

MASH Suite Pro: A Comprehensive Software Tool for Top-Down Proteomics*[§]

Wenxuan Cai^{‡§§§}, Huseyin Guner^{‡¶§§}, Zachery R. Gregorich^{‡§}, Albert J. Chen[‡], Serife Ayaz-Guner[‡], Ying Peng[‡], Santosh G. Valeja[‡], Xiaowen Liu^{**}, and Ying Ge^{‡¶¶¶¶¶}

Top-down mass spectrometry (MS)-based proteomics is arguably a disruptive technology for the comprehensive analysis of all proteoforms arising from genetic variation, alternative splicing, and posttranslational modifications (PTMs). However, the complexity of top-down high-resolution mass spectra presents a significant challenge for data analysis. In contrast to the well-developed software packages available for data analysis in bottom-up proteomics, the data analysis tools in top-down proteomics remain underdeveloped. Moreover, despite recent efforts to develop algorithms and tools for the deconvolution of top-down high-resolution mass spectra and the identification of proteins from complex mixtures, a multifunctional software platform, which allows for the identification, quantitation, and characterization of proteoforms with visual validation, is still lacking. Herein, we have developed MASH Suite Pro, a comprehensive software tool for top-down proteomics with multifaceted functionality. MASH Suite Pro is capable of processing high-resolution MS and tandem MS (MS/MS) data using two deconvolution algorithms to optimize protein identification results. In addition, MASH Suite Pro allows for the characterization of PTMs and sequence variations, as well as the relative quantitation of multiple proteoforms in different experimental conditions. The program also provides visualization components for validation and correction of the computational outputs. Furthermore, MASH Suite Pro facilitates data reporting and presentation via direct output of the graphics. Thus, MASH Suite Pro significantly simplifies and speeds up the interpretation of high-resolution

top-down proteomics data by integrating tools for protein identification, quantitation, characterization, and visual validation into a customizable and user-friendly interface. We envision that MASH Suite Pro will play an integral role in advancing the burgeoning field of top-down proteomics. *Molecular & Cellular Proteomics* 15: 10.1074/mcp.O115.054387, 703–714, 2016.

With well-developed algorithms and computational tools for mass spectrometry (MS)¹ data analysis, peptide-based bottom-up proteomics has gained considerable popularity in the field of systems biology (1–9). Nevertheless, the bottom-up approach is suboptimal for the analysis of protein posttranslational modifications (PTMs) and sequence variants as a result of protein digestion (10). Alternatively, the protein-based top-down proteomics approach analyzes intact proteins, which provides a “bird’s eye” view of all proteoforms (11), including those arising from sequence variations, alternative splicing, and diverse PTMs, making it a disruptive technology for the comprehensive analysis of proteoforms (12–24). However, the complexity of top-down high-resolution mass spectra presents a significant challenge for data analysis. In contrast to the well-developed software packages available for processing data from bottom-up proteomics experiments, the data analysis tools in top-down proteomics remain underdeveloped.

The initial step in the analysis of top-down proteomics data is deconvolution of high-resolution mass and tandem mass spectra. Thorough high-resolution analysis of spectra by horn (THRASH), which was the first algorithm developed for the deconvolution of high-resolution mass spectra (25), is still widely used. THRASH automatically detects and evaluates individual isotopomer envelopes by comparing the experimental isotopomer envelope with a theoretical envelope and reporting those that score higher than a user-defined threshold. Another commonly used algorithm, MS-Deconv, utilizes a

From the [‡]Department of Cell and Regenerative Biology, [§]Molecular Pharmacology Training Program, [¶]Human Proteomics Program, ^{‡‡}Department of Chemistry, University of Wisconsin-Madison, 1300 University Ave., Madison, WI 53706, [§]Department of BioHealth Informatics, Indiana University-Purdue University Indianapolis, 719 Indiana Ave., Indianapolis, IN 46202, ^{**}Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, 410 West 10th Street, Indianapolis, IN 46202

Received August 5, 2015, and in revised form, November 11, 2015
 Published November 23, 2015, MCP Papers in Press, DOI 10.1074/mcp.O115.054387

Author contributions: W.C., H.G., and Y.G. designed the research; W.C., H.G., Z.R.G., A.J.C., S.A., Y.P., S.G.V., X.L., and Y.G. performed the research; X.L. contributed new reagents or analytic tools; W.C. analyzed data; and W.C., Z.R.G., A.J.C., and Y.G. wrote the paper.

¹ The abbreviations used are: MS, mass spectrometry; MS/MS, tandem mass spectrometry; LC-MS, liquid chromatography-mass spectrometry; FT-ICR, Fourier transform ion cyclotron resonance; S/N, signal-to-noise ratio; PTMs, posttranslational modifications; THRASH, thorough high-resolution analysis of spectra by horn; Top-PIC, Top-Down Mass Spectrometry Based Proteoform Identification and Characterization; PrSM, proteoform spectrum match.

combinatorial approach to address the difficulty of grouping MS peaks from overlapping isotopomer envelopes (26). Recently, UniDec, which employs a Bayesian approach to separate mass and charge dimensions (27), can also be applied to the deconvolution of high-resolution spectra. Although these algorithms assist in data processing, unfortunately, the deconvolution results often contain a considerable amount of misassigned peaks as a consequence of the complexity of the high-resolution MS and MS/MS data generated in top-down proteomics experiments. Errors such as these can undermine the accuracy of protein identification and PTM localization and, thus, necessitate the implementation of visual components that allow for the validation and manual correction of the computational outputs.

Following spectral deconvolution, a typical top-down proteomics workflow incorporates identification, quantitation, and characterization of proteoforms; however, most of the recently developed data analysis tools for top-down proteomics, including ProSightPC (28, 29), Mascot Top Down (also known as Big-Mascot) (30), MS-TopDown (31), and MS-Align+ (32), focus almost exclusively on protein identification. ProSightPC was the first software tool specifically developed for top-down protein identification. This software utilizes “shotgun annotated” databases (33) that include all possible proteoforms containing user-defined modifications. Consequently, ProSightPC is not optimized for identifying PTMs that are not defined by the user(s). Additionally, the inclusion of all possible modified forms within the database dramatically increases the size of the database and, thus, limits the search speed (32). Mascot Top Down (30) is based on standard Mascot but enables database searching using a higher mass limit for the precursor ions (up to 110 kDa), which allows for the identification of intact proteins. Protein identification using Mascot Top Down is fundamentally similar to that used in bottom-up proteomics (34), and, therefore, it is somewhat limited in terms of identifying unexpected PTMs. MS-TopDown (31) employs the spectral alignment algorithm (35), which matches the top-down tandem mass spectra to proteins in the database without prior knowledge of the PTMs. Nevertheless, MS-TopDown lacks statistical evaluation of the search results and performs slowly when searching against large databases. MS-Align+ also utilizes spectral alignment for top-down protein identification (32). It is capable of identifying unexpected PTMs and allows for efficient filtering of candidate proteins when the top-down spectra are searched against a large protein database. MS-Align+ also provides statistical evaluation for the selection of proteoform spectrum match (PrSM) with high confidence. More recently, Top-Down Mass Spectrometry Based Proteoform Identification and Characterization (TopPIC) was developed (<http://proteomics.informatics.iupui.edu/software/toppic/index.html>). TopPIC is an updated version of MS-Align+ with increased spectral alignment speed and reduced computing requirements. In addition, MSPathFinder, developed by Kim *et al.*, also allows for

the rapid identification of proteins from top-down tandem mass spectra (<http://omics.pnl.gov/software/mspathfinder>) using spectral alignment. Although software tools employing spectral alignment, such as MS-Align+ and MSPathFinder, are particularly useful for top-down protein identification, these programs operate using command line, making them difficult to use for those with limited knowledge of command syntax.

Recently, new software tools have been developed for proteoform characterization (36, 37). Our group previously developed MASH Suite, a user-friendly interface for the processing, visualization, and validation of high-resolution MS and MS/MS data (36). Another software tool, ProSight Lite, developed recently by the Kelleher group (37), also allows characterization of protein PTMs. However, both of these software tools require prior knowledge of the protein sequence for the effective localization of PTMs. In addition, both software tools cannot process data from liquid chromatography (LC)-MS and LC-MS/MS experiments, which limits their usefulness in large-scale top-down proteomics. Thus, despite these recent efforts, a multifunctional software platform enabling identification, quantitation, and characterization of proteins from top-down spectra, as well as visual validation and data correction, is still lacking.

Herein, we report the development of MASH Suite Pro, an integrated software platform, designed to incorporate tools for protein identification, quantitation, and characterization into a single comprehensive package for the analysis of top-down proteomics data. This program contains a user-friendly customizable interface similar to the previously developed MASH Suite (36) but also has a number of new capabilities, including the ability to handle complex proteomics datasets from LC-MS and LC-MS/MS experiments, as well as the ability to identify unknown proteins and PTMs using MS-Align+ (32). Importantly, MASH Suite Pro also provides visualization components for the validation and correction of the computational outputs, which ensures accurate and reliable deconvolution of the spectra and localization of PTMs and sequence variations.

EXPERIMENTAL PROCEDURES

Software Design and Implementation—The default algorithm for spectral deconvolution in MASH Suite Pro is a modified version of THRASH (25) that we developed in-house based on the Decon2LS open source code (38). MS-Align+ (32) has been integrated into the program and is used for top-down protein identification. The program is written on the Microsoft .NET framework. The main scientific algorithms are written in C++, and the visual development was written with C#. The windows were developed using Qios Devsuite, and the graphs and spectrum charts were developed using Microsoft Chart Controls. The graphical user interface of MASH Suite Pro was designed using tabbed document interface. The software architecture pattern was based on model view controller, and the data tables were constructed based on Listview and Datagrid structures (36).

Extraction of Myofilament Proteins—Approximately 20 mg of swine cardiac left ventricular tissue was homogenized thoroughly in 200 μ l

of HEPES extraction buffer (25 mM HEPES, pH 7.5, 50 mM NaF, 0.25 mM Na_3VO_4 , 0.25 mM PMSF, 2.5 mM EDTA) at 4 °C using a Teflon pestle (1.5 ml tube flat tip, Scienceware, Pequannock, NJ). The resulting homogenate was centrifuged at 17,000 relative centrifugal force for 15 min at 4 °C, and the supernatant was removed (19). The pellet was subsequently homogenized in 100 μl TFA extraction solution (1% TFA, 1 mM Tris (2-carboxyethyl) Phosphine) to extract the myofilament proteins. The homogenate was centrifuged at 17,000 relative centrifugal force for 15 min at 4 °C, and the supernatant, which is enriched in myofilament proteins, was transferred to a new 1.5 ml microfuge tube and centrifuged for an additional 60 min at 17,000 relative centrifugal force and 4 °C. The resulting supernatant was subject to LC-MS analysis.

Online LC-MS Analysis and Manual Fraction Collection—LC separation of the myofilament proteins and low-resolution MS analysis were performed as previously described (19) with minor modifications. Briefly, 3 μl of the myofilament extract (equivalent to 600 μg of tissue per injection) were injected, and the proteins in the mixture were eluted at a flow rate of 12.5 $\mu\text{l}/\text{min}$ with a gradient going from 20% mobile phase B to 95% mobile phase B in 57 min (mobile phase A: 0.10% formic acid in water; mobile phase B: 0.10% formic acid in 1:1 acetonitrile:ethanol). The flow was split after LC separation with ~10% of the eluting sample being ionized via electrospray ionization through a 50 μm inner diameter tip, and analyzed directly by a linear ion trap mass spectrometer (Thermo Scientific, Bremen, Germany). The remaining ~90% of the sample was simultaneously collected as fractions on ice for subsequent high-resolution MS and MS/MS analyses.

Offline High-Resolution Fourier Transform Ion Cyclotron Resonance (FT-ICR) MS and MS/MS Analyses—Fractions containing unknown proteins were analyzed using a 7 Tesla linear ion trap/FT-ICR mass spectrometer (LTQ/FT Ultra, Thermo Scientific) equipped with an automated chip-based nanoelectrospray ionization source (Triversa NanoMate, Advion Bioscience, Ithaca, NY) as described previously (19). The sample was introduced into the mass spectrometer using a spray voltage of 1.3 to 1.5 kV versus the inlet of the mass spectrometer. The resolving power of the FT-ICR was set at 200,000 at 400 m/z . The automatic gain control for a full scan in the linear ion trap, FT-ICR cell, MS^n FT-ICR cell, and electron capture dissociation were 3E4, 5E5, 5E5, and 8E5, respectively. For MS/MS experiments, the protein molecular ions of the individual charge states were first isolated and then fragmented using 1.5% to 4.5% electron energy for electron capture dissociation (corresponding to 0.6 to 3.5 eV) with a 70 ms duration without additional delay.

Multidimensional LC-MS/MS Analyses—The detailed experimental procedures were described previously (39). Human embryonic kidney (HEK) 293 cell lysate was subject to ion exchange chromatography followed by reverse phase chromatography coupled to a Q Exactive benchtop Orbitrap mass spectrometer (Thermo Scientific). LC-MS/MS data were acquired with eight microscans at a resolving power of 70,000 (at 200 m/z) with automatic gain control set to 5E5 ions. A 10 V offset in the source was used for all of the experiments. In the top two data-dependent MS/MS scans, the intact protein ions were injected into the collision cell for higher energy collision dissociation (25 V) with a 10 s dynamic exclusion window.

Top-Down MS Data Analysis—The tandem mass spectra collected for unknown proteins were used to test MASH Suite Pro in terms of protein identification. Deconvolution was performed using enhanced-THRASH with a signal-to-noise ratio (S/N) threshold of 3 and a minimum fit of 60%. All deconvoluted masses were manually validated prior to identification with the targeted database from NCBI (SsCrofa10.2, containing 24,476 protein sequences) using MS-Align+. Alternatively, the raw MS/MS data were converted to mzXML files in centroid mode and deconvoluted using MS-Deconv (26) with

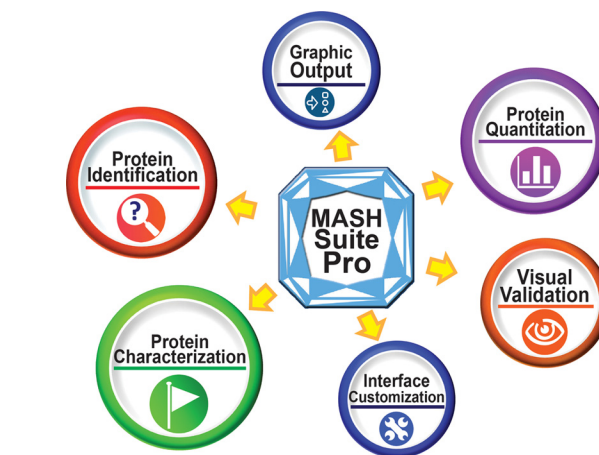


FIG. 1. Schematic summarizing the various functions and features of MASH Suite Pro. The major functions of MASH Suite Pro include protein identification, quantitation, and characterization of protein PTMs from top-down MS and MS/MS experiments. The program is equipped with various visualization components for the validation of the deconvolution results, identification results, and fragment ion assignments. Additional features include direct output of the graphics and customization of the program interface.

an S/N of 1. The m/z , charge state, and mass for each of the precursor ions were manually validated, and the list of reported monoisotopic masses of the fragment ions was aligned against the targeted database (SsCrofa10.2, containing 24,476 protein sequences) using MS-Align+. To determine whether MASH Suite Pro can process LC-MS/MS data, the dataset reported in a previous publication was used (39). Similarly, deconvolution was performed using either enhanced-THRASH with an S/N threshold of 3 and a minimum fit of 60% or MS-Deconv with an S/N of 1. The precursor list was automatically retrieved by MASH Suite Pro, and the dataset was searched against a human protein database (Uniprot-Swissprot database, released January 2013, containing 20,232 protein sequences).

RESULTS

MASH Suite Pro is an integrated software tool combining protein identification, quantitation, and characterization with visual displays for validation and correction of the computational outputs, as well as features enabling interface customization (Fig. 1). All tasks are conducted in a user-friendly interface that guides the user through the process of data analysis (Fig. 2). The zoom-in and extended views of the individual windows of the software are shown in [Supplemental Figs. S1-S3](#). MASH Suite Pro can identify proteins with unexpected modifications, provide relative quantitation of proteoforms from different experimental/biological conditions, and localize PTMs and sequence variations. MASH Suite Pro also provides visualization components for validation and correction of the deconvoluted mass list and fragment ion assignments. It also allows users to output spectra and protein sequences directly. Moreover, MASH Suite Pro provides users with flexibility in terms of customizing the program interface, which can significantly improve the user experience.

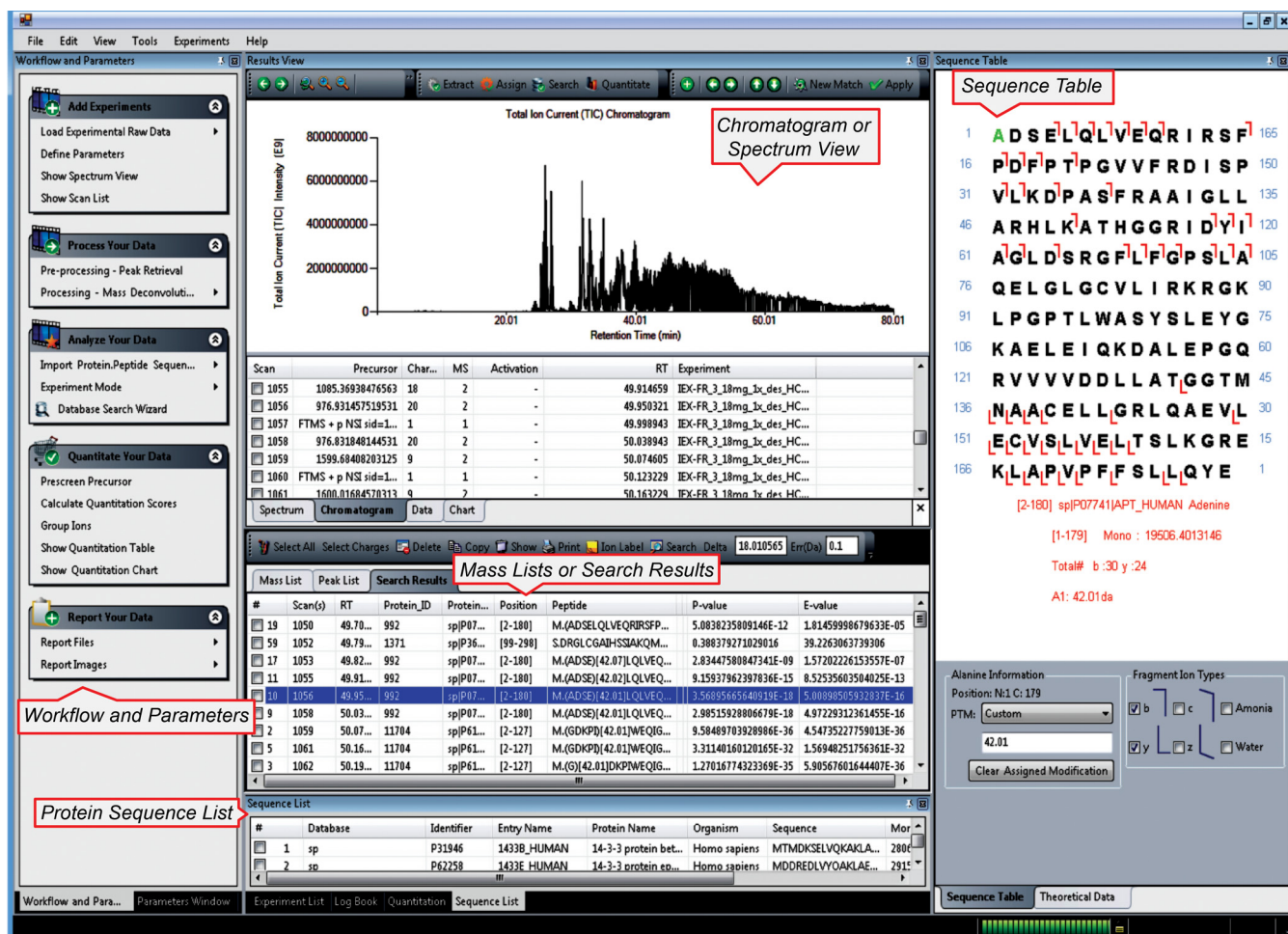


FIG. 2. Overview of the MASH Suite Pro interface. The intuitive “Workflow and Parameters” window guides users through the various steps of high-resolution data analysis. Different windows allow for clear visualization of the data, including chromatograms and mass spectra, as well as the sequence table, which displays the fragmentation map of the selected protein with detailed bond cleavage information. The deconvoluted masses and search results are organized into lists, providing detailed information of each item. The complete and extended views of the “Workflow and Parameters,” “Chromatogram/Spectrum,” and “Sequence Table” windows are shown in Supplemental Figs. S1–S3.

MASH Suite Pro for Top-Down Protein Identification—MASH Suite Pro utilizes the MS-Align+ algorithm to identify proteins from top-down LC-MS/MS and MS/MS data (32). With MS-Align+, MASH Suite Pro can identify truncated proteoforms, as well as those with unexpected PTMs and sequence variations. The protein identification results are output as PrSMs with statistical evaluation (p value and e -value) (32), allowing users to determine the confidence of the identification results. The workflow for protein identification in MASH Suite Pro is outlined in Fig. 3A. The raw data (e.g. .raw files from Thermo Scientific instruments) can be directly read by the program and all the steps for protein identification can be performed in an integrated search wizard (Fig. 3B). The *Database Search Wizard* allows users to select different deconvolution methods (MS-Deconv or enhanced-THRASH) and fragmentation types. For LC-MS/MS data, MASH Suite Pro also allows users to select a certain scan range to minimize the duration of data processing and database search. The

precursor m/z , charge states, and monoisotopic masses can also be automatically retrieved by the program. For MS/MS data lacking the precursor information, users are required to manually input the precursor information before searching the database. Additionally, as reported previously (36), users are able to perform deconvolution using MASH Suite Pro and validate the program-determined precursor ion monoisotopic masses prior to the search.

MASH Suite Pro implements a *Log Book* (Supplemental Fig. S4), allowing for tracking of the search process. Once the database search is finished, the search results are automatically imported into the program (Fig. 2B) with detailed information, including the name and identifier (ID) of the identified protein, the scan number, information pertaining to the identity and localization of any PTMs, p value, and e -value, among others (Supplemental Fig. S5). The identified proteins can be individually selected to view the protein sequence information, along with the localization of any mass discrepancies, in

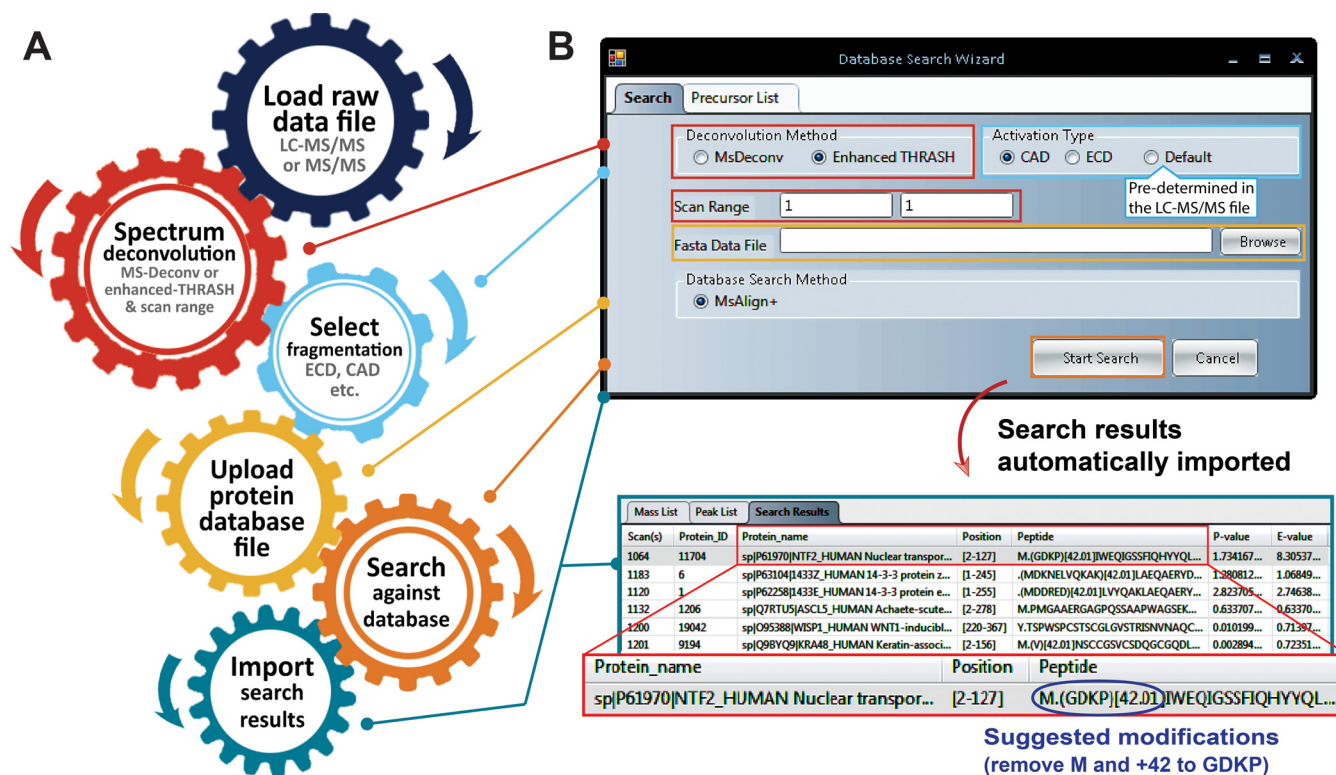


FIG. 3. **MASH Suite Pro for top-down protein identification.** (A) The workflow for protein identification in MASH Suite Pro. The raw data (e.g. .raw file from the instruments of Thermo Scientific) can be directly read by the program and all of the steps for protein identification can be performed in an integrated search wizard. (B) The highly integrated search wizard. MASH Suite Pro accommodates various options for deconvolution and database searching. The search results are automatically imported to the program for easy review and validation of the identification results.

the *Sequence Table*. In addition, bond cleavages within the sequence are also shown (Fig. 2, [Supplemental Fig. S3](#)).

It should be noted that the accuracy of the fragment ion monoisotopic masses resulting from deconvolution directly affects the identification results. To improve the accuracy of protein identification, MASH Suite Pro accommodates different algorithms for spectral deconvolution. Currently, MS-Deconv and enhanced-THRASH have been successfully implemented. Protein identification using enhanced-THRASH and MS-Deconv for deconvolution ([Supplemental Fig. S6](#)) reveals that the two methods yield similar results for the same dataset. For proteins that are identified with high confidence (e-value < 1E-30), MS-Deconv usually results in smaller e-values ([Supplemental Fig. S6](#)), indicating higher confidence, compared with enhanced-THRASH. When the e-values are between 1E-10 and 1E-30, the reported e-values/p values from MS-Deconv and enhanced-THRASH are similar ([Supplemental Fig. S6](#)). When the e-values are above 1E-10, the PrSMs are considered less confident. In these cases, enhanced-THRASH performs better than MS-Deconv ([Supplemental Fig. S6](#)). Nevertheless, both deconvolution methods can be optimized by adjusting parameters such as S/N, mass tolerance, and/or fit scores, depending on the datasets to be analyzed.

MASH Suite Pro for Top-Down Protein Quantitation—MASH Suite Pro significantly simplifies and accelerates the

quantitation of various proteoforms from top-down mass spectra and can present the quantitation results in both table and chart formats. Unlike the previous version (36), MASH Suite Pro also allows for the quantitation of several proteoforms from mass spectra using multiple charge states (note that quantitation is based on S/N values of the most abundant isotopomers from the mass spectra rather than the tandem mass spectra). When multiple charge states are selected for quantitation, the program automatically normalizes the proteoform abundance to the charge state to ensure accurate quantitation. In addition, it enables rapid quantitation and comparison of the selected proteoforms between different biological/experimental conditions (e.g. healthy *versus* diseased). Furthermore, it provides the option for averaging of the quantitation results among replicates, allowing accurate quantitation with minimal manual intervention.

The workflow for quantitation in MASH Suite Pro is outlined in Fig. 4A. Users can upload several data files and define the ions to be quantitated. For each raw data file, MASH Suite Pro calculates the relative abundances of the selected ions by summing the S/N values of the most abundant isotopomers. Users can also define the number of charge states to be calculated and the number of isotopomers of each charge state to be considered in the quantitation. Once the target ions are defined, the program will automatically search and

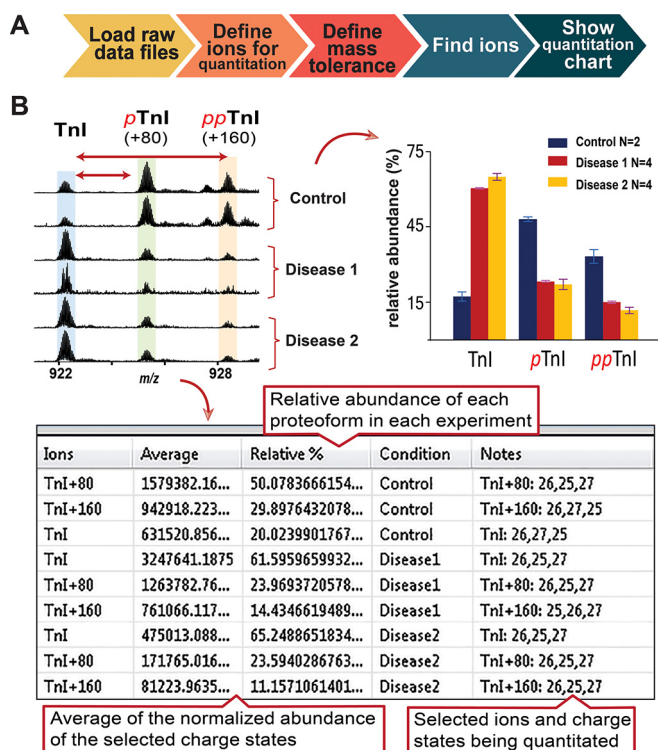


FIG. 4. MASH Suite Pro for top-down protein quantitation. (A) The workflow for protein quantitation in MASH Suite Pro. Several raw data files corresponding to the same experimental/biological condition can be quantitated and averaged to ensure accurate quantitation. The relative abundances of proteoforms in different conditions can also be directly compared in MASH Suite Pro. (B) Quantitation of Troponin I (TnI) proteoforms in cardiac left ventricular tissue from healthy (Control) and disease (Disease 1 and Disease 2) pigs. The Control, Disease 1, and Disease 2 conditions have two, four, and four biological replicates, respectively. The relative abundance of TnI, mono-phosphorylated TnI (pTnI), and bis-phosphorylated TnI (ppTnI) were determined using MASH Suite Pro. The quantitation chart provides a visual comparison of the relative percentage of each proteoform in different groups, and the quantitation table shows the exact relative percentage of the each proteoform in each experiment (only one replicate from each condition is shown). A complete quantitation table for all the experiments analyzed is shown in [Supplemental Fig. S8](#).

carry out quantitation of the selected ions, generating a quantitation table and quantitation chart (Fig. 4). Detailed procedures on the protein quantitation involving multiple charge states are outlined in [Supplemental Fig. S7](#).

An example of quantitation in MASH Suite Pro is shown in Fig. 4B. The relative quantities of troponin I (TnI) proteoforms, including un-phosphorylated TnI, mono-phosphorylated TnI (pTnI), and bis-phosphorylated TnI (ppTnI), in the left ventricular tissue of control swine (Control) and swine with myocardial infarction (Disease) were compared. Two groups of samples from the diseased animals were analyzed (Disease 1 and Disease 2) (Fig. 4B). For quantitation, the S/N ratios of the most abundant isotopomers within each isotopomer envelope were summed, and the resulting values were normalized to

charge state and defined as the normalized abundance of the selected ions. Once the proteoforms and charge state(s) are defined, MASH Suite Pro automatically normalizes the proteoform abundance to the corresponding charge states, finds the ions, and calculates the relative percentage of each proteoform. MASH Suite Pro can also average the relative abundance of the same proteoform in different biological replicates and presents the quantitation chart with error bars (standard error of the mean). Most importantly, users can directly compare the relative abundance of various proteoforms in different biological conditions (e.g. healthy *versus* diseased) (Fig. 4B). MASH Suite Pro also provides users with a quantitation table showing the exact relative abundance of each proteoform in each experiment (Fig. 4B, [Supplemental Fig. S8](#)). The quantitation feature in MASH Suite Pro significantly facilitates the analysis of molecular changes in different biological/experimental conditions.

MASH Suite Pro for Comprehensive Protein Characterization—MASH Suite Pro can facilitate the characterization of PTMs and sequence variations by allowing assignment of modifications to individual amino acids in the protein sequence (Fig. 5). Once the protein is identified, the amino acid sequence will be displayed in the “Sequence Table” window with bond cleavages (from the tandem mass spectrum) indicated. The identified PTMs are also automatically incorporated into the protein sequence to provide more accurate assignment of the fragment ions. For example, human nuclear transport factor 2 was identified in our LC-MS/MS experiment (39) using MS-Align+, and the identification results indicate that the N-terminal Met is removed and Gly1 is acetylated (Fig. 5). Without taking into account acetylation at the N terminus, only 5 *b* ions and 38 *y* ions could be assigned (Fig. 5A); however, when N-terminal acetylation was taken into account, an additional 36 *b* ions were assigned (41 *b* ions and 38 *y* ions total). MASH Suite Pro also allows users to manually change the modifications. Moreover, users can choose to display different types of fragment ions, including *b/y* and *c/z* pairs, as well as *b/y* ions with water loss (-H₂O) and neutral ammonium loss (-NH₃) ([Supplemental Fig. S9](#)) and also visualize the fragment ions to validate and refine the fragment ion assignments (Fig. 5C).

MASH Suite Pro for Identification, Characterization, and Quantitation of an Unknown Protein: An Example—To demonstrate the capabilities of MASH Suite Pro, we fraction collected an unknown protein that we observed during an LC-MS run. The fraction-collected protein was subsequently analyzed offline by high-resolution MS (Fig. 6A). High-resolution MS analysis showed that the unknown protein has multiple proteoforms (P1, P2, and P3) and that the relative abundances of these proteoforms change under different experimental conditions (Condition 1 and Condition 2) (Fig. 6A). The most abundant proteoform (P2) was selected for fragmentation using electron capture dissociation, and the tandem

A Display fragmentation maps

```

1  G D K P I W E Q I G S S F I Q H Y Y 109
19 Q L F D N D R T Q L G A I Y I D A S 91
37 C L T W E L G Q Q F L Q G K A A I V E K 73
55 L S S L P F Q K I Q H S I T A Q D H 55
73 Q P T P D S C I I S M V V G Q L K A 37
91 D E D P I M G F H Q M F L L K N I N 19
109 D A W V C L T N D M F R L A L L H N F G 1

```

[2-127] sp|P61970|NTF2_HUMAN Nuclear transport factor 2

**b ions: 5, y ions: 38****B Incorporate suggested modifications****+ 42 (Acetylation)**

```

1  G D K P I W E Q I G S S F I Q H Y Y 109
19 Q L F D N D R T Q L G A I Y I D A S 91
37 C L T W E L G Q Q F L Q G K A A I V E K 73
55 L S S L P F Q K I Q H S I T A Q D H 55
73 Q P T P D S C I I S M V V G Q L K A 37
91 D E D P I M G F H Q M F L L K N I N 19
109 D A W V C L T N D M F R L A L L H N F G 1

```

[2-127] sp|P61970|NTF2_HUMAN Nuclear transport factor 2

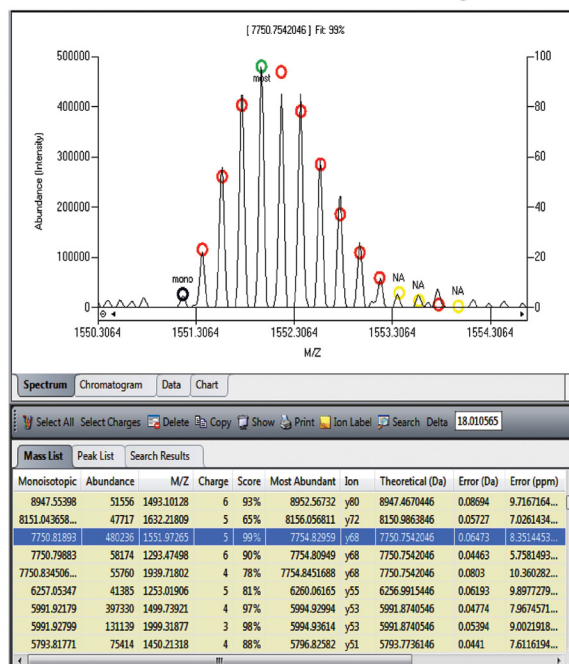
b ions: 41, y ions: 38**C Visualization and validation of the fragment ions**

FIG. 5. MASH Suite Pro for comprehensive characterization of protein PTMs and sequence variations. (A) Once the protein is identified (as shown in Fig. 2), the protein sequence can be displayed in the “Sequence Table” showing bond cleavages. (B) Suggested PTMs from MS-Align+ can be incorporated into the protein sequence. (C) Users are also able to visualize and validate the fragment ions of each identified protein in the “Mass List.”

mass spectra were highly complex (Fig. 6B), necessitating a sophisticated tool for data processing and analysis. Spectral deconvolution, protein identification, and characterization were performed in MASH Suite Pro, which identified the protein as a truncated form of ATPase inhibitor (Fig. 6C). MASH Suite Pro can also identify unexpected PTMs and sequence variations, facilitating protein characterization. In this case, the ATPase inhibitor has one unexpected sequence variation at position 37 (Ala37Val). Once P2 was identified, according to the protein sequence, the other proteoforms (P1 and P3) were subsequently deduced (Fig. 6C). The relative abundance of each proteoform under two experimental conditions was also determined by MASH Suite Pro and presented with a quantitation table for visual comparison (Fig. 6D). This demonstrates the power of MASH Suite Pro for proteoform identification, comprehensive characterization, and quantitative analysis.

Data Visualization and Validation—For very complex spectra, spectral deconvolution methods sometimes fail to accurately deconvolute the isotopomer envelopes, resulting in miscalculation of the mass value or charge state (Fig. 7, Supplemental Fig. S10). Charge state miscalculation and mass shifts are common primarily due to overlapping isotopomer envelopes (Fig. 7). For high molecular weight ions and high charge state ions, even without the involvement of overlapping peaks, the current deconvolution methods can still result in charge state calculation errors and mass shifts.

MASH Suite Pro provides users with visualization features for validation of the computational outputs. Users are able to visualize the spectrum of each ion assignment together with the automatically generated theoretical isotopomer envelope (based on averagine) in the *Spectrum View* (Supplemental Fig. S11). MASH Suite Pro allows users to correct the charge states (Fig. 7A) and shift the theoretical isotopomer envelope incrementally by 1 Dalton to fit the real spectrum (Fig. 7B) in order to obtain an accurate mass value for the selected ion (Fig. 7A) (36).

Interface Customization and Table/Graphic Output—Similar to the previous MASH Suite (36), MASH Suite Pro provides users with flexibility in the arrangement of various windows and tabs in the program interface (Fig. 2, Supplemental Fig. S12). Users can show or hide any window by pinning/unpinning it. The pinned windows will be shown in the program, and the unpinned windows will be hidden. When users hover over the hidden window, the window will toggle and appear temporarily, allowing the user to repin the window if desired. MASH Suite Pro arranges different views in the form of tabs, allowing users to switch between tabs to access different data or parameters (Supplemental Fig. S12). For example, as shown in Supplemental Fig. S2, users can switch between the “Spectrum” and “Chromatogram” to visualize different aspects of the data. Moreover, each window can be relocated by dragging the window to a different position, which allows for easy access to different features throughout the process

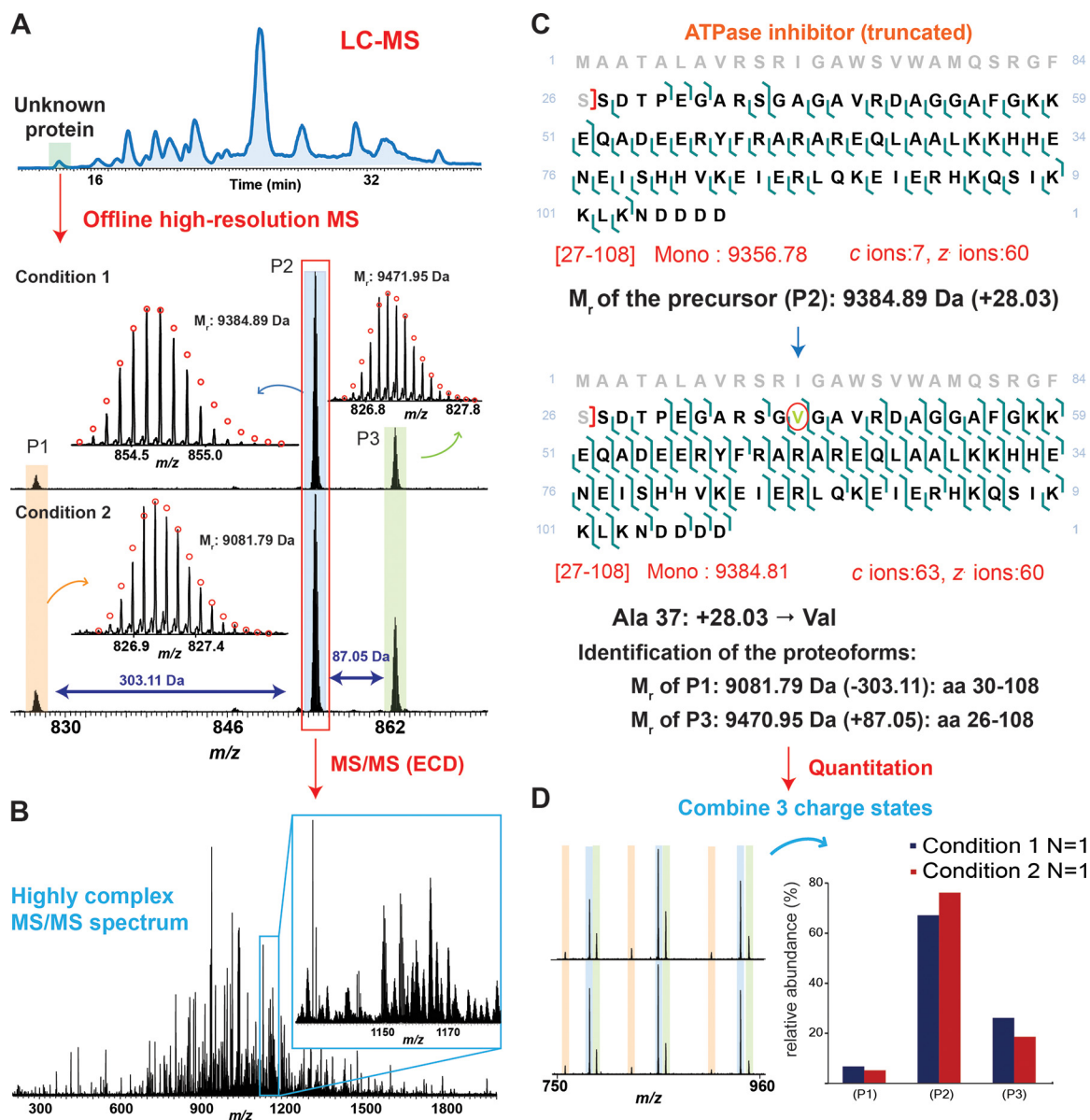


FIG. 6. Identification and characterization of an unknown protein followed by quantitation of the proteoforms using MASH Suite Pro. (A) An unknown protein was detected in LC-MS and the fraction containing the protein was collected and analyzed by high-resolution MS. Top-down MS revealed three proteoforms (P1, P2 and P3). The relative abundance of each proteoform varied in two different experimental conditions. (B) The proteoform (P2) was selected for fragmentation using electron capture dissociation, resulting in a highly complex MS/MS spectrum. (C) MASH Suite Pro was used for spectral deconvolution and protein identification. The proteoform P2 was identified to be a truncated form of the ATPase inhibitor containing amino acids (aa) 27–108. Characterization of the protein sequence using MASH Suite Pro identified a sequence variation (Ala37Val). Based on the protein sequence, the proteoforms (P1) and (P3) were deduced to be aa 30–108 and aa 26–108 of ATPase inhibitor. (D) MASH Suite Pro provided rapid quantitation to determine the relative abundances of the different proteoforms in different experimental conditions.

of data analysis. To facilitate reporting of the analyzed data, MASH Suite Pro also supports direct output of the tables and graphics (Supplemental Fig. S13), enabling further analysis of the data and editing/publishing of the figures.

DISCUSSION

Top-down proteomics is increasingly recognized as the premier method for the comprehensive analysis of proteo-

forms (12–24, 40–44). Nevertheless, the implementation and practice of the top-down approach still face significant challenges, including solubilizing hydrophobic proteins, separating intact proteins in complex mixtures, detecting large (>50 kDa) and low-abundance proteins, and analyzing highly complex high-resolution mass and tandem mass spectra (13). Recent advancements, such as the development of MS-compatible surfactants for solubilizing hydrophobic proteins (45),

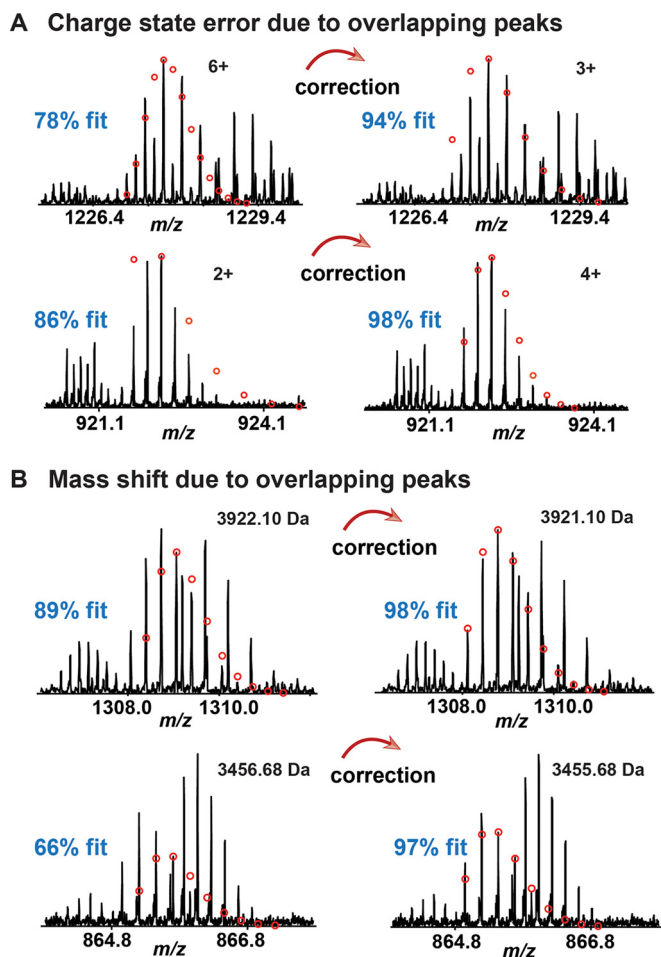


FIG. 7. **Visual validation of the computational outputs.** (A) Charge state correction. Users can adjust/correct the charge state to better fit the data and acquire accurate mass values. (B) Mass shift correction. Users can shift the theoretical isotopomer envelope incrementally by 1 Dalton to fit the real spectrum for obtaining accurate mass values.

multidimensional LC separation methodologies for the separation of complex protein mixtures (39, 46), and functionalized nanoparticles for phosphoprotein enrichment (47), have begun to address these challenges. Nevertheless, challenges still remain in terms of analyzing highly complex top-down LC-MS and LC-MS/MS data. Herein, we have developed MASH Suite Pro to address the difficulties in analyzing complex spectra arising from LC-MS and LC-MS/MS experiments in large-scale high-resolution MS-based top-down proteomics studies. It is the first comprehensive software package that fully integrates spectral deconvolution, top-down protein identification, quantitation, and characterization, together with visual validation, into a user-friendly customizable interface.

Spectral deconvolution is a critical step in top-down data analysis because the accuracy of the deconvolution results directly affects the downstream data interpretation. MASH Suite Pro currently has accommodated two commonly used algorithms, MS-Deconv and THRASH, for the deconvolution

of high-resolution mass and tandem mass spectra. These two algorithms use different methods to evaluate an experimental isotopomer envelope by matching it to the theoretical isotopomer envelope generated based on averagine (48): THRASH computes a figure of merit value that measures the similarity of peak intensities while MS-Deconv uses a scoring function that involves intensities and m/z values of peaks. In addition, these two algorithms employ different approaches for envelope selection: a greedy approach by THRASH (25) and a dynamic programming approach by MS-Deconv (26). Specifically, THRASH reports all identified masses with a reliability value above a user-specified threshold (referred to as a fit score in MASH Suite Pro), whereas MS-Deconv determines the number of masses reported from a spectrum by estimating the length of the target protein using the precursor mass. Therefore, MS-Deconv is relatively conservative in selecting and reporting isotopomer envelopes. As a consequence, MS-Deconv usually reports fewer overall masses than THRASH but also fewer false positives. It is important to note that the statistical evaluation for PrSMs takes into account both the matched and unmatched masses, which offers an accurate estimation of the confidence of protein identification. For large-scale top-down proteomics studies, the majority of identified proteins have e -values between $1E-30$ and $1E-10$ (39, 46). In this confidence range, MS-Deconv and THRASH yield very similar e -values (usually within three orders of magnitude) for the same PrSMs. For tandem mass spectra with relatively high complexity (more product ions), low noise level, and high signal intensity, high confidence in protein identification (e -values $< 1E-30$) can be achieved using either MS-Deconv or THRASH (Supplemental Fig. S6). However, under these conditions, MS-Deconv usually results in better (lower) e -values than THRASH due to the fact that more stringent criteria are used for envelope reporting in MS-Deconv (*i.e.* fewer false positives will be produced leading to fewer unmatched masses and an overall increase in the confidence of identification). On the other hand, for tandem mass spectra with relatively low complexity (fewer product ions), high noise level, and low signal intensity, the identified PrSMs are generally less confident (e -values $> 1E-10$) when deconvolution is performed using either MS-Deconv or THRASH. Yet, in these cases, THRASH tends to yield better e -values than MS-Deconv due to the fact that more masses will be matched to the sequence (Supplemental Fig. S6) as a consequence of the lower stringency for envelope reporting. Therefore, as different deconvolution algorithms may perform better for different datasets, the choice of deconvolution method should be decided on a case-by-case basis. It should also be noted that the use of appropriate parameter settings (*e.g.* S/N), regardless of the algorithm used, is also important for accurate and reliable spectral deconvolution. Besides MS-Deconv and THRASH, additional deconvolution algorithms can also be added to MASH Suite Pro with the permission of the developers. For example, the recently developed UniDec (27) can

perform mass and charge separation of complex spectra, which can also be used for the deconvolution of high-resolution MS data.

Due to the complexity of high-resolution top-down mass and tandem mass spectra, which contain many overlapping isotopomer envelopes, currently available deconvolution methods tend to generate a considerable amount of misassigned peaks (Fig. 7, Supplemental Fig. S10). Peak assignment errors can prevent confident identification of the proteins, as well as accurate localization of PTMs and sequence variations. Therefore, visual validation and manual correction of the automatically processed top-down MS data are necessary to reduce false positives and correct mass shift and charge state errors if they occur. There are software packages that provide visual components for comparing the experimental isotopomer envelopes with theoretical isotopomer envelopes generated using averagine (48). For example, Decon2LS uses a variation of THRASH for high-resolution spectral analysis and visualization (38). DataAnalysis (Bruker Daltonics) utilizes SNAP2, a THRASH-based algorithm, for spectral deconvolution and allows for direct visual comparison of the experimental isotopomer envelopes with theoretical envelope patterns (termed SNAP pattern). Nevertheless, these software packages do not allow for manual correction of the computational outputs. In contrast, MASH Suite Pro provides easy access to the real spectra and visualization of each deconvoluted mass, with additional features allowing users to delete false positives and correct the mass values or charge states when mistakes occur. The visualization features and flexibility for data correction make MASH Suite Pro advantageous in analyzing top-down spectra with high accuracy and reliability.

MASH Suite Pro utilizes MS-Align+ algorithm (32) for protein identification. As demonstrated previously (14, 19, 49, 50), MS-Align+ is highly effective for the identification of proteins and unexpected PTMs from top-down tandem mass spectra. A previous study has shown that MS-Align+ completes matching 1000 spectra in 18 min *versus* 22 min using the biomarker search mode of ProSightPC (32). This corresponds to an improvement of ~20% in the speed of protein identification. In addition, the time needed for ProSightPC to complete the search in the advanced search mode (searching against annotated top-down database) was an order of magnitude longer (32). Moreover, MS-Align+ can identify unexpected PTMs whereas ProSightPC cannot. This demonstrated the advantages of MS-Align+ in protein identification. By incorporating MS-Align+, we have developed a user-friendly interface for protein identification in top-down proteomics with high simplicity and speed. Instead of operating protein identification via command lines in the previous MS-Align+, users are able to finish all the necessary steps for protein identification in a single intuitive platform. In addition to MS-Align+, MASH Suite Pro can also accommodate other search algorithms in the future, such as the very recently developed

TopPIC (<http://proteomics.informatics.iupui.edu/software/toppic/index.html>) and MSPathFinder (<http://omics.pnl.gov/software/mspathfinder>). Similar to MS-Align+, TopPIC identifies proteoforms with unexpected sequence variations and PTMs and estimates statistical confidence of the identified PrSMs. TopPIC improves on MS-Align+ in terms of search efficiency and computing requirements. MSPathFinder also employs spectral alignment algorithm for top-down protein identification. It requires the users to input the spectrum file, a protein sequence file (.fasta), and a list of potential modifications. MSPathFinder matches the spectra against the user-specified protein sequences containing the defined modifications and reports the search results as PrSMs with their scores, which indicates the degree of confidence in the identification. When considering only one modification (acetylation), MSPathFinder is faster than MS-Align+. However, MSPathFinder only includes the user-defined modifications, so it is somewhat limited for the identification of unexpected PTMs. Therefore, MS-Align+ is favored for large-scale proteomics studies and PTM discovery.

MASH Suite Pro enables rapid and accurate relative quantitation of proteoforms in different biological/experimental conditions and, thus, has great potential in the dissection of disease-associated proteoform alterations (13, 19, 22–24, 40–42, 51, 52). MASH Suite Pro allows for the quantitation of different proteoforms with multiple charge states from multiple top-down spectrum data files. The program also automatically generates a quantitation table and chart for a direct comparison of the relative abundances of various proteoforms in different conditions. When disease-related changes in the abundance of proteoforms are observed, MASH Suite Pro can also facilitate the characterization of the proteoforms, including the localization of sequence alterations and PTMs (Fig. 6). This highlights the potential of MASH Suite Pro in unraveling proteoform-associated disease mechanisms.

Regarding the data format, MASH Suite Pro can directly process and analyze the raw data format (.raw) generated by the instruments of Thermo Scientific without conversion. Furthermore, MASH Suite Pro can also process mzXML files (53), which are an open data format for the storage and exchange of MS data. Proprietary file formats from most vendors can be converted to the open mzXML format, allowing the use of MASH Suite Pro for processing and analyzing data files generated from the instruments of most vendors.

In summary, MASH Suite Pro is a comprehensive, user-friendly, and freely available program tailored for large-scale top-down proteomics data analysis, including spectral deconvolution, protein identification, quantitation, and characterization. Distinguished from the previous version for single protein characterization, MASH Suite Pro has implemented important new features for large-scale proteomic data analysis, including the ability to process LC-MS and LC-MS/MS data, identification of unknown proteins and PTMs, as well as multiplex proteoform quantitation involving various biological/experi-

mental conditions. MASH Suite Pro also allows for visual validation and correction of the computational outputs to ensure the accuracy of data interpretation. With a user-friendly and customizable interface, MASH Suite Pro greatly simplifies and speeds up the interpretation of high-resolution data and, therefore, will play an integral role in advancing the field of top-down proteomics.

Acknowledgments—We thank David Horn, Ziqing Lin, Tania M. Guardado-Alvarez, and Yutong Jin for helpful discussions and Nicole Lane for critical reading of this manuscript.

* This work was supported by NIH R01HL096971 and R01HL109810 (to Y.G.). We acknowledge American Heart Association Scientist Development Grant 0735443Z and the Wisconsin Partnership Fund for the establishment of the Human Proteomics Program Mass Spectrometry Facility. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

§ This article contains [supplemental material](#) Supplemental Figs. S1-S13.

¶¶ To whom correspondence should be addressed: 1300 University Ave., SMI 130, Madison, WI, 53706. Tel.: 608-263-9212; Fax: 608-265-5512; E-mail: ge2@wisc.edu.

§§ These two authors contributed equally.

REFERENCES

- Gehlenborg, N., O'Donoghue, S. I., Baliga, N. S., Goesmann, A., Hibbs, M. A., Kitano, H., Kohlbacher, O., Neuweger, H., Schneider, R., Tenenbaum, D., and Gavin, A. C. (2010) Visualization of omics data for systems biology. *Nat. Methods* **7**, S56–S68
- Cox, J., and Mann, M. (2011) Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **80**, 273–299
- Larance, M., and Lamond, A. I. (2015) Multidimensional proteomics for cell biology. *Nat. Rev. Mol. Cell Biol.* **16**, 269–280
- Bensimon, A., Heck, A. J., and Aebersold, R. (2012) Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* **81**, 379–405
- Yates, J. R., Ruse, C. I., and Nakorchevsky, A. (2009) Proteomics by mass spectrometry: Approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **11**, 49–79
- Nesvizhskii, A. I., Vitek, O., and Aebersold, R. (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods* **4**, 787–797
- Cravatt, B. F., Simon, G. M., and Yates, J. R., 3rd (2007) The biological impact of mass-spectrometry-based proteomics. *Nature* **450**, 991–1000
- Huang, P. H., and White, F. M. (2008) Phosphoproteomics: Unraveling the signaling web. *Mol. Cell* **31**, 777–781
- Miteva, Y. V., and Cristea, I. M. (2014) A proteomic perspective of Sirtuin 6 (SIRT6) phosphorylation and interactions and their dependence on its catalytic activity. *Mol. Cell. Proteomics* **13**, 168–183
- Chait, B. T. (2006) Chemistry. Mass spectrometry: Bottom-up or top-down? *Science* **314**, 65–66
- Smith, L. M., Kelleher, N. L., and Proteomics, C. f. T. D. (2013) Proteoform: A single term describing protein complexity. *Nat. Methods* **10**, 186–187
- Siuti, N., and Kelleher, N. L. (2007) Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **4**, 817–821
- Gregorich, Z. R., and Ge, Y. (2014) Top-down proteomics in health and disease: Challenges and opportunities. *Proteomics* **14**, 1195–1210
- Ansong, C., Wu, S., Meng, D., Liu, X., Brewer, H. M., Deatherage Kaiser, B. L., Nakayasu, E. S., Cort, J. R., Pevzner, P., Smith, R. D., Heffron, F., Adkins, J. N., and Pasa-Tolic, L. (2013) 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.* **110**, 10153–10158
- Savaryn, J. P., Catherman, A. D., Thomas, P. M., Abecassis, M. M., and Kelleher, N. L. (2013) The emergence of top-down proteomics in clinical research. *Genome Med.* **5**, 53
- Zhang, Z., Wu, S., Stenoien, D. L., and Paša-Tolić, L. (2014) High-throughput proteomics. *Annu. Rev. Anal. Chem.* **7**, 427–454
- Zhang, H., and Ge, Y. (2011) Comprehensive analysis of protein modifications by top-down mass spectrometry. *Circ. Cardiovasc. Genet.* **4**, 711
- Tran, J. C., Zamdborg, L., Ahlf, D. R., Lee, J. E., Catherman, A. D., Durbin, K. R., Tipton, J. D., Vellaichamy, A., Kellie, J. F., Li, M., Wu, C., Sweet, S. M., Early, B. P., Siuti, N., LeDuc, R. D., Compton, P. D., Thomas, P. M., and Kelleher, N. L. (2011) Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254–258
- Peng, Y., Gregorich, Z. R., Valeja, S. G., Zhang, H., Cai, W., Chen, Y.-C., Guner, H., Chen, A. J., Schwahn, D. J., Hacker, T. A., Liu, X., and Ge, Y. (2014) Top-down proteomics reveals concerted reductions in myofibrillar and Z-disc protein phosphorylation after acute myocardial infarction. *Mol. Cell. Proteomics* **13**, 2752–2764
- Whitelegge, J. (2013) Intact protein mass spectrometry and top-down proteomics. *Expert Rev. Proteomics* **10**, 127–129
- Ryan, C. M., Souda, P., Bassilian, S., Ujwal, R., Zhang, J., Abramson, J., Ping, P., Durazo, A., Bowie, J. U., Hasan, S. S., Baniulis, D., Cramer, W. A., Faull, K. F., and Whitelegge, J. P. (2010) Post-translational modifications of integral membrane proteins resolved by top-down Fourier transform mass spectrometry with collisionally activated dissociation. *Mol. Cell. Proteomics* **9**, 791–803
- Mazur, M. T., Cardasis, H. L., Spellman, D. S., Liaw, A., Yates, N. A., and Hendrickson, R. C. (2010) Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 7728–7733
- Chamot-Rooke, J., Mikaty, G., Malosse, C., Soyer, M., Dumont, A., Gault, J., Imhaus, A. F., Martin, P., Trellet, M., Clary, G., Chafey, P., Camoin, L., Nilges, M., Nassif, X., and Duménil, G. (2011) Posttranslational modification of pili upon cell contact triggers *N. meningitidis* dissemination. *Science* **331**, 778–782
- Dang, X., Scotcher, J., Wu, S., Chu, R. K., Tolić, N., Ntai, I., Thomas, P. M., Fellers, R. T., Early, B. P., Zheng, Y., Durbin, K. R., Leduc, R. D., Wolff, J. J., Thompson, C. J., Pan, J., Han, J., Shaw, J. B., Salisbury, J. P., Easterling, M., Borchers, C. H., Brodbelt, J. S., Agar, J. N., Paša-Tolić, L., Kelleher, N. L., and Young, N. L. (2014) The first pilot project of the consortium for top-down proteomics: A status report. *Proteomics* **14**, 1130–1140
- Horn, D. M., Zubarev, R. A., and McLafferty, F. W. (2000) Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* **11**, 320–332
- Liu, X., Inbar, Y., Dorrestein, P. C., Wynne, C., Edwards, N., Souda, P., Whitelegge, J. P., Bafna, V., and Pevzner, P. A. (2010) Deconvolution and database search of complex tandem mass spectra of intact proteins: A combinatorial approach. *Mol. Cell. Proteomics* **9**, 2772–2782
- Marty, M. T., Baldwin, A. J., Marklund, E. G., Hochberg, G. K., Benesch, J. L., and Robinson, C. V. (2015) Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376
- LeDuc, R. D., Taylor, G. K., Kim, Y. B., Januszky, T. E., Bynum, L. H., Sola, J. V., Garavelli, J. S., and Kelleher, N. L. (2004) ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res.* **32**, W340–W345
- Zamdborg, L., LeDuc, R. D., Glowacz, K. J., Kim, Y. B., Viswanathan, V., Spaulding, I. T., Early, B. P., Bluhm, E. J., Babai, S., and Kelleher, N. L. (2007) ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* **35**, W701–W706
- Karabacak, N. M., Li, L., Tiwari, A., Hayward, L. J., Hong, P., Easterling, M. L., and Agar, J. N. (2009) Sensitive and specific identification of wild type and variant proteins from 8 to 669 kDa using top-down mass spectrometry. *Mol. Cell. Proteomics* **8**, 846–856
- Frank, A. M., Pesavento, J. J., Mizzen, C. A., Kelleher, N. L., and Pevzner, P. A. (2008) Interpreting top-down mass spectra using spectral alignment. *Anal. Chem.* **80**, 2499–2505
- Liu, X., Sirotkin, Y., Shen, Y., Anderson, G., Tsai, Y. S., Ting, Y. S., Goodlett, D. R., Smith, R. D., Bafna, V., and Pevzner, P. A. (2012) Protein identification using top-down. *Mol. Cell. Proteomics* **11**, M111.008524
- Pesavento, J. J., Kim, Y. B., Taylor, G. K., and Kelleher, N. L. (2004) Shotgun annotation of histone modifications: A new approach for

- streamlined characterization of proteins by top down mass spectrometry. *J. Am. Chem. Soc.* **126**, 3386–3387
34. Mazur, M. T., and Fyhr, R. (2011) An algorithm for identifying multiply modified endogenous proteins using both full-scan and high-resolution tandem mass spectrometry data. *Rapid Commun. Mass Spectrom.* **25**, 3617–3626
 35. Pevzner, P. A., Dancik, V., and Tang, C. L. (2000) Mutation-tolerant protein identification by mass spectrometry. *J. Comput. Biol.* **7**, 777–787
 36. Guner, H., Close, P. L., Cai, W., Zhang, H., Peng, Y., Gregorich, Z. R., and Ge, Y. (2014) MASH Suite: A user-friendly and versatile software interface for high-resolution mass spectrometry data interpretation and visualization. *J. Am. Soc. Mass Spectrom.* **25**, 464–470
 37. Fellers, R. T., Greer, J. B., Early, B. P., Yu, X., LeDuc, R. D., Kelleher, N. L., and Thomas, P. M. (2014) ProSight Lite: Graphical software to analyze top-down mass spectrometry data. *Proteomics* **15**, 1253–1238
 38. Jaitly, N., Mayampurath, A., Littlefield, K., Adkins, J. N., Anderson, G. A., and Smith, R. D. (2009) Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics* **10**, 87
 39. Valeja, S. G., Xiu, L., Gregorich, Z. R., Guner, H., Jin, S., and Ge, Y. (2015) Three dimensional liquid chromatography coupling ion exchange chromatography/hydrophobic interaction chromatography/reverse phase chromatography for effective protein separation in top-down proteomics. *Anal. Chem.* **87**, 5363–5371
 40. Zhang, J., Guy, M. J., Norman, H. S., Chen, Y. C., Xu, Q., Dong, X., Guner, H., Wang, S., Kohmoto, T., Young, K. H., Moss, R. L., and Ge, Y. (2011) Top-down quantitative proteomics identified phosphorylation of cardiac troponin I as a candidate biomarker for chronic heart failure. *J. Proteome Res.* **10**, 4054–4065
 41. Dong, X., Sumandea, C. A., Chen, Y. C., Garcia-Cazarin, M. L., Zhang, J., Balke, C. W., Sumandea, M. P., and Ge, Y. (2012) Augmented phosphorylation of cardiac troponin I in hypertensive heart failure. *J. Biol. Chem.* **287**, 848–857
 42. Peng, Y., Ayaz-Guner, S., Yu, D., and Ge, Y. (2014) Top-down mass spectrometry of cardiac myofilament proteins in health and disease. *Proteomics Clin. Appl.* **8**, 554–568
 43. Ge, Y., Rybakova, I. N., Xu, Q., and Moss, R. L. (2009) 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.* **106**, 12658–12663
 44. Chen, Y. C., Ayaz-Guner, S., Peng, Y., Lane, N. M., Locher, M. R., Kohmoto, T., Larsson, L., Moss, R. L., and Ge, Y. (2015) Effective top-down LC/MS+ method for assessing actin isoforms as a potential cardiac disease marker. *Anal. Chem.* **87**, 8399–8406
 45. Chang, Y. H., Gregorich, Z. R., Chen, A. J., Hwang, L., Guner, H., Yu, D., Zhang, J., and Ge, Y. (2015) New mass-spectrometry-compatible degradable surfactant for tissue proteomics. *J. Proteome Res.* **14**, 1587–1599
 46. Xiu, L., Valeja, S. G., Alpert, A. J., Jin, S., and Ge, Y. (2014) Effective protein separation by coupling hydrophobic interaction and reverse phase chromatography for top-down proteomics. *Anal. Chem.* **86**, 7899–7906
 47. Hwang, L., Ayaz-Guner, S., Gregorich, Z. R., Cai, W., Valeja, S. G., Jin, S., and Ge, Y. (2015) Specific enrichment of phosphoproteins using functionalized multivalent nanoparticles. *J. Am. Chem. Soc.* **137**, 2432–2435
 48. Senko, M. W., Beu, S. C., and McLafferty, F. W. (1995) Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass Spectrom.* **6**, 229–233
 49. Wu, S., Brown, R. N., Payne, S. H., Meng, D., Zhao, R., Tolić, N., Cao, L., Shukla, A., Monroe, M. E., Moore, R. J., Lipton, M. S., and Paša-Tolić, L. (2013) Top-down characterization of the post-translationally modified intact periplasmic proteome from the bacterium *Novosphingobium aromaticivorans*. *Int. J. Proteomics* **2013**, 279590
 50. Wu, S., Brown, J. N., Tolić, N., Meng, D., Liu, X., Zhang, H., Zhao, R., Moore, R. J., Pevzner, P., Smith, R. D., and Paša-Tolić, L. (2014) Quantitative analysis of human salivary gland-derived intact proteome using top-down mass spectrometry. *Proteomics* **14**, 1211–1222
 51. LaVoie, M. J., Ostaszewski, B. L., Weihofen, A., Schlossmacher, M. G., and Selkoe, D. J. (2005) Dopamine covalently modifies and functionally inactivates parkin. *Nat. Med.* **11**, 1214–1221
 52. Rakhit, R., Crow, J. P., Lepock, J. R., Kondejewski, L. H., Cashman, N. R., and Chakrabarty, A. (2004) Monomeric Cu, Zn-superoxide dismutase is a common misfolding intermediate in the oxidation models of sporadic and familial amyotrophic lateral sclerosis. *J. Biol. Chem.* **279**, 15499–15504
 53. Pedrioli, P. G., Eng, J. K., Hubley, R., Vogelzang, M., Deutsch, E. W., Raught, B., Pratt, B., Nilsson, E., Angeletti, R. H., Apweiler, R., Cheung, K., Costello, C. E., Hermjakob, H., Huang, S., Julian, R. K., Kapp, E., McComb, M. E., Oliver, S. G., Omenn, G., Paton, N. W., Simpson, R., Smith, R., Taylor, C. F., Zhu, W., and Aebersold, R. (2004) A common open representation of mass spectrometry data and its application to proteomics research. *Nat. Biotechnol.* **22**, 1459–1466