analytical chemistry

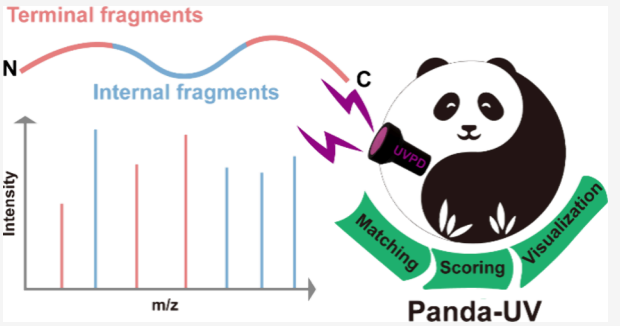
pubs.acs.org/ac
Article

Panda-UV Unlocks Deeper Protein Characterization with Internal Fragments in Ultraviolet Photodissociation Mass Spectrometry

Yinlong Zhu,[▽] Zheyi Liu,[▽] Jialiang Liu, Heng Zhao, Rui Feng, Kunxian Shu, Fangjun Wang,* and Cheng Chang*

Cite This: *Anal. Chem.* 2024, 96, 8474−8483
Read Online

ABSTRACT: Ultraviolet photodissociation (UVPD) mass spectrometry unlocks insights into the protein structure and sequence through fragmentation patterns. While N- and C-terminal fragments are traditionally relied upon, this work highlights the critical role of internal fragments in achieving near-complete sequencing of protein. Previous limitations of internal fragment utilization, owing to their abundance and potential for random matching, are addressed here with the development of Panda-UV, a novel software tool combining spectral calibration, and Pearson correlation coefficient scoring for confident fragment assignment. Panda-UV showcases its power through comprehensive benchmarks on three model proteins. The inclusion of internal fragments boosts identified fragment numbers by 26% and enhances average protein sequence coverage to a remarkable 93% for intact proteins, unlocking the hidden region of the largest protein carbonic anhydrase II in model proteins. Notably, an average of 65% of internal fragments can be identified in multiple replicates, demonstrating the high confidence of the fragments Panda-UV provided. Finally, the sequence coverages of mAb subunits can be increased up to 86% and the complementary determining regions (CDRs) are nearly completely sequenced in a single experiment. The source codes of Panda-UV are available at https://github.com/PHOENIXcenter/Panda-UV.



INTRODUCTION

Precise characterization of intact proteins, encompassing sequence, modifications, and structure, unlocks critical insights into biological processes from fundamental cellular events to intricate disease mechanisms. In recent years, top-down mass spectrometry (TDMS) has emerged as a powerful tool for bypassing the information loss inherent in enzymatic digestion, allowing direct analysis of entire proteins and their diverse forms (proteoforms) arising from splicing, mutations, and posttranslational modifications (PTMs).[1−7]

The core of successful TDMS lies in efficient protein dissociation. While fragmentation via collision-induced dissociation (CID)[8] typically yields limited coverage, particularly for larger proteins, ultraviolet photodissociation (UVPD) shines.[9−11] UVPD (193 nm) boasts unparalleled efficiency, generating near-complete sequence cleavage for proteins below 30 kDa and offering promising potential for larger ones.[12,13]

However, maximizing sequence coverage requires venturing beyond just the traditional N- and C-terminal fragments (a/x, b/y, and c/z ions). Internal fragments hold vast untapped potential.[14−17] Representing the majority of MS/MS signals, they offer access to protein regions often invisible to terminal fragments, especially for larger proteins and regions inaccessible due to disulfide bonds.[18,19] Their inclusion can dramatically increase sequence coverage, reaching over 75% for monoclonal antibodies with >150 kDa mass.[19] Internal fragments can even provide structural information when characterizing protein complexes using native top-down MS (nTDMS).[20−22]

Despite their promising potential, internal fragments remain largely underutilized due to the computational matching challenges and low confidence in assignments.[23] The large number of theoretical internal fragments compared to terminal fragments dramatically expands the search space, leading to exponential complexity and ambiguity. Additionally, traditional search engines are ill-equipped to handle the diverse array of terminal and internal fragment ions generated by UVPD, including a/x, b/y, c/z, ax, ay, az, bx, by, bz, cx, cy, and cz ions, further complicated by hydrogen migration during the photofragmentation.[24,25] Previous attempt using ClipsMS,[26] a dedicated internal fragment assignment tool, demonstrated the improvements in protein sequence coverages but suffered from low confidence for UVPD spectra due to the low signal-to-

Received: January 14, 2024
Revised: May 6, 2024
Accepted: May 6, 2024
Published: May 13, 2024

https://doi.org/10.1021/acs.analchem.4c00253
Anal. Chem. 2024, 96, 8474−8483

noise and congestion.[27] For all model proteins of Dunham's study, only an average of 18% of UVPD internal fragments was consistently identified in two out of three replicates, compared to 63% for HCD internal fragments.[27] This low confidence discourages the utilization of internal fragments in the UVPD analysis.

Herein, we introduce Panda-UV, a novel software tool specifically designed for confident photofragmentation of UVPD data. Panda-UV tackles the key challenges by implementing spectral calibration[28−30] to improve mass accuracy and a Pearson correlation coefficient (PCC) scoring system to filter the low-confidence matches. Through rigorous benchmarking on UVPD spectra of model proteins, we demonstrated that Panda-UV significantly increased the confidence of internal fragment matching, paving the way for their routine inclusion in UVPD-based protein characterization and unlocking the potential of this powerful technique.

EXPERIMENTAL SECTION

Materials and Reagents. Three model proteins (carbonic anhydrase II, myoglobin, and ubiquitin) and infliximab were used in this study. Model proteins were purchased from Merck KGaA (Darmstadt, Germany). Infliximab (Remicade) was purchased from TopScience (Shanghai, China) with 98% purity. All deionized water used in the experiment was purified by a Milli-Q system (Millipore Inc., Milford, Massachusetts, USA). IdeS protease was purchased from Yeasen Biotechnology (Shanghai, China). All solvents and mobile phases were purchased with liquid chromatography−mass spectrometry (LC-MS) grade purity. Acetonitrile (ACN) was purchased from Merck (Darmstock, Germany). Other solvents were purchased from Sigma-Aldrich. Micro Bio-Spin $P\overline{6}$ Gel Columns were purchased from Bio-Rad (California, USA). PLRP-S chromatographic packing (1000A, 5 μm) purchased from Agilent (California, USA).

Sample Preparation and Mass Spectrometry. Model proteins was dissolved in 1 mM ammonium acetate (Aladdin, Shanghai, China) with a concentration of 10 μM. Solutions were introduced into MS by direct infusion with a flow rate of 0.3 μL/min. The spray voltage was set at 1.6 kV. All MS data sets were collected by an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher, San Jose, CA, USA) implemented with a 193 nm ArF excimer laser (Gam laser, Orlando, FL, USA) as described in our previous works.[31−33] The instrument was calibrated before acquiring the data. The full mass spectra of model proteins were collected with a mass resolution of 240 K. The protein ions were isolated individually and subjected to a 5 ns single pulse of laser irradiation (1.1 mJ) for UVPD analysis. The fragment ions were analyzed with a mass resolution of 240 K. The automatic gain target (AGC) of MS2 was set at 1E6.

The Infliximab sample was dissolved in PBS (50 mM sodium phosphate, 150 mM NaCl, pH 6.6) at 5 μg/μL, and then IdeS was added at 1 unit per microgram of infliximab sample, and two fragments of $F(ab')_2$ and Fc were obtained by specifically cutting the hinge region for 30 min at 37 °C. After digestion, 6 M urea was added to denature these two fragments for 20 min. Then, 5 mM TCEP chemical reduction was added at room temperature for 30 min to open the disulfide bond to reduce $F(ab')_2$ into Lc (light chain) and Fd (heavy chain variable domain) fragments at room temperature for 30 min, followed by adding 10 mM IAA in a dark environment for alkylation for 30 min. Micro Bio-Spin $P\overline{6}$ Gel Columns were then used to replace the buffer with ammonium acetate. Then, the sample was freeze-dried. The sample was dissolved with 0.1% FA to 1 μg/μL and then introduced into LC-MS. The temperature of the ion transfer capillary was 305 °C, the electric spray voltage was 2000 V, and the RF lens was set to 55%. Before data collection, the retention time and charge state distribution of infliximab subunits were determined through LC-MS operation. The full mass spectrum was set to a resolution of 15,000 and a mass range of 400−2000 m/z. Then, targeted LC-MS/MS data collection was performed through selected ion monitor (SIM) mode. The MS2 spectrum was set with a 240 K resolution, a quality range of 350−2000 m/z, and an RF lens of 60%. The AGC of MS2 was set at 1000 with a maximum injection time of 500 ms. A single pulsed laser irradiation of 5 ns was used for UVPD analysis of subunit fragments, with an energy of about 0.6−1.4 mJ.

Data Analysis. These raw files were then converted into an mzML format using MSConvert, followed by deconvolution with TopFD GUI v 1.5.4[34] at a S/N of 2. The deconvoluted masses of UVPD fragment ions were transformed into a .csv file using a custom python script and calibrated by Panda-UV, which was then used as input for Panda-UV and ClipsMS. Panda-UV and ClipsMS were used to search for the following ion types within ±5 ppm: a, a+H, b, c, x, x+H, y, y-H, y-2H, z, ax, ay, az, bx, bz, cx, cy. Hydrogen gains and losses in ClipsMS searches were considered using an unlocalized modification file, and then improbable fragments were deleted by a python script. Panda-UV turned on mass calibration and spectral calibration functions. The error for matching fragment experimental envelope peaks is set to 10 ppm. Both Panda-UV and ClipsMS match fragments in MH+ mode and terminal fragment biased mode as described in a previous study.[26] Panda-UV only retains terminal fragments with PCC scores greater than 0.6 and internal fragments with PCC scores greater than 0.8 to ensure confidence.

Code Availability. All the codes are programmed in Python v3.7.16. Numpy v1.16.4, pandas v1.3.5, and pyteomics v4.5.6 are used to preprocess data. GUI are constructed based on PyQt5 v5.15.9. All the source codes of Panda-UV are freely available from GitHub (https://github.com/PHOENIXcenter/Panda-UV) under GNU General Public License version 3.0.

RESULTS AND DISCUSSION

Graphical User Interface (GUI) and Parameters of Panda-UV. Panda-UV is a software tool specifically designed for fragment matching of UVPD spectra, while the MS spectra obtained by using common dissociation methods are also compatible. It can match all known types of terminal and internal fragment ions generated by UVPD, scoring the matched fragments to yield highly confident results. With its user-friendly graphical interface (Figure S1) and result visualization (Figure S2), Panda-UV is easy to use even for researchers with few programing skills. Additionally, its command-line interface allows professional programmers to integrate Panda-UV as a module into other software tools for batch data processing.

Panda-UV can be run either by using a parameter file or by configuring parameters directly through its GUI (Figure S1). For interpreting UVPD spectra, the protein sequence and fixed modification are used as input of Panda-UV to generate all possible theoretical fragments according to the user-defined ion types to match the deconvoluted masses in a preset error

https://doi.org/10.1021/acs.analchem.4c00253
Anal. Chem. 2024, 96, 8474−8483

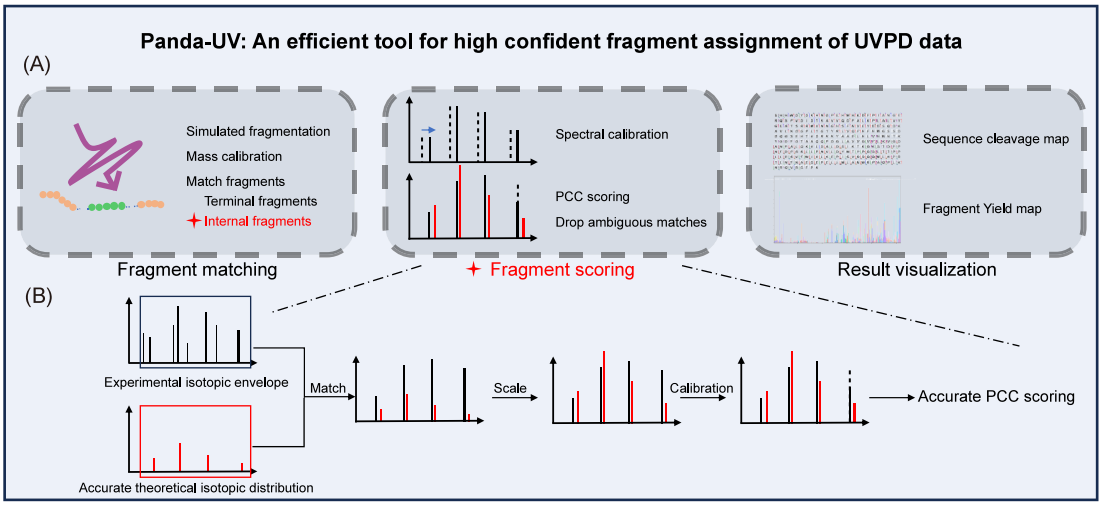


Figure 1. Workflow of Panda-UV. (A) Main steps of Panda-UV include fragment matching, fragment scoring, and result visualization. (B) Similarity between experimental isotopic envelopes extracted from spectrum and theoretical isotopic envelopes of matched fragments are calculated based on PCC to ensure confidence.

tolerance. For UVPD spectra of the protein−ligand complex, the monoisotopic masses of ligand are used as an unlocalized modification added to the theoretical fragments to assign the holo fragment ions.

The confidence of the assignment is estimated by calculating the similarity between experimental and theoretical isotopic envelopes through PCC scoring. The R package enviPat[35] is utilized for calculations of the theoretical isotopic envelopes based on the molecular formula and charge state of the fragment ion. Thus, the directory of the R environment is also used as an input parameter. We recommend installing R version 4.2.1 and the latest version of enviPat. The experimental isotopic envelopes of fragment ions are extracted from the specified mzML file within a m/z tolerance, defined as the "peaks match error". The functions named "Mass Calibration" and "MS Calibration" are also developed to calibrate the deconvoluted mass and m/z of experimental data, improving fragment matching and scoring.

After setting up the fragment matching error, ion types, and workspace path, the user can initiate program execution by clicking on the "Run" button. Upon completion, a fragment matching the result file along with four figures will be generated in your workspace path. These figures include fragment cleavage maps and bar plots of residual fragment yield for both terminal fragments and internal fragments (Figure S2). These figures are HTML files created using Plotly (v 2.24.1), and users can open them in a browser and save them in other formats if needed.

Workflow of Panda-UV. If the user has configured all parameters or uploaded a saved parameter file, then clicking the "Run" button will initiate Panda-UV, which will run as described in Figure 1A.

1. Fragment matching. Low mass error tolerance is contributed to decrease the false positive in fragment matching. However, the measured mass may shift from theoretical mass due to instrument systematic shift despite with the dedicated instrument calibration.[29] Panda-UV generates all possible theoretical fragments based on the protein sequence, modifications, and ion types (Figure 1A). It then matches these theoretical fragments with the deconvoluted mass list in two stages to eliminate the mass error. Initially, Panda-UV only matches the terminal ion types chosen by the user within a large error tolerance (20 ppm) and then calculates the average error (ppm) of all matched fragments (Figure 1A). This error (ppm level) is used to calibrate deconvoluted masses using the following calibration formula:

$$\text{mass\_calibrated} = \text{mass}/(1 + \text{error} \times 10^{-6}) \quad (1)$$

After calibration, a second round of fragment matching is performed on the calibrated deconvoluted masses by using user-defined error tolerance. Since the masses have been calibrated, Panda-UV can accurately match fragments within a low error tolerance, thereby preventing misassignments caused by an instrument systematic shift.

2. Fragment scoring. Generating internal fragments greatly expanded the search space in fragment matching. To accurate match fragments merely depending on mass error is insufficient despite low mass error Panda-UV provided with mass calibration. Here, we developed the similarity between experimental and theoretical isotopic envelopes based on PCC as auxiliary quality control to ensure positive fragment matching (Figure 1B). The theoretical isotopic envelopes of fragment ions are calculated by R package enviPat using the chemical formula and charge state of fragment ions as input. The experimental isotopic envelope is directly extracted from the mzML file within a user-defined m/z tolerance (named "peak match error" in GUI). The calibration of experimental m/z is recommended to increase the success rate and accuracy of extraction. The averaged deviations (ppm) of deconvoluted masses and m/z are found comparable for UVPD spectra collected by the orbitrap analyzer. Thus, the averaged mass error in fragment matching is also applied for m/z calibration similarly to that of mass in eq 1. The intensities of matched isotopic peaks are used for PCC scoring. While the isotopic peaks in theoretical envelop that missing in the UVPD spectrum, the intensity of this peak is set as 0 (Figure 1B). The theoretical isotopic peak intensities

https://doi.org/10.1021/acs.analchem.4c00253
Anal. Chem. 2024, 96, 8474−8483

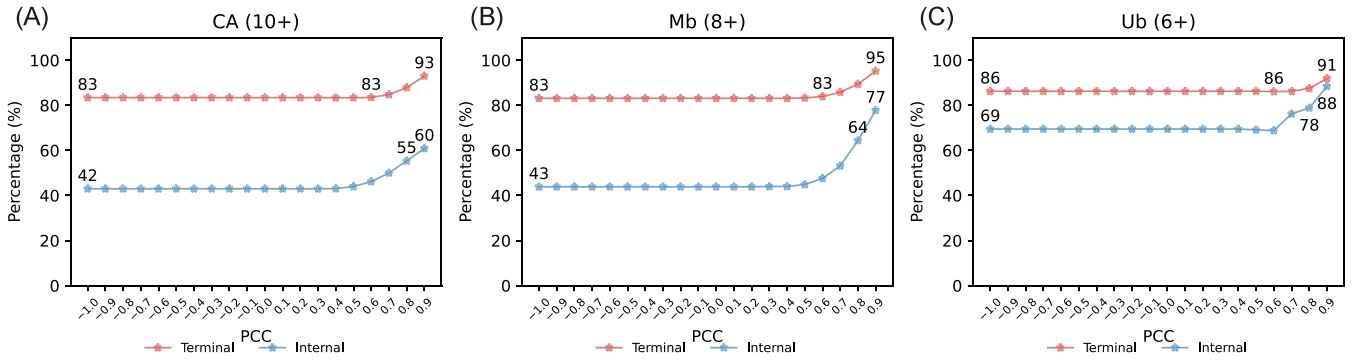


Figure 2. Effectiveness of PCC on percentage of retained in replicates was tested by Panda-UV for three model proteins, CA (10+), Mb (8+), and Ub (6+) described in the Experimental Section.

calculated by enviPat are relative values between 0 and 1, which need to be scaled to absolute intensity. The scale factor is the ratio of the sum of the three peak intensities in the experimental envelope to of the corresponding three top highest peak in the theoretical envelope (Figure 1B).

Despite the extraction of accurate experimental envelopes of matched fragments, the experimental peaks in the envelope may overlap with other envelopes of different fragments, leading to inaccurate PCC scoring. To alleviate this problem, Panda-UV calibrates the intensity of the experimental peak of matched fragment to ensure accurate PCC scoring (Figure 1B). The calibration formula is shown in following (eq 2), which based on the L-Score method:[36]

$$y'_{r}\_\text{calibrated} = \begin{cases} y'_{r}, |y'_{r} - y_{r}| \le t \\ y'_{r} + t, |y'_{r} - y_{r}| > t \text{ and } y'_{r} > y_{r} \\ y'_{r} - t, |y'_{r} - y_{r}| > t \text{ and } y'_{r} \le y_{r} \end{cases} \quad (2)$$

where t is the fixed relative intensity difference between matched theoretical and experimental peaks, in this study we set as 0.5 following a previous study.[36] y_r and y'_r are matched theoretical and experimental relative peak intensities, respectively. y'_r is obtained by scaling of experimental envelope peaks by the most abundant peak. The main idea of this method is that, when there is a large difference in relative intensity between a theoretical peak and experimental peak, the relative intensity difference between the two peaks is fixed as t.

The fitting scores of calibrated experimental envelopes and theoretical envelopes are calculated using PCC as the reliability score of matched fragments. The fragments with higher PCC scores will be proven with higher confidence in the following section.

3. Drop ambiguous matches. Due to the large theoretical search space, an experimental fragment may match multiple theoretical fragments within a user-defined mass error tolerance, in which the fragments with the lowest mass error are regarded as correct one. However, incorrect determination may occur with the instrument systematic shift caused expanding mass error for a correct match. Mass calibration helps reduce these incorrect matches to obtain more confident results. To disambiguate fragments with different masses or same masses but different formulas, mass error and PCC score are combined to drop the ambiguous matches. For example, if an experimental fragment matches both multiple terminal ions and internal ions at the same time, Panda-UV first keeps the terminal fragments with lower mass errors and larger PCC scores.
4. Result visualization. Panda-UV usually plots fragment cleavage maps and bar plots of the residual fragment yield for terminal and internal fragments based on the sequence and fragment matching result (Figure S2). If a protein backbone is cleaved and the fragments were matched by Panda-UV, a special symbol is inserted between the two amino acids at the cleavage site when plotting the fragmentation site map (Figure S2A,C). This symbol signifies that this site has been fragmented. Hovering over this symbol will display information about the type, mass, intensity, and charge of the matched fragment. If a backbone has been fragmented multiple times, all information about matched fragments will be fully displayed. This makes it convenient for users to verify whether the fragments have been correctly matched.

The horizontal axis of the bar plot of residual fragment yield represents potential fragmentation sites within the protein, while the vertical axis represents total ion intensity produced at each cleavage site (Figure S2B,D). Different fragments are represented by different colors. Similarly, hovering over these can display specific ion information for manual verification. Fragment matching results and all maps can be saved in a user-specified workplace.

Panda-UV Unlocks the Hidden Depth of Protein UVPD Spectra Using Internal Fragments. The low confidence of internal fragments caused by random matching limited the utilization of internal fragments in the UVPD analysis. Dunham et al. found only an average of 18% of UVPD internal fragments that were identified in two of three replicates based on several model proteins, compared to 63% for HCD internal fragments.[27] Panda-UV tackled this challenge by implementing spectral calibration and a PCC scoring system to filter low-confidence matches. To validate the effectiveness of PCC scoring and mass spectral calibration, Panda-UV was evaluated on the data sets of three model proteins, carbonic anhydrase II (CA, 29 kDa, 259 amino acids, 10+ charge state), holo-myoglobin (Mb, 17 kDa, 153 amino acids, 8+ charge state), and ubiquitin (Ub, 8 kDa, 76 amino acids, 6+ charge state) as described in the Experimental Section, in which all data sets contained three spectra obtained by replicated UVPD experiments. The percentages of fragments retained in multiple replicates were plotted as a function of PCC for three model proteins (Figure 2). The results

https://doi.org/10.1021/acs.analchem.4c00253
Anal. Chem. 2024, 96, 8474−8483

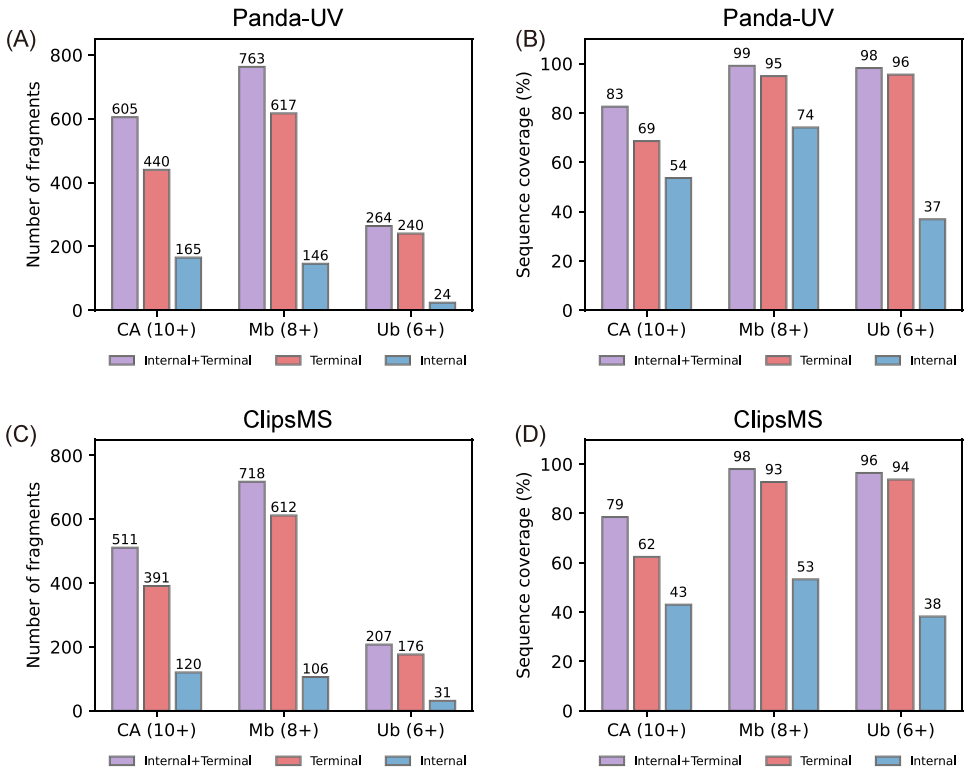


Figure 3. Number of fragments and sequence coverage provided by (A, B) Panda-UV and (C, D) ClipsMS for three model proteins, CA (10+), Mb (8+), and Ub (6+) described in the Experimental Section. Panda-UV only retained terminal fragments with PCC scores greater than 0.6 and internal fragments with PCC scores greater than 0.8 to ensure confidence.

demonstrated that, as the PCC increases, the percentages of UVPD fragments retained in replicates continued to rise for all model proteins, suggesting that the high-scoring fragments identified by Panda-UV had a higher confidence. Up to 60, 77, and 88% of internal fragments and 93, 95, and 91% of terminal fragments with PCC scores greater than 0.9 were identified in the three-replicated UVPD experiments of model proteins, respectively (Figure 2). The generation of a specific internal fragment among different experiments requires two identical fragmentation sites and ion types, while a terminal fragment needs only one identical fragmentation site and ion type; therefore, it is reasonable that the percentages of internal fragments retained in multiple replicates are lower than terminal fragments. The confidence in internal fragments was guaranteed by the high similarity between theoretical and experimental isotopic envelopes. To ensure the confidence of matched fragments, only terminal fragments with PCC greater than 0.6 and internal fragments with PCC greater than 0.8 were retained in the following analysis.

Mass accuracy is crucial for confident fragment matching. The instrument's systematic shift caused mass error is inevitable when acquiring mass spectra.[29] It is difficult to constantly maintain the instrument at an ideal 0 ppm shift. Spectral calibration compensates for the mass error caused by the instrument's systematic shift, allowing a strict mass error tolerance to improve the accuracy.[28,30] Systematic Most fragments were matched near to 0 ppm, so ±5 ppm error tolerance was allowed in fragment matching with spectral calibration (Figure S3). Although a strict mass tolerance error (±5 ppm in this study) was used, only an average of 42.7% experimental fragments is one-to-one matched (Figure S4A–C) before dropping ambiguous matches on three model proteins. The method for dropping ambiguous matches in Panda-UV ensures that 92.2% experimental fragments were uniquely matched (Figure S4D–F).

To further assess the robustness of Panda-UV, data sets of three model proteins, aldolase (35+), CA (25+), and Mb (15+) with three experimental replicates used in Dunham's study, were downloaded and tested using Panda-UV.[27] The results demonstrated that Panda-UV also guaranteed the confidence and accuracy of matched fragments (Figure S5). The percentage of internal fragments retained in multiple replicates increased to 61% for CA (25+) used in Dunham's study, by filtering out fragments with PCC scores lower than 0.9 (Figure S5B). This was significantly higher than Dunham's 14%.[27]

With the high confidence and accuracy provided by Panda-UV, including internal fragments boosts the number of identified fragments by 26% (reaching an average of 544 fragments) and enhances protein sequence coverage to an average of 93% on model proteins described in the Experimental Section (Figure 3A,B). ClipsMS was also tested on model proteins as a benchmark for Panda-UV (Figure 3C,D). The summary of the number of fragments, the percentage of fragments retained in multiple replicates, and the search time provided by Panda-UV and ClipsMS are listed in Table S1. Attributed to the abundant ion types that bypassed improbable fragment matching and deleting, Panda-UV significantly reduced the search space, thus obtaining more than 10% (432 vs 394) terminal fragments and 32% (112 vs 85) internal fragments (Table S1). The percentage of fragments retained in replicates increased from 82 to 84% for terminal fragments and from 49 to 65% for internal fragments (Table S1). Panda-UV increased both the number

https://doi.org/10.1021/acs.analchem.4c00253
Anal. Chem. 2024, 96, 8474−8483

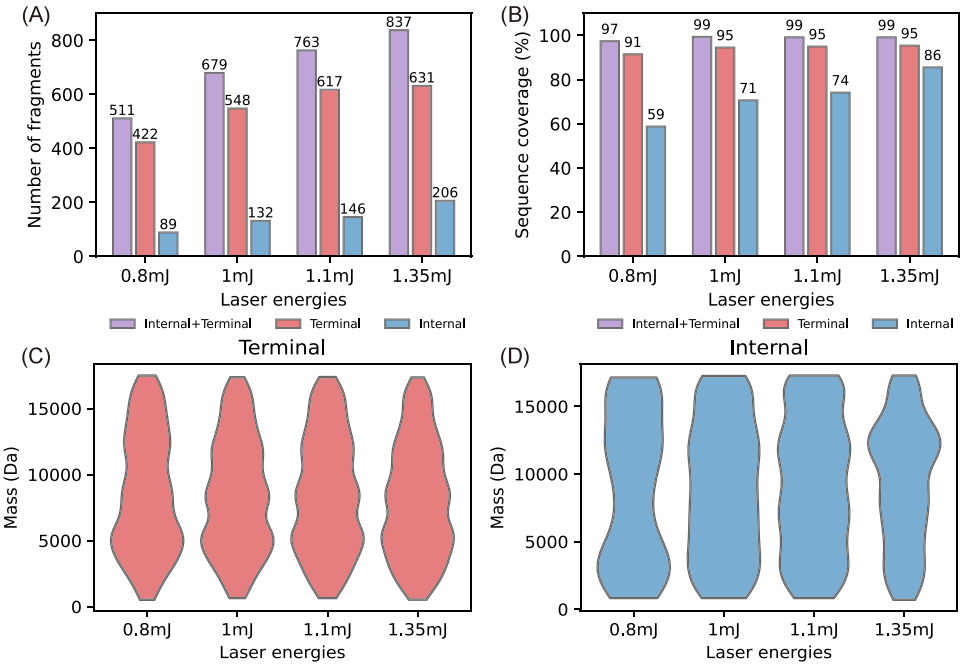


Figure 4. Model protein Mb (8+) described in the Experimental Section was fragmented by UPVD at different energies. The spectra were analyzed by a Panda-UV. (A) Fragment number, (B) sequence coverage, and (C, D) mass distribution of terminal fragments and internal fragments were plotted as a function of different laser energies for Mb (8+). Only the terminal fragments with PCC scores greater than 0.6 and the internal fragments with PCC scores greater than 0.8 were retained to ensure confidence.

and percentage of fragments retained in replicates, allowing for extensive exploration of the UVPD data. Panda-UV and ClipsMS performed similarly on HCD data regardless of whether the fragments were terminal or internal (Table S2).

Additionally, with the optimization of the code framework, Panda-UV took no more than 9 min to search one spectrum of a model protein, which is 50 times faster than ClipsMS (Table S1). Therefore, Panda-UV has a significant advantage when searching multiple spectra of large proteins simultaneously. An average of 11% signals (deconvoluted fragments) can be additionally explained by internal fragments using Panda-UV (Figure S6). For the largest protein, CA (10+), internal fragments cover most regions that cannot be reached by terminal fragments (Figure S2), increasing sequence coverage from 69 to 83% and explained signals from 50 to 67%. This suggests that internal fragments have the potential to deepen the exploration of the "dark matter" that has never been explored, aiding in the disambiguation of homologous proteoforms that have minor differences and cannot be identified solely with terminal fragments.

Impacts of Laser Energy on Protein Photofragmentation. Laser energy is a crucial parameter that influences the fragmentation of proteins in the UVPD experiments. High energy promotes secondary fragmentation of proteins,[37] leading to spectral congestion that decreases the reliability of fragment matching.[38,39] Mb (8+), a model protein described in the Experimental Section, was fragmented by UVPD at different energies and analyzed using Panda-UV. With increasing laser energy, the number of identified terminal fragments increased from 422 to 631, and internal fragments increased from 89 to 206 (Figure 4A). The total sequence coverage, provided by both terminal and internal fragments, increased from 97 to 99% (Figure 4B). On average, the percentage of internal fragments retained in multiple replicates exceeded 60%, indicating that Panda-UV ensures the confidence of fragment matching (Table S3). By using high-confidence fragment matching with Panda-UV, sequence coverage can be further improved by increasing the laser energy.

However, the number of large fragment masses, especially internal fragments, increased with laser energy (Figure 4C,D). This conflicted with the well-known transformation of large fragments to small ones at high energies.[27,40] Mb (17+) was fragmented by UVPD at different energies and analyzed using Panda-UV. The number of fragments and sequence coverage increased to a lesser extent with laser energy for Mb (17+) compared to Mb (8+) (Figure S7A,B). The number of large fragments also increased with laser energy, which corresponded with the findings of the previous study (Figure S7C,D).[27,40] The Orbitrap analyzer has a higher detection sensitivity of fragment ions with low molecular weight (MW) compared to those with high MW. This inconsistency may be partially explained that the low Mb (8+) depletion (Figure S8A) produced low intensity fragments and the complete Mb (17+) depletion (Figure S8B) produced high intensity fragments.

Panda-UV Improves Monoclonal Antibody Sequencing Coverages. Monoclonal antibodies (mAbs) were frequently used as highly specific targeted therapeutic drugs.[41−43] The therapeutic effect of mAbs depends on their primary structure, making it crucial to characterize their structure for quality control.[44,45] mAbs are composed of two identical heavy chains (Hc, ∼50 kDa) and two identical light chains (Lc, ∼25 kDa).[46] The heavy and light chains were connected by disulfide bonds, with additional disulfide bonds present within these chains, making the structure of monoclonal antibodies highly complex. The most popular bottom-up strategy usually has low sequence coverage and is not suitable for sequencing of mAbs.[47] Currently, no MS fragmentation method can effectively fragment and sequence

https://doi.org/10.1021/acs.analchem.4c00253
Anal. Chem. 2024, 96, 8474−8483

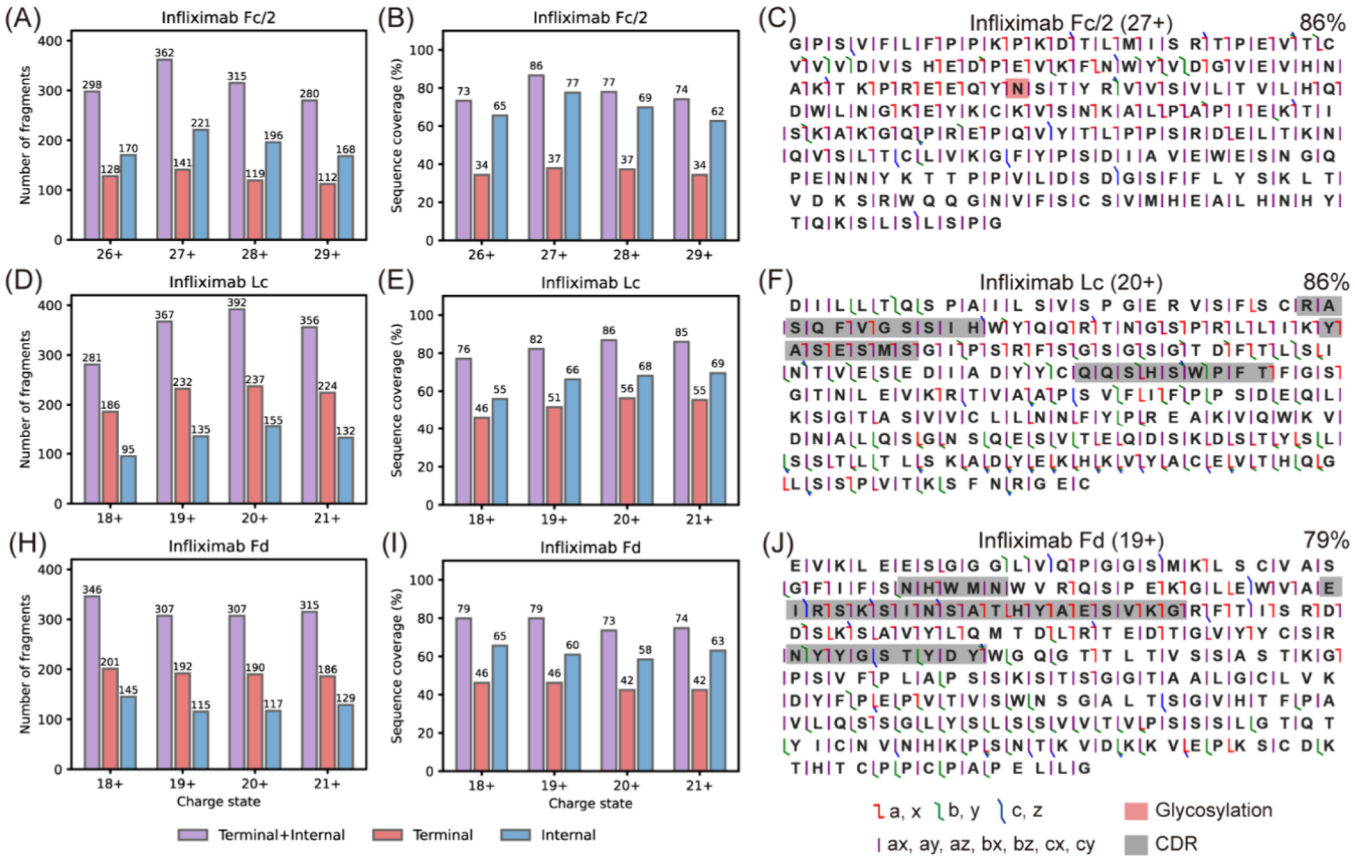


Figure 5. Number of fragments (A, D, H) and sequence coverage (B, E, I) for Fc/2, Lc, and Fd of 4 charge states of infliximab provided by terminal and internal fragments. Fragmentation site maps provided (C, F, J) by terminal fragments and internal fragments for Fc/2 (27+), Lc (20+), and Fd (19+) of infliximab. Only the terminal fragments with PCC scores greater than 0.6 and the internal fragments with PCC scores greater than 0.8 were retained to ensure confidence.

mAbs, making the top-down strategy unsuitable for antibody identification as well.[48] Middle-down compensates for the shortcomings of bottom-up and top-down to identify proteins with complex structures, making it a suitable strategy for mAb sequencing.[49]

Infliximab, a therapeutic monoclonal IgG1-type antibody used to treat autoimmune diseases, has a thoroughly studied structure. By using the middle-down strategy to obtain three subunits of infliximab described in the Experimental Section, we obtained raw spectra of 4 charge states ranging from 26+ to 29+ for Fc/2 (heavy chain Fc monomer), 18+ to 21+ for Lc (light chain), and Fd (heavy chain variable domain). UVPD fragmented the subunits to obtain an MS/MS spectrum for all target charge states. Each spectrum was obtained by averaging five hundred times. The process of the infliximab sample is illustrated in Figure S9, following a previous study.[50]

Spectra across all charge states of subunits of infliximab were analyzed by Panda-UV. Number of fragments and sequence coverage for Fc/2, Lc, and Fd of 4 charge states and fragmentation site maps for Fc/2 (27+), Lc (20+), and Fd (19+) are plotted in Figure 5. After inclusion of internal fragments, the average number of identified fragments increased from 178 to 326 (Figure 5A,D,H), and the sequence coverage increased from 44 to 78% (Figure 5B,E,I). In terms of quantity, internal fragments can even outnumber terminal fragments for Fc/2, indicating that UVPD generated abundant internal fragments (Figure 5A). Internal fragments increased the sequence coverage from 37 to 86% (Figure 5B) and accurately bracketed the glycosylation modification that cannot be localized by terminal fragments alone for Fc/2 (27+) (Figure 5C). For Lc (20+) and Fd (19+), internal fragments increased the sequence coverage from 56 to 86% and 46 to 79%, respectively (Figure 5E,I), in which six CDRs (complementary determining regions)[51,52] that are crucial for the characterization of mAbs were nearly completely sequenced (totally on 87%) (Figure 5F,J). Only terminal fragments with PCC scores greater than 0.6 and terminal fragments with PCC scores greater than 0.8 were retained to ensure confidence. Researchers can increase the PCC to 0.9 to get more confident results.

Previously, due to concerns about the confidence of fragment assignment, internal fragments in UVPD were ignored. However, Panda-UV can accurately and reliably match both terminal and internal fragments produced by UVPD. By including internal fragments, it significantly improved subunit sequence coverages. Existing methods that use multiple UVPD experiments of different parameters increased the subunit sequence coverages up to 85%.[49] This process is time-consuming and costly, and it also involves complex subsequent data processing. However, by including internal fragments with Panda-UV, the sequence coverages of subunits can be increased up to 86% and the CDRs were nearly completely sequenced in a single experiment. This approach fully leveraged the ability of UVPD to produce rich fragment ions, achieving high sequence identification coverage

https://doi.org/10.1021/acs.analchem.4c00253
Anal. Chem. 2024, 96, 8474−8483

at the algorithm level while improving the sequencing throughput.

CONCLUSIONS

UVPD mass spectrometry revolutionized protein analysis by offering near-complete sequence cleavages for proteins <30 kDa. However, the vast potential of internal fragments, the majority of MS/MS signals, remained largely untapped due to computational challenges and unreliable assignments. This study introduced Panda-UV, a novel software tool specifically designed to conquer these hurdles and unlocked the power of internal fragments in UVPD data analysis. Panda-UV tackled the issue of low confidence through a two-pronged approach: spectral calibration for enhanced mass accuracy and a Pearson correlation coefficient (PCC) scoring system for filtering low-confidence matches. This strategy increased the confidence of matched internal fragments up to remarkable 65% at the algorithm level. Benchmarked against ClipsMS, the existing internal fragment assignment software, revealed advantages of Panda-UV. Panda-UV consistently identified 65 more fragments on average, with more than 16% internal fragments retained in replicates and 50 times faster computation speed. Moreover, its efficacy extended beyond our data sets, improving fragment confidence in data sets used in Dunham's study as well.[27] Finally, Panda-UV shined in real-world applications. When analyzing monoclonal antibodies, it doubled the number of identified fragments and boosted sequence coverage to 86%, significantly enhancing the identification accuracy of mAbs. In conclusion, Panda-UV empowered researchers to unleash the full potential of internal fragments in UVPD data analysis. With its ability to deliver highly reliable fragment assignments, Panda-UV unlocked the hidden depth of protein characterization, paving the way for transformational advancements in proteomics research.

However, there are several challenges that should be addressed in the future. In this study, about 7.8% of the experimental fragments matched multiple theoretical fragments that have the same chemical formula. These matches may be disambiguated by "frameshift assignments"[23] and fragmentation propensity.[53−55] Furthermore, this study focuses on fragment matching at the algorithmic level. A comprehensive method should be developed by integrating advanced MS techniques and bioinformatic algorithms, for instance, PTCR[56−58] and new deconvolution algorithm to explore the complex UVPD spectra in the future.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.4c00253.

Graphical user interface (GUI) of Panda-UV, result visualization of Panda-UV, effectiveness of PCC and spectral calibration, mass error density plots of the fragments matched by ClipsMS, explained signal provided by Panda-UV, result summary of model proteins tested upon Panda-UV, result summary of model proteins tested upon ClipsMS, result summary of Mb (8+) at different energies tested upon Panda-UV, number of fragments and masses distribution of Mb (17+) at different energies tested upon Panda-UV, depletions of Mb (8+) and Mb (17+) at different energies, and process steps of infliximab using middle-down strategy (PDF)

AUTHOR INFORMATION

Corresponding Authors

Fangjun Wang − *CAS Key Laboratory of Separation Sciences for Analytical Chemistry, Dalian Institute of Chemical Physics and State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China; University of Chinese Academy of Sciences, Beijing 100049, China;* orcid.org/0000-0002-8118-7019; Email: wangfj@dicp.ac.cn

Cheng Chang − *State Key Laboratory of Medical Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China;* orcid.org/0000-0002-0361-2438; Email: changchengbio@163.com

Authors

Yinlong Zhu − *Chongqing Key Laboratory on Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; State Key Laboratory of Medical Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China; CAS Key Laboratory of Separation Sciences for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China*

Zheyi Liu − *CAS Key Laboratory of Separation Sciences for Analytical Chemistry, Dalian Institute of Chemical Physics and State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China; University of Chinese Academy of Sciences, Beijing 100049, China*

Jialiang Liu − *CAS Key Laboratory of Separation Sciences for Analytical Chemistry, Dalian Institute of Chemical Physics and State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China; School of Pharmacy, China Medical University, Shenyang 110122, China*

Heng Zhao − *CAS Key Laboratory of Separation Sciences for Analytical Chemistry, Dalian Institute of Chemical Physics and State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China*

Rui Feng − *State Key Laboratory of Medical Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China*

Kunxian Shu − *Chongqing Key Laboratory on Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.4c00253

Author Contributions

[▽]Y.Z. and Z.L. contributed equally to this paper. F.W and C.C. designed and cosupervised the study. Y.Z. wrote the first draft of the manuscript, developed the Panda-UV, and analyzed data. Z.L. conducted the experiments and assisted the data analysis. H.Z. prepared the sample of model proteins. J.L.

https://doi.org/10.1021/acs.analchem.4c00253
Anal. Chem. 2024, 96, 8474−8483

prepared the sample of mAbs. Y.Z., Z.L., R.F., K.S., F.W., and C.C. revised the manuscript. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work has been supported by the National Key Research and Development Program of China (2022YFA1304601 and 2021YFA1301603), the National Natural Science Foundation of China (32088101, 92253304, and 22288201), and the grant from DICP (DICPI202242 and DICPI202228). The authors acknowledge the technological support of the biological mass spectrometry station of Dalian Coherent Light Source.

■ REFERENCES

(1) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. *Biochem. Biophys. Res. Commun.* **2014**, *445* (4), 683−693.
(2) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121* (4), 806−812.
(3) Toby, T. K.; Fornelli, L.; Kelleher, N. L. *Annu. Rev. Anal. Chem.* **2016**, *9*, 499−519.
(4) Bogdanov, B.; Smith, R. D. *Mass Spectrom. Rev.* **2005**, *24* (2), 168−200.
(5) Chait, B. T. *Science* **2006**, *314* (5796), 65−66.
(6) Hollas, M. A.; Robey, M. T.; Fellers, R. T.; LeDuc, R. D.; Thomas, P. M.; Kelleher, N. L. *Nucleic Acids Res.* **2022**, *50* (D1), D526−D533.
(7) Drown, B. S.; Jooß, K.; Melani, R. D.; Lloyd-Jones, C.; Camarillo, J. M.; Kelleher, N. L. *J. Proteome. Res.* **2022**, *21* (5), 1299−1310.
(8) Mitchell Wells, J.; McLuckey, S. A. *Methods Enzymol.* **2005**, *402*, 148−185.
(9) O'Brien, J. P.; Li, W.; Zhang, Y.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2014**, *136* (37), 12920−12928.
(10) Brodbelt, J. S. *Anal. Chem.* **2016**, *88* (1), 30−51.
(11) Brodbelt, J. S.; Morrison, L. J.; Santos, I. *Chem. Rev.* **2020**, *120* (7), 3328−3380.
(12) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *Anal. Chem.* **2014**, *86* (4), 2185−2192.
(13) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646−12651.
(14) Lyon, Y. A.; Riggs, D.; Fornelli, L.; Compton, P. D.; Julian, R. R. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (1), 150−157.
(15) Zenaidee, M. A.; Lantz, C.; Perkins, T.; Jung, W.; Loo, R. R. O.; Loo, J. A. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (9), 1896−1902.
(16) Cobb, J. S.; Easterling, M. L.; Agar, J. N. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (6), 949−959.
(17) Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (5), 782−787.
(18) Wei, B.; Zenaidee, M. A.; Lantz, C.; Williams, B. J.; Totten, S.; Ogorzalek Loo, R. R.; Loo, J. A. *Analyst* **2023**, *148* (1), 26−37.
(19) Wei, B.; Lantz, C.; Liu, W.; Viner, R.; Ogorzalek Loo, R. R.; Campuzano, I. D.; Loo, J. A. *Anal. Chem.* **2023**, *95* (24), 9347−9356.
(20) Wongkongkathep, P.; Han, J. Y.; Choi, T. S.; Yin, S.; Kim, H. I.; Loo, J. A. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (9), 1870−1880.
(21) Nshanian, M.; Lantz, C.; Wongkongkathep, P.; Schrader, T.; Klärner, F. G.; Blümke, A.; Despres, C.; Ehrmann, M.; Smet-Nocca, C.; Bitan, G.; Loo, J. A. *J. Am. Soc. Mass Spectrom.* **2018**, *30* (1), 16−23.
(22) Lantz, C.; Wei, B.; Zhao, B.; Jung, W.; Goring, A. K.; Le, J.; Miller, J.; Loo, R. R. O.; Loo, J. A. *J. Am. Chem. Soc.* **2022**, *144* (48), 21826−21830.
(23) Schmitt, N. D.; Berger, J. M.; Conway, J. B.; Agar, J. N. *Anal. Chem.* **2021**, *93* (16), 6355−6362.
(24) Morrison, L. J.; Chai, W.; Rosenberg, J. A.; Henkelman, G.; Brodbelt, J. S. *Phys. Chem. Chem. Phys.* **2017**, *19* (30), 20057−20074.
(25) Morrison, L. J.; Rosenberg, J. A.; Singleton, J. P.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (9), 1443−1453.
(26) Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A. *J. Proteome. Res.* **2021**, *20* (4), 1928−1935.
(27) Dunham, S. D.; Wei, B.; Lantz, C.; Loo, J. A.; Brodbelt, J. S. *J. Proteome. Res.* **2023**, *22* (1), 170−181.
(28) Solntsev, S. K.; Shortreed, M. R.; Frey, B. L.; Smith, L. M. *J. Proteome. Res.* **2018**, *17* (5), 1844−1851.
(29) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. *Mol. Cell. Proteomics* **2005**, *4* (12), 2010−2021.
(30) Cox, J.; Michalski, A.; Mann, M. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (8), 1373−1380.
(31) Zhou, L.; Liu, Z.; Guo, Y.; Liu, S.; Zhao, H.; Zhao, S.; Xiao, C.; Feng, S.; Yang, X.; Wang, F. *J. Am. Chem. Soc.* **2023**, *145* (2), 1285−1291.
(32) Liu, Z.; Qin, Z.; Cui, C.; Luo, Z.; Yang, B.; Jiang, Y.; Lai, C.; Wang, Z.; Wang, X.; Fang, X.; Li, G.; Wang, F.; Xiao, C.; Yang, X. *Sci. China Chem.* **2022**, *65* (6), 1196−1203.
(33) Sun, B.; Liu, Z.; Fang, X.; Wang, X.; Lai, C.; Liu, L.; Xiao, C.; Jiang, Y.; Wang, F. *Anal. Chim. Acta* **2021**, *1155*, No. 338340.
(34) Basharat, A. R.; Zang, Y.; Sun, L.; Liu, X. *Anal. Chem.* **2023**, *95* (21), 8189−8196.
(35) Loos, M.; Gerber, C.; Corona, F.; Hollender, J.; Singer, H. *Anal. Chem.* **2015**, *87* (11), 5738−5744.
(36) Kou, Q.; Wu, S.; Liu, X. *BMC Genomics* **2014**, *15*, 1−10.
(37) Greer, S. M.; Sidoli, S.; Coradin, M.; Schack Jespersen, M.; Schwämmle, V.; Jensen, O. N.; Garcia, B. A.; Brodbelt, J. S. *Anal. Chem.* **2018**, *90* (17), 10425−10433.
(38) Sanders, J. D.; Mullen, C.; Watts, E.; Holden, D. D.; Syka, J. E.; Schwartz, J. C.; Brodbelt, J. S. *Anal. Chem.* **2020**, *92* (1), 1041−1049.
(39) Dunham, S. D.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2024**, *35* (2), 255−265.
(40) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. *Anal. Chem.* **1994**, *66* (18), 2809−2815.
(41) Beck, A.; Wurch, T.; Bailly, C.; Corvaia, N. *Nat. Rev. Immunol.* **2010**, *10* (5), 345−352.
(42) Liu, J. K. *Ann. Med. Surg.* **2014**, *3* (4), 113−116.
(43) Weiner, L. M.; Surana, R.; Wang, S. *Nat. Rev. Immunol.* **2010**, *10* (5), 317−327.
(44) Harding, F. A.; Stickler, M. M.; Razo, J.; DuBridge, R. The immunogenicity of humanized and fully human antibodies: residual immunogenicity resides in the CDR regions. In *MAbs*; Taylor & Francis: 2010 Vol. 2, pp 256−265.
(45) De Groot, A. S.; Scott, D. W. *Trends in immunology* **2007**, *28* (11), 482−490.
(46) Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. *Anal. Chem.* **2014**, *86* (6), 3005−3012.
(47) Schulte, D.; Peng, W.; Snijder, J. *Anal. Chem.* **2022**, *94* (29), 10391−10399.
(48) Shaw, J. B.; Liu, W.; Vasil'ev, Y. V.; Bracken, C. C.; Malhan, N.; Guthals, A.; Beckman, J. S.; Voinov, V. G. *Anal. Chem.* **2019**, *92* (1), 766−773.
(49) Cotham, V. C.; Brodbelt, J. S. *Anal. Chem.* **2016**, *88* (7), 4004−4013.
(50) Sun, M. C.; Chen, C. D.; Huang, Y. S.; Wu, Z. S.; Ho, Y. P. *J. Sep. Sci.* **2008**, *31* (3), 538−547.
(51) Alhazmi, H. A.; Albratty, M. *Pharmaceuticals* **2023**, *16* (2), 291.
(52) Wang, W.; Singh, S.; Zeng, D. L.; King, K.; Nema, S. *J. Pharm. Sci.* **2007**, *96* (1), 1−26.
(53) Ives, A. N.; Su, T.; Durbin, K. R.; Early, B. P.; dos Santos Seckler, H.; Fellers, R. T.; LeDuc, R. D.; Schachner, L. F.; Patrie, S. M.; Kelleher, N. L. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (7), 1398−1409.

https://doi.org/10.1021/acs.analchem.4c00253
Anal. Chem. 2024, 96, 8474−8483

(54) Haverland, N. A.; Skinner, O. S.; Fellers, R. T.; Tariq, A. A.; Early, B. P.; LeDuc, R. D.; Fornelli, L.; Compton, P. D.; Kelleher, N. L. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (6), 1203−1215.
(55) Macias, L. A.; Sipe, S. N.; Santos, I. C.; Bashyal, A.; Mehaffey, M. R.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2021**, *32* (12), 2860−2873.
(56) Weisbrod, C. R.; Anderson, L. C.; Hendrickson, C. L.; Schaffer, L. V.; Shortreed, M. R.; Smith, L. M.; Shabanowitz, J.; Hunt, D. F. *Anal. Chem.* **2021**, *93* (26), 9119−9128.
(57) Holden, D. D.; McGee, W. M.; Brodbelt, J. S. *Anal. Chem.* **2016**, *88* (1), 1008−1016.
(58) Kline, J. T.; Mullen, C.; Durbin, K. R.; Oates, R. N.; Huguet, R.; Syka, J. E. P.; Fornelli, L. *J. Am. Soc. Mass Spectrom.* **2021**, *32* (9), 2334−2345.

https://doi.org/10.1021/acs.analchem.4c00253
Anal. Chem. 2024, 96, 8474−8483