Electronic supplementary information (ESI)

Real-Time Bottom-Up Characterization of Protein Mixtures Enabled by Online Microdroplet-Assisted Enzymatic Digestion (MAED)

Cheng-Hua Ma, Chih-Lin Chen, and Cheng-Chih Hsu*

Department of Chemistry, National Taiwan University, Taipei, 106216, Taiwan
*Correspondence: ccrhsu@ntu.edu.tw

Table of contents

Materials and methods

Chemicals and reagents

Horse cytochrome C (C2506), horse myoglobin (M0630), bovine hemoglobin (contains subunit α and β, H2500), bovine carbonic anhydrase (C2624), bovine serum albumin (P5619), SILu™Lite SigmaMAb Universal Antibody Standard (MSQC4), sodium chloride (NaCl), sodium deoxycholate, ammonium hydroxide solution (NH_4OH), ammonium bicarbonate (ABC), and iodoacetamide (IAM) were purchased from Sigma-Aldrich. Acetonitrile (ACN), urea, chloroform ($CHCl_3$), and sodium dodecyl sulfate (SDS) were purchased from J.T. Baker. Formic acid (FA) was purchased from Honeywell Fluka. Porcine pepsin was purchased from Roche. Triton X-100 was purchased from Emperor Chemical. Tris was purchased from Bio basic. OneQuant ProteaseArrest and ethylenediaminetetraacetic acid (EDTA) were purchased from G-Biosciences. Methanol (MeOH) was purchased from Duksan reagents. Dithiothreitol (DTT) was purchased from Cyrus Bioscience. Deionized water (H_2O) was purified from ELGA purification system. Cycs standard was prepared in H_2O/ACN/FA (100-*x*:*x*:0.1), and the percentage of ACN (*x*) ranges from 0 to 50 as described in the figures. Pepsin was prepared in H_2O/FA (100:0.1). The samples for online MAED and conventional top-down analysis were prepared in H_2O/ACN/FA (95:5:0.1). Urea, DTT, and IAM were prepared in H_2O solution of 25 mM ABC. The samples for conventional bottom-up analysis were prepared/reconstituted in H_2O /FA (100:0.1) before digestion.

Homemade ESSI sprayer

The homemade ESSI sprayer (**Figure S1**) was built similarly as described elsewhere.[1] Teflon tubing sleeve (ID 0.0155 in., OD 1/16 in., IDEX Health & Science) was used to fix the fused-silica capillary (ID 250 μm, OD 360 μm, Polymicro Technologies) to the metal tee and seal the back end of it, whereas a stainless-steel capillary (ID 0.020 in., OD 1/16 in., IDEX Health & Science) was attached to the front end, allowing the fused-silica capillary to go through. Coaxial N_2 sheath gas could be delivered through the spacing between the stainless-steel capillary and the fused-silica capillary.

Direct infusion setup

The experiment setup for direct infusion experiments was shown in **Figure S2**. Briefly, 500 ppm Cycs and 25 ppm pepsin were infused with 2 separate syringes. The two solutions were mixed in IDEX micro static mixing tee and sprayed from the tip of a homemade ESSI sprayer with 120 psi of N_2 sheath gas. The flow rate of Cycs was fixed at 5 μL/min. The flow rate of the pepsin solution and the distance between the ESSI sprayer and MS inlet were varied as stated in the article and figures. High voltage was applied by attaching the alligator clip to a stainless-steel union. Spectra were recorded with an Orbitrap Elite mass spectrometer (Thermo Scientific). The source parameters were set as follows: source voltage +3 kV, capillary temperature 320℃, S-lens RF level 60%. The m/z range for full mass scan was 350 to 2000. Data-dependent MS/MS (ddMS2) spectra were acquired for the 10 most intense ions in a full mass scan by 3 m/z isolation width. Ions were fragmented via collision-induced dissociation (CID) with a normalized collision energy (NCE) of 35. Ions with unidentified charge or charge of +1 were excluded for ddMS2. Dynamic exclusion duration was 36 sec with a repeat count of 1. The auto gain control (AGC) targets were 1E6 and 2E5 for full mass scan and ddMS2, while the maximum injection time for full mass scan and ddMS2 were 300 ms (1 microscan) and 500 ms (1

microscan). The resolution for full mass scan and ddMS2 were 60000 and 30000. Data were acquired for 1 min, and the full mass spectra were averaged for further analysis.

Online MAED setup

The experiment setup for online MAED was depicted in **Figure S3**. Reversed-phased liquid chromatography (RPLC) was performed on nanoACQUITY UPLC system (Waters). Samples were injected (1 μL) by partial-loop mode and separated using Waters nanoEase M/Z Protein BEH C4 Column (300 μm × 5 cm, 1.7 μm, 300 Å pore size). The flow rate was set at 5 μL/min, and the column temperature was maintained at 45°C. Mobile phase A was H_2O/ACN/FA (95:5:0.1), mobile phase B was H_2O/ACN/FA (5:95:0.1). The gradient was set as follows: 5% B at 5 min, 15% B at 7.5 min, 55% B at 27.5 min, 95% B at 28.5 min, 95% B at 33.5 min, 5% B at 35 min, 5% B at 40 min. The eluent of RPLC was mixed with pepsin (infused at a flowrate of 15 μL/min) in IDEX micro static mixing tee, sprayed from ESSI sprayer with 120 psi of nitrogen nebulizing gas, and analyzed directly with Orbitrap Elite mass spectrometer (Thermo Scientific). The distance between the ESSI sprayer and the MS inlet was fixed at 0.5 cm. High voltage was applied by attaching the alligator clip to a stainless-steel union. The instrument parameters for MS analysis were the same as stated in the previous section. To prevent backflow when switching the valve upon sample injection, pepsin infused at the first and last 30 sec of the gradient was diverted to waste.

Conventional bottom-up and top-down analysis setup

The column was directly connected to the ESSI sprayer without additional infusion as shown in **Figure S4**. For conventional bottom-up analysis, the gradient was set as follows: 5% B at 5 min, 25% B at 22.5 min, 55% B at 37.5 min, 95% B at 38.5 min, 95% B at 43.5 min, 5% B at 45 min, 5% B at 50 min. The instrument parameters for MS analysis were the same as stated in the previous section. For conventional top-down analysis, the conditions and parameters for reversed-phase separation remained the same in the previous section. There were some modifications in the parameters regarding MS analysis. ddMS2 were acquired for the 5 most intense ions in a full mass scan by 3 m/z isolation width. The maximum injection time was 500 ms (2 microscans) and 1000 ms (1 microscan) for full mass scan and ddMS2 respectively. The resolution for full mass scan and ddMS2 were 120000 and 60000. The parameters not mentioned remained the same.

Bulk digestion

Fifty microliters of pepsin solution was added to 100 μL protein solution (protein-enzyme ratio of 20:1) and reacted at 37°C for bulk digestion. Twenty-five microliters of 10% NH_4OH solution was later on added for quenching (pH > 10), followed by the addition of 25% FA for reacidification (pH < 4). Since the sample volume had doubled, the original sample concentration for bulk digestion was twice the concentration of MAED and online MAED.

Mouse brain extract preparation

Mouse brain (purchased from BioLASCO) was snap-froze in liquid nitrogen and ground into a fine powder with a ceramic mortar and pestle. Ground tissue was resuspended with RIPA lysis buffer (NaCl 150 mM, Tris-HCl 50 mM, Triton X-100 1%, SDS 0.1%, Sodium deoxycholate 0.5%, EDTA 5 mM, OneQuant ProteaseArrest 1%) with 5 times the volume of sample weight (1 mg : 5 μL). The resuspended sample was vortexed for 2 hr (4°C) and the supernatant was collected after centrifugation at 14000 g for 30 min (4°C).

The total protein concentration was 15.8 μg/μL determined by BCA assay. The sample was then cleaned by precipitation using MeOH, $CHCl_3$, and ddH_2O described elsewhere.[2, 3] The sample was first diluted to a concentration of approximately 6 μg/μL (100 μL). Four hundred microliters of MeOH, 100 μL of $CHCl_3$, and 300 μL of ddH_2O were sequentially added to the sample with vortexing after each solvent addition. The protein pellet was formed in the middle of 2 solvent layers, and the upper layer was aspirated. The pellet was then washed 2 times with 400 μL of MeOH, dried briefly in a fume hood, and prechilled under -20 ℃. The pellet was resolubilized by adding 20 μL ice-cold 80% FA and incubating the sample under -20℃ for 15 min.[3, 4] The sample was diluted 10 times with H_2O/ACN/FA (95:5:0.1) and stored under -20℃ before analysis. For the sample subjected to conventional bottom-up analysis, the pellet was resolubilized with 10 μL ice-cold 80% FA and diluted 10 times with H_2O/FA (100:0.1).

Reduction and alkylation of BSA and MAb

One hundred microliters of sample was buffer exchanged to 25 mM ABC with 6M urea using a 10 kDa Amicon Ultra-0.5 centrifugal filter. For reduction, the samples were treated with 20 mM of DTT and reacted at 30℃ for 45 min. Alkylation was carried out by treating the samples with 60 mM of IAM and reacted at room temperature for 60 min in the dark. The reaction was quenched by adding DTT to approximately 60 mM and reacted at 30℃ for 45 min. The alkylated samples were then buffer exchanged back to H_2O/ACN/FA (95:5:0.1).

Data processing

Data was viewed with Xcalibur Qual Browser (version 3.0.63, Thermo Scientific). Mass spectrum deconvolution was performed with UniDec.[5] Boxcar averaging (3 points) was applied to all chromatograms.

MaxQuant[6] (version 2.0.3.1) was used for the identification of peptides in digested samples. Raw files were directly imported and searched against horse proteome (289 entries, downloaded on 2022-09-23), bovine proteome (6035 entries, downloaded on 2022-09-23), MAb sequence, and mouse proteome (17138+1 entries, porcine pepsin appended, downloaded on 2022-10-18). The FASTA proteome databases were downloaded from Uniprot, and the entries contained have all been reviewed. The false discovery rate (FDR) cutoff for peptide spectrum match (PSM) and protein match were both set at 0.01, estimated by target-decoy approach with decoy mode set as "revert". Acetylation (N-terminal) and oxidation (M) were set as variable modification. Carbamidomethylation (C) was set as fixed modification for alkylated samples, and pyroglutamic acid (E) was set as variable modification only for MAb. Other search parameters were set as follows: digestion mode unspecific, first search peptide tolerance 20 ppm, main search peptide tolerance 5 ppm, minimum peptide length 7, maximum peptide length 85, maximum peptide mass 9000 Da. "Match between runs" was enabled for mouse brain extract with a match time window set at 0.5 min. The search parameters not mentioned were set as default. For the identification results of mouse brain extract, the protein group of pepsin and the peptides belonging to that group were filtered out. The first entry of the gene name was used as the label for the protein group.

The identification of intact proteins in mouse brain extract (both undigested and online digested) and peptides of Cycs was conducted with TOPPIC suite[7] (version 1.4). Raw files were converted to mzml files by MSConvert[8] and then imported into TopFD for deconvolution and feature detection. The parameters for TopFD were set as default. The msalign files generated were input into TOPPIC for proteoform identification. Data were searched against horse proteome (289 entries) and mouse proteome (17138 entries) mentioned in

the previous paragraph. The false discovery rates (FDR) for spectrum and proteoform level were both set at 0.01, which were estimated by target-decoy approach. The search parameters were set as follows: maximum number of mass shift 1, maximum mass shift 1000, minimum mass shift -500. The search parameters not mentioned were set as default. The +615 mass shift for Cycs within the range of Cys15 to Cys18 was identified as heme. Proteoforms/peptides with other unexpected mass shifts or E values lower than 0.01 were filtered out.

For the identification results for online MAED of mouse brain extract (MaxQuant), the undigested proteins were picked out by matching the identification results from undigested and online digested conditions. The overlapping entries were the proteins that remained undigested under online MAED. The identifications of undigested proteins in MaxQuant were filtered out from the downstream analysis. Venn diagrams were generated via EVenn.[9]

Data availability

Raw spectrum files were uploaded to the MassIVE repository with the identifier MSV000092201 (password: Ntu33661681).

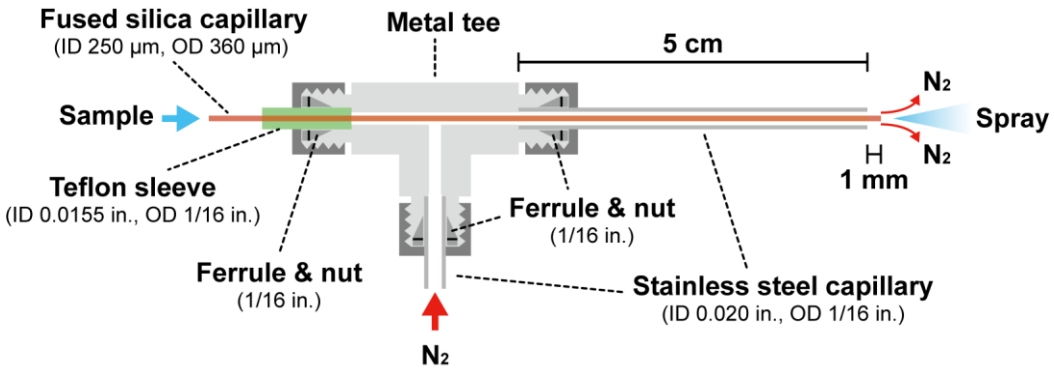

Figure S1. The detailed structure of the homemade ESSI sprayer.

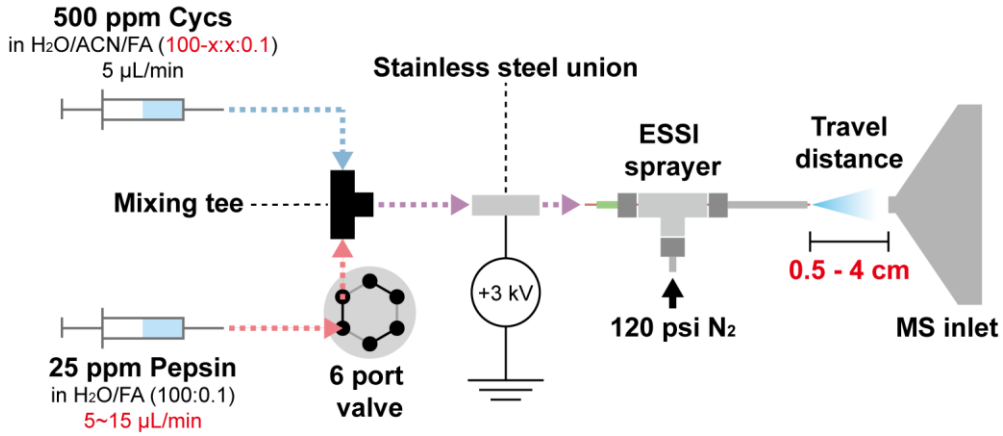

Figure S2. Instrument setup for direct infusion experiments. The experimental conditions changed for optimization were labeled in red.

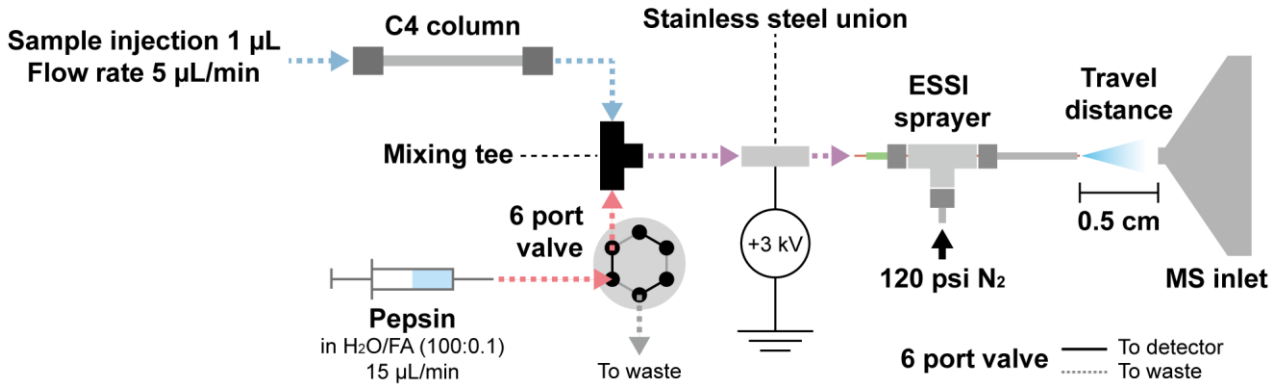

Figure S3. Instrument setup for online MAED.

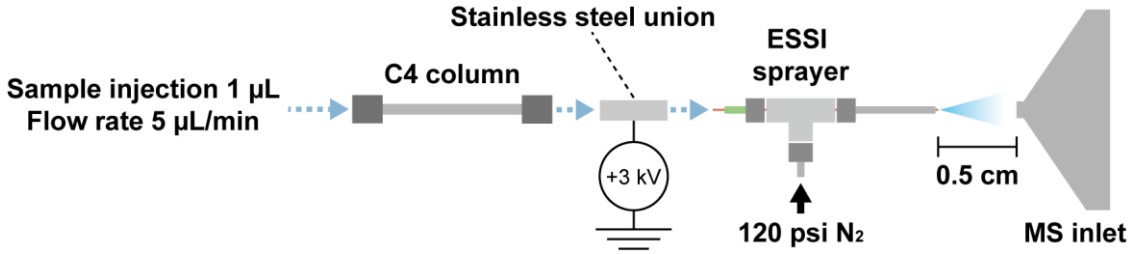

Figure S4. Instrument setup for conventional top-down and bottom-up analysis.

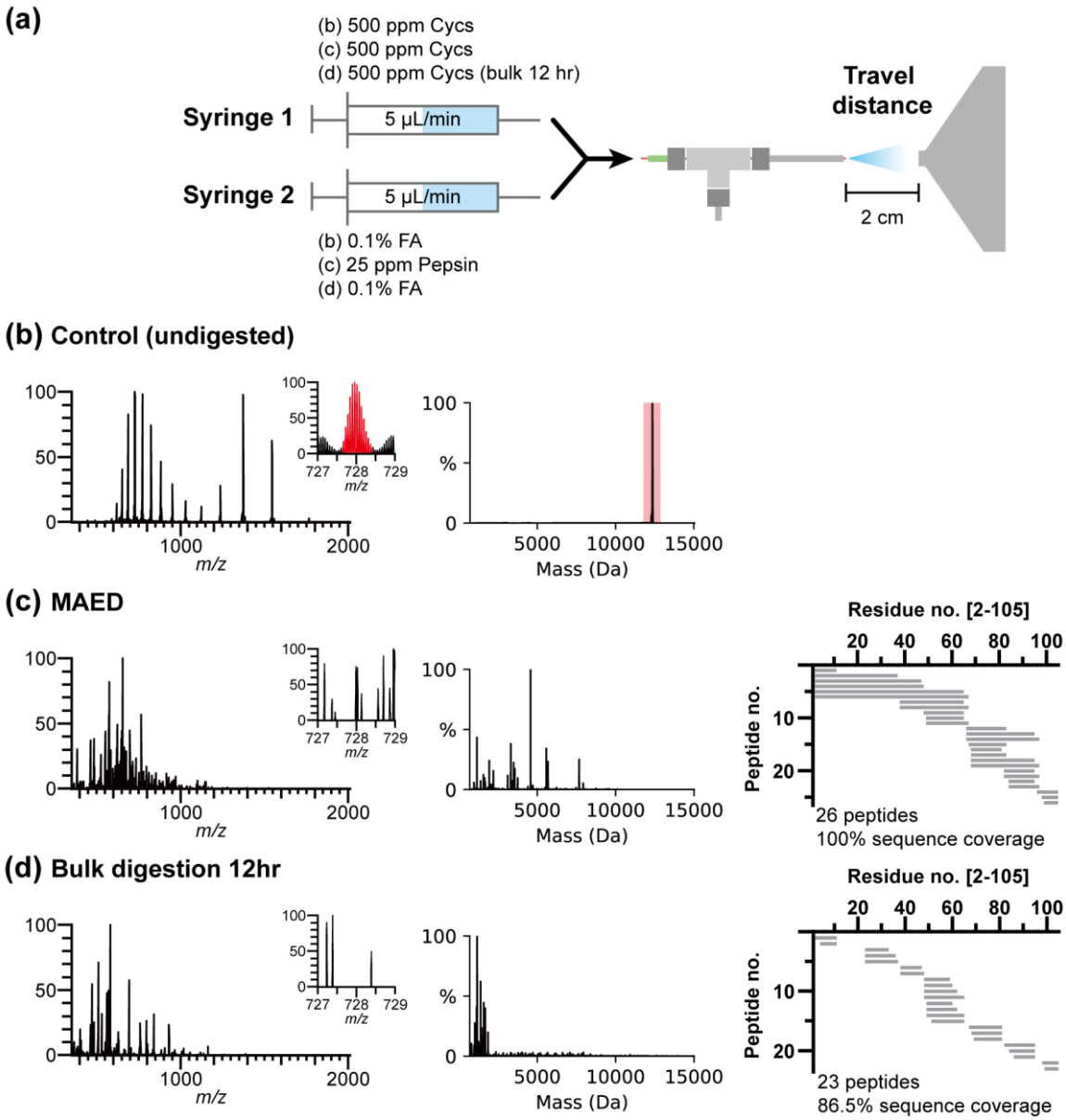

Figure S5. Assessing the feasibility of MAED under conventional RPLC buffer conditions. **(a)** Summary of experimental conditions. **(b, c, d)** Cycs undergoing no digestion **(a)**, MAED **(b)**, and 12 hr bulk digestion **(c)**. The left panel shows the raw spectra. The inset shows the zoomed-in range at m/z 727-729. The signal of intact Cycs (m/z 728, +17) was labeled in red. The middle panel shows the deconvoluted spectrum in which the red bar marks the deconvoluted signal of intact Cycs (~12.3 kDa). The right panel shows the peptide map from either MAED or overnight digestion.

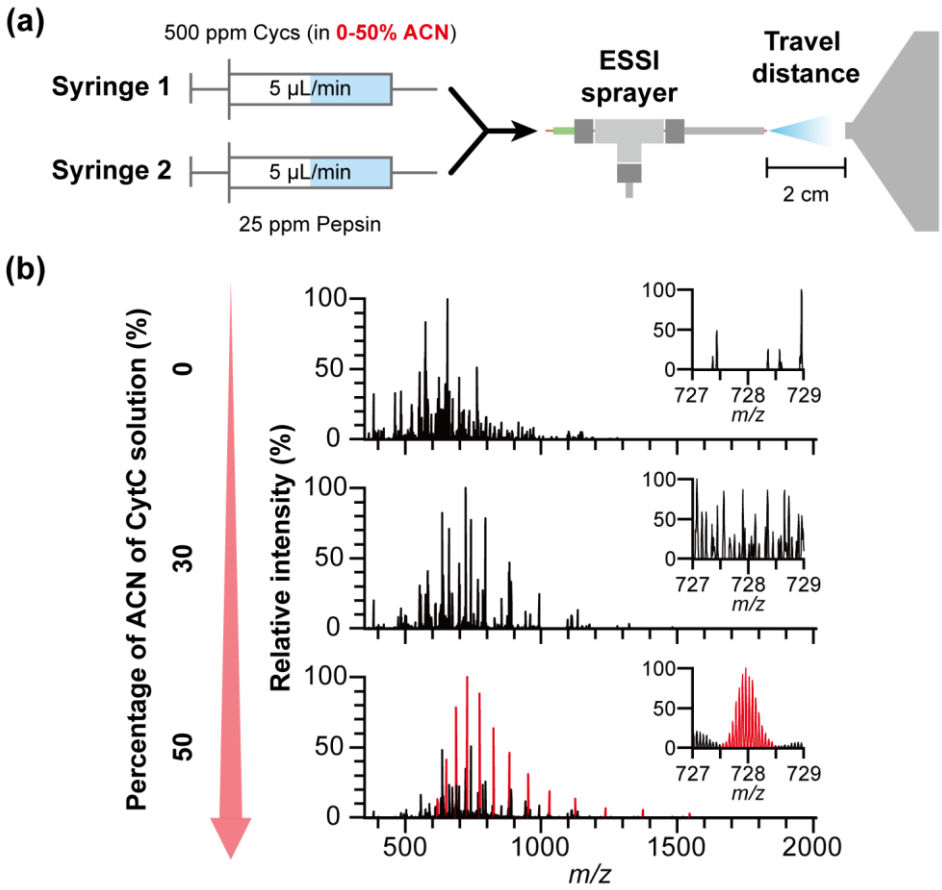

Figure S6. The effect of organic phase percentage on MAED performance. **(a)** Summary of experimental conditions. The percentage of ACN in Cycs solution was varied, which was labeled in red. **(b)** MAED of 500 ppm Cycs dissolved in 0, 30, and 50% ACN (from top to bottom). The inset shows the zoomed-in range of m/z 727 to 729. The signal of intact Cycs (m/z 728, +17) was labeled in red.

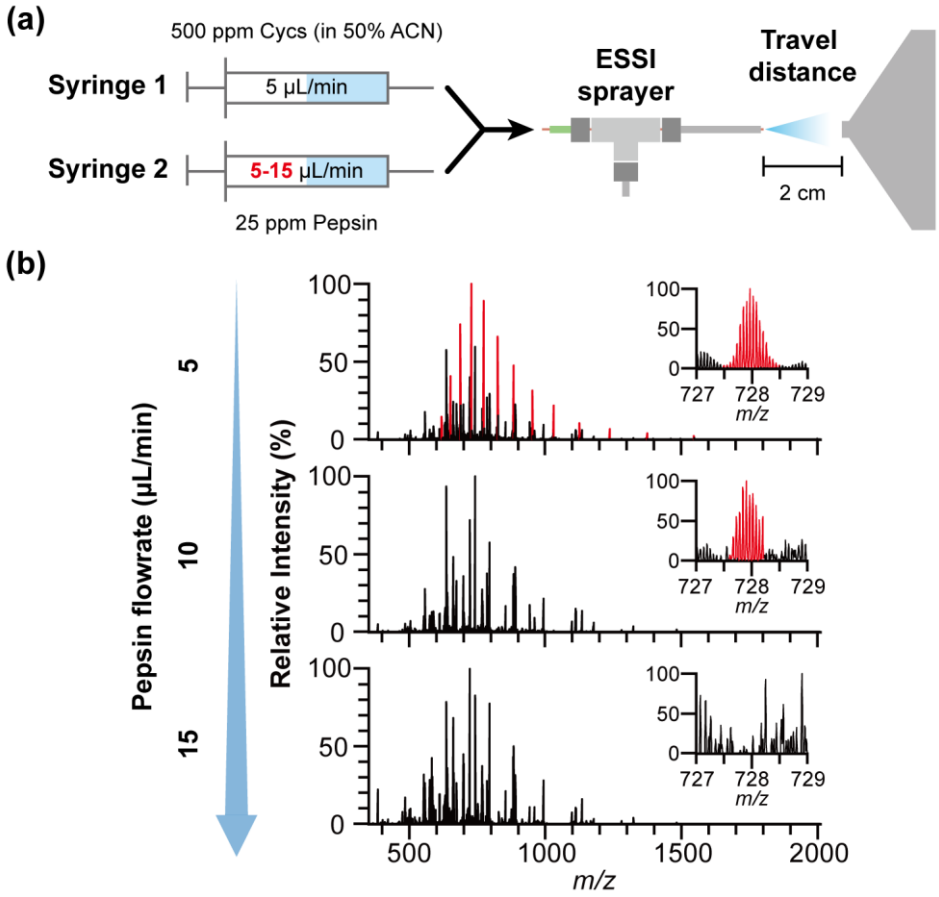

Figure S7. Optimization of pepsin flow rate. **(a)** Summary of experimental conditions. The flow rate of pepsin was varied, which was labeled in red. **(b)** MAED of 500 ppm Cycs dissolved in 50% ACN with pepsin infused at a flow rate of 5, 10, and 15 µL/min (from top to bottom). The inset shows the zoomed-in range of m/z 727 to 729. The signal of intact Cycs (m/z 728, +17) was labeled in red.

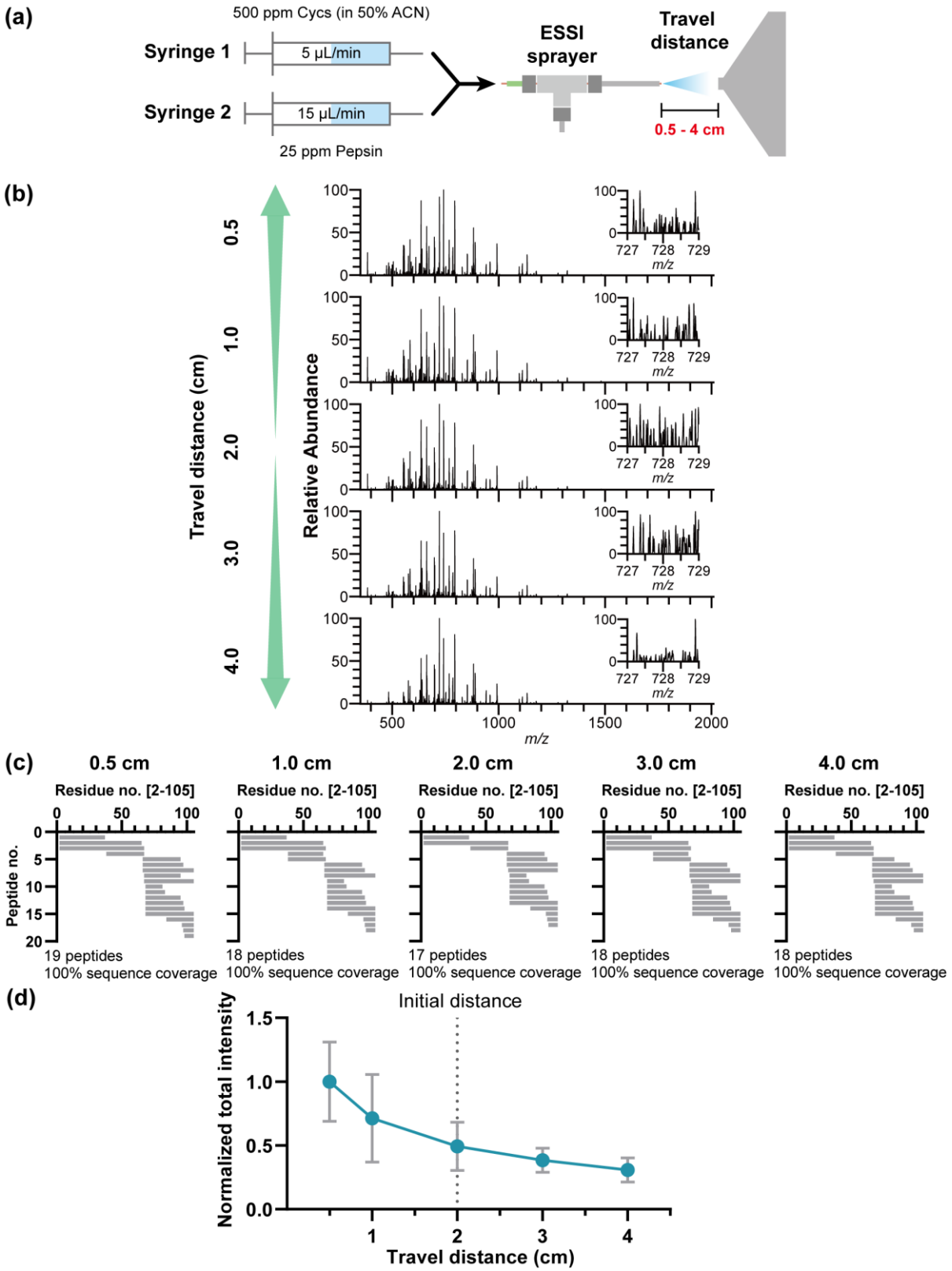

Figure S8. Optimization of microdroplet travel distance. **(a)** Summary of experimental conditions. The travel distance was varied, which was labeled in red. **(b)** MAED of 500 ppm Cycs dissolved in 50% ACN with the travel distance of 0.5 to 4.0 cm (from top to bottom). The inset shows the zoomed-in range of m/z 727 to 729. **(c)** Peptide map of Cycs undergoing MAED at varying travel distances in **(b)**. **(d)** The normalized total ion intensities of peptide signals from MAED of Cycs with varied travel distances. The error bars represent the standard deviations of normalized total ion intensities.

Feasibility assessment and optimization of MAED

The feasibility of MAED under conventional RPLC buffer (0.1% formic acid (FA)) was assessed by simulating the digestion condition via direct infusion experiments shown in **Figure S5a**. The mass profile of Cycs undergoing MAED (**Figure S5c**) largely differs from undigested Cycs (**Figure S5b**). Moreover, the signal of intact Cycs wasn't observed in the raw and deconvoluted spectra of Cycs undergoing MAED, indicating complete digestion of Cycs (≥ 1 cleavage). Full sequence coverage was obtained, which was comparable with the results of conventional overnight digestion using pepsin (**Figure S5d, Table S1-2**) and previous studies of MAED using trypsin or pepsin under different buffer composition.[10, 11]

The performance of MAED was also evaluated under varying percentages of acetonitrile (ACN) (**Figure S6a**), which is the case for RPLC gradient elution. MAED using pepsin appeared to be tolerant to small portions of organic phase, for the signal of intact Cycs was not observed when the percentage of ACN of Cycs solution increased from 0% to 30% (**Figure S6b**). However, the signal of intact Cycs became exceedingly significant as the percentage of ACN was raised to 50%, suggesting most Cycs remained undigested. This emphasized the need for optimizing spraying conditions for complete digestion of proteins eluting at a higher percentage of organic phase. Our results showed good agreement with previous studies that the water percentage in microdroplets may largely affect the catalytic property of microdroplets.[12-14]Since porcine pepsin still retains most of its activity at ACN percentage up to 60%,[15] we hypothesize that the loss in the catalytic activity of microdroplets mainly contributed from the different levels in water autoionization[16, 17] or H_2O_2 generation[14, 18, 19] rather than the inactivation of pepsin.

Regarding this issue, our approach was to increase the flow rate of pepsin (**Figure S7a**) to dilute the organic phase and increase the water content. Since most proteins elute before 50% ACN using the C4 column, the spraying condition at a high percentage of organic phase was optimized using Cycs dissolved in H_2O/ACN/FA (50:50:0.1). Complete digestion was achieved when the flowrate of pepsin was increased to 15 μL/min as intact Cycs signal wasn't observed (**Figure S7b**).

To obtain the optimal total peptide signal intensity and digestion efficiency, we varied the microdroplet travel distance (**Figure 8a**). Extending the travel distance to 4.0 cm yielded no significant differences in mass profile, number of peptides, and sequence coverage, but led to a decrease in total signal intensity (**Figure S8b-d, Table S3c-e**). We then considered reducing the cross-section of ESSI spray to enhance signal intensity by shortening the travel distance.[1] Given that digestion takes place when microdroplets travel from the sprayer to the heated MS inlet,[20, 21] the digestion time would decrease with a smaller travel distance, which may result in incomplete digestion. Therefore, the signals of intact Cycs were also monitored while shortening travel distance. Shortening the travel distance to 0.5 cm yielded a similar mass profile compared to 2.0 cm (**Figure S8b**). Intact Cycs was not detected, and full sequence coverage could still be achieved (**Figure S8c, Table S3a-c**). A 2-fold increase in total ion intensity was observed after reducing the travel distance (**Figure S8d**). Consequently, a travel distance of 0.5 cm and pepsin flow rate of 15 μL/min were adopted for further online MAED experiments.

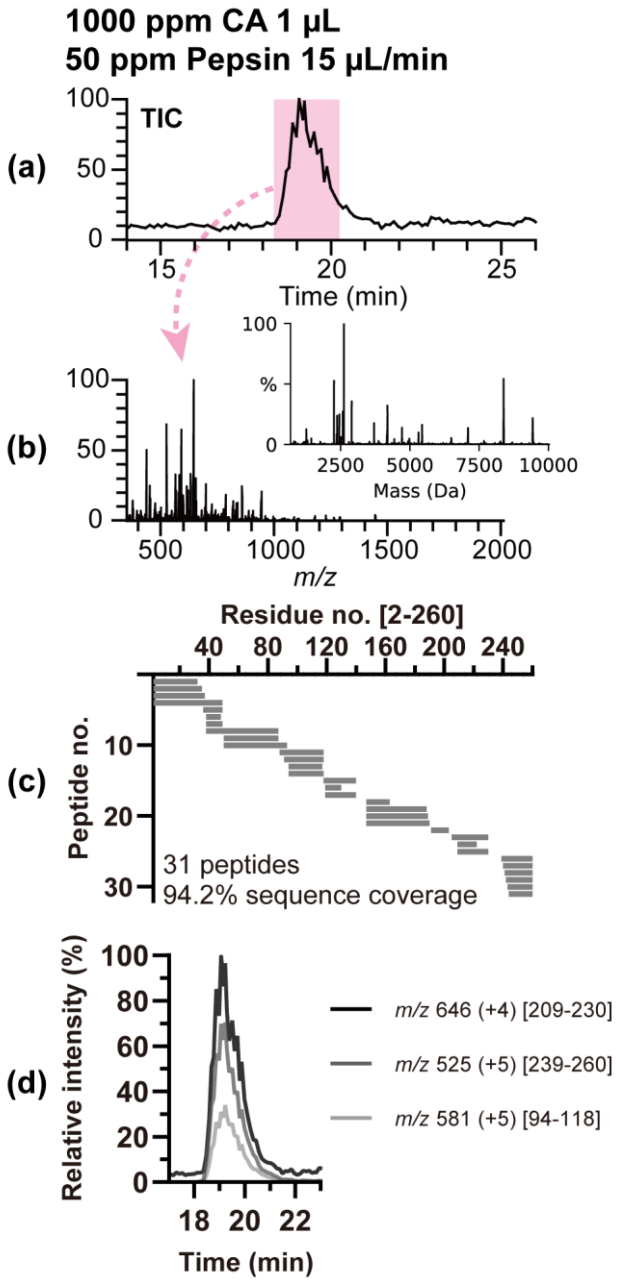

Figure S9. Characterization of CA via online MAED. **(a)** TIC of online digested CA. **(b)** Raw spectrum at selected retention time range. The inset of the raw spectrum shows the deconvoluted spectrum. **(c)** Peptide map of online digested CA. **(d)** EICs of 3 top intensity peptides for CA.

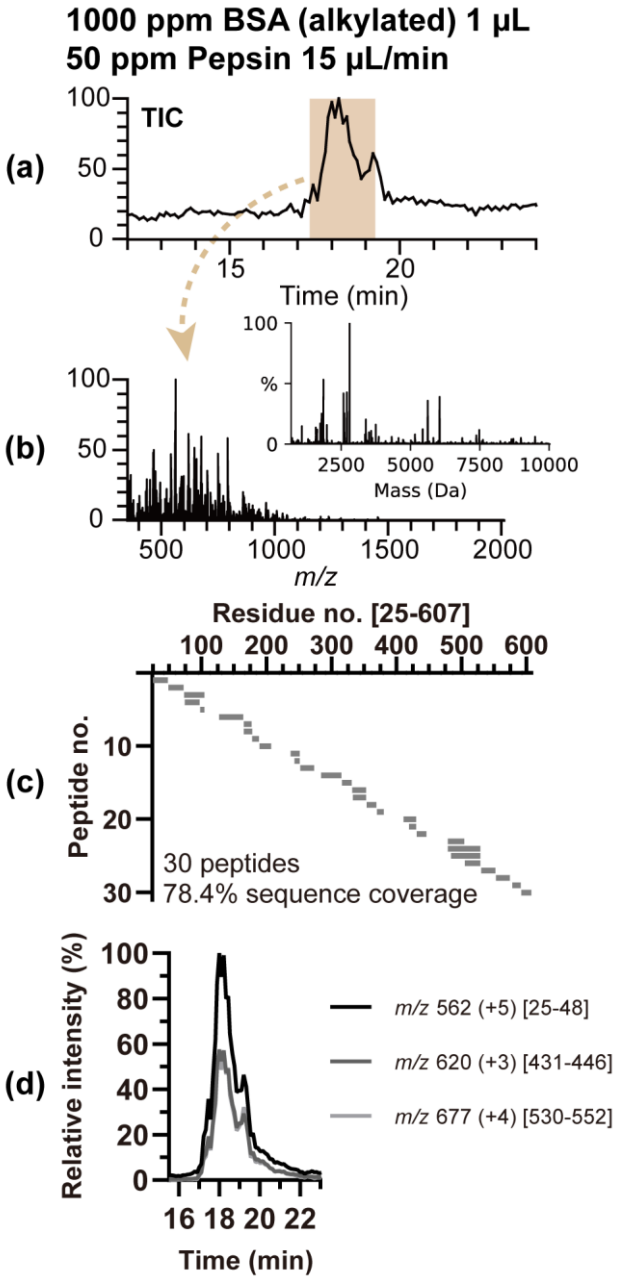

Figure S10. Characterization of alkylated BSA via online MAED. **(a)** TIC of online digested BSA. **(b)** Raw spectrum at selected retention time range. The inset of the raw spectrum shows the deconvoluted spectrum. **(c)** Peptide map of online digested BSA. **(d)** EICs of 3 top intensity peptides for BSA.

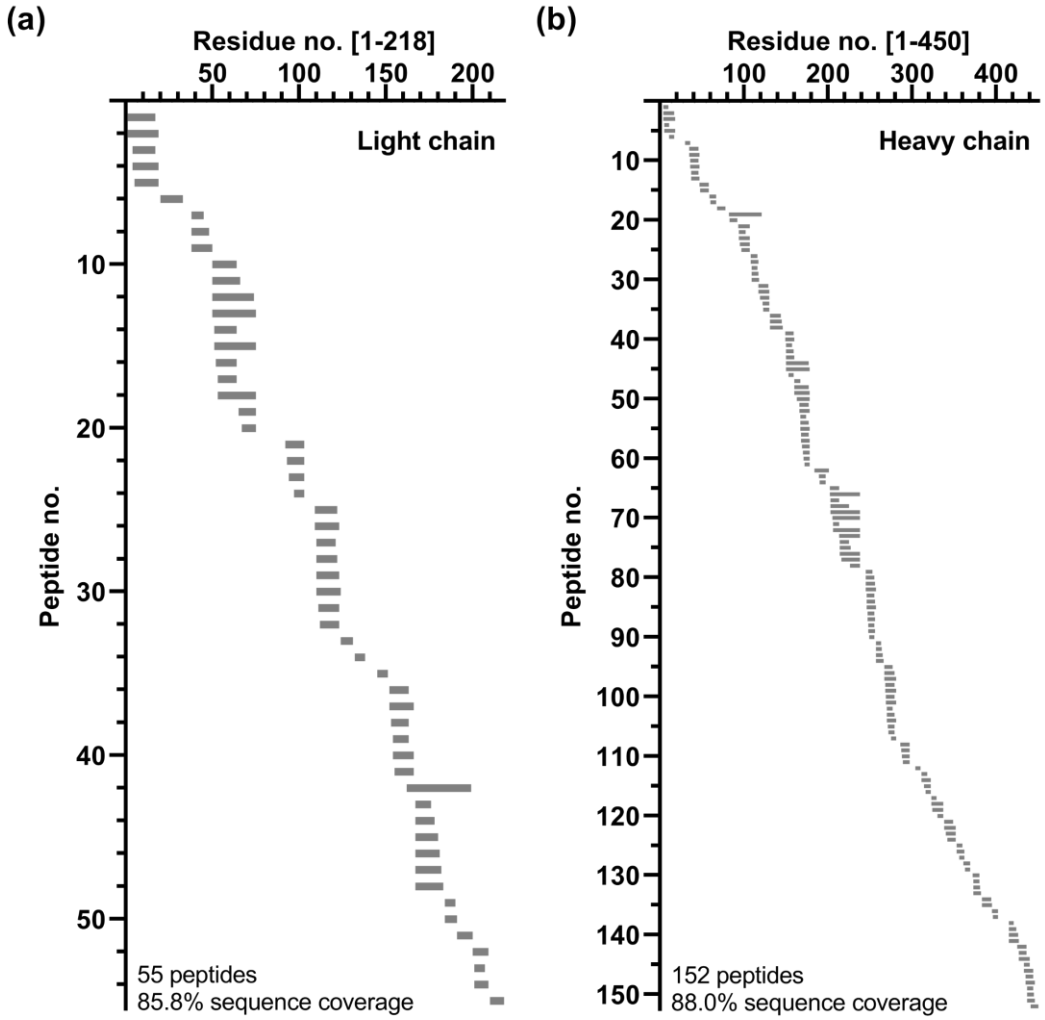

Figure S11. Characterization of MAb via the conventional bottom-up approach (12 hr bulk digestion). Peptide maps of MAb light chain **(a)** and heavy chain **(b)**.

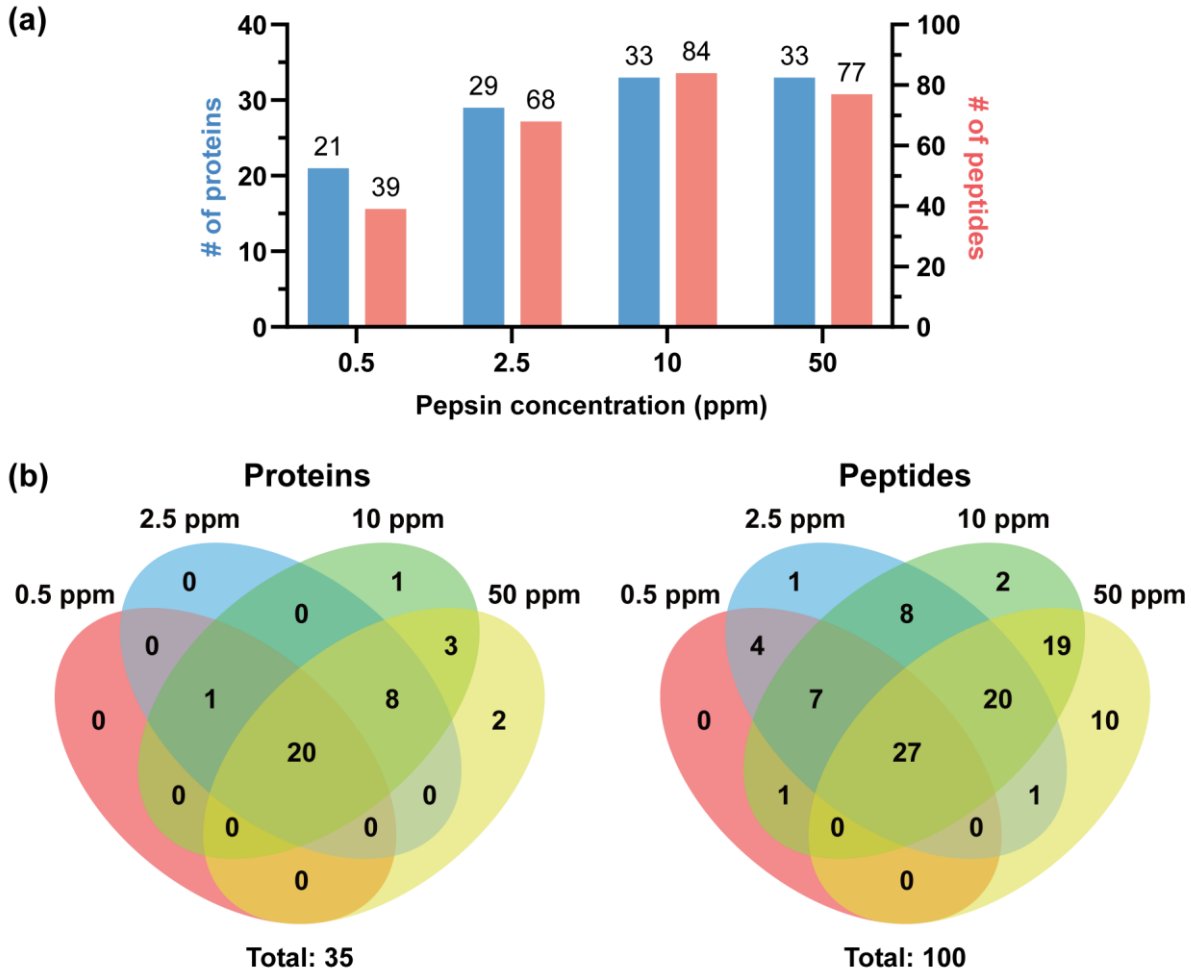

Figure S12. Characterization of mouse brain extract using online MAED with 4 concentrations of pepsin. **(a)** The number of proteins and peptides identified. **(b)** Venn diagram showing the overlap of proteins and peptides identified at different pepsin concentrations.

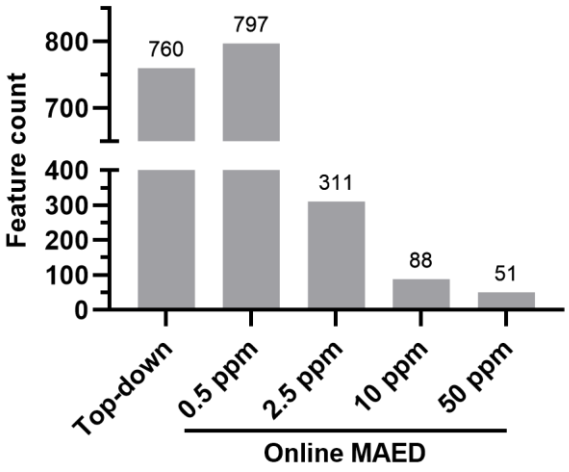

Figure S13. The number of features above 9 kDa detected in conventional top-down analysis and online MAED. Features were detected using the TOPPIC suite.

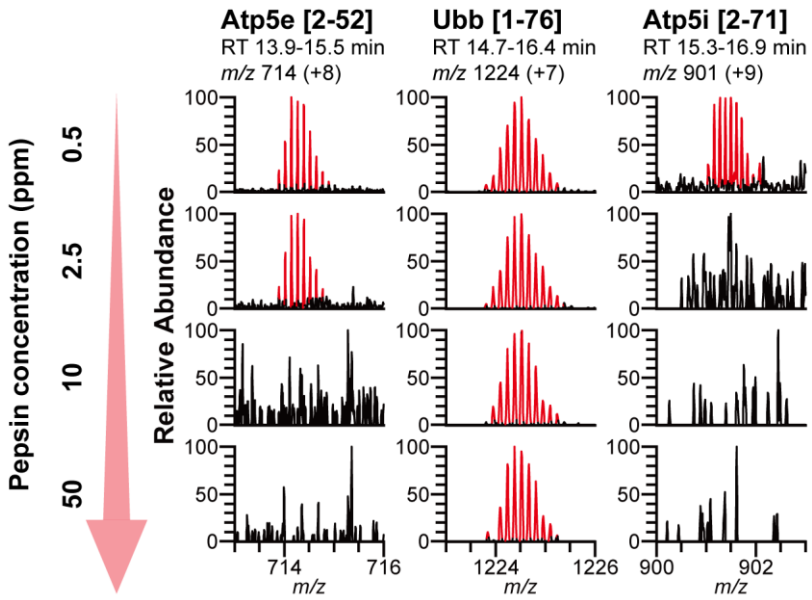

Figure S14. Intact protein signals of mouse brain proteins undergoing online MAED at varying concentrations of pepsin. Intact protein signals of ATP synthase subunit epsilon (Atp5e, left column), ubiquitin (Ubb, mid column), and ATP synthase subunit e (Atp5i, right column) were labeled in red.

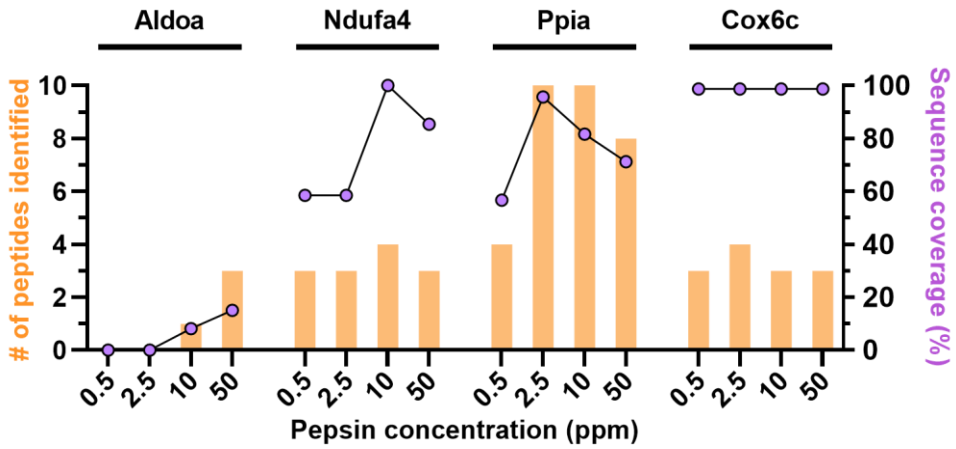

Figure S15. Digestion performance of online MAED on mouse brain proteins at varying concentrations of pepsin. The sequence coverages and the numbers of peptides identified for fructose-bisphosphate aldolase A (Aldoa), cytochrome C oxidase subunit NDUFA4 (Ndufa4), peptidyl-prolyl cis-trans isomerase A (Ppia), and cytochrome C oxidase subunit 6C (Cox6C) were shown.

The effect of pepsin concentration on complex samples

Considering the sample complexity in terms of the number of protein species, their properties, and concentrations, online MAED on 3 μg of mouse brain proteins was tested using 4 concentrations of pepsin ranging from 0.5 to 50 ppm. The highest number of identifications in both protein and peptide was yielded using 10 ppm of pepsin, while online MAED using 0.5 ppm pepsin rendered the least number of identification (**Figure S12a**). As inferred from the number of features above 9 kDa detected compared with that from conventional top-down analysis (**Figure S13**), most proteins may remain undigested at the pepsin concentration of 0.5 ppm.

Some changes in the species of proteins and peptides identified could be observed at different pepsin concentrations (**Figure S12b**). Online MAED using 10 and 50 ppm of pepsin could cover most of the proteins and peptides detected across all pepsin concentrations, but there remain some proteins and peptides uniquely detected within certain concentrations of pepsin. Therefore, thorough characterization of proteins via online MAED using a single concentration of pepsin might not apply to complex samples. The pepsin concentration(s) for performing online MAED should be determined case by case based on sample complexity and protein amount.

We further examined the effect of pepsin concentration on the digestion efficiencies of different mouse brain proteins. **Figure S14** shows the online MAED of 3 proteins detected in the mouse brain, ATP synthase epsilon (Atp5e), ubiquitin (Ubb), and ATP synthase subunit e (Atp5i), using 4 concentrations of pepsin. Full digestion (intact protein signal not observed) of Atp5e and Atp5i could be achieved with pepsin concentrations above 2.5 and 0.5 ppm. Yet, Ubb wasn't fully digested even at a pepsin concentration of up to 50 ppm. This suggested that the minimum concentration required for each protein to achieve complete digestion could vary due to the diverse properties and concentrations in the sample.

A higher concentration of pepsin generally guarantees complete digestion, yet the sequence coverage for each protein isn't necessarily proportional to pepsin concentration. **Figure S15** displays the sequence coverages and the number of peptides detected for cytochrome C oxidase subunit 6C (Cox6C), cytochrome C oxidase subunit NDUFA4 (Ndufa4), peptidyl-prolyl cis-trans isomerase A (Ppia), and fructose-bisphosphate aldolase A (Aldoa) under different pepsin concentrations. While the sequence coverage and the number of peptides detected for Cox6C showed little dependence on pepsin concentration, those 2 values for the other 3 proteins reached the maximum at different pepsin concentrations. This indicated that the optimal pepsin concentration range for the maximum sequence coverage of each protein in complex samples may vary. The difference in optimal pepsin concentration for each protein could explain the observations in **Figure S12**. Therefore, when performing online MAED on complex samples, it is recommended to either reduce sample complexity by pre-fractionation or test with different pepsin concentrations.

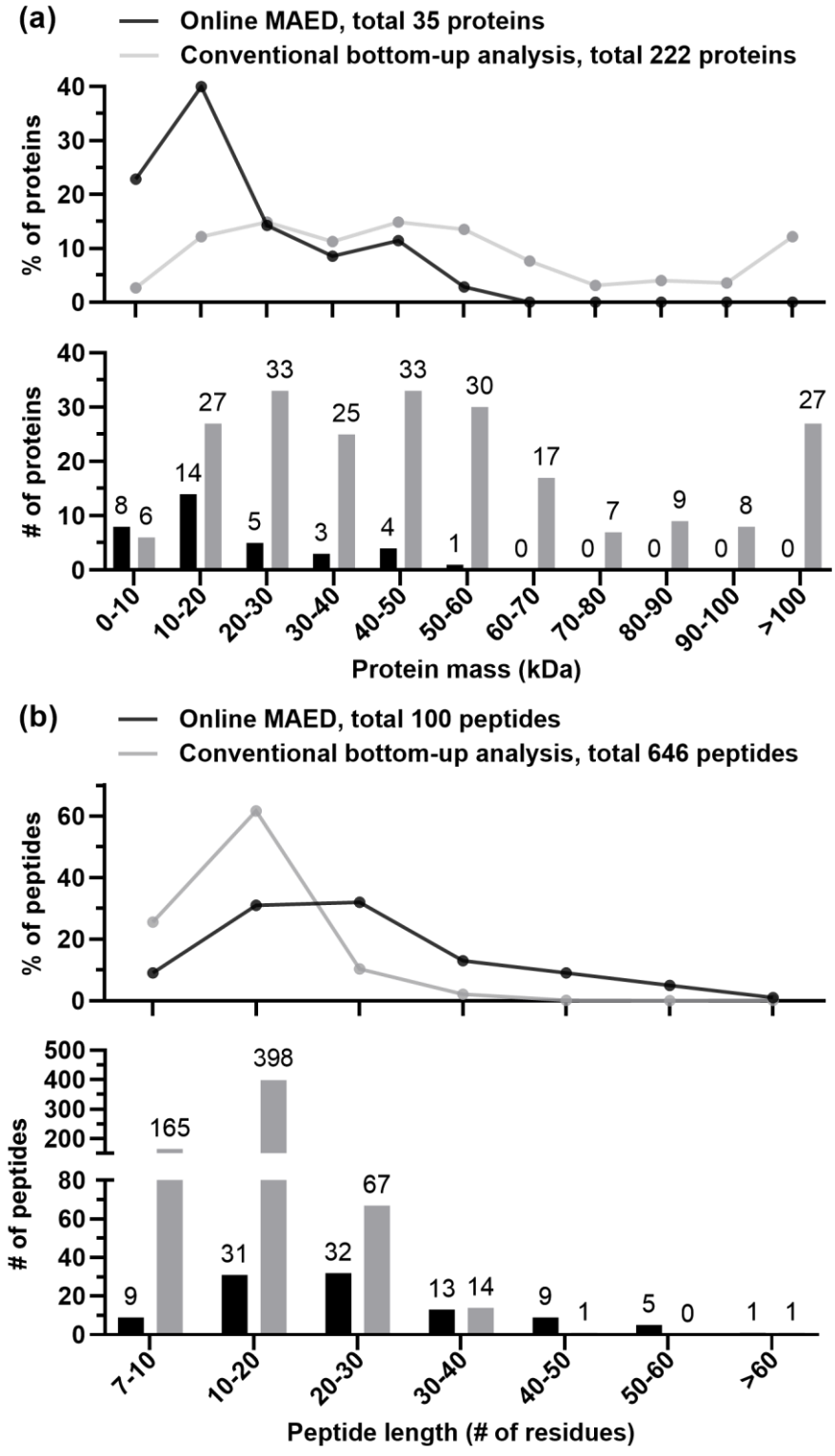

Figure S16. The comparison of the characterization results of mouse brain extract from conventional bottom-up analysis and online MAED (results from 4 concentrations of pepsin merged). **(a)** Mass distribution of proteins identified. **(b)** Length distribution of peptides identified.

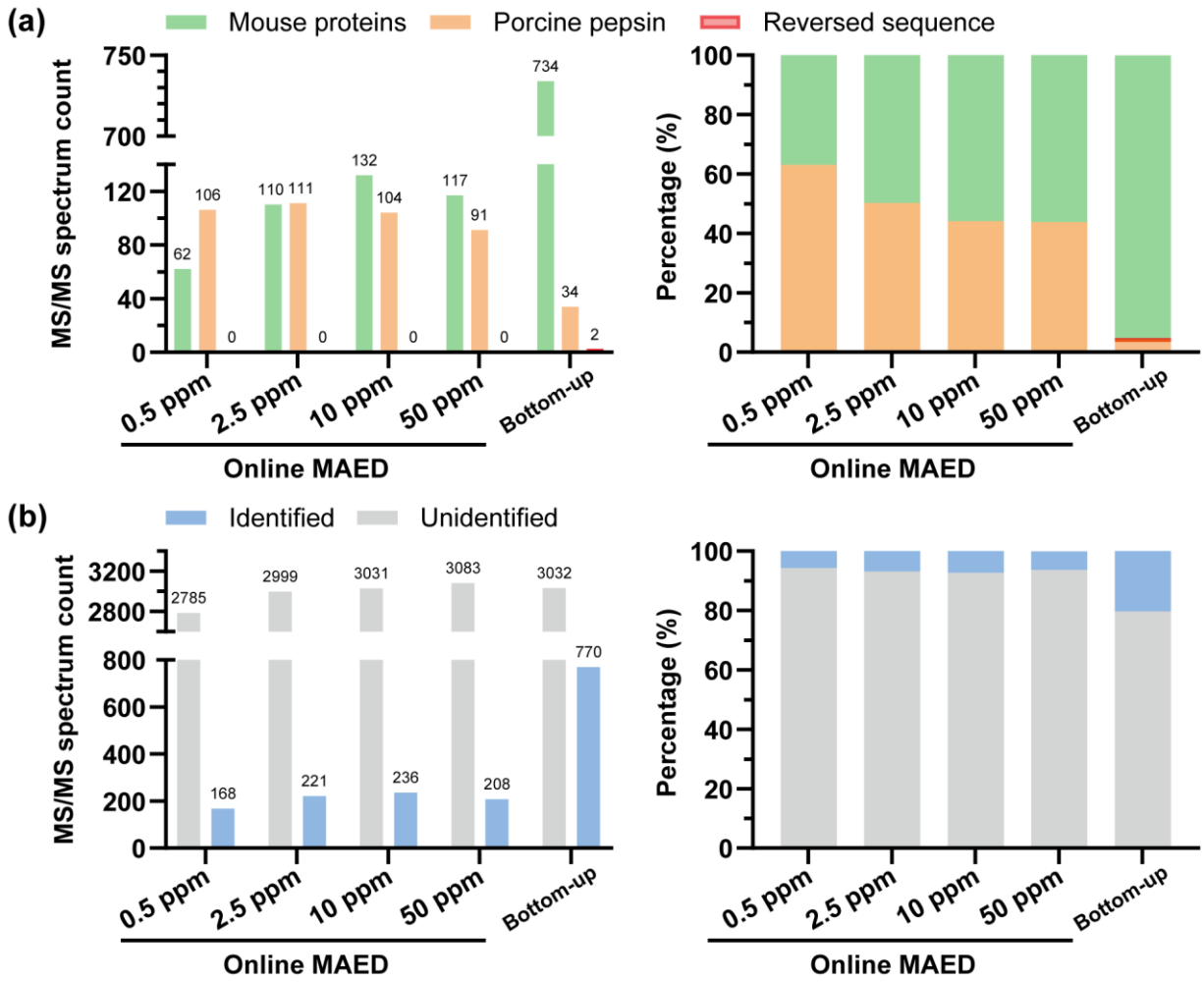

Figure S17. Identification status of MS/MS spectra. (a) The number (left panel) and percentage (right panel) of identified MS/MS spectra matched to mouse proteins, porcine pepsin, and reversed sequences of mouse proteins. **(b)** The number (left panel) and percentage (right panel) of identified and unidentified MS/MS spectra.

Possible reasons for the lower identification numbers of online MAED

Digestion efficiency

The conventional bottom-up approach (12 hr bulk digestion) results in more thorough digestion of proteins, as evidenced by the shorter peptide lengths obtained compared to online MAED (**Figure S5, S11, and S16**). Additionally, the minimum pepsin concentration required to achieve complete digestion varies for each protein (**Figure S12**). Some proteins may remain undigested in the concentration range of 0.5 to 50 ppm.

In the demonstration of online MAED on mouse brain extract, the disulfide bonds of mouse brain proteins have not been reduced nor alkylated. We have previously found that online MAED of proteins containing disulfide bonds yielded poor digestion efficiencies if the disulfide bonds were not reduced and alkylated (data not shown). As we searched "mouse brain proteins" on Uniprot (accessed on 05/10/2023), 588 out of 3664 reviewed proteins (16%) contain at least 1 disulfide bond. Compared to the assumed complete digestion of 12 hr bulk digestion, the characterization results of the 588 proteins from online MAED may be undesirable and could be improved if alkylation is performed. However, the proteins containing disulfide bonds only account for a small portion and might not be the main reason for the lower number of identifications.

Separation at the protein level

In general, intact proteins are harder to separate than peptides. Several proteins may co-elute in a similar time range and each protein may yield multiple peptides upon online digestion. Consequently, more peptides could appear within a certain time range compared to the conventional bottom-up approach, which could be viewed as a less efficient separation of peptides. As peptides of multiple proteins appear simultaneously, collecting MS/MS spectra for all peptide ions may be challenging due to the instrument scan speed limit and the selection of ions for ddMS2 acquisition. Moreover, the excessive presence of peptide ions can cause ion suppression effects, reducing the sensitivity of peptide detection. With better protein separation conditions, it is expected that MS/MS spectra could be collected for more peptide ions, leading to an increased identification number.

Post-column infusion of pepsin

The post-column infusion of pepsin is essential for online digestion. However, one straightforward drawback of this approach is the dilution of proteins. With the sample flowrate at 5 μL/min and pepsin flowrate at 15 μL/min, the eluted proteins were diluted 4-fold. Autolytic products of pepsin can also interfere with peptide detection and cause ion suppression. MS/MS spectra were collected for the autolysis products, which occupy the instrument duty cycle, hindering the acquisition of usable MS/MS spectra available for protein and peptide identification. The situation is evidenced by the number and percentage of MS/MS spectra matched to pepsin (**Figure S16a**). Using online MAED, around 100 MS/MS spectra were matched to pepsin at each pepsin concentration, which was about 3 times the amount in the conventional bottom-up approach. Around 40-60% of the identified MS/MS spectra in online MAED were matched to pepsin, compared to 5% for the conventional bottom-up approach. The constant infusion of pepsin leads to greater interference in online MAED than in the conventional bottom-up approach. By setting an exclusion list for pepsin autolysis products, the interference of pepsin regarding MS/MS spectra collection and the occupation of instrument duty cycle could be reduced. However, ion suppression from pepsin autolysis products still exists and currently could not

be relieved unless a different protease is used. In such case, it is essential to adjust the pH value of the mobile phase or consider employing other separation techniques since most proteases function at neutral or alkalic conditions.

The lower number of identified MS/MS spectra

The number of identified MS/MS spectra in online MAED was found to be approximately 3 times lower than that of the conventional bottom-up approach at each pepsin concentration, despite having a similar number of unidentified spectra (**Figure S16b**). For online MAED, around 5% of MS/MS spectra were identified, whereas the value was approximately 20% for the conventional bottom-up approach.

The quality of MS/MS spectra and unexpected modifications are two factors affecting the identification rate in proteomics experiments,[22] and these factors may have been enhanced in online MAED. Ion suppression effects from both pepsin autolysis products and other peptides could decrease the signal-to-noise ratio of peptide signals, leading to lower-quality MS/MS spectra. The occurrence of peptides containing unexpected modifications is possibly increased in online MAED. Pepsin is considered unspecific relative to trypsin, and online MAED tends to yield longer peptides compared to 12 hr bulk digestion. Assuming peptides are randomly cleaved, longer peptides have a higher probability of including unexpected modifications in the case of proteins with such modifications. Consequently, these peptides are not identified using restricted search engines (only a few known modifications included for database search), resulting in a lower identification rate.

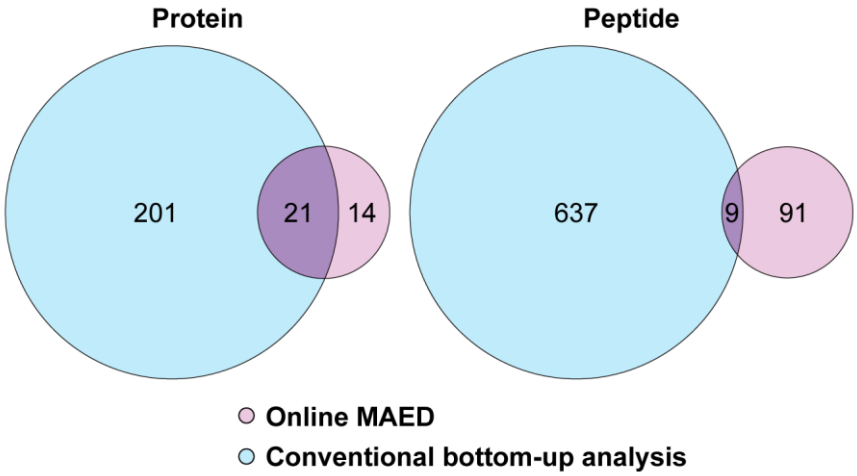

Figure S18. Venn diagrams showing the overlap of proteins and peptides identified using conventional bottom-up analysis and online MAED (results from 4 concentrations of pepsin merged). The identification of undigested proteins, porcine pepsin, and reversed sequences were excluded.

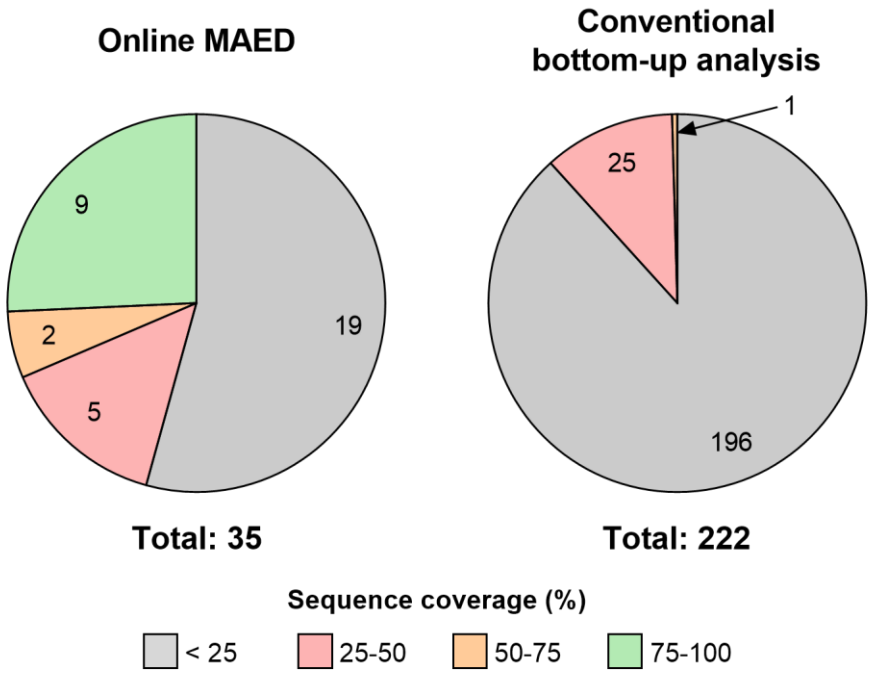

Figure S19. Pie charts showing the distribution of protein sequence coverages from conventional bottom-up analysis and online MAED (results from 4 concentrations of pepsin merged).

Properties and challenges of online MAED

Separation at the protein level

The separation at the protein level preserves the retention times (RT) of proteins upon online digestion. Peptides from the same protein possess identical RT, ensuring the connectivity between protein and peptide. Owing to this feature, integration of data obtained via both top-down and bottom-up approaches is made possible. Ultimately, the discrimination of protein isoforms in bottom-up manner could be derived from this application. Isoforms of protein could be produced in various including alternative splicing and truncation.[23] Therefore, they could share a considerable portion of the sequence with their canonical form. For peptides from the shared sequence, we couldn't differentiate the contribution of each isoform due to the loss of connectivity in the conventional bottom-up approach. However, since peptides could be directly assigned to their corresponding proteins with online MAED, the discrimination of isoform from the shared sequence could be achieved as long as the isoforms could be chromatographically separated.

Despite the potential applications mentioned above, several challenges still exist or could arise from this feature. Intact proteins are harder to separate than peptides, which could be one of the reasons for the lower proteome coverage compared to the bottom-up approach. Therefore, highly complex samples composed of numerous proteins may require prefractionation. As peptides from the same protein are all detected at the same RT, the effect of ion suppression could be severe. Moreover, to collect MS2 spectra for all peptides of an eluted protein, the instrument scan rate and dynamic exclusion should be adjusted depending on the peak width.

Detection at the peptide level

One of the advantages of detecting peptides than proteins is that peptides are more easily ionized with fewer charge states, giving better sensitivity. This could be evidenced by the identification of Ppia via online MAED, whereas intact Ppia was not detected in the conventional top-down analysis. However, full sequence coverage is hardly achieved using conventional bottom-up methods in most cases. This could lead to difficulties in identifying the presence/connectivity of post-translational modifications.[24]

Rapidness and simplicity

The rapidness and simplicity are the major features of online MAED. Proteins could be digested within a millisecond scale upon elution. Thus, bottom-up characterization of protein mixtures could be achieved without digestion required before MS analysis, saving a considerable amount of time. Besides, online MAED could be performed merely with the additional infusion of pepsin. Little modifications to conventional LC-MS setup are required. The parts for the ESSI sprayer are also cheap and commercially available.

Post-column infusion of pepsin

Pepsin is a double-edged sword: essential for digestion but also suppresses ion signals with its autolysis products, which is more severe compared to conventional bottom-up methods as pepsin was constantly infused. Excessive pepsin concentration would result in ion suppression, while insufficient pepsin concentration would lead to incomplete digestion of proteins. As a consequence, the pepsin concentration for performing online MAED should be carefully determined. In conventional bottom-up methods, the protease concentration is determined by the "total amount" of proteins in the sample. However, in online MAED, the pepsin

concentration should be determined by the "individual amount" of each protein since digestion is performed after separation. For complex samples in which the concentration dynamic range of proteins contained could reach up to 6 orders,[25, 26] the concentration of pepsin could be hard to determine. The thorough characterization of such samples may be challenging. Complex samples may require prefractionation and/or enrichment or be tested with various concentrations of pepsin.

Sensitivity

Despite the successful identification of proteins and peptides in mouse brain extract using online MAED, the number of identifications was far lower compared to conventional shotgun proteomics in previous studies.[27-30] To eliminate the dissimilarity from using different ionization sources or instrument setups when compared with the conventional workflow, we have conducted conventional bottom-up analysis using the same ESSI setup. It turned out that the number of identifications obtained was still 1-2 orders lower. Hence, aside from the factors mentioned in the previous additional discussion, we have speculated that a major contributing factor limiting the number of proteins and peptides identified could be the lower sampling rate of our ESSI setup compared to nano-electrospray ionization (nanoESI),[1] resulting in lower sensitivity. Given that microdroplet accelerated reactions also work with nanoESI,[11, 31-34] it is anticipated that performing online MAED with nanoESI could improve sensitivity and lead to increased proteome coverage.

References

1. Z. Takáts, J. M. Wiseman, B. Gologan and R. G. Cooks, *Anal. Chem.*, 2004, **76**, 4050-4058.
2. D. Wessel and U. I. Flügge, *Anal. Biochem.*, 1984, **138**, 141-143.
3. D. P. Donnelly, C. M. Rawlins, C. J. DeHart, L. Fornelli, L. F. Schachner, Z. Lin, J. L. Lippens, K. C. Aluri, R. Sarin, B. Chen, C. Lantz, W. Jung, K. R. Johnson, A. Koller, J. J. Wolff, I. D. G. Campuzano, J. R. Auclair, A. R. Ivanov, J. P. Whitelegge, L. Pasa-Tolic, J. Chamot-Rooke, P. O. Danis, L. M. Smith, Y. O. Tsybin, J. A. Loo, Y. Ge, N. L. Kelleher and J. N. Agar, *Nat. Methods*, 2019, **16**, 587-594.
4. A. A. Doucette, D. B. Vieira, D. J. Orton and M. J. Wall, *J. Proteome Res.*, 2014, **13**, 6001-6012.
5. M. T. Marty, A. J. Baldwin, E. G. Marklund, G. K. A. Hochberg, J. L. P. Benesch and C. V. Robinson, *Anal. Chem.*, 2015, **87**, 4370-4376.
6. S. Tyanova, T. Temu and J. Cox, *Nat. Protoc.*, 2016, **11**, 2301-2319.
7. Q. Kou, L. Xun and X. Liu, *Bioinformatics*, 2016, **32**, 3495-3497.
8. M. C. Chambers, B. Maclean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T. A. Baker, M.-Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S. L. Seymour, L. M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E. W. Deutsch, R. L. Moritz, J. E. Katz, D. B. Agus, M. MacCoss, D. L. Tabb and P. Mallick, *Nat. Biotechnol.*, 2012, **30**, 918-920.
9. T. Chen, H. Zhang, Y. Liu, Y.-X. Liu and L. Huang, *J. Genet. Genomics*, 2021, **48**, 863-866.
10. T. Rainer, R. Eidelpes, M. Tollinger and T. Müller, *J. Am. Soc. Mass. Spectrom.*, 2021, **32**, 1841-1845.
11. X. Zhong, H. Chen and R. N. Zare, *Nat. Commun.*, 2020, **11**, 1049.
12. D. B. Eremin and V. V. Fokin, *J. Am. Chem. Soc.*, 2021, **143**, 18374-18379.
13. L. Feng, X. Yin, S. Tan, C. Li, X. Gong, X. Fang and Y. Pan, *Anal. Chem.*, 2021, **93**, 15775-15784.
14. D. Gao, F. Jin, J. K. Lee and R. N. Zare, *Chem. Sci.*, 2019, **10**, 10974-10978.
15. L. M. Simon, M. Kotormán, A. Szabó, J. Nemcsók and I. Laczkó, *Process Biochem.*, 2007, **42**, 909-912.
16. K.-H. Huang, Z. Wei and R. G. Cooks, *Chem. Sci.*, 2021, **12**, 2242-2250.
17. H. Xiong, J. K. Lee, R. N. Zare and W. Min, *J. Phys. Chem. Lett.*, 2020, **11**, 7423-7428.
18. J. K. Lee, K. L. Walker, H. S. Han, J. Kang, F. B. Prinz, R. M. Waymouth, H. G. Nam and R. N. Zare, *Proc. Natl. Acad. Sci. U. S .A.*, 2019, **116**, 19294-19298.
19. M. A. Mehrgardi, M. Mofidfar and R. N. Zare, *J. Am. Chem. Soc.*, 2022, **144**, 7606-7609.
20. J. K. Lee, S. Kim, H. G. Nam and R. N. Zare, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 3898-3903.
21. Z. Wei, Y. Li, R. G. Cooks and X. Yan, *Annu. Rev. Phys. Chem.*, 2020, **71**, 31-51.
22. J. Griss, Y. Perez-Riverol, S. Lewis, D. L. Tabb, J. A. Dianes, N. del-Toro, M. Rurik, M. Walzer, O. Kohlbacher, H. Hermjakob, R. Wang and J. A. Vizcaíno, *Nat. Methods*, 2016, **13**, 651-656.
23. R. E. Breitbart, A. Andreadis and B. Nadal-Ginard, *Annu. Rev. Biochem*, 1987, **56**, 467-495.
24. L. Switzar, M. Giera and W. M. Niessen, *J. Proteome Res.*, 2013, **12**, 1067-1077.
25. M. Wang, C. J. Herrmann, M. Simonovic, D. Szklarczyk and C. von Mering, *PROTEOMICS*, 2015, **15**, 3163-3168.
26. R. A. Zubarev, *PROTEOMICS*, 2013, **13**, 723-726.

27. D. M. Walther and M. Mann, *Mol. Cell. Proteom.*, 2011, **10**, S1-S7.
28. K. Sharma, S. Schmitt, C. G. Bergner, S. Tyanova, N. Kannaiyan, N. Manrique-Hoyos, K. Kongi, L. Cantuti, U.-K. Hanisch, M.-A. Philips, M. J. Rossner, M. Mann and M. Simons, *Nat. Neurosci.*, 2015, **18**, 1819-1831.
29. H. Wang, W.-J. Qian, M. H. Chin, V. A. Petyuk, R. C. Barry, T. Liu, M. A. Gritsenko, H. M. Mottaz, R. J. Moore, D. G. Camp, A. H. Khan, D. J. Smith and R. D. Smith, *J. Proteome Res.*, 2006, **5**, 361-369.
30. K. Artemis G, A. Athanasios K, P. Vasileios, S. Dimitrios J and T. George Th, *Cancer Genom. Proteom.*, 2020, **17**, 757.
31. X. Yan, R. M. Bain and R. G. Cooks, *Angew. Chem. Int. Ed.*, 2016, **55**, 12960-12972.
32. H. Kafeenah, H.-H. Jen and S.-H. Chen, *ELECTROPHORESIS*, 2022, **43**, 74-81.
33. P. Basuri, L. E. Gonzalez, N. M. Morato, T. Pradeep and R. G. Cooks, *Chem. Sci.*, 2020, **11**, 12686-12694.
34. R. M. Bain, S. Sathyamoorthi and R. N. Zare, *Angew. Chem. Int. Ed.*, 2017, **56**, 15083-15087.