteomics - proteomics in genomeland. *Science*. 2001; 291:1221–1224. [PubMed: 11233445]
10. McLafferty FW, Breuker K, Jin M, Han XM, Infusini G, Jiang H, Kong XL, Begley TP. Top-down ms, a powerful complement to the high capabilities of proteolysis proteomics. *Febs J*. 2007; 274:6256–6268. [PubMed: 18021240]
11. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol*. 2003; 21:255–261. [PubMed: 12610572]
12. Yuan C, Solaro RJ. Myofilament proteins: From cardiac disorders to proteomic changes. *Proteomics Clin Appl*. 2008; 2:788–799. [PubMed: 21136879]
13. Jin WH, Brown AT, Murphy AM. Cardiac myofilaments: From proteome to pathophysiology. *Proteomics Clin Appl*. 2008; 2:800–810. [PubMed: 21136880]
14. Biesiadecki BJ, Elder BD, Yu ZB, Jin JP. Cardiac troponin T variants produced by aberrant splicing of multiple exons in animals with high instances of dilated cardiomyopathy. *J Biol Chem*. 2002; 277:50275–50285. [PubMed: 12377784]
15. Messer AE, Jacques AM, Marston SB. Troponin phosphorylation and regulatory function in human heart muscle: Dephosphorylation of ser23/24 on troponin I could account for the contractile defect in end-stage heart failure. *J Mol Cell Cardiol*. 2007; 42:247–259. [PubMed: 17081561]
16. Marston SB, de Tombe PP. Troponin phosphorylation and myofilament Ca<sup>2+</sup> -sensitivity in heart failure: Increased or decreased? *J Mol Cell Cardiol*. 2008; 45:603–607. [PubMed: 18691597]
17. Chait BT. Mass spectrometry: Bottom-up or top-down? *Science*. 2006; 314:65–66. [PubMed: 17023639]
18. Han XM, Aslanian A, Yates JR. Mass spectrometry for proteomics. *Curr Opin Chem Biol*. 2008; 12:483–490. [PubMed: 18718552]
19. Bogdanov B, Smith RD. Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrom Rev*. 2005; 24:168–200. [PubMed: 15389855]
20. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: Approaches, advances, and applications. *Annu Rev Biomed Eng*. 2009; 11:49–79. [PubMed: 19400705]
21. Steen H, Mann M. The abc's (and xyz's) of peptide sequencing. *Nat Rev Mol Cell Biol*. 2004; 5:699–711. [PubMed: 15340378]
22. Kelleher NL, Lin HY, Valaskovic GA, Aaserud DJ, Fridriksson EK, McLafferty FW. Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J Am Chem Soc*. 1999; 121:806–812.
23. Ge Y, Lawhorn BG, ElNaggar M, Strauss E, Park JH, Begley TP, McLafferty FW. Top down characterization of larger proteins (45 kda) by electron capture dissociation mass spectrometry. *J Am Chem Soc*. 2002; 124:672–678. [PubMed: 11804498]
24. Zabrouskov V, Han XM, Welker E, Zhai HL, Lin C, van Wijk KJ, Scheraga HA, McLafferty FW. Stepwise deamidation of ribonuclease a at five sites determined by top down mass spectrometry. *Biochemistry*. 2006; 45:987–992. [PubMed: 16411774]
25. Ayaz-Guner S, Zhang J, Li L, Walker JW, Ge Y. In vivo phosphorylation site mapping in mouse cardiac troponin I by high resolution top-down electron capture dissociation mass spectrometry: Ser22/23 are the only sites basally phosphorylated. *Biochemistry*. 2009; 48:8161–8170. [PubMed: 19637843]

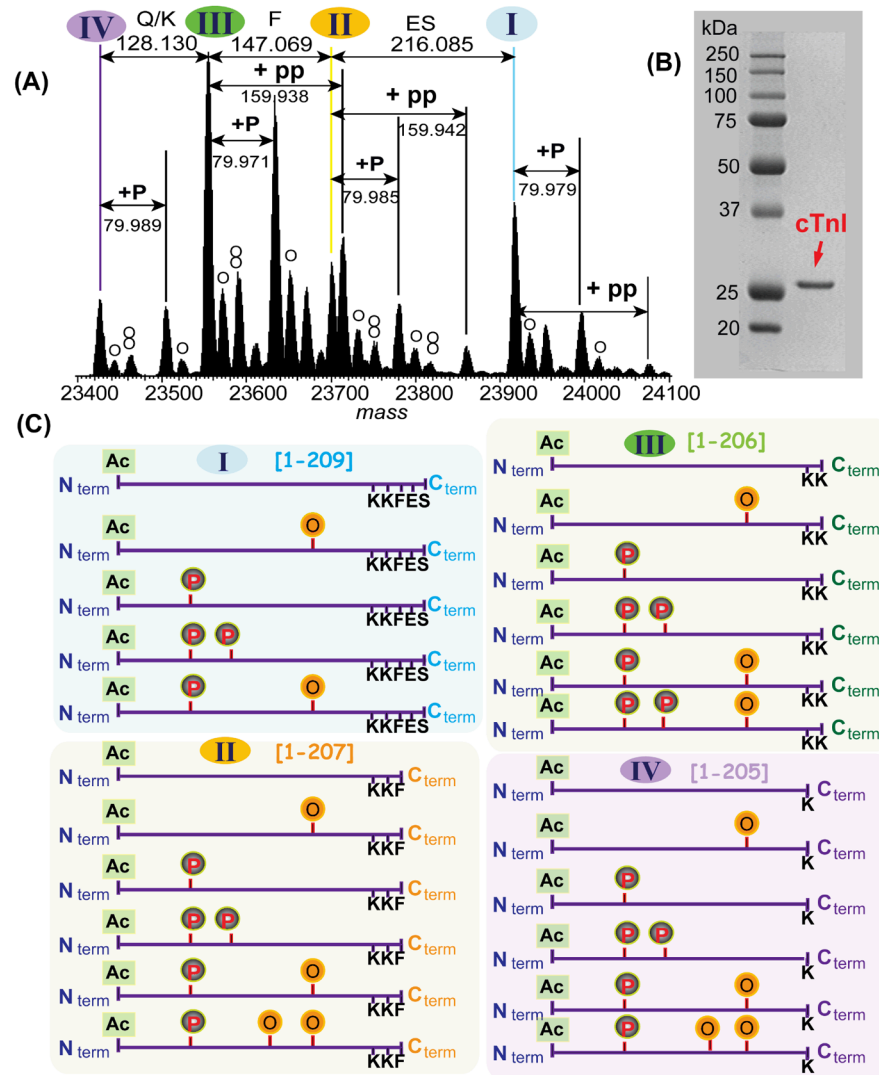
26. Ge Y, Rybakova IN, Xu QG, Moss RL. Top-down high-resolution mass spectrometry of cardiac myosin binding protein c revealed that truncation alters protein phosphorylation state. *Proc Natl Acad Sci USA*. 2009; 106:12658–12663. [PubMed: 19541641]
27. Solis RS, Ge Y, Walker JW. Single amino acid sequence polymorphisms in rat cardiac troponin revealed by top-down tandem mass spectrometry. *J Muscle Res Cell Motil*. 2008; 29:203–212. [PubMed: 19165611]
28. Zabrouskov V, Ge Y, Schwartz J, Walker JW. Unraveling molecular complexity of phosphorylated human cardiac troponin I by top down electron capture dissociation/electron transfer dissociation mass spectrometry. *Mol Cell Proteomics*. 2008; 7:1838–1849. [PubMed: 18445579]
29. Zhang JA, Dong XT, Hacker TA, Ge Y. Deciphering modifications in swine cardiac troponin I by top-down high-resolution tandem mass spectrometry. *J Am Soc Mass Spectrom*. 2010; 21:940–948. [PubMed: 20223681]
30. Xu F, Xu Q, Dong X, Guy M, Guner H, Hacker TA, Ge Y. Top-down high-resolution electron capture dissociation mass spectrometry for comprehensive characterization of post-translational modifications in rhesus monkey cardiac troponin I. *Int J Mass Spectrom*. 2010 Epub ahead of print.
31. Han XM, Jin M, Breuker K, McLafferty FW. Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science*. 2006; 314:109–112. [PubMed: 17023655]
32. Siuti N, Kelleher NL. Decoding protein modifications using top-down mass spectrometry. *Nat Methods*. 2007; 4:817–821. [PubMed: 17901871]
33. Ryan CM, Souda P, Bassilian S, Ujwal R, Zhang J, Abramson J, Ping PP, Durazo A, Bowie JU, Hasan SS, Baniulis D, Cramer WA, Faull KF, Whitelegge JP. Post-translational modifications of integral membrane proteins resolved by top-down fourier transform mass spectrometry with collisionally activated dissociation. *Mol Cell Proteomics*. 2010; 9:791–803. [PubMed: 20093275]
34. Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, McLafferty FW. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem*. 2000; 72:563–573. [PubMed: 10695143]
35. Cooper HJ, Hakansson K, Marshall AG. The role of electron capture dissociation in biomolecular analysis. *Mass Spectrom Rev*. 2005; 24:201–222. [PubMed: 15389856]
36. Kelleher NL. Top-down proteomics. *Anal Chem*. 2004; 76:196A–203A.
37. Kellie JF, Tran JC, Lee JE, Ahlf DR, Thomas HM, Ntai I, Catherman AD, Durbin KR, Zamdborg L, Vellaichamy A, Thomas PM, Kelleher NL. The emerging process of top down mass spectrometry for protein analysis: Biomarkers, protein-therapeutics, and achieving high throughput. *Mol BioSyst*. 2010; 6:1532–1539. [PubMed: 20711533]
38. Garcia BA. What does the future hold for top down mass spectrometry? *J Am Soc Mass Spectrom*. 2010; 21:193–202. [PubMed: 19942451]
39. Armirotti A, Damonte G. Achievements and perspectives of top-down proteomics. *Proteomics*. 2010; 10:3566–3576. [PubMed: 20859958]
40. Chen EI, McClatchy D, Park SK, Yates JR. Comparisons of mass spectrometry compatible surfactants for global analysis of the mammalian brain proteome. *Anal Chem*. 2008; 80:8694–8701. [PubMed: 18937422]
41. Meng FY, Cargile BJ, Patrie SM, Johnson JR, McLoughlin SM, Kelleher NL. Processing complex mixtures of intact proteins for direct analysis by mass spectrometry. *Anal Chem*. 2002; 74:2923–2929. [PubMed: 12141648]
42. Zong CG, Young GW, Wang YJ, Lu HJ, Deng N, Drews O, Ping PP. Two-dimensional electrophoresis-based characterization of post-translational modifications of mammalian 20S proteasome complexes. *Proteomics*. 2008; 8:5025–5037. [PubMed: 19003867]
43. Ping PP, Zhang J, Pierce WM, Bolli R. Functional proteomic analysis of protein kinase C epsilon signaling complexes in the normal heart and during cardioprotection. *Circ Res*. 2001; 88:59–62. [PubMed: 11139474]
44. Labugger R, McDonough JL, Neverova I, Van Eyk JE. Solubilization, two-dimensional separation and detection of the cardiac myofibrillar protein troponin T. *Proteomics*. 2002; 2:673–678. [PubMed: 12112847]

45. Fridriksson EK, Baird B, McLafferty FW. Electrospray mass spectra from protein electroeluted from sodium dodecylsulfate polyacrylamide gel electrophoresis gels. *J Am Soc Mass Spectrom.* 1999; 10:453–455. [PubMed: 10222598]
46. Tran JC, Doucette AA. Gel-eluted liquid fraction entrapment electrophoresis: An electrophoretic method for broad molecular weight range proteome separation. *Anal Chem.* 2008; 80:1568–1573. [PubMed: 18229945]
47. Vellaichamy A, Tran JC, Catherman AD, Lee JE, Kellie JF, Sweet SMM, Zamdborg L, Thomas PM, Ahlf DR, Durbin KR, Valaskovic GA, Kelleher NL. Size-sorting combined with improved nanocapillary liquid chromatography-mass spectrometry for identification of intact proteins up to 80 kda. *Anal Chem.* 2010; 82:1234–1244. [PubMed: 20073486]
48. Neverova I, Van Eyk JE. Role of chromatographic techniques in proteomic analysis. *J Chromatogr B.* 2005; 815:51–63.
49. Qian WJ, Jacobs JM, Liu T, Camp DG, Smith RD. Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications. *Mol Cell Proteomics.* 2006; 5:1727–1744. [PubMed: 16887931]
50. Hage DS. Affinity chromatography: A review of clinical applications. *Clin Chem.* 1999; 45:593–615. [PubMed: 10222345]
51. Gundry RL, Fu Q, Jelinek CA, Van Eyk JE, Cotter RJ. Investigation of an albumin-enriched fraction of human serum and its albuminome. *Proteomics Clin Appl.* 2007; 1:73–88. [PubMed: 20204147]
52. Sharma S, Simpson DC, Tolic N, Jaitly N, Mayampurath AM, Smith RD, Pasa-Tolic L. Proteomic profiling of intact proteins using WAX-RPLC 2-D separations and FTICR mass spectrometry. *J Proteome Res.* 2007; 6:602–610. [PubMed: 17269717]
53. Simpson DC, Ahn S, Pasa-Tolic L, Bogdanov B, Mottaz HM, Vilkov AN, Anderson GA, Lipton MS, Smith RD. Using size exclusion chromatography-RPLC and RPLC-CIEF as two-dimensional separation strategies for protein profiling. *Electrophoresis.* 2006; 27:2722–2733. [PubMed: 16732621]
54. Sheng S, Chen D, Van Eyk JE. Multidimensional liquid chromatography separation of intact proteins by chromatographic focusing and reversed phase of the human serum proteome - optimization and protein database. *Mol Cell Proteomics.* 2006; 5:26–34. [PubMed: 16188874]
55. Marshall AG, Hendrickson CL. High-resolution mass spectrometers. *Annu Rev Anal Chem.* 2008; 1:579–599.
56. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature.* 2003; 422:198–207. [PubMed: 12634793]
57. Scherperel G, Reid GE. Emerging methods in proteomics: Top-down protein characterization by multistage tandem mass spectrometry. *Analyst.* 2007; 132:500–506. [PubMed: 17525804]
58. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA.* 2004; 101:9528–9533. [PubMed: 15210983]
59. Ge Y, ElNaggar M, Sze SK, Bin Oh H, Begley TP, McLafferty FW, Boshoff H, Barry CE. Top down characterization of secreted proteins from mycobacterium tuberculosis by electron capture dissociation mass spectrometry. *J Am Soc Mass Spectrom.* 2003; 14:253–261. [PubMed: 12648932]
60. McAlister GC, Phanstiel D, Good DM, Berggren WT, Coon JJ. Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer. *Anal Chem.* 2007; 79:3525–3534. [PubMed: 17441688]
61. Bunger MK, Cargile BJ, Ngunjiri A, Bundy JL, Stephenson JL. Automated proteomics of E-Coli via top-down electron-transfer dissociation mass spectrometry. *Anal Chem.* 2008; 80:1459–1467. [PubMed: 18229893]
62. 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; 11:320–332. [PubMed: 10757168]

63. Meng FY, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL. Informatics and multiplexing of intact protein identification in bacteria and the archaea. *Nat Biotechnol.* 2001; 19:952–957. [PubMed: 11581661]
64. Tsai YS, Scherl A, Shaw JL, MacKay CL, Shaffer SA, Langridge-Smith PRR, Goodlett DR. Precursor ion independent algorithm for top-down shotgun proteomics. *J Am Soc Mass Spectrom.* 2009; 20:2154–2166. [PubMed: 19773183]
65. Karabacak NM, Li L, Tiwari A, Hayward LJ, Hong PY, Easterling ML, Agar JN. Sensitive and specific identification of wild type and variant proteins from 8 to 669 kda using top-down mass spectrometry. *Mol Cell Proteomics.* 2009; 8:846–856. [PubMed: 19074999]
66. Takeda S, Yamashita A, Maeda K, Maeda Y. Structure of the core domain of human cardiac troponin in the  $Ca^{2+}$ -saturated form. *Nature.* 2003; 424:35–41. [PubMed: 12840750]
67. Solaro RJ, Rosevear P, Kobayashi T. The unique functions of cardiac troponin I in the control of cardiac muscle contraction and relaxation. *Biochem Biophys Res Commun.* 2008; 369:82–87. [PubMed: 18162178]
68. Kuhn P, Xu QG, Cline E, Zhang D, Ge Y, Xu W. Delineating anopheles gambiae coactivator associated arginine methyltransferase 1 automethylation using top-down high resolution tandem mass spectrometry. *Protein Sci.* 2009; 18:1272–1280. [PubMed: 19472346]
69. Solaro RJ, van der Velden J. Why does troponin I have so many phosphorylation sites? Fact and fancy. *J Mol Cell Cardiol.* 2010; 48:810–816. [PubMed: 20188739]
70. Huang L, Wolska BM, Montgomery DE, Burkart EM, Buttrick PM, Solaro RJ. Increased contractility and altered  $Ca^{2+}$  transients of mouse heart myocytes conditionally expressing pkc beta. *Am J Physiol Cell Physiol.* 2001; 280:C1114–C1120. [PubMed: 11287324]
71. Goldspink PH, Montgomery DE, Walker LA, Urboniene D, McKinney RD, Geenen DL, Solaro RJ, Buttrick PM. Protein kinase C epsilon overexpression alters myofilament properties and composition during the progression of heart failure. *Circ Res.* 2004; 95:424–432. [PubMed: 15242976]
72. Scruggs SB, Walker LA, Lyu T, Geenen DL, Solaro RJ, Buttrick PM, Goldspink PH. Partial replacement of cardiac troponin I with a non-phosphorylatable mutant at serines 43/45 attenuates the contractile dysfunction associated with pkc epsilon phosphorylation. *J Mol Cell Cardiol.* 2006; 40:465–473. [PubMed: 16445938]
73. Whitelegge JP, Zabrouskov V, Halgand F, Souda P, Bassiliana S, Yan W, Wolinsky L, Loo JA, Wong DTW, Faull KF. Protein-sequence polymorphisms and post-translational modifications in proteins from human saliva using top-down fourier-transform ion cyclotron resonance mass spectrometry. *Int J Mass Spectrom.* 2007; 268:190–197. [PubMed: 19050733]
74. Sze SK, Ge Y, Oh H, McLafferty FW. Top-down mass spectrometry of a 29-kda protein for characterization of any posttranslational modification to within one residue. *Proc Natl Acad Sci USA.* 2002; 99:1774–1779. [PubMed: 11842225]
75. Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol.* 2005; 1:252–262. [PubMed: 16408053]
76. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: A critical review. *Anal Bioanal Chem.* 2007; 389:1017–1031. [PubMed: 17668192]
77. 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; 78:4271–4280. [PubMed: 16808433]
78. Gordon EF, Muddiman DC. Quantification of singly charged biomolecules by electrospray ionization fourier transform ion cyclotron resonance mass spectrometry utilizing an internal standard. *Rapid Commun Mass Spectrom.* 1999; 13:164–171.
79. Jiang L, Smith JN, Anderson SL, Ma P, Mizzen CA, Kelleher NL. Global assessment of combinatorial post-translational modification of core histones in yeast using contemporary mass spectrometry. *J Biol Chem.* 2007; 282:27923–27934. [PubMed: 17652096]
80. Pesavento JJ, Bullock CR, Leduc RD, Mizzen CA, Kelleher NL. Combinatorial modification of human histone h4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. *J Biol Chem.* 2008; 283:14927–14937. [PubMed: 18381279]

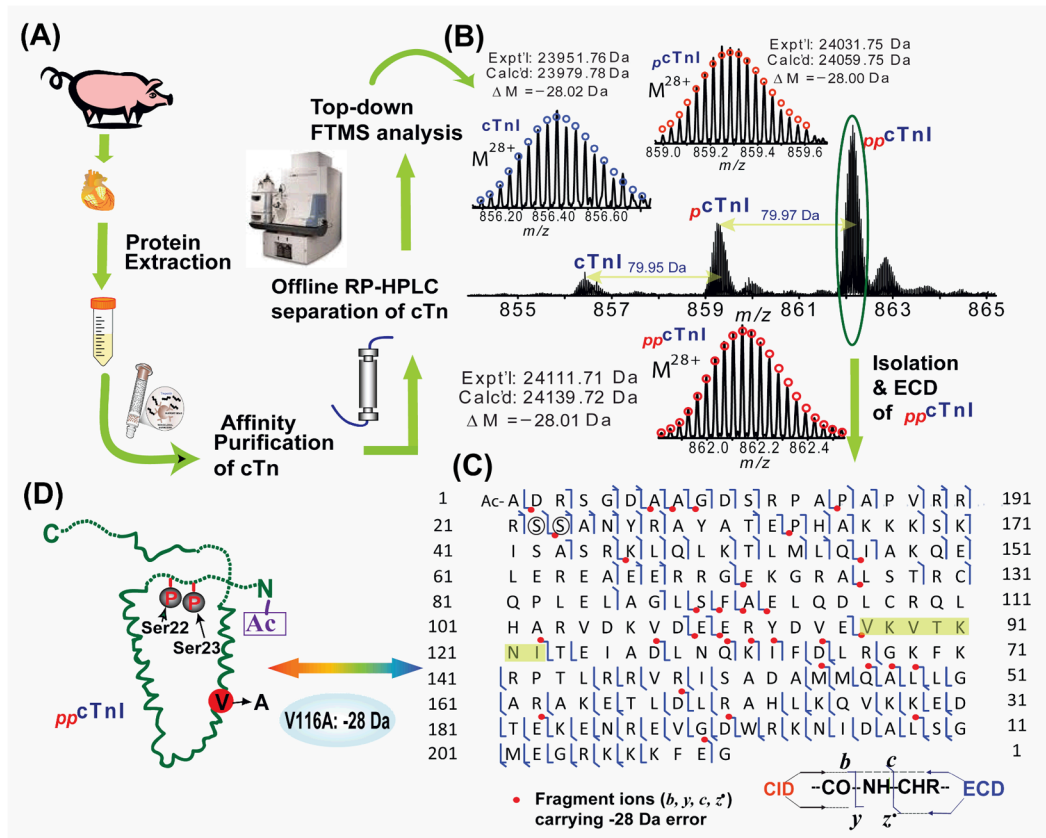
81. Waanders LF, Hanke S, Mann M. Top-down quantitation and characterization of silac- labeled proteins. *J Am Soc Mass Spectrom.* 2007; 18:2058–2064. [PubMed: 17920290]
82. Collier TS, Sarkar P, Rao B, Muddiman DC. Quantitative top-down proteomics of silac labeled human embryonic stem cells. *J Am Soc Mass Spectrom.* 2010; 21:879–889. [PubMed: 20199872]
83. Flashman E, Redwood C, Moolman-Smook J, Watkins H. Cardiac myosin binding protein C - its role in physiology and disease. *Circ Res.* 2004; 94:1279–1289. [PubMed: 15166115]
84. Winegrad S. Cardiac myosin binding protein C. *Circ Res.* 1999; 84:1117–1126. [PubMed: 10347086]
85. Moss RL, Razumova M, Fitzsimons DP. Myosin crossbridge activation of cardiac thin filaments - implications for myocardial function in health and disease. *Circ Res.* 2004; 94:1290–1300. [PubMed: 15166116]
86. de Tombe PP. Myosin binding protein C in the heart. *Circ Res.* 2006; 98:1234–1236. [PubMed: 16728667]
87. Gautel M, Zuffardi O, Freiburg A, Labeit S. Phosphorylation switches specific for the cardiac isoform of myosin binding protein-c - a modulator of cardiac contraction. *EMBO J.* 1995; 14:1952–1960. [PubMed: 7744002]
88. Reid GE, Shang H, Hogan JM, Lee GU, McLuckey SA. Gas-phase concentration, purification, and identification of whole proteins from complex mixtures. *J Am Chem Soc.* 2002; 124:7353–7362. [PubMed: 12071744]
89. Roth MJ, Parks BA, Ferguson JT, Boyne MT, Kelleher NL. “Proteotyping”: Population proteomics of human leukocytes using top down mass spectrometry. *Anal Chem.* 2008; 80:2857–2866. [PubMed: 18351787]
90. Wu S, Lourette NM, Tolic N, Zhao R, Robinson EW, Tolmachev AV, Smith RD, Pasa-Tolic L. An integrated top-down and bottom-up strategy for broadly characterizing protein isoforms and modifications. *J Proteome Res.* 2009; 8:1347–1357. [PubMed: 19206473]
91. Fu Q, Schoenhoff FS, Savage WJ, Zhang PB, Van Eyk JE. Multiplex assays for biomarker research and clinical application: Translational science coming of age. *Proteomics Clin Appl.* 2010; 4:271–284. [PubMed: 21137048]





**Fig. 2. Top-down MS for unraveling molecular complexity of commercially available human cTnI**

(A) Deconvoluted high resolution ESI/MS spectrum ( $m/z$  is converted to mass) of multiply charged intact cTnI molecular ions revealed a total of thirty-six modified forms resulting from acetylation, phosphorylation and C-terminal proteolytic truncations. Roman numerals indicate the full-length (I) and three different C-terminal truncated forms of cTnI (II, III, IV), respectively. +80 Da and +160 Da correspond to mono- and bis-phosphorylation (+P, +PP), respectively. Oxidized species (+16 Da) are indicated by “O”. (B). 1-D SDS PAGE analysis of human cTnI stained with Coomassie Blue. (C) Schematic representation of the major modified forms of human cTnI detected in (A) with a combinatorial modifications of one acetylation (Ac-), one or two phosphorylation (P) sites, one or two oxidation (O) sites, and three possible C-terminal proteolytic truncations. N<sub>term</sub> and C<sub>term</sub> stand for N- and C-termini of cTnI, respectively. Modified based on Ref.<sup>28</sup> with permission.



**Fig. 3. Top-down MS for discovery of unexpected modifications**

(A) The workflow of extraction and purification of cTn from domestic swine hearts for MS analysis. (B) FTMS spectrum of swine cTnI for the charge state 28<sup>+</sup> precursor ions, showing its distribution in un-, mono- and bis-phosphorylated forms. Inset: isotopically resolved molecular ions of un-, mono-, and bis-phosphorylated cTnI (M<sup>28+</sup>) with high accuracy molecular weight measurements. p cTnI and pp cTnI represent mono-(+79.95 Da) and bis-phosphorylated (+159.92 Da) cTnI, respectively. Circles represent the theoretical isotopic abundance distribution of the isotopomer peaks corresponding to the assigned mass. Calc'd, calculated most abundant mass; Expt'l, experimental most abundant mass. (C) MS/MS fragmentation and product ion map from ECD and CID spectra for bis-phosphorylated swine cTnI (pp cTnI). Bis-phosphorylated cTnI was isolated and fragmented by ECD. Fragments assignments were made to the swine cTnI (UnitProtKB/Swiss-Prot A5X5T5, TNNI3\_pig) with the removal of N-terminal Met and acetylation at the new terminus. Bis-phosphorylation sites of Ser22/23 were highlighted by circles. The fragmentation ions carrying the mass discrepancy (-28 Da) was 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 bis-phosphorylated swine cTnI. Modified based on Ref.<sup>29</sup> with permission.

