Published in final edited form as:

Anal Chem. 2019 March 19; 91(6): 3835–3844. doi:10.1021/acs.analchem.8b04082.

A Top-Down Proteomics Platform Coupling Serial Size Exclusion Chromatography and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

Trisha Tucholski^{1,2}, Samantha Knott^{1,2}, Bifan Chen^{1,2}, Paige Pistono³, Ziqing Lin⁴, and Ying Ge^{1,2,4,*}

¹Department of Chemistry, University of Wisconsin-Madison, Madison, WI, 53706

²Human Proteomics Program, University of Wisconsin, Madison, WI, 53705

³Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, 53706

⁴Department of Cell and Regenerative Biology, University of Wisconsin-Madison, Madison, WI, 53705

Abstract

Mass spectrometry (MS)-based top-down proteomics provides rich information about proteoforms arising from combinatorial amino acid sequence variations and post-translational modifications (PTMs). Fourier transform ion cyclotron resonance (FT-ICR) MS affords ultra-high resolving power and provides high-accuracy mass measurements, presenting a powerful tool for top-down MS characterization of proteoforms. However, detection and characterization of large proteins from complex mixtures remain challenging due to the exponential decrease in S:N with increasing molecular weight (MW) and co-eluting low-MW proteins; thus, size-based fractionation of complex protein mixtures prior to MS analysis is necessary. Here, we directly combine MS-compatible serial size exclusion chromatography (sSEC) fractionation with 12 T FT-ICR MS for targeted top-down characterization of proteins from complex mixtures extracted from the human and swine heart proteome. Benefiting from the ultra-high resolving power of FT-ICR, we isotopically resolved 31 distinct proteoforms (30–50 kDa) simultaneously in a single mass spectrum within a 100 m/z window. Notably, within a 5 m/z window, we obtained baseline isotopic resolution for 6 distinct large proteoforms (30–50 kDa). The ultra-high resolving power of FT-ICR MS combined with sSEC fractionation enabled targeted top-down analysis of large proteoforms (>30 kDa) from the human heart proteome without extensive chromatographic separation or protein purification. Further separation of proteoforms inside of the mass spectrometer (in-MS) allowed for isolation of individual proteoforms for targeted electron capture dissociation (ECD) for high sequence coverage. sSEC/FT-ICR ECD facilitated identification and sequence characterization of important metabolic enzymes. This platform, which facilitates deep

*Correspondence: Prof. Ying Ge, Department of Cell and Regenerative Biology, Department of Chemistry, University of Wisconsin-Madison, 1111 Highland Ave., WIMR II 8551, Madison, WI 53705, Tel: 608-265-4744, Fax: 608-265-8745, ying.ge@wisc.edu.

The **Supporting Information** is available free of charge on the ACS Publications website. Supplementary Experimental Procedures and Results and Discussion, Intact Proteoform Mass List for Fractions 7 and 8 (Table S3), Human and Swine Protein Sequences, and Topic Results for 82 kDa Protein Identification.

interrogation of proteoform primary structure, is highly tunable, allows for adjustment of MS and MS/MS parameters in real-time, and can be utilized for a variety of complex protein mixtures.

Graphical Abstract

Introduction

The ability to accurately measure intact protein mass prior to amino acid sequence characterization is a coveted facet of mass spectrometry (MS)-based top-down proteomics.¹⁻³ Accurate intact mass measurements combined with the high-sequence coverage accessed by top-down proteomics provides rich information about the proteoforms which arise from a combination of genetic mutations, RNA-editing, alternative splicing events, and post-translational modifications (PTMs) of a single gene product.⁴⁻⁶ Thus, the top-down approach has become an indispensable tool for comprehensive characterization of proteoform primary structure to investigate important biological questions.⁷⁻¹⁴

With unmatched resolving power and high mass accuracy, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) is well-suited to accurately measure the mass of large biomacromolecules and resolve complex MS/MS spectra, making it a powerful technique for top-down protein characterization.^{10, 15-20} FT-ICR mass spectrometers are compatible with a wide variety of tandem MS (MS/MS) strategies, both energy- and electron-based, which can be used individually or in combination to obtain high sequence coverage.^{16, 19, 21-25} The high resolving power afforded by FT-ICR MS also allows for isotopic resolution of many protein components from complex mixtures without extensive chromatographic separation.^{26, 27} However, detection and characterization of large proteins remain challenging in top-down proteomics because of the exponential decay in S:N with increasing molecular weight (MW), making it difficult to detect larger proteins (30–200 kDa), especially in the presence of low-MW proteins (5–20 kDa).²⁸ Hence, size-based fractionation, such as gel elution liquid fraction entrapment electrophoresis (GELFrEE), is commonly used at the front-end of top-down proteomics workflows to improve the analysis of high-MW proteins from complex mixtures.²⁹⁻³² However, gel-based methods such as GELFrEE require the use of MS-incompatible detergents (e.g. SDS), necessitating additional detergent-removal steps prior to MS analysis. In contrast, size exclusion chromatography (SEC) is an attractive alternative for size-based separation because of its compatibility with MS-friendly solvents.^{33, 34} Recently, our group introduced serial (s)SEC, a high-resolution size-based separation strategy which combines SEC columns with different pore sizes in series to enhance fractionation power for complex mixtures across a wide MW range. sSEC employs an MS-compatible eluent (1% formic acid in water), making it highly compatible with top-down proteomics workflows.³⁵

Herein, we sought to combine MS-compatible sSEC fractionation directly with the ultra-high resolving power of FT-ICR MS for the top-down characterization of proteins without extensive purification or liquid chromatography. We fractionated cytosolic proteomes, extracted from human and swine heart tissue, by MS-compatible sSEC. Benefiting from the use of a highly-MS compatible sSEC eluent (1% formic acid (FA) in water), this platform requires little sample preparation between fractionation and top-down MS analysis. Direct infusion FT-ICR MS analysis of the sSEC fractions allowed for ultra-high resolution separation of ions inside of the mass spectrometer (in-MS). This permitted intact mass measurement of multiple species from a single spectrum and sequence characterization by MS/MS analysis (Scheme 1). From a single spectrum, we were able to detect 31 distinct proteoforms (30–50 kDa) within a 100 m/z window. Notably, in a 5 m/z window, we achieved baseline isotopic resolution for 6 distinct species. From the sSEC fractions, we identified large metabolic enzymes with good mass accuracy using FT-ICR MS. With this platform, we also isolated single proteoforms in-MS and performed electron capture dissociation (ECD) to obtain high sequence coverage for metabolic enzymes (>30 kDa). The versatility afforded by this sSEC/FT-ICR MS platform offers a valuable tool for top-down proteomics and for interrogation of proteoform primary structure.

Experimental Procedures

Materials

All chemicals and reagents were purchased from Sigma-Aldrich Inc. (St. Louis, MO, USA) unless otherwise noted. HPLC grade water and acetonitrile (ACN) were purchased from Fisher Scientific (Fair Lawn, NJ, USA). Amicon ultracentrifugal 10 kDa molecular weight cutoff (MWCO) filters were purchased from Millipore Sigma (Burlington, MA, USA). TGX Stain Free 8–16% gradient gels were purchased from Bio-Rad (Hercules, CA, USA).

Extraction of the Soluble Heart Proteome

Donor human heart tissue was obtained from the University of Wisconsin Hospital and Clinics according to the protocol approved by the Institutional Review Board at University of Wisconsin-Madison as reported previously.³⁶ A detailed protein extraction procedure and list of reagents is provided in the Supporting Information. Briefly, heart tissue was homogenized in 4 volumes (mL/g tissue) of HEPES extraction buffer (pH=7.4) using an electric homogenizer to extract the cytosolic portion of the heart proteome. The resulting homogenate was centrifuged at 16,100 *g* for 15 min at 4 °C. The supernatant was collected and the pellet was resuspended in HEPES extraction buffer and centrifuged at 16,100 *g* for 15 min at 4 °C. The supernatants from both wash steps were combined, resulting in a protein extract with a total protein concentration of approximately 5.6 mg/mL.

Serial Size Exclusion Fractionation

Heart protein extract was exchanged from HEPES extraction buffer to 1% FA in water (pH = 2) using ultracentrifugal 10 kDa MWCO filters prior to sSEC fractionation. A detailed procedure regarding buffer exchange is provided in the Supporting Information. sSEC fractionation was performed using an ACQUITY H-class UPLC system equipped with a UV detector and an automated fraction collector (Waters, Milford, MA, USA). A series of

PolyHYDROXYETHYL A (PolyHEA) columns (9.4×200 mm, $3 \mu\text{m}$) from PolyLC Inc. (Columbia, MD, USA) were assembled in order of decreasing pore sizes ($1000 \text{ \AA}$ - $500 \text{ \AA}$ - $300 \text{ \AA}$), as previously described.³⁵ Proteins were eluted isocratically with 1% FA in water at a flow rate of 0.5 mL/min . Protein elution was monitored with UV detection at $\lambda=280 \text{ nm}$. $150 \mu\text{g}$ of total protein ($25 \mu\text{L}$ injection, 6 mg/mL) was loaded per sSEC run. Automatic fraction collection was performed based on the protein elution window. Twelve fractions were collected over 12 min ($24\text{--}36 \text{ min}$). Fractions from 18 replicate injections were pooled and concentrated to a final volume of $100 \mu\text{L}$ using ultracentrifugal 10 kDa MWCO filters. Prior to FT-ICR MS, SDS-PAGE analysis was performed to evaluate the protein content and MW range for each fraction. For sSEC fractionation of soluble proteome extracted from swine heart tissue, the same procedure was followed, except $1000 \text{ \AA}$ - $500 \text{ \AA}$ - $500 \text{ \AA}$ series was used. Further details regarding this fractionation are available in the Supporting Information.

FT-ICR MS Analysis

Fractions collected and concentrated following sSEC fractionation were diluted with 0.1% FA in ACN (1:1 for sSEC fractions, 1:3 for unfractionated protein mixture) to aid nano-electrospray ionization (nanoESI). Protein fractions were directly infused to the Bruker 12 Tesla Solarix FT-ICR mass spectrometer using an Advion TriVersa NanoMate with gas pressure and spray voltage set between $0.3\text{--}0.5 \text{ psi}$ and $1.3\text{--}1.5 \text{ kV}$, respectively. In source energy of 70 V was applied at skimmer 1 to aid the desolvation of protein ions. For MS/MS experiments, collisionally activated dissociation (CAD) or ECD are used. A detailed list of MS and MS/MS parameters corresponding to data shown throughout manuscript can be found in the Supporting Information. FT-ICR mass spectra were calibrated externally using sodium trifluoroacetic acid (1 mg/mL) in 50:50 water:ACN.

Data Analysis

MS and MS/MS data were analyzed using Bruker DataAnalysis (version 4.3.110) and in-house developed MASH Suite Pro. First, data were externally calibrated using DataAnalysis Calibrate function. Mass lists were then extracted from raw Bruker .baf files in MASH Suite Pro using the Extract function (with THRASH deconvolution algorithm) in MASH Suite Pro with a quality fit factor of 60 and S:N of 3.³⁷ TopPIC (version 1.1.1) search algorithm run against the human or swine Uniprot database was used to identify proteins from extracted mass lists.³⁸ For all searches, a mass error tolerance of 15 ppm was used with 0.01 as an E-value cut-off. Matches were manually validated using MASH Suite Pro.

Results and Discussion

sSEC Fractionation of the Heart Proteome

We used a $1000\text{--}500\text{--}300 \text{ \AA}$ pore size series for fractionation of proteins extracted from the cytosolic proteome of human heart tissue ($10\text{--}250 \text{ kDa}$). The combination of different pore sizes in series provided an extension of MW fractionation range and allows for high-resolution fractionation of complex mixtures.³⁵ The protein elution time window and signal intensity for the sSEC runs were highly reproducible, evident from the overlay of replicate UV chromatograms (Figure S3). The UV chromatogram was used to determine the elution

window for the proteins (12 min, Figure S4A) and automatic fraction collection was performed in 1-min intervals (12 total fractions) from the time proteins began eluting from the column series. We found that 1-min intervals provided efficient MW-fractionation of our specific mixture of cytosolic proteins extracted from heart tissue, based on SDS-PAGE visualization (Figure S4B).

Visualization of the sSEC fractions with SDS-PAGE analysis showed a good distribution of protein size across the fractions (Figure S4B). Fractions 1–5 primarily contained larger proteins (75–200 kDa), which is conceivable since the largest components of a mixture have the least access to the column pore volume and elute first in SEC.³⁹ Fraction 6 contained proteins mainly in the MW range of 50–75 kDa whereas fractions 7 and 8 contained primarily proteins between 25 – 50 kDa as indicated in the SDS-gel (Figure S4B). The absence of proteins less than 25 kDa in these sSEC fractions significantly improved MS detection of proteins within this size range (*vide infra*). Gel visualization showed that fraction 8 and 9 had similar composition between the ranges of 25 – 50 kDa, except that fraction 9 contained proteins less than 25 kDa. This difference in composition between fraction 8 and 9 may seem minor, however, we later showed that the presence of low-MW components significantly impacted FT-ICR MS analysis of the intermediate-MW protein components in fraction 9. Fraction 10 contained primarily lower MW proteins (<25 kDa) with some remnants of higher-abundance proteins over 25 kDa. This was likely a result of the higher-abundance proteins in the mixture eluting from the column across a wider window than the lower-abundance components. The later eluting fractions (11 and 12) contained primarily proteins less than 25 kDa, which are easily analyzed with top-down MS. Subsequently, fractions 11/12, 8/9, and 7/8 were analyzed by FT-ICR MS which demonstrated that minor differences in protein composition (i.e. the presence of low-MW proteins) had a significant impact on the mass spectrum acquired (*vide infra*).

FT-ICR MS Analysis of sSEC Fractions

We compared the mass spectra for later-eluted fractions 9–12, with that of the unfractionated protein mixture (LM, Figure 1A). Since the sSEC fractions were eluted in highly MS-compatible solvent, 1 % FA in water, the fractions were ready for direct FT-ICR MS analysis. Following concentration with MWCO filters, sSEC fractions were diluted 1:1 with 0.1% FA in ACN to aid in nanoESI efficiency and ionization of the proteins. Overall mass spectra for the range of 850–1150 m/z were obtained in segments by isolating 100 m/z windows inside of the quadrupole region of the instrument (prior to the ICR detection cell), with window centers 900, 1000, and 1100 m/z . The segmented spectra were combined (Figure 1) to represent the 850–1150 m/z range. For fractions 9–12, 100 transients were acquired for each segment spectra (transient length 1.13 s).

The unfractionated protein mixture contained protein species in the MW range of 10 – 250 kDa (Figure 1). The most abundant protein species in each mass spectrum are indicated with a shape corresponding to the protein MW (Figure 1A, **lower left-hand corner**). Comparison of the mass spectrum with SDS-PAGE visualization shows that large protein signal was significantly suppressed by the presence of the low-MW proteins in the mixture. From sSEC fraction 10, the proteins which dominated the MS signal in the LM were targeted for

MS/MS and identified as myoglobin (yellow, MB, 17042 Da); fatty acid binding protein, heart and muscle isoform (green, FABP3, 14760 Da); and α -crystallin B, CRA isoform (magenta, CRYAB, 20188 Da) (Figure S5). MB was expected to be a dominant species in the heart cytosolic proteome since it is a cytoplasmic hemoprotein which is highly expressed in cardiomyocytes.⁴⁰ FABP3 is also highly abundant in the cytoplasm of cardiomyocytes,⁴¹ playing an important role in fatty-acid metabolism and has been proposed as an early biomarker of myocardial injury.⁴¹⁻⁴³ CRYAB is an important chaperone protein which is responsible for binding misfolded proteins in the cytoplasm, and also appeared to be highly abundant in the cytosolic protein mixture.⁴⁴ The other larger metabolic enzymes which are present in the LM (based on gel visualization) were undetected in the LM mass spectrum due to signal suppression by the lower MW protein species.

The above comparison of mass spectra for LM and the sSEC fractions 9–12 (Figure 1B) showed that the most abundant low-MW proteins were detected in the unfractionated sample, but after size-based fractionation, signal of other proteins present in the soluble proteome was enriched. Fractions 9–11 shared signal from the same low-MW, high-abundance species (MB, FABP3, and CRYAB). MB (yellow, 17042 Da) was spread across fractions 9–11, which corresponds well with the gel visualization of these fractions. As shown by the MS data, the high abundance of these low-MW species affected the detection of other proteins present in lower abundance in the fractions. Notably, in fraction 12, signal for many other smaller proteins in the sample was enriched by depletion of the high-abundance MB from this fraction. Zoom-in mass spectra of 961–977 m/z for each fraction mass spectrum shows proteins other than the most abundant species are detected by FT-ICR MS (Figure 1B). This is especially evident by the many small proteins that are enriched in fraction 12 compared to the LM and fractions 9–11. More specifically, the 21+ charge state of CRYAB (magenta, 20188 Da) interfered with signal from the 38+ charge state of an intermediate-sized protein (red, 36526 Da) (Figure 1B, inset). As CRYAB elutes in fraction 9, but minimally in fraction 10, we were able to detect the signal of the larger protein, which was carried over from the earlier-eluting fractions. This result demonstrates that sSEC fractionation not only helps reduce co-elution of large and small proteins, but also helps to enrich signal from low-abundance proteins in the fraction or mixture by depletion of high-abundance proteins.

The comparison of MS spectra of fractions 8 and 9 reinforced the purpose of size-based fractionation prior to MS analysis for improved detection of high-MW proteins and demonstrated that the presence of low-MW proteins interfered with the detection of high-MW species. SDS-PAGE gel visualization showed (Figure 2A) that the composition and complexity of the two fractions are very similar above MW of 25 kDa. However, sSEC 9 contains proteins less than 25 kDa which are present in high-abundance (in this case, MB and CRYAB). This small difference in fraction contents significantly affected the overall mass spectra of the two fractions (Figure 2). The mass spectrum for fraction 9 was dominated by the small proteins present in the fraction, specifically MB (yellow, 17042 Da). This is an interesting case to demonstrate the suppression of signal for high-MW proteins by low-MW proteins, reinforcing the importance of size-based fractionation in the analysis of large proteins from complex mixtures.

With lower MW proteins (<25 kDa) almost completely depleted from the sSEC fractions 7 and 8 (Figure S2), the suppression of high-MW protein signal due to these species was negligible. With this, the signal of proteins larger than 25 kDa was enriched, specifically those within the range 30–50 kDa. Comparison of the gel lanes of fractions 7 and 8 with the mass spectra for these protein mixtures shows that the two fractions are qualitatively similar in composition and complexity (Figure 3A). From fraction 8, we detected and resolved 31 distinct proteoforms from 30–50 kDa within a 100 m/z window (Table S3). A zoom-in of the fraction 8 mass spectrum between 997–1001 m/z shows 6 distinct proteoforms isotopically resolved in a 5 m/z window (Figure 3B). Creatine kinase, M-type (CKM; pink, 42942 Da) was quite abundant in fraction 7, whereas other metabolic enzymes, including glyceraldehyde 3-phosphate dehydrogenase (GAPDH; dark red, 35899 Da) and malate dehydrogenase (MDH; purple, 32980 Da) were enriched in fraction 8. Narrowing the quadrupole isolation window enriched the signal for species that are lower abundance by excluding more abundant protein ions (*vide supra*). With the high resolution afforded by the FT-ICR MS, we were able to isotopically resolve these species and obtain accurate intact mass measurements. With measurements of intact mass, protein identifications were possible for many proteins involved in cardiomyocytes metabolism, specifically glycolysis (Figure 3C). The ability to obtain tentative protein identifications using their known accurate intact mass guided downstream analysis for targeted MS/MS. For instance, we targeted the proteoform tentatively identified as MDH for MS/MS and confirmed its identity with CAD (Figure S6).

When matched to the reported Uniprot sequence, the muscle-type and cytosolic creatine kinases (CKM and CKS, respectively), both measurements differed by 1 Da. We postulated that this mass mis-match could be caused by an amidation event, such as the switch from aspartic acid to asparagine or glutamic acid to glutamine.⁴⁵ We evaluated appropriate matches for the theoretical monoisotopic mass (based on reported Uniprot protein sequence) to the experimentally measured monoisotopic mass by mass error and quality fit of the theoretical isotopomer envelope to the experimental isotopomer envelope (Figure S7). When we made a change of one amino acid in the theoretical sequence (D → N), we were able to match the experimentally measured mass and theoretical mass of CKM and CKS more confidently, based on the aforementioned metrics (Figure S7). For the experimentally measured masses of both CKM and CKS, the D → N switch showed the lowest mass error and the highest quality fit to the theoretical isotopomer envelope. Nevertheless, in some instances, due to the difficulty in determining monoisotopic mass for large proteins because of increased relative abundance of heavier isotopomers with increasing MW, it is possible to find a 1-Da mass discrepancy in the assignment of monoisotopic mass of large proteins.⁴⁶

Moving up in MW, detection of high-MW proteins from complex mixtures becomes challenging. For one, longer transient lengths are required to resolve isotopomer peaks of high-MW proteins due to increased charge. In addition, signal spreading across an increased number of charge states decreases spectral S:N (*vide supra*). Often times, increased number of summed transients is also required to accumulate a higher MS signal. To test whether we could perform top-down proteomics on a large protein from our sSEC fractions without extensive liquid chromatography or purification, we analyzed fraction 4 (Figure S8) from an sSEC fractionation for a mixture of cytosolic proteins extract from *sus scrofa* (swine) heart

tissue. We successfully isolated and isotopically resolved the 75+ charge state of an 82 kDa proteoform (Figure S8). To identify this protein, we performed CAD on the isolated species and used the Toppic search algorithm to identify the protein.³⁸ Though sequence coverage was relatively low, we were able to identify the protein as a proteoform of the mitochondrial form of aconitate hydratase (ACOT2, E-value = 1.18×10^{-15}), which is an important enzyme involved in energy metabolism. To obtain higher sequence coverage for this high-MW protein, electron-based methods (ECD/ ETD) could be used.

Sensitivity of this sSEC/FT-ICR analysis platform could be further improved by decreasing sample complexity and dynamic range during the protein extraction, prior to size-based fractionation. The protein mixtures used in the above experiments were highly complex tissue extracts containing cytosolic proteins with a high dynamic range. Dynamic range of detection for any mixture will depend on the highest- and lowest-abundance species detectable in the complex mixture. Here we sought to improve detection of low-abundance proteoforms by narrowing the m/z window of isolation in the quadrupole prior to detection in the ICR cell. By excluding high-abundance ions, we reduce complexity of the ion populations in the ICR cell. As a result, this minimizes collisional damping and space charge effects, and thereby improves detection of low-abundance ions.^{15, 47, 48}

Sequence Characterization of Metabolic Enzymes

FT-ICR MS is well suited for protein sequence characterization, especially in offline mode. For one, ultra-high resolving power allows for accurate intact mass measurement of closely related species and can therefore resolve highly complex protein MS/MS spectra with overlapping fragment ions. Second, FT-ICR mass spectrometers are equipped with a variety of MS/MS options, allowing for a combination of dissociation techniques to be used concurrently, in tandem, or by combining mass spectra post-acquisition.^{24, 25, 49, 50} Direct infusion analysis of sSEC fractions with the FT-ICR MS allowed us to tune MS/MS parameters in real-time to optimize the acquisition for individual proteins of interest. Here, we have demonstrated offline ECD MS/MS characterization of larger metabolic enzymes extracted from swine heart tissue.

We analyzed fraction 8 collected from sSEC fractionation of pig heart tissue extract and isolated intermediate size (32–47 kDa) proteins for top-down analysis (Figure 4A). For MDH and CKM, we selected the highest abundance ions with the highest charge (34+ and 44+, respectively), since ECD fragmentation is more efficient for higher-charged protein ions.⁵¹⁻⁵³ For the mitochondrial malate dehydrogenase (MDH), we found that using a 1.0 V electron energy with 20 ms pulse was adequate to achieve efficient fragmentation of the N- and C-termini, however it has been previously demonstrated that the combination of mass spectra from ECD experiments using a both high and low ECD electron energy improves achievable protein sequence coverage.²⁵ With 500 transients summed, we observed 55 c and 47 z• ions (Figure 4B) for MDH. For creatine kinase, muscle-type (CKM), a larger protein, we observed 44 c and 41 z• ions by using 0.6 V electron energy, 50 ms pulse, and 1000 summed transients (Figure 4B). In real-time, we could assess the effect of different ECD electron energy and pulse time by observing the formation of reduced-charge precursor ions.

This allowed evaluation of whether the MS/MS acquisition would provide efficient fragmentation and sequence coverage prior to acquisition of mass spectra.

Using our offline top-down proteomics platform, ECD parameters used for each protein could be tuned in real-time to select the most appropriate parameters for fragmentation and acquire a larger number of scans, allowing for deep sequence coverage. The direct infusion of the sSEC fractions allows for the selection of appropriate ions for in-MS isolation and fragmentation, such as those that are high-abundance and highly charged. Electron-based fragmentation methods (e.g. ETD and ECD) cleave non-specifically along the protein backbone and preserve labile PTMs, such as phosphorylation, and thus are a preferred option for deep sequencing and mapping PTMs.^{51, 52, 54, 55} However, these dissociation methods are best utilized offline, since online methods often provide limited sequence coverage due to the limited elution time of proteins on an LC-timescale, and can preclude localization of PTMs which are located deep within a protein backbone.²⁵ We have demonstrated greatly improved sequence coverage for the swine proteins hemoglobin subunit β and subunit α (targeted from sSEC fractions) using offline CAD and ECD MS/MS, compared to relatively low sequence coverage achieved using an online MS/MS approach (Figure S9 and Figure S10). We have also previously shown high protein sequence coverage is achievable using offline FT-ICR MS/MS.^{22, 23, 25, 56} Similarly, sequence coverage generated by ultraviolet photodissociation (UVPD) has shown great improvements for offline MS/MS over online platforms.⁵⁷⁻⁵⁹

The MS-compatible nature of sSEC makes it an ideal method for separation of large proteins from small proteins, which interfere with their analysis by MS. Because high-MW species are eluted first and low-MW species last, contamination of smaller proteins in early-eluted fractions which contain large proteins is inherently limited. As seen in Figure S4, there is some degree of carry-over for higher-abundance large and intermediate-sized proteins into later fractions containing small proteins. However, the suppression of small protein signal due by large protein contamination is minimal (*vide infra*). The use of different pore sizes can be optimized based on the MW range of the protein mixture to be separated. Combination of large and small pore sizes will improve the separation of high- and low-MW proteoforms, respectively. In this study, we use both a 1000 – 500 – 300 Å and 1000 – 500 – 500 Å pore series to separate the soluble heart proteome from human and swine heart tissue, respectively. The observed difference provided by using 300 Å vs. 500 Å pore size for fractionation of our mixture is minimal, though 300 Å provides slightly improved separation of low-MW proteoforms.

Here, we have shown that it is possible to achieve reasonably good level of MW-fractionation for a highly complex protein mixture with a wide dynamic range (Figure S4). We demonstrated sSEC fractionation for extremely complex mixtures of proteins extracted from the cytosolic portion of the human and swine heart proteomes (which contain high levels of serum albumin and myoglobin). Optimal sSEC protein loading amount, time-based fraction collection, and number of replicates pooled are sample-dependent parameters and should be tuned appropriately depending on level of mixture complexity and dynamic range. For example, for the specific protein mixture used in this study, which contains high-dynamic range with high-abundance protein species (e.g. serum albumin and myoglobin),

we chose to load less total protein per sSEC run to avoid carry-over of these high-abundance species into later-eluted fractions (Figure S2, Supplemental Information pg. S3-S4). In cases like these, loading less protein per run may require pooling a higher number of replicates in order to obtain the desired protein concentration in each sSEC fraction. However, it is worth noting the mixture used in this study represents a case on the upper extreme of complexity and dynamic range. In fact, there is a wide range of applications for this high-resolution, MS-compatible strategy for size-based fractionation of other mixtures, such as affinity purified or immunoprecipitation mixtures. Moreover, this fractionation method is amenable to MS-compatible native conditions (aqueous ammonium acetate mobile phase, data not shown). The elution solvent for sSEC is highly customizable, depending on column choice and sample composition.

FT-ICR MS can be used as a stand-alone tool for top-down protein characterization from complex mixtures, such as cell lysates which has been demonstrated previously.²⁶ However, most of the proteins detected and characterized in this early study were < 10 kDa, which perfectly demonstrated a common interference problem seen for mass spectrometry analysis of large proteins from complex protein mixtures.²⁶ Later studies utilized SEC fractionation of complex protein mixtures prior to top-down analysis using FT-ICR MS.^{27, 33} However, these works utilized salt-based SEC eluents which often require desalting of fractions prior to MS-analysis. In contrast, our sSEC platform utilizes an MS-compatible eluent, there is minimal sample work-up prior to analysis with FT-ICR MS, which ultimately reduces sample loss and preserves sample integrity due to reduced handling. Our previous study, in which we coupled sSEC-RPC to a quadrupole time of flight (Q-TOF) mass spectrometer, demonstrated increased proteome coverage across a wide range of protein sizes from complex heart tissue lysate.³⁵ Although our current method and the former both utilize MS-compatible sSEC fractionation, for the first time, this current strategy combines the advantages of high-resolution, MS-compatible size-based sSEC fractionation of a complex protein mixture with the benefits of ultra-high resolving power of FT-ICR MS for top-down protein characterization. This has allowed for isolation of individual proteoforms “in-MS” for MS/MS to ultimately achieve in-depth sequence coverage for proteins ranging from low-MW to high-MW. Also for the first time, we have demonstrated that the ultra-high resolving power of FT-ICR MS allowed for the base-line isotopic resolution of over thirty 30 distinct proteoforms (30–50 kDa) in a single mass spectrum in a 100 m/z window. Here, ultra-high resolving power FT-ICR MS acted as a dimension of separation and allowed for in-MS separation of proteoforms, including those >30 kDa, for subsequent isolation and MS/MS experiments toward deeper proteoform sequence coverage.

We envision that further modifications and improvements could be made to this top-down proteomics platform to address individual needs. For instance, to enhance the proteome coverage, an additional dimension of separation could be added between sSEC fractionation and FT-ICR MS analysis. Previously, we have shown that proteome coverage was improved by using the two-dimensional sSEC-RPC strategy coupled to a Q-TOF mass spectrometer, but this 2DLC strategy has yet to be implemented with FT-ICR MS.³⁵ Additionally, Anderson et al. demonstrated impressive coverage of the human colorectal cancer proteome by implementing GELFrEE-RPC-FT ICR MS.³² Together these two studies demonstrate the power of size-based fractionation coupled with online RPC for improved proteome

coverage. Recently, Sun and co-workers showed that capillary zone electrophoresis (CZE) coupled to top-down MS yielded high proteome coverage, providing opportunities for coupling sSEC fractionation with CZE and FT-ICR MS.⁶⁰ On the other hand, should protein characterization and proteoform sequence coverage be the primary goal, it is best to implement offline fractionation followed by direct infusion into the FT-ICR MS to eliminate the limitations of LC-MS timescale for MS/MS, especially when conducting ECD experiments which often require a high number of summed transients for high sequence coverage.

Conclusion

We have introduced a versatile top-down proteomics platform, which combines MS-compatible sSEC fractionation with FT-ICR MS. We analyzed an sSEC-fractionated complex mixture extracted from the soluble heart proteome and reduced signal suppression of large proteins by reducing co-elution with small proteins. Specifically, we detected and resolved 31 distinct proteoforms (30–50 kDa) from a single FT-ICR mass spectrum in a 100 m/z window. As demonstrated, isotopic resolution at the baseline level were achieved for 6 high-MW proteoforms simultaneously in one FT-ICR mass spectrum (within a 5 m/z window). From these fractions, we accurately measured the mass of large metabolic enzymes. This demonstrates the power of combining MS-compatible sSEC fractionation with ultra-high resolving power of FT-ICR MS without a need for extensive liquid chromatography separation or protein purification. Additionally, we isolated protein ions in-MS for subsequent deep sequence characterization by ECD. We envision this top-down proteomics platform can also be adapted for simpler mixtures, such as those generated by immunoprecipitation or affinity purification procedures. Since sSEC is compatible with variety of MS-compatible eluents, this platform is amenable to native MS applications. The real-time tunability of direct infusion FT-ICR MS analysis allows for users to adjust parameters as need to customize top-down analysis for individual proteins from the same mixture.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to acknowledge Dr. Andrew Alpert of PolyLC Inc. for continued dialogue about intact protein separations and for providing SEC columns for use in this work. We would like to acknowledge the NIH grants NIH R01 HL109810, R01 HL096971, R01 GM117058, R01 GM125085, and S10 OD018475 (Y.G.). T.T. would like to acknowledge support from the NIH Chemistry-Biology Interface Training Program NIH T32GM008505. S.K. was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1747503 (S.K.) and the WiscAMP-BD program under Grant No. HRD-1612530 (S.K.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Toby TK; Fornelli L; Kelleher NL, Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu Rev Anal Chem* (Palo Alto Calif) 2016, 9 (1), 499–519. [PubMed: 27306313]

2. Cai W; Tucholski TM; Gregorich ZR; Ge Y, Top-down Proteomics: Technology Advancements and Applications to Heart Diseases. *Expert Rev Proteomics* 2016, 13 (8), 717–30. [PubMed: 27448560]
3. Chen B; Brown KA; Lin Z; Ge Y, Top-Down Proteomics: Ready for Prime Time? *Anal Chem* 2018, 90 (1), 110–127. [PubMed: 29161012]
4. Smith LM; Kelleher NL; Consortium for Top Down, P., Proteoform: a single term describing protein complexity. *Nat Methods* 2013, 10, (3), 186–7. [PubMed: 23443629]
5. Aebersold R; Agar JN; Amster IJ; Baker MS; Bertozzi CR; Boja ES; Costello CE; Cravatt BF; Fenselau C; Garcia BA; Ge Y; Gunawardena J; Hendrickson RC; Hergenrother PJ; Huber CG; Ivanov AR; Jensen ON; Jewett MC; Kelleher NL; Kiessling LL; Krogan NJ; Larsen MR; Loo JA; Ogorzalek Loo RR; Lundberg E; MacCoss MJ; Mallick P; Mootha VK; Mrksich M; Muir TW; Patrie SM; Pesavento JJ; Pitteri SJ; Rodriguez H; Saghatelian A; Sandoval W; Schluter H; Sechi S; Slavoff SA; Smith LM; Snyder MP; Thomas PM; Uhlen M; Van Eyk JE; Vidal M; Walt DR; White FM; Williams ER; Wohlschlagel T; Wysocki VH; Yates NA; Young NL; Zhang B, How many human proteoforms are there? *Nat Chem Biol* 2018, 14 (3), 206–214. [PubMed: 29443976]
6. Smith LM; Kelleher NL, Proteoforms as the next proteomics currency. *Science* 2018, 359 (6380), 1106–1107. [PubMed: 29590032]
7. Dong X; Sumandea CA; Chen YC; Garcia-Cazarin ML; Zhang J; Balke CW; Sumandea MP; Ge Y, Augmented phosphorylation of cardiac troponin I in hypertensive heart failure. *J Biol Chem* 2012, 287 (2), 848–57. [PubMed: 22052912]
8. Peng Y; Gregorich ZR; Valeja SG; Zhang H; Cai W; Chen Y-C; Guner H; Chen AJ; Schwahn DJ; Hacker TA; Liu X; Ge Y, Top-down proteomics reveals concerted reductions in myofilament and Z-disc protein phosphorylation after acute myocardial infarction. *Mol Cell Proteomics* 2014, 13, (10), 2752–2764. [PubMed: 24969035]
9. Zhang J; Guy MJ; Norman HS; Chen YC; Xu Q; Dong X; Guner H; Wang S; Kohmoto T; Young KH; Moss RL; Ge Y, Top-down quantitative proteomics identified phosphorylation of cardiac troponin I as a candidate biomarker for chronic heart failure. *J Proteome Res* 2011, 10 (9), 4054–65. [PubMed: 21751783]
10. Ge Y; Rybakova IN; Xu Q; 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 U S A* 2009, 106 (31), 12658–63. [PubMed: 19541641]
11. Ansong C; Wu S; Meng D; Liu X; Brewer HM; Deatherage Kaiser BL; Nakayasu ES; Cort JR; Pevzner P; Smith RD; Heffron F; Adkins JN; Pasa-Tolic L, Top-down proteomics reveals a unique protein S-thiolation switch in *Salmonella Typhimurium* in response to infection-like conditions. *Proc Natl Acad Sci U S A* 2013, 110 (25), 10153–8. [PubMed: 23720318]
12. Chamot-Rooke J; Mikaty G; Malosse C; Soyer M; Dumont A; Gault J; Imhaus AF; Martin P; Trellet M; Clary G; Chafey P; Camoin L; Nilges M; Nassif X; Dumenil G, Posttranslational modification of pili upon cell contact triggers *N. meningitidis* dissemination. *Science* 2011, 331 (6018), 778–82. [PubMed: 21311024]
13. Ntai I; Fornelli L; DeHart CJ; Hutton JE; Doubleday PF; LeDuc RD; van Nispen AJ; Fellers RT; Whiteley G; Boja ES; Rodriguez H; Kelleher NL, Precise characterization of KRAS4b proteoforms in human colorectal cells and tumors reveals mutation/modification cross-talk. *Proc Natl Acad Sci U S A* 2018, 115 (16), 4140–4145. [PubMed: 29610327]
14. Carel C; Marcoux J; Reat V; Parra J; Latge G; Laval F; Demange P; Burlet-Schiltz O; Milon A; Daffe M; Tropis MG; Renault MAM, Identification of specific posttranslational O-mycoloylations mediating protein targeting to the mycomembrane. *Proc Natl Acad Sci U S A* 2017, 114 (16), 4231–4236. [PubMed: 28373551]
15. Marshall AG; Hendrickson CL; Jackson GS, Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* 1998, 17, (1) 1–35. [PubMed: 9768511]
16. 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 (4), 672–8. [PubMed: 11804498]
17. Han X; Jin M; Breuker K; McLafferty FW, Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science* 2006, 314 (5796), 109–12. [PubMed: 17023655]

18. Gregorich ZR; Cai W; Lin Z; Chen AJ; Peng Y; Kohmoto T; Ge Y, Distinct sequences and post-translational modifications in cardiac atrial and ventricular myosin light chains revealed by top-down mass spectrometry. *J Mol Cell Cardiol* 2017, 107, 13–21. [PubMed: 28427997]
19. Li H; Nguyen HH; Ogorzalek Loo RR; Campuzano IDG; Loo JA, An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat Chem* 2018, 10 (2), 139–148. [PubMed: 29359744]
20. Tipton JD; Tran JC; Catherman AD; Ahlf DR; Durbin KR; Lee JE; Kellie JF; Kelleher NL; Hendrickson CL; Marshall AG, Nano-LC FTICR tandem mass spectrometry for top-down proteomics: routine baseline unit mass resolution of whole cell lysate proteins up to 72 kDa. *Anal Chem* 2012, 84 (5), 2111–7. [PubMed: 22356091]
21. 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 U S A* 2002, 99 (4), 1774–9. [PubMed: 11842225]
22. Peng Y; Yu D; Gregorich Z; Chen X; Beyer AM; Gutteran DD; Ge Y, In-depth proteomic analysis of human tropomyosin by top-down mass spectrometry. *J Muscle Res Cell Motil* 2013, 34, (3–4), 199–210. [PubMed: 23881156]
23. Peng Y; Chen X; Zhang H; Xu Q; Hacker TA; Ge Y, Top-down targeted proteomics for deep sequencing of tropomyosin isoforms. *J Proteome Res* 2013, 12 (1), 187–98. [PubMed: 23256820]
24. Bourgoin-Voillard S; Leymarie N; Costello CE, Top-down tandem mass spectrometry on RNase A and B using a Qh/FT-ICR hybrid mass spectrometer. *Proteomics* 2014, 14 (10), 1174–84. [PubMed: 24687996]
25. Lin Z; Guo F; Gregorich ZR; Sun R; Zhang H; Hu Y; Shanmuganayagam D; Ge Y, Comprehensive Characterization of Swine Cardiac Troponin T Proteoforms by Top-Down Mass Spectrometry. *J Am Soc Mass Spectrom* 2018, 29 (6), 1284–1294. [PubMed: 29633223]
26. Ge Y; El-Naggar M; Sze SK; Oh HB; Begley TP; McLafferty FW; Boshoff H; Barry CE 3rd, Top down characterization of secreted proteins from *Mycobacterium tuberculosis* by electron capture dissociation mass spectrometry. *J Am Soc Mass Spectrom* 2003, 14 (3), 253–61. [PubMed: 12648932]
27. Zabrouskov V; Giacomelli L; van Wijk KJ; McLafferty FW, A new approach for plant proteomics: characterization of chloroplast proteins of *Arabidopsis thaliana* by top-down mass spectrometry. *Mol Cell Proteomics* 2003, 2 (12), 1253–60. [PubMed: 14506204]
28. Compton PD; Zamdborg L; Thomas PM; Kelleher NL, On the scalability and requirements of whole protein mass spectrometry. *Anal Chem* 2011, 83 (17), 6868–74. [PubMed: 21744800]
29. Kellie JF; Catherman AD; Durbin KR; Tran JC; Tipton JD; Norris JL; Witkowski CE 2nd; Thomas PM; Kelleher NL, Robust analysis of the yeast proteome under 50 kDa by molecular-mass-based fractionation and top-down mass spectrometry. *Anal Chem* 2012, 84 (1), 209–15. [PubMed: 22103811]
30. Tran JC; Doucette AA, Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. *Anal Chem* 2008, 80 (5), 1568–73. [PubMed: 18229945]
31. Tran JC; Zamdborg L; Ahlf DR; Lee JE; Catherman AD; Durbin KR; Tipton JD; Vellaichamy A; Kellie JF; Li M; Wu C; Sweet SM; Early BP; Siuti N; LeDuc RD; Compton PD; Thomas PM; Kelleher NL, Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* 2011, 480 (7376), 254–8. [PubMed: 22037311]
32. Anderson LC; DeHart CJ; Kaiser NK; Fellers RT; Smith DF; Greer JB; LeDuc RD; Blakney GT; Thomas PM; Kelleher NL; Hendrickson CL, Identification and Characterization of Human Proteoforms by Top-Down LC-21 Tesla FT-ICR Mass Spectrometry. *J Proteome Res* 2017, 16 (2), 1087–1096. [PubMed: 27936753]
33. 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 (13), 2722–33. [PubMed: 16732621]
34. Chen X; Ge Y, Ultrahigh pressure fast size exclusion chromatography for top-down proteomics. *Proteomics* 2013, 13 (17), 2563–6. [PubMed: 23794208]

35. Cai W; Tucholski T; Chen B; Alpert AJ; McIlwain S; Kohmoto T; Jin S; Ge Y, Top-Down Proteomics of Large Proteins up to 223 kDa Enabled by Serial Size Exclusion Chromatography Strategy. *Anal Chem* 2017, 89 (10), 5467–5475. [PubMed: 28406609]
36. Yi-Chen C; Ayaz-Guner S; Peng Y; Lane NM; Locher MR; Kohmoto T; Larsson L; Moss RL; Ge Y, Effective top-down LC/MS+ method for assessing actin isoforms as a potential cardiac disease marker. *Anal Chem* 2015, 87, (16), 8399–406. [PubMed: 26189812]
37. Cai W; Guner H; Gregorich ZR; Chen AJ; Ayaz-Guner S; Peng Y; Valeja SG; Liu X; Ge Y, MASH Suite Pro: A Comprehensive Software Tool for Top-Down Proteomics. *Mol Cell Proteomics* 2016, 15 (2), 703–14. [PubMed: 26598644]
38. Kou Q; Xun L; Liu X, TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* 2016, 32 (22), 3495–3497. [PubMed: 27423895]
39. Alpert AJ, 8 - Size Exclusion High-Performance Liquid Chromatography of Small Solutes A2 - Wu, Chi-san In *Column Handbook for Size Exclusion Chromatography*, Academic Press: San Diego, 1999; pp 249–266.
40. Ordway GA; Garry DJ, Myoglobin: an essential hemoprotein in striated muscle. *J Exp Biol* 2004, 207 (Pt 20), 3441–6. [PubMed: 15339940]
41. Kleine AH; Glatz JF; Van Nieuwenhoven FA; Van der Vusse GJ, Release of heart fatty acid-binding protein into plasma after acute myocardial infarction in man. *Mol Cell Biochem* 1992, 116 (1–2), 155–62. [PubMed: 1480144]
42. Tanaka T; Hirota Y; Sohmiya K; Nishimura S; Kawamura K, Serum and urinary human heart fatty acid-binding protein in acute myocardial infarction. *Clin Biochem* 1991, 24 (2), 195–201. [PubMed: 2040092]
43. Catalucci D; Latronico MVG; Condorelli G, FABP3 as Biomarker of Heart Pathology In *General Methods in Biomarker Research and their Applications*, Preedy VR; Patel VB, Eds. Springer Netherlands: Dordrecht, 2015; pp 439–454.
44. Bhat SP; Horwitz J; Srinivasan A; Ding L, Alpha B-crystallin exists as an independent protein in the heart and in the lens. *Eur J Biochem* 1991, 202 (3), 775–81. [PubMed: 1765091]
45. Wood TD; Chen LH; White CB; Babbitt PC; Kenyon GL; McLafferty FW, Sequence verification of human creatine kinase (43 kDa) isozymes by high-resolution tandem mass spectrometry. *Proc Natl Acad Sci U S A* 1995, 92 (25), 11451–5. [PubMed: 8524781]
46. Senko MW; Beu SC; McLaffertycor FW, Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J Am Soc Mass Spectrom* 1995, 6 (4), 229–33. [PubMed: 24214167]
47. Ledford EB Jr.; Rempel DL; Gross ML, Space charge effects in Fourier transform mass spectrometry. *Mass calibration. Anal Chem* 1984, 56 (14), 2744–8. [PubMed: 6524653]
48. Masselon C; Tolmachev AV; Anderson GA; Harkewicz R; Smith RD, Mass measurement errors caused by "local" frequency perturbations in FTICR mass spectrometry. *J Am Soc Mass Spectrom* 2002, 13 (1), 99–106. [PubMed: 11777206]
49. Mikhailov VA; Cooper HJ, Activated Ion Electron Capture Dissociation (AI ECD) of proteins: synchronization of infrared and electron irradiation with ion magnetron motion. *J Am Soc Mass Spectrom* 2009, 20 (5), 763–71. [PubMed: 19200749]
50. Riley NM; Westphall MS; Coon JJ, Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal Chem* 2015, 87 (14), 7109–16. [PubMed: 26067513]
51. Zubarev RA; Kelleher NL; McLafferty FW, Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society* 1998, 120, (13), 3265–66.
52. Zubarev RA; Kruger NA; Fridriksson EK; Lewis MA; Horn DM; Carpenter BK; McLafferty FW, Electron Capture Dissociation of Gaseous Multiply-Charged Proteins Is Favored at Disulfide Bonds and Other Sites of High Hydrogen Atom Affinity. *Journal of the American Chemical Society* 1999, 121 (12), 2857–2862.
53. Zhurov KO; Fornelli L; Wodrich MD; Laskay UA; Tsybin YO, Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chem Soc Rev* 2013, 42 (12), 5014–30. [PubMed: 23450212]

54. Syka JE; Coon JJ; Schroeder MJ; Shabanowitz J; Hunt DF, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 2004, 101 (26), 9528–33. [PubMed: 15210983]
55. Siuti N; Kelleher NL, Decoding protein modifications using top-down mass spectrometry. *Nat Methods* 2007, 4 (10), 817–21. [PubMed: 17901871]
56. Jin Y; Wei L; Cai W; Lin Z; Wu Z; Peng Y; Kohmoto T; Moss RL; Ge Y, Complete Characterization of Cardiac Myosin Heavy Chain (223 kDa) Enabled by Size-Exclusion Chromatography and Middle-Down Mass Spectrometry. *Anal Chem* 2017, 89 (9), 4922–4930. [PubMed: 28366003]
57. Brodbelt JS, Photodissociation mass spectrometry: New tools for characterization of biological molecules. *Chem Soc Rev* 2014, 43 (8), 2757–83. [PubMed: 24481009]
58. Cleland TP; DeHart CJ; Fellers RT; VanNispen AJ; Greer JB; LeDuc RD; Parker WR; Thomas PM; Kelleher NL; Brodbelt JS, High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer. *J Proteome Res* 2017, 16 (5), 2072–2079. [PubMed: 28412815]
59. Shaw JB; Robinson EW; Pasa-Tolic L, Vacuum Ultraviolet Photodissociation and Fourier Transform-Ion Cyclotron Resonance (FT-ICR) Mass Spectrometry: Revisited. *Anal Chem* 2016, 88 (6), 3019–23. [PubMed: 26882021]
60. McCool EN; Lubeckyj RA; Shen X; Chen D; Kou Q; Liu X; Sun L, Deep Top-Down Proteomics Using Capillary Zone Electrophoresis-Tandem Mass Spectrometry: Identification of 5700 Proteoforms from the Escherichia coli Proteome. *Anal Chem* 2018, 90 (9), 5529–5533. [PubMed: 29620868]

Figure 1: Direct FT-ICR MS analysis of sSEC Fractions 9-12.

A) Top left: SDS-PAGE visualization of unfractionated soluble human heart protein extract (loading mixture, LM) and sSEC fractions 9-12. MW markers in decreasing order: 250, 150, 100, 75, 50, 37, 25, 20, 15, 10 kDa (corresponds to SDS-PAGE in Figure S1). Center: FT-ICR mass spectra acquired for LM and sSEC fractions 9-12. Segmented spectra were acquired in 100 m/z -wide quadrupole isolation windows, (window center 900, 1000, and 1100 m/z) and combined to generate a continuous overall spectrum for the LM and each sSEC fraction 9-12. Intensity for all spectra were normalized to the most intense peak in the spectrum. The protein species with the most intense signal in each mass spectrum are indicated by colored shapes (lower left-hand corner). *Ions with charge less than 3+. **B)** A zoom-in mass spectra (961-977 m/z) to show signal from lower-abundance species. Insets are shown for LM and fractions 9-11 to demonstrate the difference between visible protein components in each fraction.

Figure 2: sSEC enabled detection of larger proteins after separation from the high abundance, low-MW proteins.

A) FT-ICR mass spectra for fractions 8 and 9. SDS-PAGE visualization (corresponding to Figure S1). **B)** A zoom-in mass spectra between 995–1012 m/z of fractions 8 and 9. Colored labels correspond to proteins with specific MWs, respectively. *oxidation products.

Figure 3: Accurate mass measurements of metabolic enzymes by FT ICR MS following sSEC separation.

A) SDS-PAGE visualization of fractions 7 and 8 with corresponding FT-ICR mass spectra.

B) A zoom-in mass spectra of 997-1001 m/z for fractions 7 and 8 showing the difference between the mass spectra of the two fractions. **C)** Proteins detected by intact mass from top-left to bottom-right: creatine kinase, M-type (CKM); creatine kinase, S-type (CKS); isocitrate dehydrogenase (IDH); glyceraldehyde-3-phosphate dehydrogenase (GAPDH); malate dehydrogenase, mitochondrial (MDH); aspartate aminotransferase (AAT). External calibration was performed using DataAnalysis and theoretical fit output is generated by MASH Suite Pro using Enhanced THRASH deconvolution.

Figure 4: ECD of Metabolic Enzymes from Soluble Heart Proteome.

A) Representative mass spectrum of 940-1040 *m/z* for sSEC fraction 8 of heart soluble proteome extracted from *sus scrofa* (swine) heart tissue. Charge states for 4 protein species are highlighted. Malate dehydrogenase, mitochondrial (MDH); creatine kinase, M-type (CKM); acetylated β -enolase (Ac β -ENO3); glyceraldehyde-3-phosphate dehydrogenase (GAPDH). The calculated and experimental monoisotopic masses with mass errors are indicated for each intact mass. **B)** Sequence maps for metabolic enzymes malate dehydrogenase, mitochondrial and creatine kinase, M-type. Each sequence table represents fragments from 1 ECD experiment. Parameters (pulse energy, pulse time, and number of transients) are indicated under each sequence table.

Scheme 1: A schematic representation of the sSEC/FT-ICR MS top-down proteomics workflow. 1) Offline sSEC fractionation of a protein mixture with automatic fraction collection, 2) FT-ICR MS analysis of sSEC fractions, 3) In-mass spectrometer (in-MS) separation of ions for MS or MS/MS analysis, 4) Measurement of intact mass for tentative protein identification, 5) MS/MS analysis for protein characterization.