HHS Public Access

Author manuscript

J Proteome Res. Author manuscript; available in PMC 2020 December 11.

Published in final edited form as:

J Proteome Res. 2020 September 04; 19(9): 3867–3876. doi:10.1021/acs.jproteome.0c00469.

MASH Explorer: A Universal Software Environment for Top-Down Proteomics

Zhijie Wu,
Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin 53705, United States

David S. Roberts,
Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin 53705, United States

Jake A. Melby,
Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin 53705, United States

Kent Wenger,
Department of Cell and Regenerative Biology and Human Proteomics Program, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin 53705, United States

Molly Wetzel,
Department of Cell and Regenerative Biology, University of Wisconsin-Madison, Madison, Wisconsin 53705, United States

Yiwen Gu,
Department of Cell and Regenerative Biology and Human Proteomics Program, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin 53705, United States

Sudharshanan Govindaraj Ramanathan,
Department of Cell and Regenerative Biology, University of Wisconsin-Madison, Madison, Wisconsin 53705, United States

Elizabeth F. Bayne,

Corresponding Authors: sean.mcilwain@wisc.edu; ying.ge@wisc.edu.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00469.

Supplementary Results and Discussion with tables and figures (PDF)

MASH Explorer User Manual v2.0 (PDF)

MASH Video Part 1 Introduction (MP4)

MASH Video Part 2 Configuration Setup (MP4)

MASH Video Part 3 Discovery Mode (MP4)

MASH Video Part 4 Targeted Mode (MP4)

MASH Video Part 5 Data Processing and Export Functions (MP4)

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.0c00469

The authors declare no competing financial interest.

Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin 53705, United States

Xiaowen Liu,
Department of BioHealth Informatics and Center for Computational Biology and Bioinformatics, Indiana University-Purdue University Indianapolis, Indianapolis, Indiana 46202, United States

Ruixiang Sun,
National Institute of Biological Sciences, Beijing 102206, China

Irene M. Ong,
Department of Biostatistics and Medical Informatics, University of Wisconsin Carbone Cancer Center, and Department of Obstetrics and Gynecology, University of Wisconsin-Madison, Madison, Wisconsin 53705, United States

Sean J. McIlwain,
Department of Biostatistics and Medical Informatics and University of Wisconsin Carbone Cancer Center, University of Wisconsin-Madison, Madison, Wisconsin 53705, United States

Ying Ge
Department of Chemistry, Department of Cell and Regenerative Biology, and Human Proteomics Program, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin 53705, United States

Abstract

Top-down mass spectrometry (MS)-based proteomics enable a comprehensive analysis of proteoforms with molecular specificity to achieve a proteome-wide understanding of protein functions. However, the lack of a universal software for top-down proteomics is becoming increasingly recognized as a major barrier, especially for newcomers. Here, we have developed MASH Explorer, a universal, comprehensive, and user-friendly software environment for top-down proteomics. MASH Explorer integrates multiple spectral deconvolution and database search algorithms into a single, universal platform which can process top-down proteomics data from various vendor formats, for the first time. It addresses the urgent need in the rapidly growing top-down proteomics community and is freely available to all users worldwide. With the critical need and tremendous support from the community, we envision that this MASH Explorer software package will play an integral role in advancing top-down proteomics to realize its full potential for biomedical research.

Graphical Abstract

Keywords

Top-down Proteomics; Data Analysis Software; Proteoform Characterization; Intact Protein Analysis

INTRODUCTION

Top-down mass spectrometry (MS)-based proteomics provides a comprehensive analysis of "proteoforms"—all protein products arising from post-translational modifications (PTMs), alternative splicing, and genetic variations originating from a single gene—with molecular specificity to achieve a proteome-wide understanding of protein functions.[1–4] Top-down MS analyzes intact proteins without proteolytic digestion and can detect various proteoforms simultaneously in a single MS experiment, thereby enabling their comprehensive molecular characterization. Specific information about proteoforms including PTM sites and sequence variations can be further characterized by tandem MS (MS/MS).[5–7] In contrast to the well-developed software packages in the peptide-based bottom-up proteomics, the data analysis tools for protein-based top-down proteomics remain under-developed due to the major challenge in handling the enormous complexity of high-resolution intact protein mass spectra.[7–9] Particularly, the lack of a universal and user-friendly software for streamlined analysis of complex top-down proteomics data is becoming increasingly recognized as a major barrier, especially for newcomers, thus limiting the broader impact of top-down proteomics in the biomedical research communities. Additionally, the relatively high cost of commercial top-down software limits the accessibility for general users and thus necessitates a freely available academic version.

Here, we have developed MASH Explorer, a universal, comprehensive, user-friendly, and freely available software environment for top-down proteomics (http://ge.crb.wisc.edu/MASH_Explorer/index.htm). This software can process high-resolution MS, MS/MS, and liquid-chromatography tandem MS (LC-MS/MS) data across multiple vendor-specific formats, with automated database searching for protein identification as well as user-friendly tools for proteoform characterization and data visualization/validation. MASH Explorer includes two major workflows: "Discovery Mode" for analysis of complex high-resolution LC-MS/MS data to achieve global protein identification and "Targeted Mode" for comprehensive proteoform characterization including PTMs and sequence variants, with user-friendly graphic user interface (GUI) support. Advancing on our previous generations of proteomics software, MASH Suite[10] and MASH Suite Pro,[11] MASH Explorer has many

new features including (1) development of a universal platform for streamlined data processing from various vendor formats to standardize the data analysis; (2) integration of multiple deconvolution and database search algorithms for significantly enhanced protein identification; (3) workflow management for high-throughput data processing using the Process Wizard and Workflow Manager; (4) comprehensive proteoform characterization tools with the capability of handling highly complex data resulting from various MS/MS techniques such as collision-induced dissociation (CID), electron capture dissociation (ECD), electron transfer dissociation (ETD), and ultraviolet photodissociation (UVPD). The universal accessibility of nonproprietary, free software solutions such as MASH Explorer will significantly bolster the growth of the top-down proteomics community and welcome newcomers to employ this powerful technology to realize its impact in biomedical research.

EXPERIMENTAL SECTION

Software Design and Algorithm Support

MASH Explorer is a multithreaded Windows application implemented in C# using the .NET framework within the Visual Studio Integrated Development Environment. The software visual components are provided by Microsoft Office Runtime Support. Importing data obtained from different MS instruments is supported using ProteoWizard,[12] DeconEngine,[13] and vendor provided libraries. Additionally, MASH Explorer supports multiple deconvolution and database search algorithms, including TopPIC suite,[14] pTop,[15] Informed-Proteomics,[16] MS-Deconv,[17] MS-Align+,[18] and a modified version of THRASH[19] (eTHRASH[11]). As of March 24th, 2020, the supported versions of the deconvolution and database search algorithms are summarized in Table S1.

Computer Setup for Data Analysis

Data analysis was performed to simulate a basic research environment. This computer has Windows 10 Student Edition operating system installed. It was equipped with an Intel i5-2400 central processing unit, which has 4 cores and 4 threads, 16 GB DDR3 2400 MHz random access memory, and 1 TB SATA hard drive.

Mass Spectrometry Data

Two LC-MS/MS data sets from two different mass spectrometer vendors, Thermo Scientific and Bruker Corporation (referred to as Thermo and Bruker, respectively, in this manuscript), were utilized to demonstrate the Discovery Mode workflow of MASH Explorer. The Thermo data set is publicly available in the MassIVE repository with identifier/username MSV000079978 (ftp://massive.ucsd.edu/MSV000079978/).[20] The data set was acquired by extracting protein from DLD-1 parental (KRas wt/G13D) human colorectal cancer cells and using a GELFrEE system for size-based separation.[21] The MS experiment was performed using reverse-phase (RP) LC-MS/MS analysis using a 21 T Fourier Transform Ion Cyclotron Resonance mass spectrometer.

The Bruker LC-MS/MS data set used was publicly available from the PRIDE repository via ProteomeXchange with identifier PXD010825.[4] Briefly, the samples from this data set were prepared by protein extraction using a photocleavable surfactant, 4-hexylphenylazosulfonate

(Azo), from human embryonic kidney 293 K stem cells. The samples were irradiated to cleave the Azo surfactant. The RPLC-MS/MS experiment was performed on a Bruker maXis II quadrupole-time-of-flight (Q-TOF) mass spectrometer. For the Bruker data set, the mass spectra were also deconvoluted using a Maximum Entropy Algorithm with 80,000 resolution from 1000,00 Da to 50,000 Da using Bruker DataAnalysis ver. 4.3.

The data set for MS/MS analysis was previously published.[22] It is publicly available through ProteomeXchange Consortium via the PRIDE partner repository with the PXD018043 identifier. Briefly, the samples were prepared by extracting proteins from nonhuman primate skeletal muscle.[23] Target sarcomeric proteins were fractionated using a Waters nano-AQUITY liquid chromatography system, and the fractionated samples were analyzed with a Bruker solariX 12 T FT-ICR instrument using an Advion Nanomate. Specifically, beta-tropomyosin (βTpm, Uniprot-Swissprot accession number P07951) with the ECD spectrum and myosin light chain 2 slow isoform (MLC-2S, Uniprot-Swissprot accession number A0A1D5RDY5) with the CID spectrum were used for demonstration of top-down protein characterization using the “Targeted Mode” of MASH Explorer.

A Bruker MS/MS data set was used for demonstrating the functions of the Targeted Mode in MASH Explorer for characterization of the antibody–drug conjugate (ADC), Adcetris (brentuximab vedotin) subunits, as previously published and it is accessible through ProteomeXchange Consortium via the PRIDE partner repository with the PXD020615 identifier.[24] Briefly, Adcetris was digested by IdeS, and the interchain disulfide bond was reduced by dithiothreitol (DTT). The subunits were analyzed by LC-MS/MS using a combination of a Waters M-Class LC system and a Bruker maXis II Q-TOF mass spectrometer. The precursor of each subunit was subject to MS/MS experiments using both CID and ETD. The MS/MS spectra for each subunit were averaged using Bruker DataAnalysis ver. 4.3 software and exported in .ascii format. The ions were extracted using eTHRASH at 60% fit, and the fragment ions were manually validated.

The MS/MS data set for demonstrating UVPD ion fragments in Figure 1 was previously published by the Brodbelt group and could be accessed through ProteomeXchange with the PXD009447 accession number.[25] This data set was acquired by applying both CID and UVPD fragmentation methods on single amino acid variants of the human mitochondrial enzyme branched-chain amino acid transferase 2 using a modified prototype of a Thermo Q Exactive UHMR instrument.

Algorithm Parameters and Database Search

For comparison of deconvolution and database search algorithms in this study, our analysis used the default parameters of each algorithm. Additionally, we attempted to use the same parameters to minimize runtime differences caused by parameters. For instance, all algorithms were set to 1000,00 Da for maximum protein mass. A standard list of modifications such as N-terminal acetylation and N-terminal methionine removal was included during database search. A human database (Uniprot-Swissprot database, release December 2019, containing 20,367 protein sequences) was used for LC-MS/MS database search.

RESULTS

MASH Explorer software is a multifaceted software, which is built upon C# programming language using Visual Studio software under a .NET framework environment. The combination of C# and Visual Studio enables the development of a user-friendly Windows-based graphical interface, which is very intuitive for users, especially newcomers, to learn streamlined routine analysis. This software development environment allows high performance, low latency, and rich data interaction for high-throughput data processing.

The core functions of MASH Explorer include spectral deconvolution, protein identification, proteoform characterization, graphical data output, data validation, and workflow automation (Figure 1). Users can choose the integrated deconvolution and database search algorithms to perform spectral deconvolution, which extracts spectral features and subsequently generates a mass list from a complex mass spectrum to search against a database for protein identification. Spectral deconvolution and protein identification tasks are supported by GUI tools in the MASH Explorer software for automation. The proteoform characterization function allows users to match fragment ions to a protein sequence for localizing PTM sites and identifying sequence variations. MASH Explorer provides GUI to visualize experimental data for LC chromatograms, mass spectra, and fragment ion maps generated from various MS/MS experiments such as CID, ECD/ETD, and UVPD.

One unique feature of MASH Explorer is its universal data processing platform for top-down proteomics with the capability to process data from multiple vendor formats. MASH Explorer currently support specific vendor raw data formats from Thermo (.raw), Bruker (.d and .ascii), and Waters (.raw) (Figure 1). Moreover, universal data formats such as mzXML and mgf can be imported. The data import function is supported by ProteoWizard,[12] DeconEngine,[13] and vendor provided libraries. To allow successful data import, codes in MASH Explorer are continuously updated to accommodate the latest version of ProteoWizard and vendor-specific data acquisition software.

For the first time, MASH Explorer integrates multiple deconvolution and database search algorithms into a single platform to maximize the performance for enhanced protein identification (Figure 1). Currently, the software incorporates various deconvolution algorithms including MS-Deconv,[17] TopFD,[14] eTHRASH,[19] pParseTD,[15] and ProMex[26] for both MS and MS/MS deconvolution. The database search algorithms such as MS-Align+,[18] TopPIC,[14] pTop,[15] and MSPathFinderT[26] were integrated in the software for protein identification. MASH Explorer implements the Process Wizard, a user-friendly GUI to allow users to easily select deconvolution and database search algorithms and to customize the parameters of the selected algorithms for data processing, which is particularly convenient for users. In contrast, some database search algorithms, such as MS-Align+, require command line inputs using the Windows terminal, which is complicated and difficult for users with limited computational experience. The Configuration tool provides an intuitive interface for the users to find the directory of the supported deconvolution and database search algorithms (Figure S1).

The main interface of MASH Explorer allows users to perform data visualization, data validation, and customized output. The panels in the main interface include Workflow and Parameters, Status Bar, Results View, Mass List, Logbook, and Sequence Table (Figure S2). In the Workflow and Parameters panel, several sections are available for users to process top-down MS data, including “Discovery Mode” for LC-MS/MS data processing, “Targeted Mode” for single protein characterization. In addition, “Data Reporting” allows users to save processed data sets in Extensible Markup Language (XML) format, which can be reopened for further analysis, and to export Microsoft object files of both mass spectra and fragment ion maps for image processing. In the Results View panel, a mass spectrum is displayed for data visualization. Users can navigate through different scans, zoom-in and zoom-out of the selected spectrum, and adjust the theoretical Gaussian distribution of the fragment ions using the buttons displayed in the panel. The Mass List panel allows users browse through the deconvoluted mass list from the mass spectra for data validation. The entries in the Mass List panel interact with the Results View and Sequence Table panels, allowing users to visualize the fragment ion mapping for different types of MS/MS techniques to characterize the protein sequence. The entries in the Mass List panel can be copied to text editing software and converted to .msalign format during data processing. In the Sequence Table panel, PTMs of the protein sequences can also be selected and analyzed. The Logbook and Status Bar panels record all data processing by the software such as the versions of the tools used for raw data import, and the parameters used in deconvolution and database search tasks. Users can copy the Logbook recordings to a text editor in the event an error occurs. Moreover, the information in the Logbook recordings can help the MASH Explorer software developers troubleshoot any problems.

MASH Explorer features a “Discovery Mode” workflow that is useful for high-throughput data processing and proteoform identification from batch LC-MS/MS raw data files without *a priori* knowledge of specific proteins (Figure 2). “Discovery Mode” integrates several top-down MS processing tools to centroid, deconvolute, and search databases against raw data sets. The software environment highlights the intuitive and user-friendly Process Wizard and Workflow Manager to enhance the efficiency of data processing.

MASH Explorer offers a user-friendly GUI, Process Wizard, for different deconvolution and database search algorithms (Figure S3). This GUI tool bundles top-down data processing steps including centroiding, deconvolution, and database search. After data import, users can choose available processing pipelines in the Process Wizard. Users can run the algorithms using default settings or change the parameters of each algorithm in the Advanced tab. Additionally, MASH Explorer implements a Workflow Manager to enhance the efficiency of processing top-down proteomics data sets (Figure S4). In the Workflow Manager, users can run a batch analysis of top-down proteomic data sets in sequence. The Workflow Manager achieves this function by reading the workflow log created during the algorithm process and gives instructions to wait and execute the next operation. Upon completion, the Workflow Manager automatically imports both the deconvolution and database search results into MASH Explorer for validation of identified proteins. It provides users with convenience in both automatic data file conversion and parameter input in algorithms without sacrificing the efficiency of the database search.

Incorporation of various deconvolution and database search algorithms enables MASH Explorer to improve global proteoform identification and characterization (Figure 3 and Figure S5). As an example, multiple deconvolution and database search workflows have been performed on both the Thermo data set from human colorectal cancer cell protein extracts[20] and the Bruker data set from surfactant-extracted protein mixtures[4] for global proteoform identification (Figure 3B and Figure S5A; detailed discussions on using "Discovery Mode" for data analysis are provided in the Supporting Information). Identified proteoforms can be further analyzed using tools provided by MASH Explorer for comprehensive proteoform characterization (Figure 3C). In addition to the current list of deconvolution and database search algorithms, MASH Explorer has the capability to incorporate more algorithms, owing to the modularity of the software. The incorporation of recently developed deconvolution algorithms such as FLASHDeconv[8] and UniDec[27,28] could increase the diversity in deconvolution methods and thus enable MASH Explorer to process data sets more effectively. Moreover, the results from multiple algorithms can be used for analysis and further implementation of machine learning algorithms. Recent algorithm development in the MASH project will enable users to run a machine learning tool on deconvolution.[23] This machine learning tool used hierarchical clustering to combine deconvoluted peak lists from different algorithms, which can effectively detect true positive peaks while filtering out false positive peaks, resulting in enhanced accuracy and confidence in protein identification during database search.

Another important feature of MASH Explorer is a complementary "Targeted Mode" workflow that is optimized for the detailed and comprehensive characterization of individual proteins, enabling users to identify site-specific PTMs within a protein target (Figure 4). The "Targeted Mode" workflow was developed for comprehensive protein characterization. It includes data import, spectral deconvolution to identify and verify isotopic distributions, database search to identify target protein, and finally protein characterization by matching the identified isotopic distribution to the target proteoform sequence. The "Targeted Mode" workflow aims to perform identification of fragment ions that help identify and localize PTMs of a target proteoform sequence.

In addition to the functions introduced in our previous generation software, MASH Suite Pro,[11] which provides tools for users to perform charge state and mass shift correction, the "Targeted Mode" in MASH Explorer introduces an Ion Finder Tool GUI that parses through generated ion lists from different fragmentation methods to find proteoform annotations and allow users to match theoretical and observed fragment ions (Figure S6). Using the Ion Finder Tool, users can input the fragment ion type and the charge state of the specific fragment ion of interest. The software will then zoom-in to the m/z region of the targeted ion and attempt to perform fragment ion matching. The Ion Finder Tool complements the existing eTHRASH algorithm in MASH Explorer to provide a more comprehensive fragment ion mapping for top-down protein analysis. As an example, we have demonstrated on a previously published data set in the characterization of cardiac sarcomeric proteins from nonhuman primate skeletal muscle such as βTpm, which was modified with N-terminal acetylation, and MLC-2S with N-terminal methionine removal and PTMs including N-terminal acetylation and deamidation at Asn13 (Figure S7).[22] Moreover, MASH Explorer can also be extended to characterize the subunits of ADCs,[24] which combine the target

specificity of monoclonal antibody and the potency of the cytotoxin drugs, gaining enormous interest in the pharmaceutical industry (Figure 5 and Figure 6). One of the analytical tasks for ADC characterization is the site localization of the drug payload. The digestion of an ADC, brentuximab vedotin, with IdeS resulted Fd1 subunits in three possible isomers, where drugs can be incorporated on three possible cysteine residues (Figure 5A; detailed discussion on using "Targeted Mode" for data analysis is provided in the Supporting Information). Using MASH Explorer, MS/MS spectra can be imported and performed by fragment ion mapping on specific Fd1 subunit (Figure 6). Additionally, fragment ions near three possible sites including Cys220, Cys226, and Cys229, which are the specific locations of interchain disulfide bonds for drug linkage, can be localized. $z^{\bullet}_{15}$, $z^{\bullet}_{16}$, $z^{\bullet}_{23}$, and $z^{\bullet}_{24}$ ions were visualized using the Ion Finder Tool to localize Cys220 as the site for the payload for an Fd1 subunit isomer (Figure 5B).

DISCUSSION

MASH Explorer is a nonproprietary and free software solution, providing a universal and comprehensive environment for processing top-down proteomics data. The major innovations of MASH Explorer include the integration of multiple deconvolution and search algorithms into a single, universal platform to process raw data from various vendor formats in a user-friendly interface. Since the development of the MASH project, the software has been downloaded and used by more than 600 users around the world (as of March 24th, 2020) (Figure 7). While the majority of users are from North America, the MASH software has continuously attracted users across the globe, including users from continents such as Europe and Asia. As the popularity of top-down MS-based proteomics grows, MASH software is increasingly becoming a vital and integral tool for users to process complex high-resolution top-down LC-MS/MS data. In addition to the case studies of protein identification from human colorectal cancer cell protein extracts[20] and surfactant-extracted protein mixture,[4] as well as the characterization of ADC,[24] many other groups have used the MASH software packages in top-down proteomics projects including analysis of the light and heavy chain connectivity of a monoclonal antibody,[29] characterization of branched ubiquitin chains,[30,31] identification and characterization of intact phosphoproteins,[32] and localization of phosphorylation sites of a phosphatase.[33]

As the burgeoning top-down proteomics community continues its rapid growth and has gained momentum through the creation of the Consortium for Top-Down Proteomics (CTDP) (http://www.topdownproteomics.org/), the need for universal, comprehensive, and globally accessible top-down proteomics software has increased tremendously. With the critical need and tremendous support from the community, we envision that this MASH Explorer software package will serve as a powerful tool to enable top-down proteomics researchers worldwide, playing an integral role in advancing top-down proteomics to realize its full potential for biomedical research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by the NIH R01 GM125085 (to Y.G). Y.G. would also like to acknowledge support by NIH R01 HL096971, GM117058, and S10 OD018475. We would like to thank Ziqing Lin, Yutong Jin, Bifan Chen, Trisha Tucholski, Kyle Brown, and Austin Carr for the helpful discussions. We also thank all the MASH users worldwide for the excellent feedback which has helped the development of the software.

REFERENCES

(1). Smith LM; Kelleher NL Proteoforms as the next proteomics currency. Science 2018, 359 (6380), 1106–1107. [PubMed: 29590032]

(2). Smith LM; Thomas PM; Shortreed MR; Schaffer LV; Fellers RT; LeDuc RD; Tucholski T.; Ge Y; Agar JN; Anderson LC; Chamot-Rooke J; Gault J; Loo JA; Pasa-Tolic L; Robinson CV; Schluter H; Tsybin YO; Vilaseca M; Vizcaino JA; Danis PO; Kelleher NL. A five-level classification system for proteoform identifications. Nat. Methods 2019, 16, 939. [PubMed: 31451767]

(3). 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; Loo RRO; 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; Wohlschlager 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]

(4). Brown KA; Chen BF; Guardado-Alvarez TM; Lin ZQ; Hwang L; Ayaz-Guner S; Jin S; Ge Y A photocleavable surfactant for top-down proteomics. Nat. Methods 2019, 16 (5), 417–420. [PubMed: 30988469]

(5). Siuti N; Kelleher NL Decoding protein modifications using top-down mass spectrometry. Nat. Methods 2007, 4 (10), 817–21. [PubMed: 17901871]

(6). 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]

(7). Chen B; Brown KA; Lin Z; Ge Y Top-Down Proteomics: Ready for Prime Time? Anal. Chem 2018, 90 (1), 110–127. [PubMed: 29161012]

(8). Jeong K; Kim J; Gaikwad M; Hidayah SN; Heikaus L; Schluter H; Kohlbacher O FLASHDeconv: Ultrafast, High-Quality Feature Deconvolution for Top-Down Proteomics. Cell Syst. 2020, 10 (2), 213–218. [PubMed: 32078799]

(9). Schaffer LV; Millikin RJ; Miller RM; Anderson LC; Fellers RT; Ge Y; Kelleher NL; LeDuc RD; Liu X; Payne SH; Sun L; Thomas PM; Tucholski T; Wang Z; Wu S; Wu Z; Yu D; Shortreed MR; Smith LM Identification and Quantification of Proteoforms by Mass Spectrometry. Proteomics 2019, 19 (10), No. 1970085.

(10). Guner H; Close PL; Cai W; Zhang H; Peng Y; Gregorich ZR; Ge Y MASH Suite: a user-friendly and versatile software interface for high-resolution mass spectrometry data interpretation and visualization. J. Am. Soc. Mass Spectrom 2014, 25 (3), 464–70. [PubMed: 24385400]

(11). Cai WX; Guner H; Gregorich ZR; Chen AJ; Ayaz-Guner S; Peng Y; Valeja SG; Liu XW; Ge Y MASH Suite Pro: A Comprehensive Software Tool for Top-Down Proteomics. Mol. Cell. Proteomics 2016, 15 (2), 703–714. [PubMed: 26598644]

(12). Kessner D; Chambers M; Burke R; Agus D; Mallick P ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 2008, 24 (21), 2534–2536. [PubMed: 18606607]

(13). Jaitly N; Mayampurath A; Littlefield K; Adkins JN; Anderson GA; Smith RD Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. BMC Bioinf. 2009, 10, 87.

(14). 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]

(15). Sun RX; Luo L; Wu L; Wang RM; Zeng WF; Chi H; Liu C; He SM pTop 1.0: A High-Accuracy and High-Efficiency Search Engine for Intact Protein Identification. Anal. Chem 2016, 88 (6), 3082–3090. [PubMed: 26844380]

(16). Park J; Piehowski PD; Wilkins C; Zhou M; Mendoza J; Fujimoto GM; Gibbons BC; Shaw JB; Shen Y; Shukla AK; Moore RJ; Liu T; Petyuk VA; Tolic N; Pasa-Tolic L; Smith RD; Payne SH; Kim S Informed-Proteomics: open-source software package for top-down proteomics. Nat. Methods 2017, 14 (9), 909–914. [PubMed: 28783154]

(17). Liu X; Inbar Y; Dorrestein PC; Wynne C; Edwards N; Souda P; Whitelegge JP; Bafna V; Pevzner PA Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. Mol. Cell. Proteomics 2010, 9 (12), 2772–82. [PubMed: 20855543]

(18). Liu X; Sirotkin Y; Shen Y; Anderson G; Tsai YS; Ting YS ; Goodlett DR; Smith RD; Bafna V; Pevzner PA Protein Identification Using Top-Down Spectra. Mol. Cell. Proteomics 2012, 11 (6), M111.008524.

(19). Horn DM; Zubarev RA; McLafferty FW Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. J. Am. Soc. Mass Spectrom 2000, 11 (4), 320–332. [PubMed: 10757168]

(20). 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 T FT-ICR Mass Spectrometry. J. Proteome Res 2017, 16 (2), 1087–1096. [PubMed: 27936753]

(21). 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]

(22). Jin Y; Diffee GM; Colman RJ; Anderson RM; Ge Y Top-down Mass Spectrometry of Sarcomeric Protein Post-translational Modifications from Non-human Primate Skeletal Muscle. J. Am. Soc. Mass Spectrom 2019, 30 (12), 2460–2469. [PubMed: 30834509]

(23). McIlwain SJ; Wu Z; Wetzel M; Belongia D; Jin Y; Wenger K; Ong IM; Ge Y Enhancing Top-Down Proteomics Data Analysis by Combining Deconvolution Results through a Machine Learning Strategy. J. Am. Soc. Mass Spectrom 2020, 31 (5), 1104–1113. [PubMed: 32223200]

(24). Chen B; Lin Z; Zhu Y; Jin Y; Larson E; Xu Q; Fu C; Zhang Z; Zhang Q; Pritts WA; Ge Y Middle-Down Multi-Attribute Analysis of Antibody-Drug Conjugates with Electron Transfer Dissociation. Anal. Chem 2019, 91 (18), 11661–11669. [PubMed: 31442030]

(25). Mehaffey MR; Sanders JD; Holden DD; Nilsson CL; Brodbelt JS Multistage Ultraviolet Photodissociation Mass Spectrometry To Characterize Single Amino Acid Variants of Human Mitochondrial BCAT2. Anal. Chem 2018, 90 (16), 9904–9911. [PubMed: 30016590]

(26). Park J; Piehowski PD; Wilkins C; Zhou M; Mendoza J; Fujimoto GM; Gibbons BC; Shaw JB; Shen Y; Shukla AK; Moore RJ; Liu T; Petyuk VA; Tolić N; Paša-Tolić L; Smith RD; Payne SH; Kim S Informed-Proteomics: open-source software package for top-down proteomics. Nat. Methods 2017, 14, 909. [PubMed: 28783154]

(27). Marty MT; Baldwin AJ; Marklund EG; Hochberg GK; Benesch JL; Robinson CV Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. Anal. Chem 2015, 87 (8), 4370–6. [PubMed: 25799115]

(28). Marty MT A Universal Score for Deconvolution of Intact Protein and Native Electrospray Mass Spectra. Anal. Chem 2020, 92 (6), 4395–4401. [PubMed: 32069030]

(29). Srzentic K; Nagornov KO; Fornelli L; Lobas AA; Ayoub D; Kozhinov AN; Gasilova N; Menin L; Beck A; Gorshkov MV; Aizikov K; Tsybin YO Multiplexed Middle-Down Mass Spectrometry as a Method for Revealing Light and Heavy Chain Connectivity in a Monoclonal Antibody. Anal. Chem 2018, 90 (21), 12527–12535. [PubMed: 30252447]

(30). Crowe SO; Rana ASJB; Deol KK; Ge Y; Strieter ER Ubiquitin Chain Enrichment Middle-Down Mass Spectrometry Enables Characterization of Branched Ubiquitin Chains in Cellulo. Anal. Chem 2017, 89 (17), 9610–9610. [PubMed: 28792741]

(31). Rana ASJB; Ge Y; Strieter ER Ubiquitin Chain Enrichment Middle-Down Mass Spectrometry (UbiChEM-MS) Reveals Cell-Cycle Dependent Formation of Lys11/Lys48 Branched Ubiquitin Chains. J. Proteome Res 2017, 16 (9), 3363–3369. [PubMed: 28737031]

(32). Roberts DS; Chen B; Tiambeng TN; Wu Z; Ge Y; Jin S Reproducible large-scale synthesis of surface silanized nanoparticles as an enabling nanoproteomics platform: Enrichment of the human heart phosphoproteome. Nano Res. 2019, 12 (6), 1473–1481. [PubMed: 31341559]

(33). Wu CG; Chen H; Guo F; Yadav VK; McIlwain SJ; Rowse M; Choudhary A; Lin Z; Li Y; Gu T; Zheng A; Xu Q; Lee W; Resch E; Johnson B; Day J; Ge Y; Ong IM; Burkard ME; Ivarsson Y; Xing Y PP2A-B' holoenzyme substrate recognition, regulation and role in cytokinesis. Cell Discovery 2017, 3, 17027. [PubMed: 28884018]

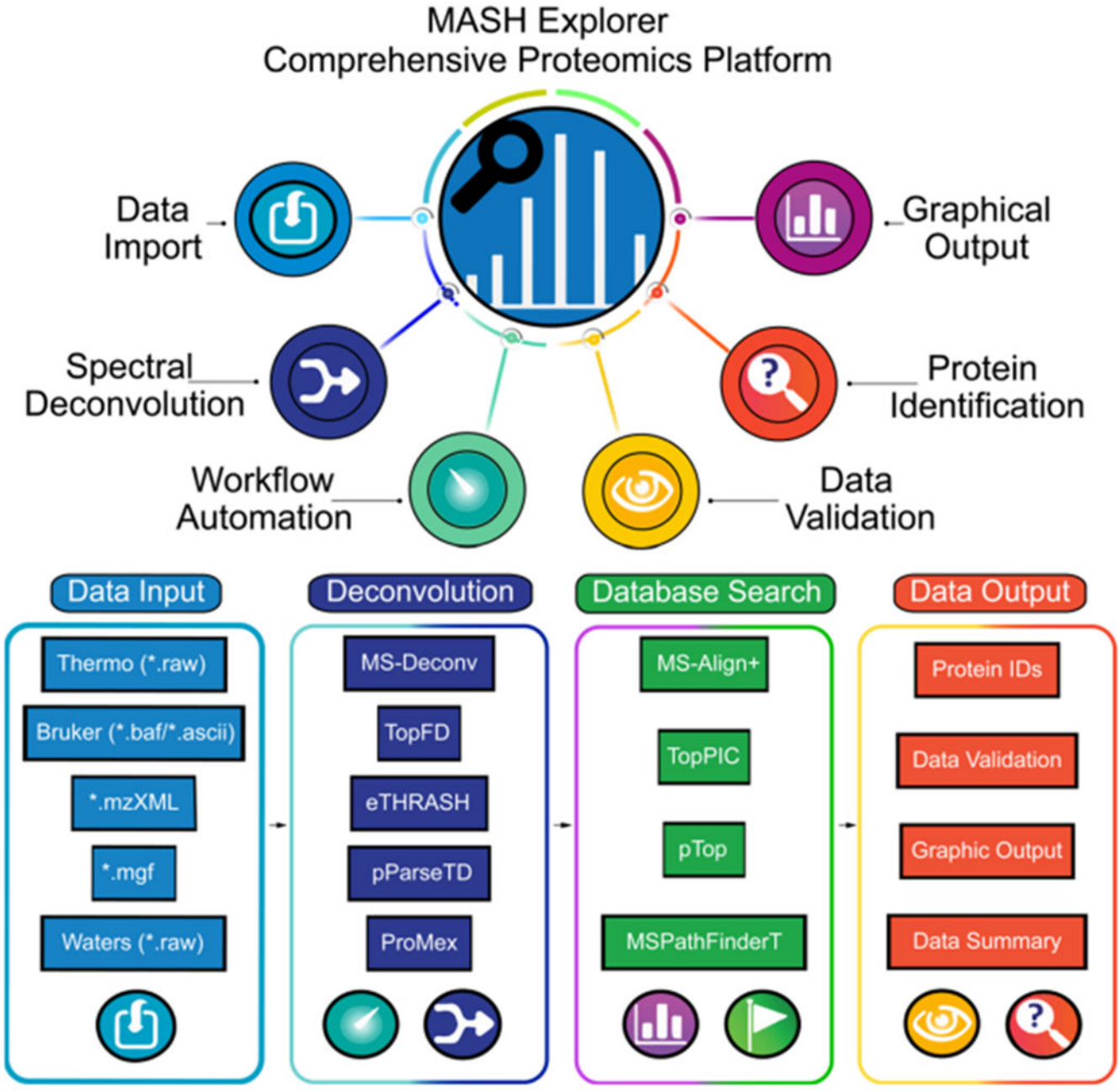

Figure 1.

Schematic of the various MASH Explorer functions for proteomics data processing. Main functions of MASH Explorer include data import, spectral deconvolution, workflow automation, data validation, protein identification, and graphical output. MASH Explorer utilizes a new data processing module based on the ProteoWizard Library to accept various data input file formats from major instrument vendors (e.g., Thermo, Bruker, and Waters). Raw MS and MS/MS data files are then processed by deconvolution algorithms (i.e., MS-Deconv, TopFD, eTHRASH, pParseTD, and ProMex), and database search algorithms (i.e.,

MS-Align+, TopPIC, pTop, and MSPathFinderT). MASH Explorer provides a user-friendly interface for data validation, proteoform identification, and characterization.

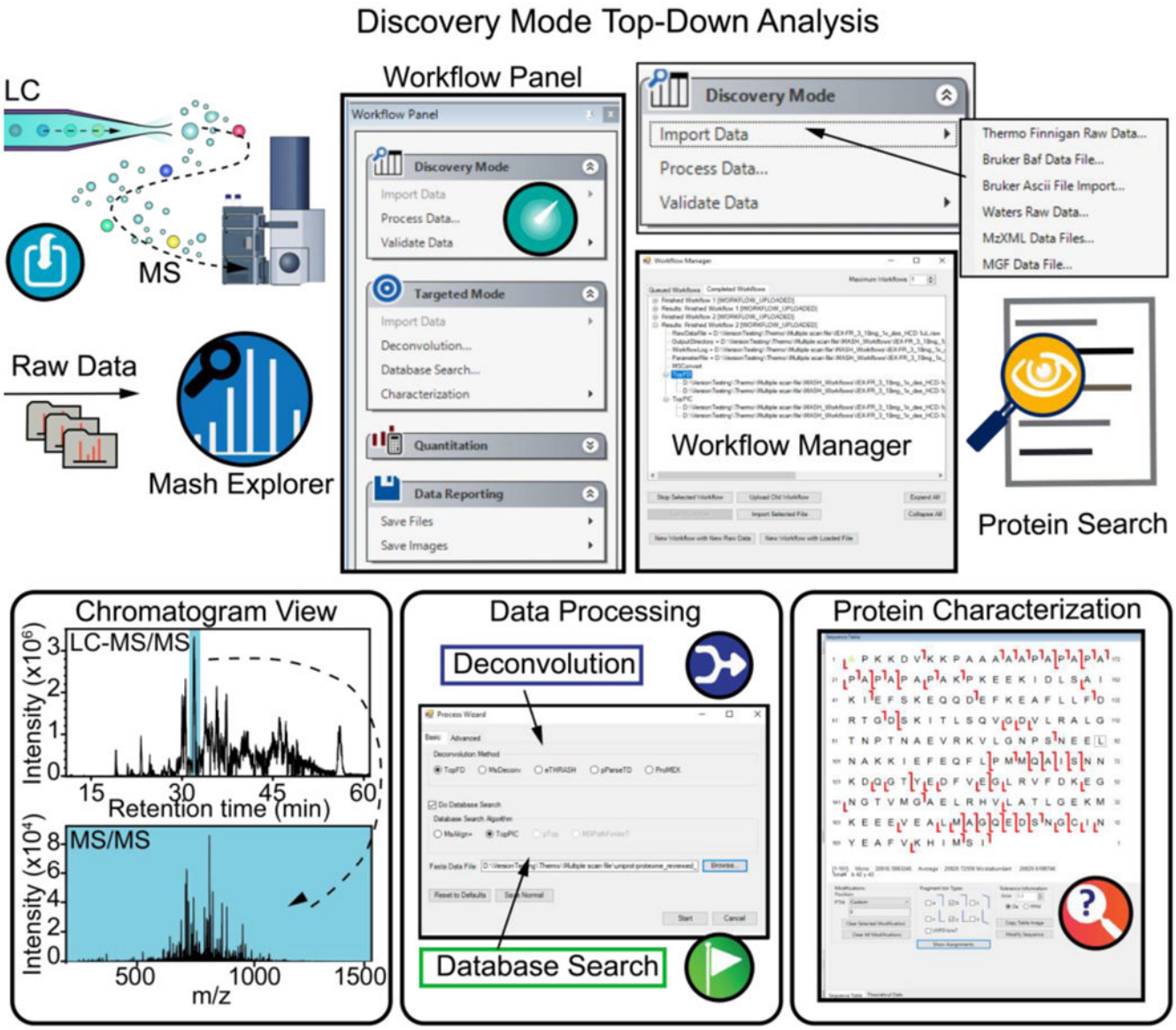

Figure 2.

Illustration of “Discovery mode” for LC-MS/MS data processing. “Discovery mode” can handle batch LC-MS/MS raw data files and includes features such as data import, data processing (deconvolution and database search), and data validation for protein identification. A simple and user-friendly Workflow Manager GUI automates the search and validation process and outputs processed data to a tabulated Mass List panel where users can view individual fragment ions and assign additional PTMs to reflect the fragment ion mapping on individual protein sequences.

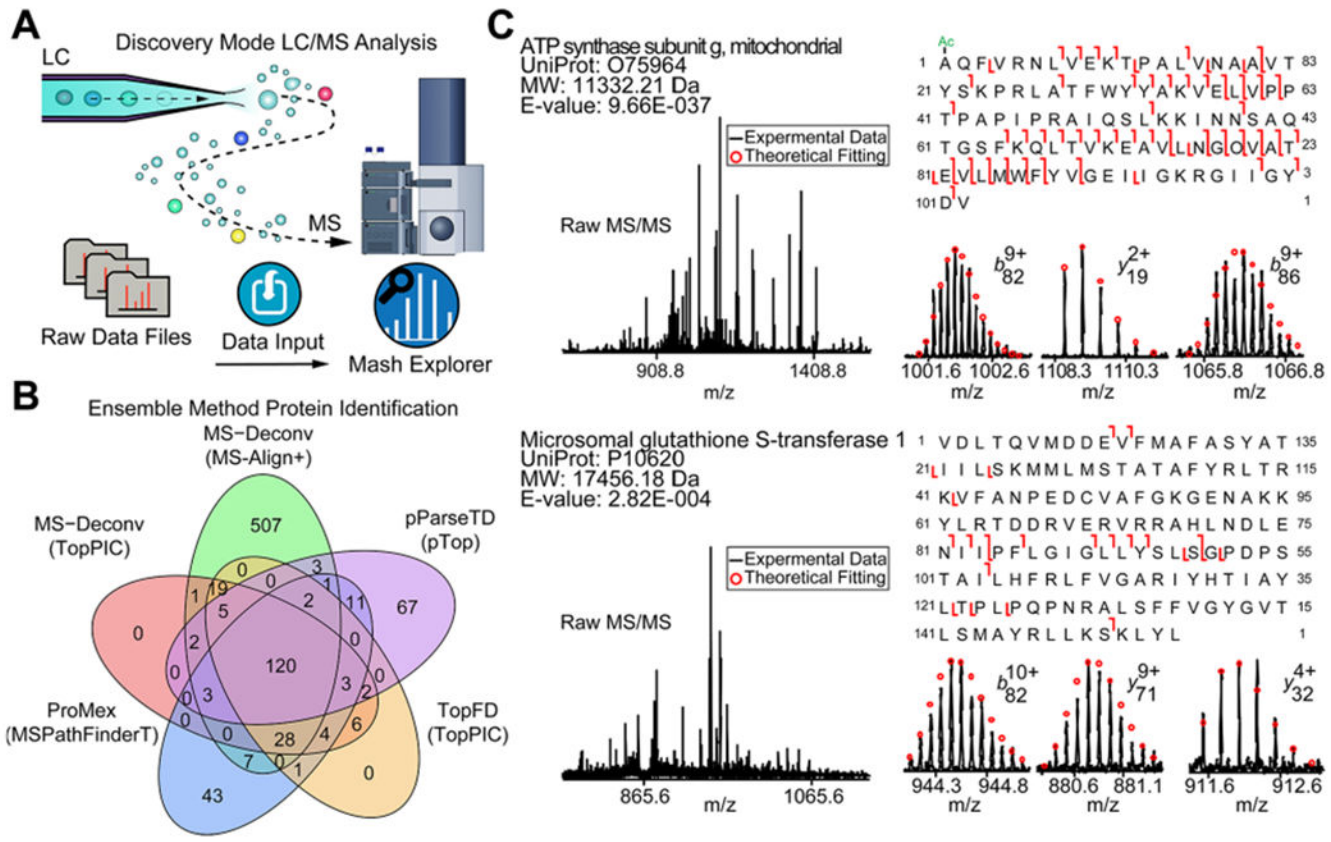

Figure 3.

Top-down proteomics data analysis using “Discovery Mode” in MASH Explorer. A. Cartoon illustration of a typical “Discovery Mode” top-down LC-MS workflow. B. Venn diagram showing the overlap of protein identifications using an ensemble of five combined deconvolution and protein search workflows using a Thermo LC-MS/MS data set. This combined deconvolution algorithm capability enables a deeper proteome coverage and enhanced protein identifications. C. Top-down MS identification and characterization using “Discovery Mode” workflow with ATP synthase subunit g, mitochondrial and microsomal glutathione S-transferase 1 shown as examples. The MS/MS spectra, sequence tables, and fragment ions were output directly from MASH Explorer. Uniprot-Swissprot accession and protein E-value score are reported for each protein.

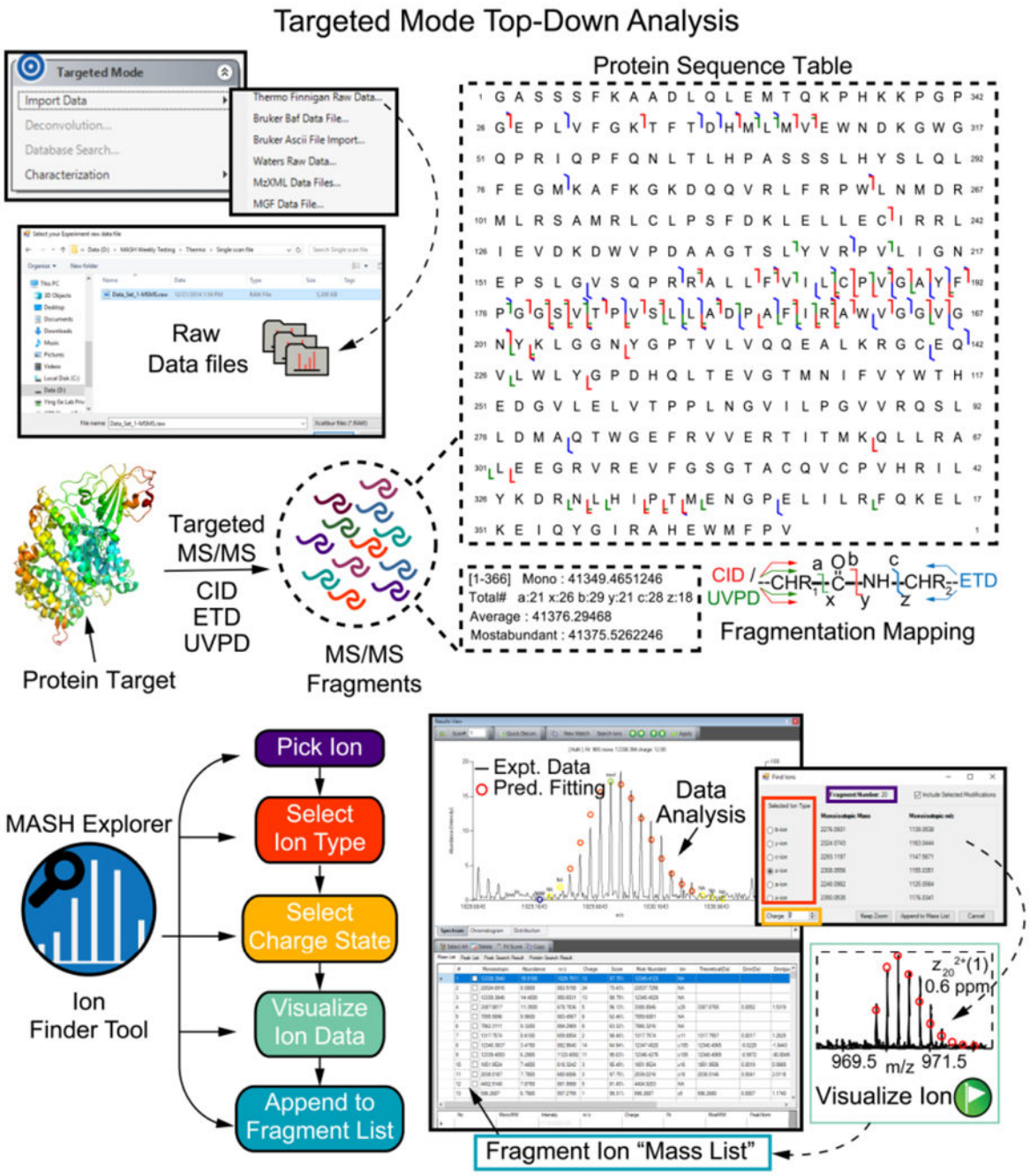

Figure 4.

Illustration of "Targeted Mode" workflow for MASH Explorer. "Targeted Mode" workflow includes data import, spectral deconvolution to identify and verify isotopic distributions, database search based on identified isotopic distributions, and proteoform characterization by matching identified isotopic distributions to the target proteoform sequence. "Targeted Mode" helps expedite PTM localization by a simple Ion Finder Tool, which searches for fragment ions to confidently localize PTMs.

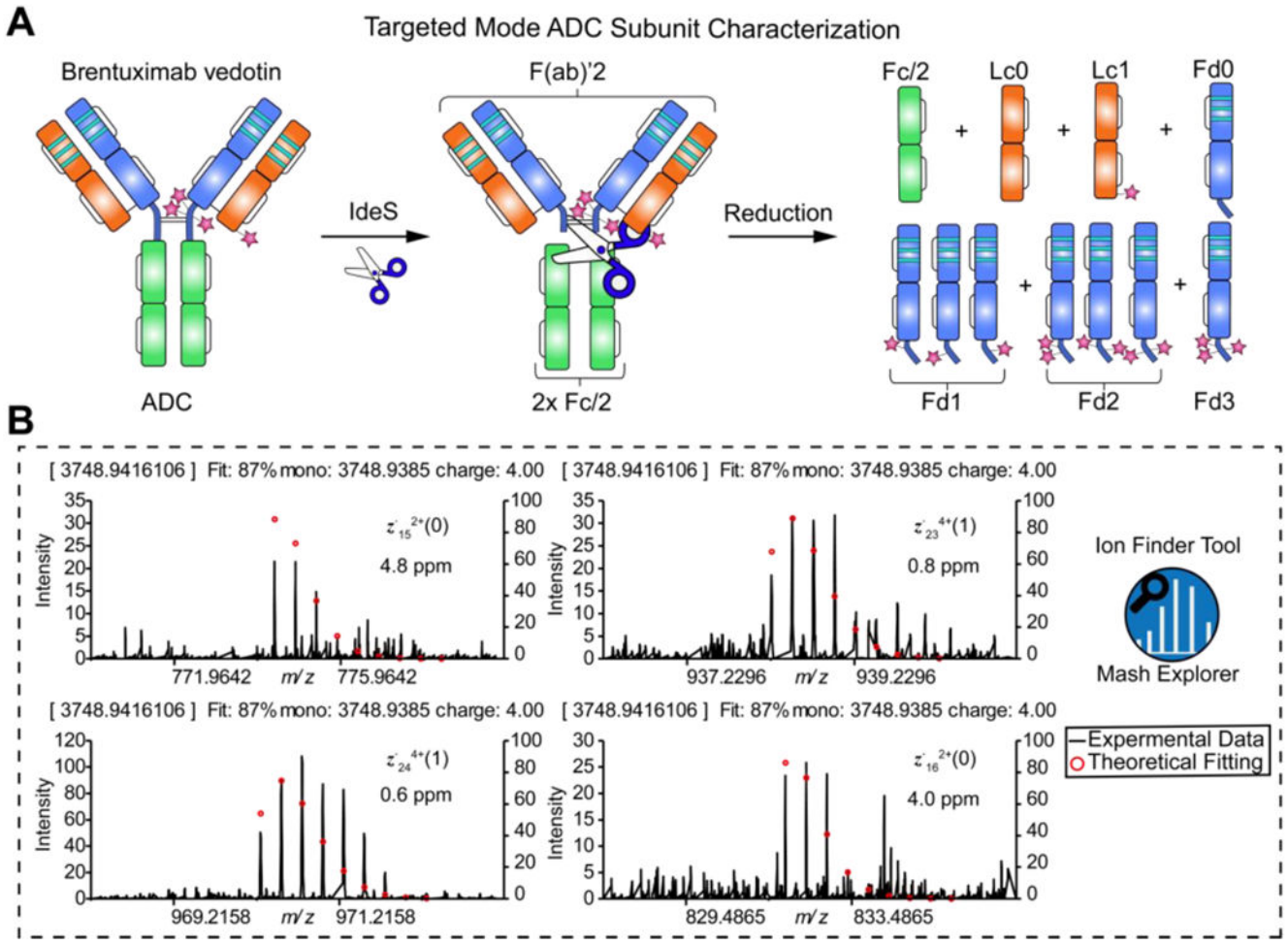

Figure 5.

Characterization of ADC subunits using "Targeted Mode" in MASH Explorer. A. Intact ADC, brentuximab vedotin (Adcetris), is first subjected to IdeS digestion to cleave the hinge region and then further reduced to generate the ADC subunits (Fc/2, Lc0, Lc1, Fd0, Fd1, Fd2, and Fd3). B. The MASH Explorer Ion Finder Tool was used to search through candidate ions and generate fragment ion maps for the identification and localization of the site-specific drug conjugation site of a positional isomer of Fd1 subunit. The number in parentheses represents the number of drug payloads included in the fragment ion.

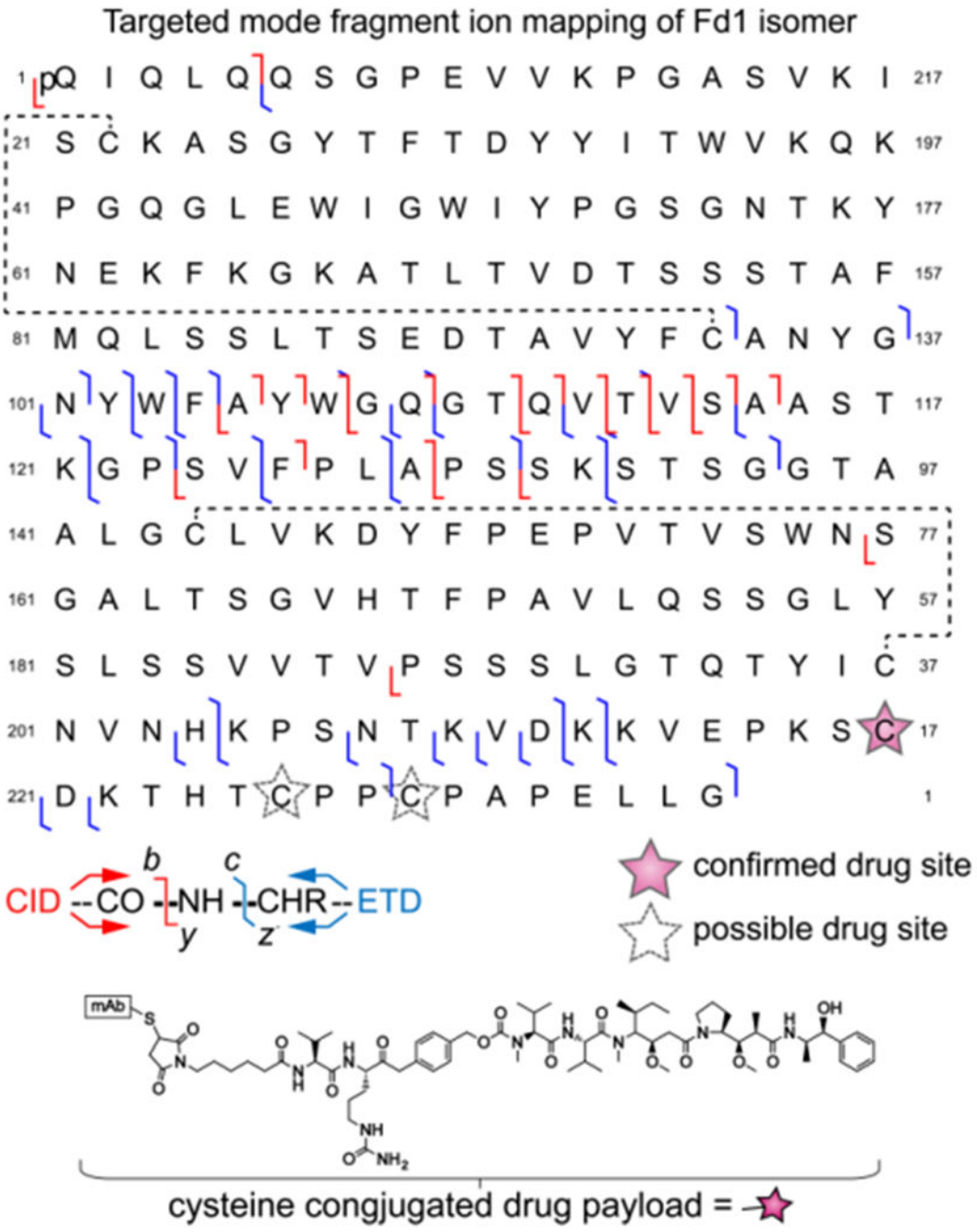

Figure 6.

Protein sequence characterization and fragment ion mapping of Fd1 isomer from an ADC. Fragment ion map shows both CID and ETD fragment ions. Fragment ions were used to confirm the specific localization of a drug site of an Fd1 isomer. The pink star represents the cysteine-conjugated drug warhead corresponding to the Adcetris drug molecule. The data shown corresponds to the ADC fragmentation data shown in Figure 5.

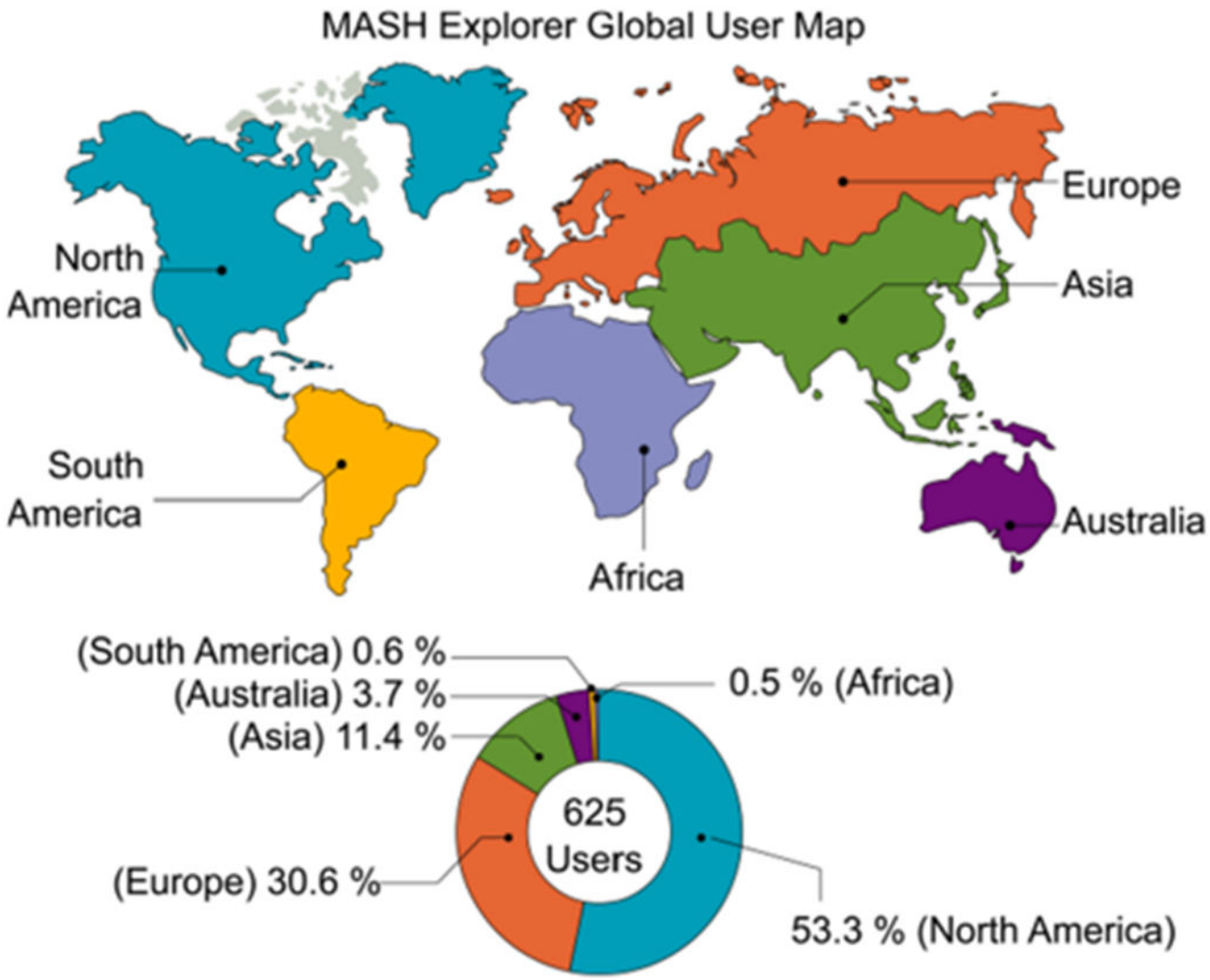

Figure 7.

Cartoon schematic of a “world map” featuring the location distribution of MASH users across the globe. There are currently 625 active users (03/24/2020) with ~53% of users from North America, ~31% from Europe, and ~11% from Asia.