HHS Public Access

Author manuscript

Anal Chem. Author manuscript; available in PMC 2024 June 27.

Published in final edited form as:

Anal Chem. 2023 June 27; 95(25): 9497–9504. doi:10.1021/acs.analchem.3c00551.

Coupling High-Field Asymmetric Waveform Ion Mobility Spectrometry with Capillary Zone Electrophoresis-Tandem Mass Spectrometry for Top-Down Proteomics

Tian Xu,
Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States

Qianjie Wang,
Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States

Qianyi Wang,
Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States

Liangliang Sun
Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States

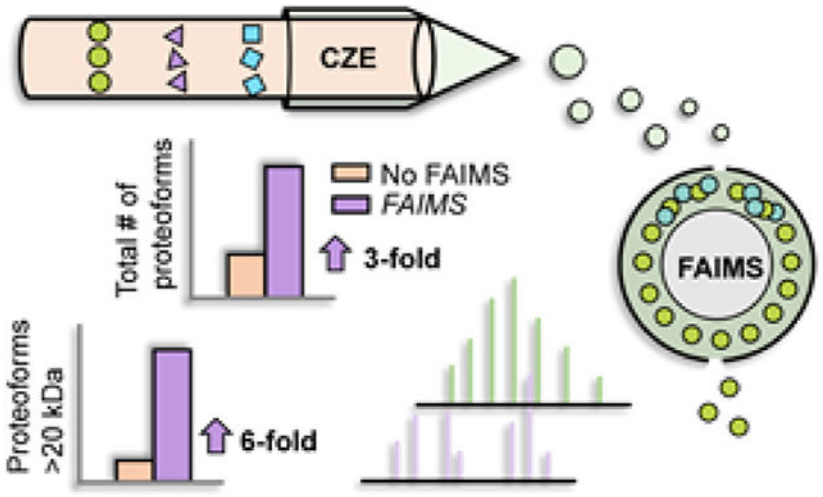
Abstract

Capillary zone electrophoresis-tandem mass spectrometry (CZE-MS/MS) has emerged as an essential technique for top-down proteomics (TDP), providing superior separation efficiency and high detection sensitivity for proteoform analysis. Here, we aimed to further enhance the performance of CZE-MS/MS for TDP via coupling online gas-phase proteoform fractionation using high-field asymmetric waveform ion mobility spectrometry (FAIMS). When the compensation voltage (CV) of FAIMS was changed from −50 to 30 V, the median mass of identified proteoforms increased from less than 10 kDa to about 30 kDa, suggesting that FAIMS can efficiently fractionate proteoforms by their size. CZE-FAIMS-MS/MS boosted the number of proteoform identifications from a yeast sample by nearly 3-fold relative to CZE-MS/MS alone. It particularly benefited the identification of relatively large proteoforms, improving the number of proteoforms in a mass range of 20–45 kDa by 6-fold compared to CZE-MS/MS alone. FAIMS fractionation gained nearly 20-fold better signal-to-noise ratios of randomly selected proteoforms than no FAIMS. We expect that CZE-FAIMS-MS/MS will be a useful tool for further advancing the sensitivity and coverage of TDP. This work shows the first example of coupling CE with ion mobility spectrometry (IMS) for TDP.

Graphical Abstract

Corresponding Author Liangliang Sun – *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States;* Phone: 1-517-353-0498; lsun@chemistry.msu.edu.

The authors declare no competing financial interest.



INTRODUCTION

Top-down proteomics (TDP) aims to characterize all proteoforms in cells and tissues via direct measurement of intact proteoforms using mass spectrometry (MS).[1,2] Sufficient separations of complex proteoform mixtures are needed to improve proteoform detection and identification. Reversed-phase liquid chromatography (RPLC) and capillary electrophoresis (CE) are predominant techniques compatible with electrospray ionization-tandem mass spectrometry (ESI-MS/MS) in TDP for resolving heterogeneous proteoforms.[1-3]

Constructing orthogonal multidimensional separation platforms in TDP is vital due to the high complexity of proteoforms in biological systems.[1-3] Offline proteoform fractionations in solution or gel, such as gel-eluted liquid fraction entrapment electrophoresis (GELFrEE), Passively Eluting Proteins from Polyacrylamide gels as Intact species (PEPPI), size exclusion chromatography (SEC), and RPLC, have been extensively hyphenated to RPLC-MS/MS or CE-MS/MS to boost the proteoform identification to thousands and even tens of thousands level from cell lysates.[4-9] For example, most recently, our laboratory integrated SEC and/or RPLC to capillary zone electrophoresis (CZE)-MS/MS for TDP of a pair of colorectal cancer cell lines and identified over 23,000 proteoforms.[9] Incorporating multi-dimensional liquid-phase fractionation is beneficial for achieving a higher proteome coverage but needs large starting materials (hundred micrograms to milligrams of the samples) and is relatively labor-intensive.

High-field asymmetric waveform ion mobility spectrometry (FAIMS), as a fractionation strategy in the gas phase, provides rapid and online filtering of ions based on their differential mobilities in oscillating high and low electric fields. The detailed mechanism of FAIMS has been explained in various studies.[10-12] Briefly, the FAIMS is composed of cylindrical inner and outer electrodes with dispersion voltage (DV) applied to deliver an asymmetric waveform. The ions which have different mobilities in high- and low-field segments of waveform eventually end up colliding with the electrodes. The fractionation of ions can be facilitated by applying compensation voltage (CV) on the inner electrode to

offset the drift of ions and selectively allow the transmission of specific groups of ions. A variety of applications of FAIMS in bottom-up proteomics (BUP) have been reported.[12-18] By offering gas-phase fractionation and improved sensitivity, the technique greatly benefited the identification of peptides carrying post-translational modifications (PTMs) and enhanced the depth of proteome coverage. Most recently, FAIMS presented attractive performances in protein complex analysis and TDP.[5,11,19-21] In particular, for TDP, controlling the CV of FAIMS showed potency to fractionate proteoforms according to masses.[11,20] The FAIMS has been coupled with RPLC-MS/MS for deep TDP of cells and tissues.[5,11,20,21] The works generally doubled the number of proteoform identifications (IDs) in contrast to the conditions without FAIMS.

CZE is one of the most popular CE approaches, which enables the differentiation of proteoforms on the basis of their charge-to-size ratios. Our group previously demonstrated the advantages of CZE-MS/MS on separation resolution and sensitivity for TDP.[3,7,22] In this study, we presented the first example of CZE-FAIMS-MS/MS for TDP. The yeast lysate extracted by an ammonium bicarbonate (ABC, pH ~ 8) buffer was used as a model sample to evaluate the performance of the platform. Different CVs were tested by performing a single CV per CZE-MS/MS run. The results were largely compared between FAIMS and no FAIMS conditions to understand the features of FAIMS and how FAIMS benefits the sensitivity of detection and proteoform IDs in CZE-MS/MS analyses. In addition, we examined how CZE-FAIMS-MS/MS benefits the decipherment of proteoform families. Finally, we investigated the CZE-FAIMS-MS/MS for the characterization of yeast proteoforms extracted using a urea buffer and compared the data with that from the ABC lysis buffer.

EXPERIMENTAL SECTION

Materials and Reagents

ABC, phosphate-buffered saline (PBS), urea, baker's yeast, yeast extract peptone dextrose, and Amicon Ultra centrifugal filter units (0.5 mL, 10 kDa cutoff size) were purchased from Millipore Sigma (Burlington, MA). Protease inhibitors (cOmplete ULTRA Tables, Roche), phosphatase inhibitors (PhosSTOP, Roche), bicinchoninic acid (BCA) kit, bare fused silica capillaries (50 μm i.d., 360 μm o.d., Polymicro Technologies), LC/MS-grade water, acetic acid, formic acid, and methanol were purchased from Fisher Scientific (Hampton, NH).

Sample Preparation

Around 0.2g of baker's yeast (*Saccharomyces cerevisiae*, strain ATCC 204508/S288c) was added in 1 L of yeast extract peptone dextrose (YPD) medium (autoclaved) and cultured at 37 °C (300 rpm shaking) overnight in an incubator shaker (Thermo Scientific MaxQ 4000). The yeast was harvested by centrifugation at 3000*g* for 5 min, followed by washing with PBS three times. The yeast pellets were suspended in cell lysis buffer containing 100 mM ABC (pH 8.0), protease inhibitor, and phosphatase inhibitor. The yeast cells were lysed (3 min, 3 times) using a homogenizer 150 (Fisher Scientific) and sonicated on ice (10 min) with Branson Sonifier 250 (VWR Scientific). The supernatant of the cell lysate was collected with centrifugation at 18,000*g* for 10 min. The same procedure was applied to

the protein extraction of yeast using a buffer consisting of 8 M urea, 100 mM ABC (pH 8.0), protease inhibitor, and phosphatase inhibitor. The concentrations of protein samples were determined using a bicinchoninic acid (BCA) kit according to the manufacturer's instructions. Before CZE-MS/MS and CZE-FAIMS-MS/MS analyses, 150 μg of the yeast lysates in ABC buffer/urea buffer was loaded onto Amicon centrifugal filters (10 kDa molecular weight cutoff) for buffer exchange. The samples were centrifuged at 14,000*g* for 15 min at 10 °C and then washed four times with 50 mM ABC. Around 60 μL of ABC-extracted protein (2.0 mg/mL) and urea-extracted protein samples (1.2 mg/mL) were recovered from each filter. The buffer-exchanged protein samples were aliquoted and stored at −20 °C. Prior to CE separation, the samples were further centrifuged at 14,000*g* for 3 min to remove potential precipitates to avoid capillary clogging or current dropping.

CZE-MS/MS and CZE-FAIMS-MS/MS Analyses

The CE-MS/MS system was set up by coupling a CESI 8000 Plus CE system (Sciex) to an Orbitrap Exploris 480 spectrometer (Thermo Fisher Scientific) with an in-house-constructed electrokinetically pumped sheath-flow CE-MS nanospray interface.[23,24] A glass spray emitter with an orifice size of 30–35 μm was installed on the interface and filled with sheath liquid consisting of 0.2% (v/v) formic acid and 10% (v/v) methanol. The spray voltage was adjusted in the range of 2.2~2.4 kV to generate stable electrospray. The capillary (100 cm length, 50 μm i.d, and 360 μm o.d) for CZE was coated with linear polyacrylamide (LPA), according to our previous protocol.[25] The inlet of the capillary was fixed in the cartridge of the CE system, and the outlet was inserted into the emitter of the interface. The distance of the capillary outlet to the emitter orifice was around 0.5 mm.

To carry out CZE, the capillary was flushed with a background electrolyte (BGE, 5% acetic acid, pH 2.4) at 10 psi for 10 min, followed by loading of 200 ng of yeast lysate (1 mg/mL, injection volume of 200 nL). Afterward, the inlet of the capillary was inserted into the BGE (5% acetic acid) for CZE separation with a separation voltage of 30 kV applied.

For the mass spectrometer, the temperature of the ion transfer tube was set to 320 °C, and the RF lens was 60%. The intact protein mode was turned on, and the low-pressure mode was selected. The MS/MS experiments were performed using data-dependent acquisition (DDA). Full MS scan was performed with the following parameters: Orbitrap resolution of 480,000 (at *m/z* of 200), *m/z* range of 500–2500, normalized AGC target of 300%, and microscan of 1. The top 6 most intense precursors in full MS spectra were isolated with a window of 2 *m/z* and fragmented using HCD collision energy of 25%. Only precursors with charge states in the range of 5–60 and intensities higher than the threshold value of 10,000 were included for fragmentation. Other parameters for MS/MS include a resolution of 60,000 (at *m/z* 200), *m/z* range of 200–2000, microscan of 3, normalized AGC target of 100%, and auto maximum injection time. The dynamic exclusion was applied with a duration of 30 s, and the exclusion of isotopes was enabled.

The FAIMS Pro Duo interface (Thermo Fisher Scientific) was installed, had auto DV tune, and was set to standard resolution for CZE-FAIMS-MS/MS analysis. The distance between the spray emitter and the FAIMS inlet was controlled to 2–3 mm. The nitrogen carrier gas was set as default (4.6 L/min). Different CV voltages ranging from −50 to 30 V with 10 V

intervals were tested for nine individual CZE-MS/MS runs to investigate the fractionation performance of the FAIMS.

Data Analysis

All the raw files were converted to mzML files using MSConvert[26] and further deconvoluted to Msalign files using TopFD (version 1.4.7). The converted data were searched against the Uniprot *S. cerevisiae* database (UP000002311_559292) using TopPIC (1.4.7).[27] Parameters for database search were set as follows: mass error of precursors and fragments of 15 ppm, the maximum number of unexpected modifications of 2, and maximum and minimum mass shifts of unknown modifications of 500 and −500 Da, respectively. The false discovery rates (FDRs) were estimated using the target-decoy approach. The spectrum level FDR cutoff was 1%, and the proteoform level FDR cutoff was 5%.

RESULTS AND DISCUSSION

CZE-FAIMS-MS/MS as an Online Two-Dimensional Technique for Substantially Better Identification of Proteoforms than CZE-MS/MS

As shown in Figure 1A, CZE-FAIMS-MS/MS provides online two-dimensional (2D) separations of proteoforms prior to MS and MS/MS. Proteoforms are first separated by CZE in the liquid phase by their charge-to-size ratios and further fractionated in the gas phase by FAIMS according to their charges and sizes. A yeast protein sample extracted by a salt buffer (ABC) was used to investigate the performance of the system. Our main purpose is to study the online proteoform fractionation performance of FAIMS for CZE-MS/MS-based TDP.

Nine different CVs ranging from −50 to 30 V with 10 V increments were tested. The same CZE-MS/MS condition was also conducted without FAIMS for comparison. We first evaluated the reproducibility of CZE-MS/MS and CZE-FAIMS-MS/MS for the characterization of the yeast sample. For CZE-FAIMS-MS/MS, we chose the −40 V CV for this purpose. As shown in Figure S1A, the CZE-MS/MS without FAIMS and CZE-FAIMS-MS/MS (−40 V) are reproducible in terms of the proteoform separation profiles, the number of proteoform IDs (326 ± 12 for without FAIMS and 342 ± 10 for FAIMS; relative standard deviations, RSDs, <4%), and the normalized level of base peak intensity (RSDs ≤ 27%). The proteoform overlaps between technical replicates of CZE-MS/MS and CZE-FAIMS-MS/MS (−40 V) ranged from 54 to 58% and 60 to 73%, respectively, Figure S1B. CZE-FAIMS-MS/MS provided better reproducibility regarding the identified proteoforms (p-value = 0.09, Student's t-test, two-tailed distribution) compared to CZE-MS/MS, most likely due to the simplified mass spectra of proteoforms after FAIMS fractionation, reducing the randomness of precursor isolation for MS/MS in DDA.

As shown in Figure 1B, CZE-FAIMS-MS/MS with different CVs produced unique separation profiles of proteoforms that are substantially different from CZE-MS/MS without FAIMS. When the CV increased from −50 to 30 V, the number of proteoform IDs per run decreased from 432 to 20–40, Table S1. CZE-FAIMS-MS/MS with a single CV of −50 V identified about 30% more proteoforms compared to CZE-MS/MS without FAIMS (432 vs

327), Table S1. CZE-FAIMS-MS/MS with the combination of 9 CVs improved the number of proteoform IDs and protein IDs by about 3-fold and 2-fold compared to the results without FAIMS (940 vs 327 proteoforms; 288 vs 126 proteins). The data suggest the good complementarity of CZE and FAIMS for proteoform separations and indicate the capability of CZE-FAIMS-MS/MS for boosting the proteome coverage from TDP.

To better understand the CZE-FAIMS-MS/MS data, we examined the proteoform overlaps between different CVs, Figure 2A. The proteoform overlap is about 30–40% or even lower between the neighboring CVs. For example, a 10% proteoform overlap was observed between the 20 and 30 V CVs. The difference in the CV value is bigger, and the proteoform overlap is smaller. Additionally, the proteoform overlap between different CVs is much smaller than the technical replicates of the same CV (≤40% vs 60–73%). The data demonstrate the nice fractionation performance of FAIMS for proteoforms in the gas phase.

We further studied the mass distribution of proteoforms identified from CZE-MS/MS and CZE-FAIMS-MS/MS with different CVs, Figure 2B. The median mass was 7 kDa without FAIMS fractionation. With FAIMS, the median mass increased from 6 to 30 kDa when the CV was increased from −50 to 30 V, Figure 2B. The results provide clear evidence that the CV value of FAIMS and proteoform mass have a positive correlation in our CZE-FAIMS-MS/MS experiment. CZE-FAIMS-MS/MS boosted the identification of relatively large proteoforms (>20 kDa). As shown in Figure 2C, CZE-FAIMS-MS/MS with 9 different CVs greatly improved the proteoform IDs in the whole mass range of <10–45 kDa compared to CZE-MS/MS, particularly in the mass ranges of 20–30 and 30–45 kDa. Nearly 6-fold larger proteoforms (>30 kDa) were identified using CZE-FAIMS-MS/MS, Figure 2C. In addition, 65% of proteoforms identified by CZE-MS/MS were covered by the CZE-FAIMS-MS/MS data, Figure 2C. The data here are important for TDP. It is challenging to identify large proteoforms (i.e., >30 kDa) in typical CZE-MS/MS and RPLC-MS/MS analyses of complex proteomes due to their drastically lower signal-to-noise ratios (S/N).[3,28] Our data suggest that the 2D CZE-FAIMS-MS/MS has the potential to advance TDP toward the characterization of large proteoforms.

To study how CV impacts the proteoform transmission and fragmentation, we selected six proteoforms with different masses and isoelectric points (pIs) and compared their charge states, feature intensity (proteoform intensity reported by TopPIC), number of fragment ions, and sequence coverage at CZE-FAIMS-MS/MS provides online two-dimensional (2D) different CVs, Figure S2. Large proteoforms were able to transmit through a wider CV voltage range than smaller proteoforms, due to their wider charge state distribution. In addition, large proteoforms favored transmission of higher charge states as CV was changed from negative toward positive values, Figure S2A-C, which is consistent with the previous observation.[20] The pIs of proteoforms could also impact the transmission of charge states. For example, H2A1 and PROF have a similar mass (~14 kDa) but different pIs (10.7 vs 5.5). PROF favored a higher charge state when CV was more negative, whereas H2A showed a different trend, Figure S2C,D. The proteoform's feature intensity had a strong impact on the number of fragment ions and sequence coverage of proteoforms, Figure S2B,D,F. This could be because higher feature intensities of precursors benefit S/N of fragment ions. The charge

state can also influence the number of fragment ions and sequence coverage of proteoforms from HCD substantially, Figure S2A.

We also investigated the sensitivity improvement of CZE-FAIMS-MS/MS compared to CZE-MS/MS. We manually extracted base peak electropherograms of 20 proteoforms identified in both CZE-FAIMS and no FAIMS conditions and compared their S/N, Figure 2D. On average, CZE-FAIMS offered a 50-fold improvement in S/N compared to no FAIMS. The median S/N improvement is about 18-fold. The improvement is mainly because of the drastically reduced background noise and nice proteoform fractionation by FAIMS, which agrees reasonably well with the observations in other studies.[11,18] We need to highlight that CZE-FAIMS-MS/MS (9 CVs) as a multi-dimensional (MD) TDP technique only requires several micrograms of protein materials (i.e., 2 μg of yeast proteins). When other MD TDP techniques (e.g., SEC-RPLC-MS/MS,[6] GELFrEE-RPLC-MS/MS,[29] and SEC-CZE-MS/MS[7,9]) were employed for TDP of complex proteomes, usually hundreds of micrograms of protein materials are needed due to the offline operations. We expect that the CZE-FAIMS-MS/MS will be a useful tool for TDP of mass-limited biological samples.

Better Decipherment of Proteoform Families by CZE-FAIMS-MS/MS

Proteoform family represents a group of proteoforms derived from the same gene.[30] A better knowledge about proteoform members in each proteoform family will undoubtedly improve our understanding of protein function. We studied the potential of CZE-FAIMS-MS/MS for better decipherment of proteoform families compared to CZE-MS/MS.

CZE-FAIMS-MS/MS with 9 CVs identified 940 proteoforms and 288 proteoform families. CZE-MS/MS identified 327 proteoforms and 126 proteoform families. By matching the UniProt protein accession number between the two data sets, 118 proteoform families were identified in both cases. As shown in Figure 3A, most of the overlapped proteoform families have proteoforms from more than one CV, and about 50% of the proteoform families have proteoforms from more than 3 different CVs. The CZE-FAIMS-MS/MS in the best CV condition produced more proteoforms than CZE-MS/MS for 56% of the overlapped proteoform families. Overall, the number of proteoforms per proteoform family was improved by over 2-fold with CZE-FAIMS-MS/MS (9 CVs) compared to CZE-MS/MS, Figure 3B. Figure 3C shows two examples of proteoform families (TPIS and CYPH) from CZE-FAIMS-MS/MS and CZE-MS/MS. The proteoforms of TPIS were detected in a wide CV range from −30 to 30 V. Four out of six CVs identified more proteoforms than no FAIMS. Similarly, for CYPH, three out of four CVs (from −40 to −10 V) produced more proteoforms than no FAIMS. For both proteoform families, the total number of proteoforms was increased by over 6-fold with FAIMS. The results demonstrate that CZE-FAIMS-MS/MS will provide much better decipherment of proteoform families than CZE-MS/MS.

CZE offers a good solution for the separation of proteoforms carrying some PTMs, i.e., phosphorylation, due to their impact on proteoforms' charge-to-size ratios. We previously reported a significant reduction of electrophoretic mobility of proteoforms carrying phosphorylation and N-terminal acetylation.[31,32] Here, we further investigated whether or not the CZE-FAIMS-MS/MS can provide additional benefits for the characterization

of proteoforms containing PTMs, i.e., phosphorylation, acetylation, and methylation. The CZE-FAIMS-MS/MS (9 CVs) identified 214 proteoforms carrying acetylation, 17 proteoforms with phosphorylation, and 26 proteoforms with methylation. In comparison, the CZE-MS/MS only identified 86 proteoforms with acetylation, 10 proteoforms with phosphorylation, and 5 proteoforms with methylation.

As shown in Figure 4A, the two peaks baseline-separated in CZE-FAIMS (0 V)-MS/MS represents two proteoforms of the Nascent polypeptide-associated complex subunit alpha (NAC-α) with N-terminal methionine excision and N-terminal acetylation but containing no phosphorylation (proteoform 1) and phosphorylation (proteoform 2), respectively. The annotated MS/MS spectra of the two proteoforms are shown in Figure S3. The fragmentation pattern in Figure 4B shows that the phosphorylation of NAC-α is located between Pro92 and Ala113. The modification site can be further confirmed to Ser93 by the information on UniProt (https://www.uniprot.org/uniprot/P38879). NAC-α can either be tethered to the cytoplasmic ribosome and function as a complex component of the nascent polypeptide-associated complex (NAC) to modulate co-translational processes and protein translocation or accumulate in nuclei to participate in transcriptional coactivation.[33-37] The phosphorylation of NAC-α was found to be regulated by the proteasome pathway and associated with its degradation.[33] Interestingly, Figure 4A shows that non-phosphorylated NAC-α has a much higher abundance than the phosphorylated proteoform. None of the two NAC-α proteoforms identified by CZE-FAIMS-MS/MS were detected in the triplicate CZE-MS/MS runs without FAIMS, most likely due to signal suppression from other co-migrating species. Another example is the non-phosphorylated and phosphorylated 60 S acidic ribosomal protein P2-beta (RPLP2-β) resolved by CZE-FAIMS (−30 V)-MS/MS (Figures S4 and S5). The phosphorylation was identified at both −30 and −40 V (Figure S4B,C) and could be localized to Ser100 by combining fragmentation and UniProt information (https://www.uniprot.org/uniprot/P02400). RPLP2-β is one of the constituents of ribosomes and is engaged with protein synthesis. The phosphorylation at Ser at the highly conserved C-terminal can be regulated by a series of kinases (e.g., multifunctional protein kinase CK II and RAP kinases) to promote their interaction with elongation factor 2.[38] In our study, the phosphorylated RPLP2-β was shown as a major form in the yeast. Only the high-abundance phosphorylated proteoform of RPLP2-β was found in no FAIMS condition, and it presented a much lower feature intensity and confidence of proteoform identification (intensity: 1.49×10^5 and *E*-value: 1.3×10^{-14}) than FAIMS at −40 V (intensity: 9.47×10^7 and *E*-value: 1.13×10^{20}).

The results mentioned above strongly suggest that the incorporation of FAIMS into CZE-MS/MS can provide better decipherment of proteoform families in complex proteomes compared to CZE-MS/MS alone.

Impact of Proteoform Extraction Buffers on Proteoform IDs Using CZE-FAIMS-MS/MS

In this study, we offered CZE-FAIMS-MS/MS as a multi-dimensional platform to promote proteoform IDs, especially large proteoforms. In TDP, sample preparation also has a strong influence on proteoform IDs. TDP typically uses a buffer containing detergents or urea for cell lysis and protein extraction, followed by desalting/buffer exchange using a filter

membrane with a specific molecular weight cutoff (i.e., 10 kDa) before ESI-MS and MS/MS analysis.[39,40] Protein precipitation could occur in this process due to a dramatic change in protein solubility, leading to loss of relatively large proteins. For our dynamic pH junction-based CZE-MS/MS, the sample buffer needs to be exchanged to a relatively basic buffer (50 mM ABC, pH~8) to enable dynamic pH junction-based proteoform stacking.[22,41] We speculate that the direct use of the ABC buffer for proteoform extraction could reduce the loss of large proteins in buffer exchange and benefit their IDs using CZE-MS/MS. To achieve a better understanding of how the protein extraction buffers influence large proteoform IDs using CZE-MS/MS, we also carried out CZE-FAIMS-MS/MS analyses of a yeast cell lysate extracted by an 8 M urea buffer and compared the identified proteoforms with that from the ABC buffer.

CZE-FAIMS-MS/MS experiment of the urea sample was performed at three single CVs (−50, −40, and −30 V) for a better throughput. Based on our experience with the ABC sample, 80% of proteoforms could be identified in those three CVs. We found that the overlaps of proteoforms between the urea and yeast samples at the same CVs or without FAIMS are generally small (20–28%), Figure S6A, suggesting that the two buffers favor the extraction of different proteoforms. The pI distribution of proteoforms (Figure S7A) presented a similar pattern between ABC and urea. However, ABC tends to extract more hydrophilic proteoforms (lower GRAVY score) than urea, Figure S7B. The gene ontology (GO) analysis of cellular components revealed that ABC was able to extract proteins from the cytoplasm, organelles, and nucleus, but urea was more effective in extracting plasma membrane proteins (Tables S2 and S3), which agrees with our findings in the GRAVY score, Figure S7B.

We identified more proteoforms from the urea sample than from the ABC sample, either without FAIMS (1217 vs 327) or with 3-CV combined (2070 vs 769), Figure S6B. However, a very low percentage of these proteoforms (~4%) in urea has masses higher than 10 kDa. For the ABC sample, the proteoforms above 10 kDa account for 22% without FAIMS and 27% for three-combined CVs (−50, −40, and −30 V). The absolute number of proteoforms larger than 10 kDa is also higher in ABC than in urea (211 vs 103, 3 CV-combined), Figure S6B. Using nine CVs for the ABC sample further boosted larger proteoform identification (>20 kDa) from 14 (3 CVs) to 114, which is 12% of total proteoform IDs. We noted that, for the urea sample, CZE-FAIMS-MS/MS (3 CVs) identified 70% more total proteoforms and nearly 200% more proteoforms larger than 10 kDa than CZE-MS/MS alone. The data agree reasonably well with the ABC sample data. Furthermore, the ABC sample tends to present more “intact” proteoforms that either cover the full protein sequences or only have N-terminal methionine removed. Our result showed that 78 (23%) and 268 (29%) proteoforms in the ABC sample were intact without and with FAIMS, respectively, Figure S6C. In contrast, the majority of proteoforms (96%) in urea were truncated forms. The results here are very interesting and could be important for TDP of large proteoforms. The reasons for the drastic differences between the ABC and urea samples in terms of proteoform mass are unclear, and we will study more biological systems to pursue a better understanding of the phenomenon in our future work.

The improvement of intact proteoform ID allows us to decipher their functions and properties in cells more accurately. We identified eight intact proteoforms of isoform cytoplasmic of glutaredoxin 2 (Grx2c, P17695-2) in the ABC sample. Grx2c is an important glutathione-dependent oxidor-eductase in the cytosol, participating in the reduction of protein disulfide bonds.[42,43] Our result showed that intact Crx2c proteoforms were either N-terminal methionine removed or reserved and contained different PTMs. Typically, four proteoforms with N-terminal methionine cleavage were identified by CZE-FAIMS(−40 V)-MS/MS, Figure S8. The representative fragmentation pattern and MS/MS spectra for assigning mass shifts and PTMs are shown in Figure S8B,C. One has a disulfide bond between Cys 27 and Cys30 (*E*-value: 1.77×10^{-11}); one contains phosphorylation at Ser57 (*E*-value: 6.49×10^{-7}), one has a mass shift of 129 Da in the range of Leu51 to Leu53 (*E*-value: 4.05×10^{10}), which might be glutamylation on Glu52; and the other one has a mass shift of 210 Da from Leu51 to Ser57 (*E*-value: 1.54×10^{-6}), which could be a combination of phosphorylation and glutamylation. The glutamylation of Glu was previously found in tubulin[44] but never reported in Grx2c. We suspect that the two glutamylated Grx2 proteoforms might result from other biological processes related to glutathione. In addition, we identified 33 proteoforms related to isomers of acidic ribosomal P proteins (RPLP1- α, RPLP1-β, RPLP2-α, and RPLP2-β). Interestingly, while intact proteoforms were found in both RPLP2-α and RPLP2-β, only truncated proteoforms were detected in RPLP1-α and RPLP1-β. The result is in good agreement with the report that P1 proteins generally have much lower half-lives than P2 in yeast cells.[38]

CONCLUSIONS

We presented the first example of coupling FAIMS to CZE-MS/MS for TDP of a complex proteome, and the results demonstrated that CZE-FAIMS-MS/MS is a useful tool for advancing TDP toward better proteome coverage and sensitivity as well as better decipherment of proteoform families.

We need to further highlight two important features of CZE-FAIMS-MS/MS. One is the identification of large proteoforms (>30 kDa). TDP of large proteoforms is challenging due to their low S/N during MS detection, the difficulty for efficient gas-phase fragmentation, and loss during the sample preparation (i.e., offline fractionation, sample cleanup, and transfer). Here, we demonstrated that CZE-FAIMS-MS/MS as an online 2D TDP technique offered efficient size-based proteoform fractionation and improved the ID of large proteoforms by 6-fold compared to CZE-MS/MS alone. The other one is sensitivity improvement. TDP of mass-limited samples (i.e., single cells) is impeded by its low sensitivity. We showed that CZE-FAIMS-MS/MS boosted the S/N of proteoforms by nearly 20-fold compared to CZE-MS/MS alone. It has been well documented that CZE-MS/MS has better sensitivity than RPLC-MS/MS for proteoform measurement.[3,22] The CZE-FAIMS-MS/MS identified thousands of proteoforms from a yeast cell lysate (ABC or urea) using only less than two μg of proteins. These features render CZE-FAIMS-MS/MS an important tool for next-generation TDP toward the characterization of large proteoforms and mass-limited samples.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The work was funded by the National Cancer Institute (NCI) through the grant R01CA247863. We also thank the support from National Institute of General Medical Sciences (NIGMS) through grants R01GM125991 and R01GM118470. We also thank the support from the National Science Foundation (CAREER Award, Grant DBI1846913).

REFERENCES

(1). Toby TK; Fornelli L; Kelleher NL Annu Rev Anal Chem (Palo Alto Calif) 2016, 9, 499–519. [PubMed: 27306313]

(2). Chen B; Brown KA; Lin Z; Ge Y Anal. Chem 2018, 90, 110–127. [PubMed: 29161012]

(3). Chen D; McCool E; Yang Z; Shen X; Lubeckyj RA; Xu T; Wang Q; Sun LA Mass Spectrom. Rev 2023, 42, 617–642. [PubMed: 34128246]

(4). Melani RD; Gerbasi VR; Anderson LC; Sikora JW; Toby TK; Hutton JE; Butcher DS; Negrão F; Seckler HS; Srzentić K; Fornelli L; Camarillo JM; LeDuc RD; Cesnik AJ; Lundberg E; Greer JB; Fellers RT; Robey MT; DeHart CJ; Forte E; Hendrickson CL; Abbatiello SE; Thomas PM; Kokaji AI; Levitsky J; Kelleher NL Science 2022, 375, 411–418. [PubMed: 35084980]

(5). Takemori A; Kaulich PT; Cassidy L; Takemori N; Tholey A Anal. Chem 2022, 94, 12815–12821. [PubMed: 36069571]

(6). Cai W; Tucholski T; Chen B; Alpert AJ; McIlwain S; Kohmoto T; Jin S; Ge Y Anal. Chem 2017, 89, 5467–5475. [PubMed: 28406609]

(7). McCool EN; Lubeckyj RA; Shen X; Chen D; Kou Q; Liu X; Sun L Anal. Chem 2018, 90, 5529–5533. [PubMed: 29620868]

(8). Xu T; Shen X; Yang Z; Chen D; Lubeckyj R; McCool EN; Sun L Anal. Chem 2020, 92, 15890–15898. [PubMed: 33263984]

(9). McCool E; Xu T; Chen W; Beller NC; Nolan SM; Hummon AB; Liu X; Sun L Sci. Adv 2022, 8, 6348.

(10). Cooper HJ J. Am. Soc. Mass Spectrom 2016, 27, 566–577. [PubMed: 26843211]

(11). Gerbasi VR; Melani RD; Abbatiello SE; Belford MW; Huguet R; McGee JP; Dayhoff D; Thomas PM; Kelleher NL Anal. Chem 2021, 93, 6323–6328. [PubMed: 33844503]

(12). Hebert AS; Prasad S; Belford MW; Bailey DJ; McAlister GC; Abbatiello SE; Huguet R; Wouters ER; Dunyach J-J; Brademan DR; Westphall MS; Coon JJ Anal. Chem 2018, 90, 9529–9537. [PubMed: 29969236]

(13). Swearingen KE; Hoopmann MR; Johnson RS; Saleem RA; Aitchison JD; Moritz RL Mol. Cell. Proteomics 2012, 11, M111.014985.

(14). Schnirch L; Nadler-Holly M; Siao S-W; Frese CK; Viner R; Liu F Anal. Chem 2020, 92, 10495–10503. [PubMed: 32643919]

(15). Adoni KR; Cunningham DL; Heath JK; Leney AC J. Proteome Res 2022, 21, 930–939. [PubMed: 35235327]

(16). Greguš M; Kostas J; Ray S; Abbatiello SE; Ivanov AR Anal. Chem 2020, 92, 14702–14712. [PubMed: 33054160]

(17). Fang P; Ji Y; Silbern I; Viner R; Oellerich T; Pan K; Urlaub HT Anal. Chem 2021, 93, 8846–8855.

(18). Johnson KR; Greguš M; Ivanov AR J. Proteome Res 2022, 21, 2453–2461. [PubMed: 36112031]

(19). Hale OJ; Illes-Toth E; Mize TH; Cooper HJ Anal. Chem 2020, 92, 6811–6816. [PubMed: 32343119]

(20). Fulcher J; Makaju A; Moore RJ; Zhou M; Bennett DA; De Jager PL; Qian WJ; Paša-Tolić L; Petyuk VA J. Proteome Res 2021, 20, 2780–2795. [PubMed: 33856812]

(21). Kaulich PT; Cassidy L; Winkels K; Tholey A Anal. Chem 2022, 94, 3600–3607. [PubMed: 35172570]

(22). Lubeckyj RA; McCool EN; Shen X; Kou Q; Liu X; Sun L Anal. Chem 2017, 89, 12059–12067. [PubMed: 29064224]

(23). Sun L; Zhu G; Zhang Z; Mou S; Dovichi NJ J. Proteome Res 2015, 14, 2312–2321. [PubMed: 25786131]

(24). Wojcik R; Dada OO; Sadilek M; Dovichi NJ Rapid Commun. Mass Spectrom 2010, 24, 2554–2560. [PubMed: 20740530]

(25). Xu T; Han L; George Thompson AM; Sun L Anal. Methods 2022, 14, 383–393. [PubMed: 34939625]

(26). Kessner D; Chambers M; Burke R; Agus D; Mallick P Bioinformatics 2008, 24, 2534–2536. [PubMed: 18606607]

(27). Kou Q; Xun L; Liu X Bioinformatics 2016, 32, 3495–3497. [PubMed: 27423895]

(28). Schaffer LV; Millikin RJ; Miller RM; Anderson LC; Fellers RT; Ge Y; Kelleher NL; LeDuc RD; Liu X; Payne SH; et al. Proteomics 2019, 19, 1800361.

(29). Vellaichamy A; Tran J; Kelleher NL; Lee JE; Kellie JF; Catherman AD; Zamdborg L; Thomas PM; Sweet SMM; Durbin KR; Valaskovic GA; Ahlf DR F. Anal. Chem 2010, 82, 1234–1244.

(30). Smith LM; Kelleher NL Nat. Methods 2013, 10, 186–187. [PubMed: 23443629]

(31). Chen D; Lubeckyj RA; Yang Z; McCool EN; Shen X; Wang Q; Xu T; Sun L Anal. Chem 2020, 92, 3503–3507. [PubMed: 32043875]

(32). Chen D; Yang Z; Shen X; Sun L Anal. Chem 2021, 93, 4417–4424. [PubMed: 33650845]

(33). Andersen KM; Semple CA; Hartmann-Petersen R Mol. Biol. Rep 2007, 34, 275–281. [PubMed: 17211518]

(34). Ott A; Locher L; Koch M; Deuerling E PLoS One 2015, 10, No. e0143457. [PubMed: 26618777]

(35). Raue U; Oellerer S; Rospert SJ Biol. Chem 2007, 282, 7809–7816.

(36). George R; Walsh P; Beddoe T; Lithgow T FEBS Lett. 2002, 516, 213–216. [PubMed: 11959135]

(37). Quélo I; Akhouayri O; Prud'homme J; St-Arnaud R Biochemistry 2004, 43, 2906–2914. [PubMed: 15005626]

(38). Tchórzewski M. Int. J. Biochem. Cell Biol 2002, 34, 911–915. [PubMed: 12007628]

(39). Yang Z; Shen X; Chen D; Sun L J. Proteome Res 2020, 19, 3315–3325. [PubMed: 32419461]

(40). Donnelly DP; Rawlins CM; DeHart CJ; Fornelli L; Schachner LF; Lin Z; Lippens JL; Aluri KC; Sarin R; Chen B; et al. Nat. Methods 2019, 16, 587–594. [PubMed: 31249407]

(41). Britz-McKibbin P; Chen DD Anal. Chem 2000, 72, 1242–1252. [PubMed: 10740866]

(42). Collinson EJ; Grant CM J. Biol. Chem 2003, 278, 22492–22497. [PubMed: 12684511]

(43). Porras P; McDonagh B; Pedrajas J; Bárcena JA; Padilla CA Biochim Biophys Acta Proteins Proteom BBA-PROTEINS PROTEOM 2010, 1804, 839–845.

(44). Eddé B; Rossier J; Le Caer J-P; Desbruyères E; Gros F; Denoulet P Science 1990, 247, 83–85. [PubMed: 1967194]

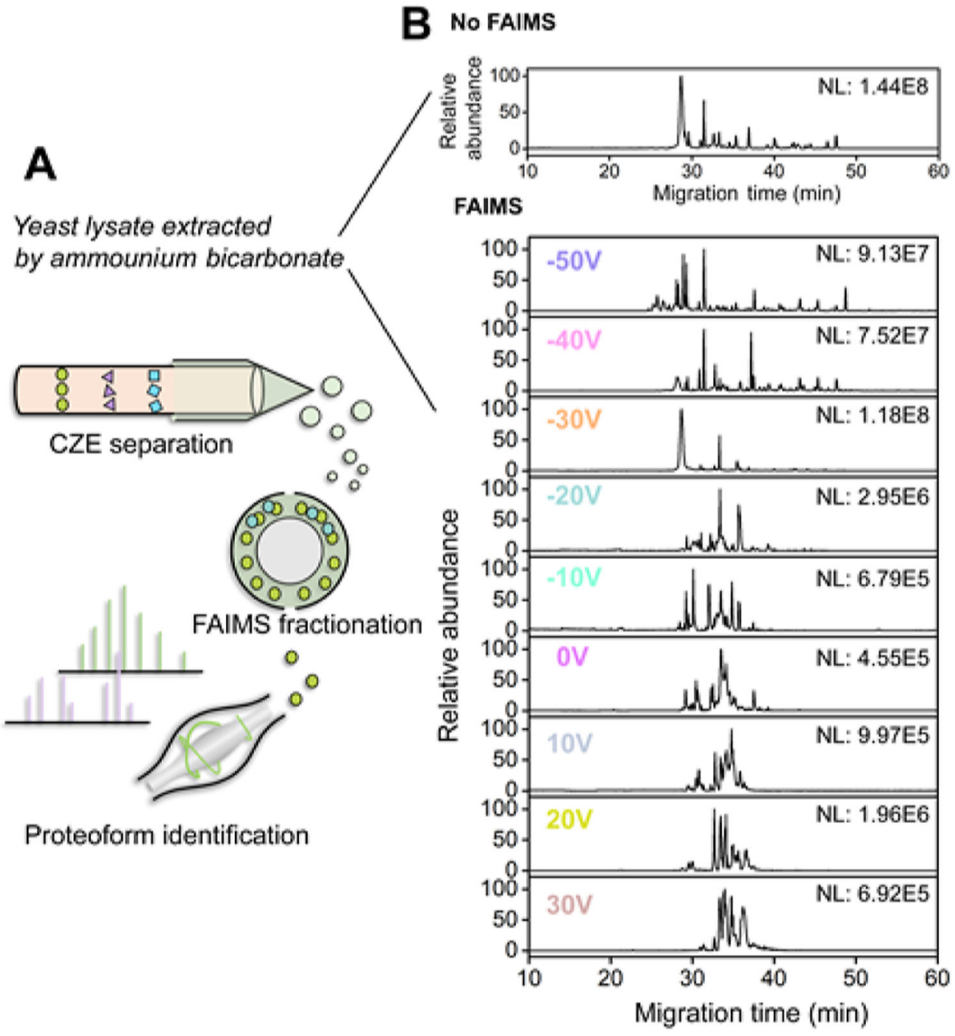


Figure 1.

Top-down proteomics of yeast lysate by CZE-FAIMS-MS. (A) Flow chart of CZE-FAIMS-MS for profiling the proteoforms of yeast lysate extracted by 100 mM ABC. (B) Base peak electropherograms of yeast lysate without FAIMS and with FAIMS at nine different CVs.

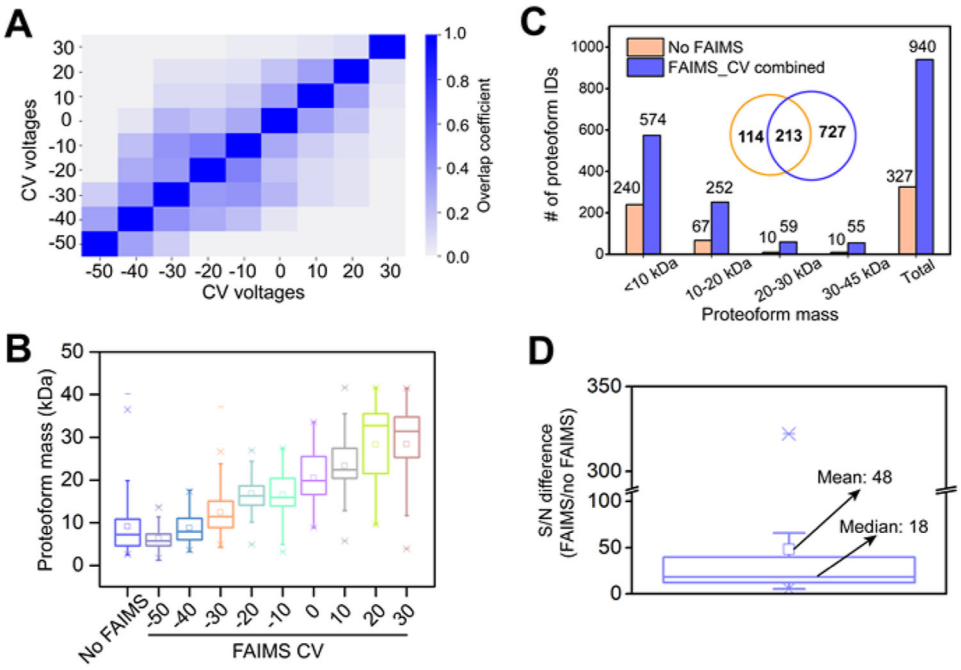


Figure 2.

Comparison of proteoform identifications without FAIMS and with FAIMS in CZE-MS analysis. (A) Overlap of identified proteoforms between FAIMS CVs. (B) Mass distributions of proteoforms identified without FAIMS and with different FAIMS CVs. (C) Comparison of the number of proteoform identifications at the different mass ranges between no FAIMS condition and FAIMS condition. (D) Improvement of the signal-to-noise ratio (S/N) of proteoforms with FAIMS relative to no FAIMS condition.

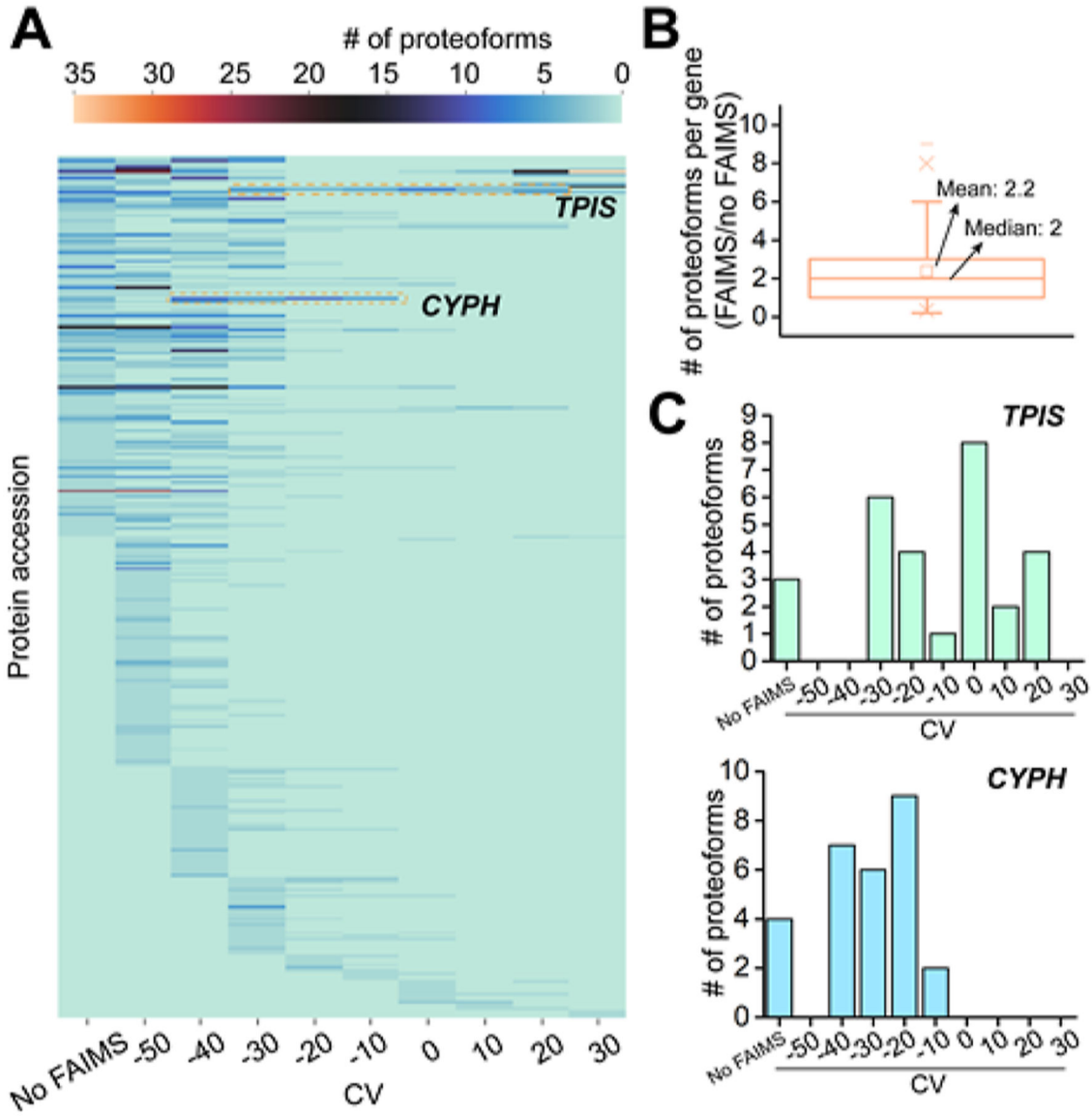


Figure 3.

Gas-phase fractionation of proteoforms originating from the same genes in CZE-FAIMS-MS analysis. (A) Heatmap of proteoform IDs at different CVs merged by their protein accessions. (B) Improvement of the number of proteoforms per gene using FAIMS relative to no FAIMS condition. (C)Number of proteoform IDs of CYPH and TPIS without FAIMS and at different CVs.

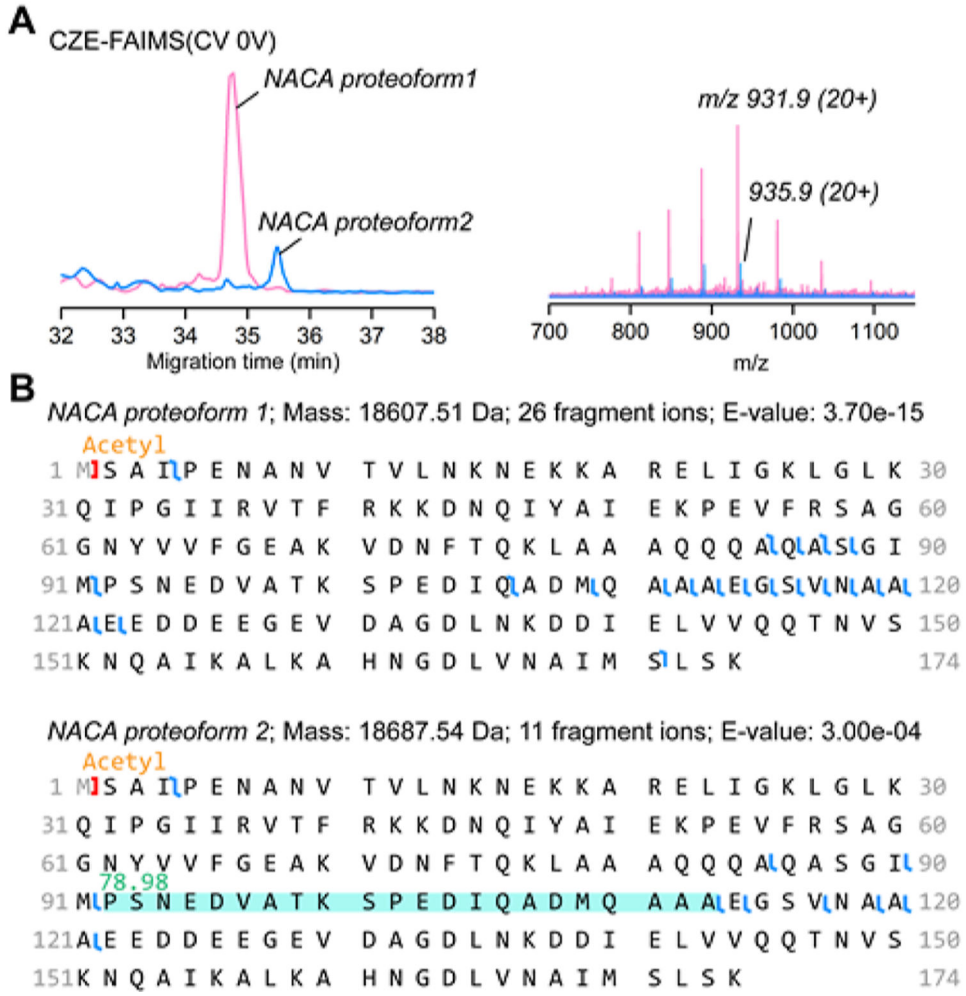


Figure 4.

Two intact proteoforms of the nascent polypeptide-associated complex subunit alpha (NAC-α) identified in CZE-FAIMS-MS analysis at a CV value of 0 V. (A) Overlapped base peak electropherograms (left) and mass spectra (right) of two NAC-α proteoforms. (B) Fragmentation patterns of NAC-α proteoform 1 (unphosphorylated intact NAC-α) and NAC-α proteoform 2 (phosphorylated NAC-α).