analytical chemistry

pubs.acs.org/ac Article

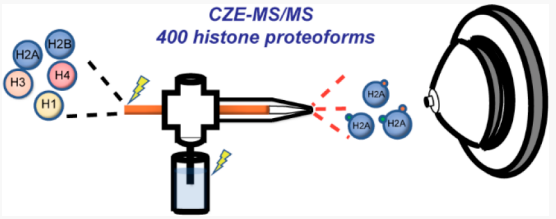
Capillary Zone Electrophoresis-Tandem Mass Spectrometry As an Alternative to Liquid Chromatography-Tandem Mass Spectrometry for Top-down Proteomics of Histones

Daoyang Chen, Zhichang Yang, Xiaojing Shen, and Liangliang Sun*

Cite This: *Anal. Chem.* 2021, 93, 4417−4424 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Top-down proteomics (TDP) is an ideal approach for deciphering the histone code and it routinely employs reversed-phase liquid chromatography (RPLC)-tandem mass spectrometry (MS/MS). Because of the extreme complexity of histones regarding the number of proteoforms, new analytical tools with high-capacity separation and highly sensitive detection of proteoforms are required for TDP of histones. Here we present capillary zone electrophoresis (CZE)-MS/MS via the electro-kinetically pumped sheath-flow CE-MS interface for large-scale top-down delineation of histone proteoforms. CZE-MS/MS identified a comparable number of proteoforms to RPLC-MS/MS from a calf histone sample with more than 30-fold less sample consumption (75-ng vs. Three μg), indicating its substantially higher sensitivity. We identified about 400 histone proteoforms from the calf histone sample using two-dimensional size-exclusion chromatography (SEC)-CZE-MS/MS with less than 300-ng proteins consumed. We identified histone proteoforms carrying various tentative post-translational modifications (PTMs), for example, acetylation, methylation (mono-, di-, and tri-), phosphorylation, and succinylation. The electrophoretic mobility (μ_{ef}) of unmodified histone proteoforms can be predicted accurately (R^2 = 0.98) with an optimized semiempirical model based on our recent work. The results render CZE-MS/MS as a useful tool for deciphering the histone code in a proteoform-specific manner and on a global scale.



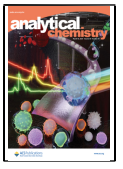
INTRODUCTION

Histones, as crucial members of the nucleosome, are of great importance in modulating chromatin structure and mediating epigenetic regulations.[1−3] The N-termini of histones are heavily modified by various post-translational modifications (PTMs) such as methylation, acetylation, and phosphorylation.[4,5] The combinations of these PTMs, "histone code", play a central role in regulating gene expression. Hundreds of PTMs have been discovered from histones, which significantly increases the complexity of the histone code.[6] Deciphering the histone code is vital for understanding the mechanism of gene expression regulation.

Mass spectrometry (MS)-based proteomics has been widely used for identifying histone PTMs and determining the combinations of various PTMs in a sensitive and high-throughput fashion.[2,7,8] Bottom-up (BU), middle-down, and top-down proteomics (TDP) have been used for studying the histone code.[2,7−10] BU and middle-down approaches have shown their robustness in studying histones and related PTMs.[11−21] However, after enzymatic digestion, the extreme basic N-termini can be fragmented into very short peptides which are undetectable in BU approach, resulting in loss of PTM information. Interesting alternative BU approaches were introduced to generate large histone peptides from trypsin digestion via chemically modifying lysine residues before digestion.[22−24] BU and middle-down approaches cannot distinguish histone variants having similar sequences and they are not ideal for delineating histones in a proteoform-specific manner. TDP directly characterizes intact histone proteoforms and can provide a bird's eye view of histone proteoforms with various PTMs. It is an ideal approach for studying the histone code.

The TDP community has made substantial efforts for characterizing histones.[25−37] In the 2000s, the Kelleher's group and the Mizzen's group pioneered the histone TDP with a series of publications.[25−31] Using weak cation exchange-hydrophilic interaction liquid chromatography (WCX-HILIC), the Kelleher's group identified and quantified 42 proteoforms from histone H4 with a concentration dynamic range of 10^4.[30] To further improve the throughput for histone TDP, the Pasa-Tolic's group established a salt-free online WCX-HILIC/RPLC-MS/MS platform, which remarkably identified over 700 histone proteoforms,[32] representing one of the largest histone TDP data sets. The Young's group quantitatively characterized proteoforms of histone H2A, H2B,

Received: October 8, 2020
Accepted: February 18, 2021
Published: March 2, 2021



https://dx.doi.org/10.1021/acs.analchem.0c04237
Anal. Chem. 2021, 93, 4417−4424

and H4 using RPLC-MS/MS.[33,34] Recently, the Brodbelt's group employed RPLC-MS/MS in the combination of different fragmentation methods, including 193 nm ultraviolet photodissociation (UVPD), for TDP of histones using a commercial calf histone sample.[35] The number of histone proteoform identifications (IDs) is still limited. We believe alternative methods with better separation capacity and higher detection sensitivity of proteoforms can further advance TDP of histones drastically.

Capillary zone electrophoresis (CZE)-MS/MS has been recognized as a useful approach for large-scale TDP.[38−42] CZE-MS/MS has also been employed for the characterization of intact histones in some early studies.[36,37] For example, the Lindner's group utilized a cationic coated capillary and CZE-MS to analyze intact histones.[37] However, there is still no report of large-scale TDP of histones using CZE-MS/MS.

Here we present CZE-MS/MS for large-scale TDP of enriched calf histones for the first time. We first performed triplicate CZE-MS/MS analyses of the histone sample and compared our data with the RPLC-MS/MS data from the Brodbelt's group. To maximize the number of proteoform IDs, we coupled offline size exclusion chromatography (SEC) to CZE-MS/MS for TDP of histones and confidently identified about 400 histone proteoforms. At the end, we optimized the semiempirical model from our previous work[42] for predicting the μ_{ef} of histone proteoforms.

EXPERIMENTAL SECTION

The details about "Material and Regents", "Capillary Coating", and "SEC Fractionation" are described in the Supporting Information I.

CZE-MS/MS. An ECE-001 CE autosampler (CMP scientific, Brooklyn, NY) was used for CZE separation. A 1-m-long linear polyacrylamide (LPA)-coated capillary was used for CZE separation and the coating was prepared according to the previous publications.[43−45] A third-generation electrokinetically pumped sheath flow CE-MS interface (an EMASS-II CE-MS interface, CMP Scientific) was employed to couple the autosampler and a Q Exactive HF mass spectrometer.[46,47] An electrospray emitter with 20−40-μm opening was pulled with a Sutter P-1000 flaming/brown micropipette puller.

For CZE separation, the background electrolyte (BGE) was 5% (v/v) acetic acid (AA, pH 2.4). The sheath liquid was 0.2% (v/v) formic acid (FA) and 10% (v/v) methanol in water. The etched end of a 1-m-long LPA-coated capillary was threaded through the tee of CE-MS interface and introduced to the electrospray emitter. The distance between the capillary end and the orifice of the emitter was ~300 μm. The distance between the emitter orifice and the inlet of mass spectrometer was ~2 mm. 2.2 kV voltage was applied to the sheath liquid reservoir for electrospray ionization.

For sample injection, 5 psi was applied for 5 s for each sample. Based on Poiseuille's law, approximately 25 nL of each sample was injected. Thirty kV voltage was applied to the injection end for 3000 and 4800 s for the analyses of SEC fractions and the histone sample without SEC fractionation, respectively. After the separation, 30 kV voltage and 10 psi pressure were applied for 600 s to push out the residue analytes in the capillary.

A Q Exactive HF mass spectrometer (Thermo Fisher Scientific) was used for all the experiments. A Top 3 data-dependent acquisition (DDA) method was used. For MS, the resolution, AGC target, and maximum injection time were set to 120,000, 1e6 and 50 ms, respectively. The scan range of MS and MS/MS were 400−1500 and 200−2000 m/z, respectively. For MS/MS, the resolution, AGC target, and maximum injection time were set to 60 000, 1e6 and 400 ms, respectively. The ions for MS/MS were isolated in the quadrupole with an isolation window of 2 m/z, followed by fragmentation employing a stepped higher energy collisional induced dissociation (HCD) method with three-step normalized collision energy as 12%, 16%, and 20%. Dynamic exclusion was set to 50 s. Ions with charge states lower than 7 were excluded for the fragmentation.

Data Analysis. Proteome Discoverer 2.4 sp software (Thermo Fisher Scientific) with the ProSight PD top-down nodes was used for database search.[48] The MS1 spectra were first averaged using the cRAWler algorithm in Proteome Discoverer. The precursor m/z tolerance was set to 0.2 m/z. For both precursor and fragmentation Xtract parameters, the signal-to-noise ratio threshold, the lowest and the highest m/z were set to 3, 200, and 2000, respectively. Then deconvolution was performed by the Xtract algorithm followed by database search against a *Bos taurus* downloaded from http://proteinaceous.net/-database-warehouse-legacy/ in April 2018. A three-prone database search was performed: (1) a search was performed with a 10-ppm mass tolerance of absolute mass for both MS1 and MS2; (2) a search was performed with a 200-Da mass tolerance for MS1, and a 10-ppm mass tolerance for MS2 for matching unexpected PTMs; (3) a subsequent search was performed to find unreported truncated proteoforms with 10 ppm tolerance for both MS1 and MS2. The target-decoy strategy was employed for evaluating the false discovery rates (FDRs). For a possible ID, FDR estimation was performed for each of three search strategies. Proteoform IDs were filtered first using single FDR threshold from each search, and then by the global FDR estimated by the best q-value in three searches. The identified proteoform-spectrum matches (PrSMs), proteoforms and proteins were filtered using a 1% FDR.

To ensure confident identification, we applied a C-score filter to all PrSMs and proteoforms (C-score >3). The C-Score (Characterization Score) is used to indicate how well proteoforms are characterized (e.g., location of PTMs). A higher C-Score indicates better proteoform characterization. Proteoforms with C-Scores higher than 3 are confidently identified and partially characterized; Proteoforms with C-Scores higher than 40 are fully characterized. Because the backbone cleavage coverage of identified histone proteoforms are limited in this work and it is challenging to accurately assign and localize PTMs on proteoforms solely based on the matched fragment ions, all the PTM assignments and localization in this work are tentative. The lists of identified proteoforms are shown in the Supporting Information II.

Experimental and Predicted Electrophoretic Mobility (μ_{ef}). Migration time (t_M, s) of each proteoform or PrSM was obtained from the database search results. The electroosmotic flow (EOF) was assumed zero in the LPA-coated capillary with 5% AA (pH 2.4) as the BGE. For calculating experimental μ_{ef}, eq 1 was used,

$$\text{experimental}\mu_{ef} = L/((30 - 2)/L \times t_M)(\text{unit of cm}^2\text{kV}^{-1}\text{s}^{-1}) \quad (1)$$

Where L is the capillary length in cm, 30 and 2 are separation voltage and electrospray voltage in kV. The eq 1 is based on the literature.[42,49] For calculating predicted μ_{ef}, eq 2 was used,

https://dx.doi.org/10.1021/acs.analchem.0c04237
Anal. Chem. 2021, 93, 4417−4424

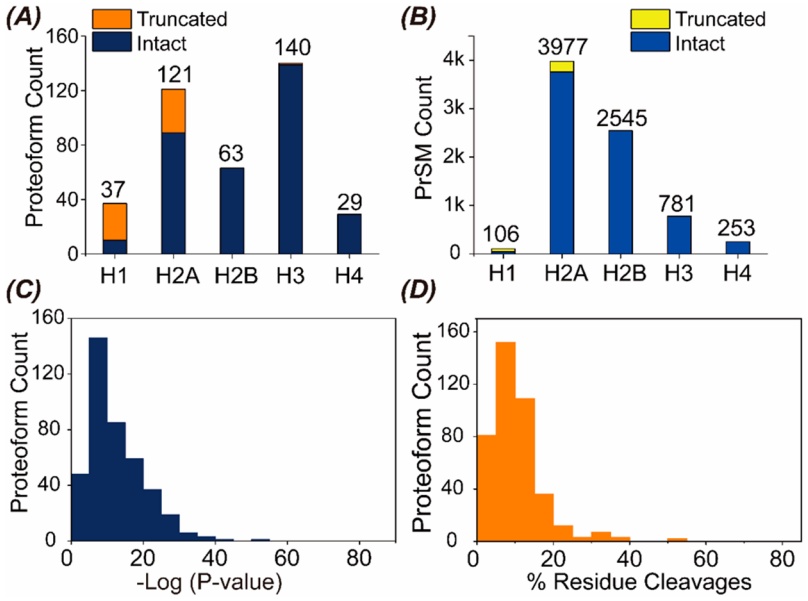


Figure 1. Summary of the SEC-CZE-MS/MS data. (A) Proteoform count distribution for histones. (B) PrSM count distribution for histones. (C) Distribution of the−log(*P*-score) of the identified proteoforms. (D) Residue cleavage coverage distribution of the identified proteoforms.

$$\text{predicted}\mu_{ef} = \ln(1 + 0.350 \times Q)/M^{0.411} \quad (2)$$

Where Q is the number of charges of each proteoform in the liquid phase, represented by the number of positively charged amino acid residues in the proteoform sequence (K, R, H, and N-terminus). M is the molecular mass equals to the mass reported by the search engine in Da. The eq 2 is based on the previous publications.[42,49−51]

The calculation of electrophoretic mobility relative difference (EMRD) was performed using the following equation,

$$\text{EMRD} = (\text{experimental}\mu_{ef}\text{-predicted}\mu_{ef}(\text{abs}))/\text{predicted}\mu_{ef}(\text{abs}) \quad (3)$$

Where the predicted μ_{ef} (abs) is the absolute value of predicted μ_{ef} which can be calculated using the following equation based on the correlation equation shown in Figure 2B,

$$\text{predicted}\mu_{ef}(\text{abs}) = ((\text{predicted}\mu_{ef}) - 0.016)/0.11 \quad (4)$$

With the assumption that our prediction model can accurately predict the μ_{ef} of unmodified proteoforms, the absolute predicted μ_{ef} of each modified proteoform calculated solely based on the proteoform sequence equals the μ_{ef} of its unmodified counterpart. Therefore, the EMRD represents the relative difference in μ_{ef} between the modified and unmodified proteoforms with the same protein sequence.

RESULTS AND DISCUSSIONS

Comparisons of CZE-MS/MS and RPLC-MS/MS for TDP of Histones. We identified 152 proteoforms from triplicate analyses of the calf histone sample with only 25-ng calf histones injected per CZE-MS/MS run. The data clearly indicates the high sensitivity of CZE-MS/MS for TDP of histones. Histone proteoforms migrated out of the capillary within 30 min, SI Figure S1A. Major histone proteins were separated by CZE except for histone H2A and H2B due to their similar charge-to-size ratios.

We compared the CZE-MS/MS data with that from one recent RPLC-MS/MS study of the same calf histone sample. Greer et al. performed TDP of calf histones using RPLC-MS/MS with comparable MS and database search conditions to our work.[35] Using the HCD fragmentation, CZE-MS/MS and RPLC-MS/MS identified a comparable number of proteoforms from the same histone sample (152 vs. 176) with FDRs less than 1% and C-Score better than 3. CZE-MS/MS consumed more than 30-fold lower amount of proteins than RPLC-MS/MS (75 ng vs. Three μg).

The data can be explained by two possible reasons. First, CZE-MS has shown 10−100-times less sample consumption compared to nanoRPLC-MS for intact protein detection.[52,53] Second, a potential sample loss on the RPLC column exists because histones are very basic and hydrophilic. SI Table S1 shows lower percentages of identified proteoforms of intact histone H1 and H2B in the RPLC data set compared to the CZE data set. Histone H1 and H2B are relatively hydrophilic in the histone sample as demonstrated by their early elution during RPLC separation.[35,54]

CZE-MS/MS and RPLC-MS/MS identified different pools of histone proteoforms, SI Table S1 and Figure S1B. CZE-MS/MS and RPLC-MS/MS identified in total 35 post-translationally modified proteoforms of Histone H3.1 (UniProt ID: P68432), and only seven of them were identified in both data sets, SI Figure S1B. The data clearly suggest the good complementarity between RPLC-MS/MS and CZE-MS/MS in TDP of histones.

CZE has shown powerful separations of histone proteoforms in the literature using UV detection.[37] Here, we did not systematically optimize the CZE-MS conditions, and we expect a drastic improvement of CZE separation of histones regarding separation resolution could be achieved after optimizations of the capillary coatings and the composition of BGEs.

Overall, the results demonstrate that CZE-MS/MS is a highly sensitive analytical tool with good complementarity with RPLC-MS/MS for large-scale delineation of histone proteoforms.

SEC-CZE-MS/MS for Comprehensive Delineation of Histone Proteoforms. We coupled SEC fractionation to CZE-MS/MS for analysis of the calf histone sample to boost the number of histone proteoform IDs. We fractionated the

https://dx.doi.org/10.1021/acs.analchem.0c04237
Anal. Chem. 2021, 93, 4417−4424

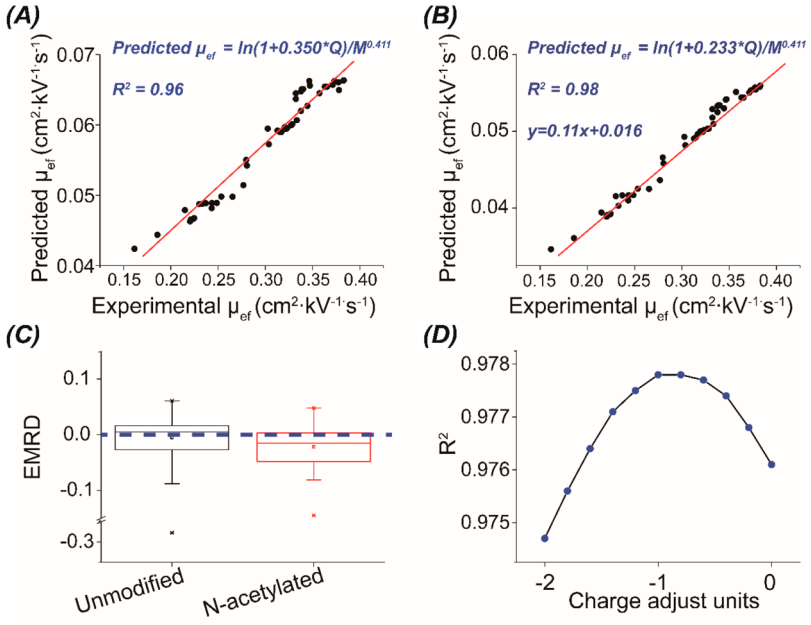


Figure 2. Correlations between predicted μ_{ef} and experimental μ_{ef} of unmodified proteoforms of calf histones before the model optimization using a prefactor of 0.350 to Q (A) and after model optimization using a prefactor of 0.233 to Q (B). (C) Box plots of EMRDs of unmodified and N-terminal acetylated proteoforms. (D) The R^2 values between predicted and observed μ_{ef} when different charge adjustment was made to N-terminal acetylated proteoforms.

histone sample into 11 fractions using SEC. The protein concentration in each fraction was determined by the BCA assay. Then each fraction was analyzed by CZE-MS/MS with a consumption of 25-ng proteins per run. The SEC-CZE-MS/MS identified 48 proteins, 405 proteoforms, and 7832 PrSMs with a total protein consumption of less than 300 ng. The data were filtered with 1% FDRs and C-Score better than 3. Among the 405 proteoforms, 173 had C-score higher than 40 (43%), 390 were histone proteoforms and 15 were proteoforms of other proteins. The SEC separated the Histone H1 (i.e., fraction 3) from the truncated proteoforms (i.e., fraction 7), SI Figure S2. The average mass of identified proteoforms become smaller as the SEC fraction number increases from 3 to 7, indicating reasonable SEC separation. The data of SEC fractions 1−2 and 8−11 are strange regarding the average proteoform mass, which might be due to the interaction between histones and beads as well as protein−protein interactions.

Figure 1 summarizes the TDP data set of histones. We identified 140 proteoforms of H3, 121 proteoforms of H2A, 63 proteoforms of H2B, 37 proteoforms of H1, and 29 H4 proteoforms, Figure 1A. We noted that only 10 out of the 37 H1 proteoforms (27%) were intact ones with masses higher than 20 kDa. For H2A, H2B, H3, and H4, 74−100% of the identified proteoforms are intact ones, for example, 139 out of the 140 H3 proteoforms (99%) are intact forms of the protein with masses around 15 kDa. The number of PrSMs could be used as a label-free approach for roughly evaluating relative abundance of similar proteins and proteoforms.[39,55] As shown in Figure 1B, the H2A is the most-abundance histone protein in the sample, followed by H2B, H3, H4, and H1.

We further analyzed the distributions of P-Score, and residue cleavage coverage of identified proteoforms, Figure 1C,D. The *P*-score used by the ProSight represents a nonlinear transformation of the number of matched fragment ions. A lower *P*-score suggests a better match between a candidate proteoform and a tandem mass spectrum. The *P*-score of the identified proteoforms centers around 10^{-10}. Most of the identified proteoforms have lower than 20% residue cleavage coverage, Figure 1D. The results indicate that integration of alternative gas-phase fragmentation methods that have shown better fragmentation performance than HCD for histones (e.g., ETD/ECD,[33,34] AI-ETD,[56] and UVPD[35]) in our CZE-MS/MS system will drastically improve the characterization of the identified histone proteoforms.

Electrophoretic Mobility Prediction of Histone Proteoforms. Recently, we have reported accurate prediction of proteoforms' μ_{ef} with a simple semiempirical model (eq 2) using *E. coli* and zebrafish brain data sets.[42] The work opened the door of evaluating the confidence of proteoform IDs via comparing their predicted and experimental μ_{ef}. Herein, we evaluated the model in predicting μ_{ef} of histone proteoforms. We employed proteoforms without any PTMs for this purpose. The details of calculating the predicted and experimental μ_{ef} of these proteoforms are described in the Experimental Section. Only proteoforms or PrSMs having C-score higher than 3, −log *P*-value higher than 11.5, and mass error less than 20 ppm were used for μ_{ef} calculation. The SEC-CZE-MS/MS data were used.

As illustrated in Figure 2A, the correlation coefficient (R^2) between the predicted and experimental μ_{ef} of unmodified histone proteoforms was 0.96, which is lower than that in our previous report (0.98).[42] The prediction model, as shown in eq 2, contains two terms: the charge, represented by the logarithmic value of the prefactored charge (Q) of a proteofrom, and the size, represented by the value of the proteoform's mass (M) to the power of 0.411. To improve the correlation between the predicted and experiemntal μ_{ef} of histone proteoforms, we optimized the prediction model via independent adjustments of the size and charge terms. Little improvement was observed by adjusting the size term. Interestingly, by reducing the prefactor of Q from 0.350 to

https://dx.doi.org/10.1021/acs.analchem.0c04237
Anal. Chem. 2021, 93, 4417−4424

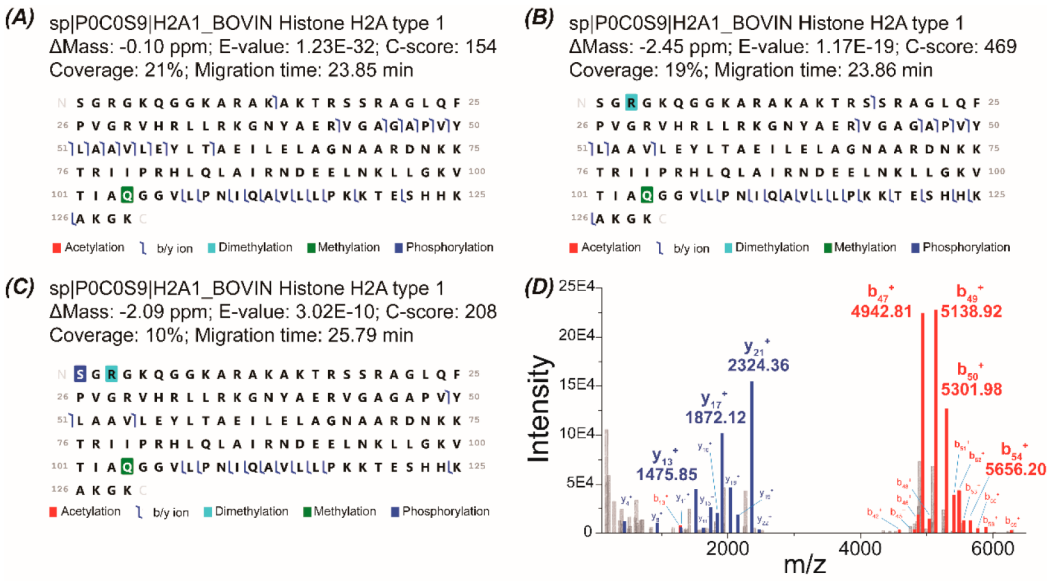


Figure 3. Sequences and fragmentation patterns of three histone H2A type 1 proteoforms (A–C) with various PTMs. (D) Deconvoluted tandem mass spectrum of the proteoform in (A) with fragment ion annotations. The *x*-axis shows the m/z values (+1) of the ions in the tandem mass spectrum after deconvolution.

0.233, the correlation coefficient (R^2) was improved to 0.98, Figure 2B. The prefactor of Q was originally introduced to compensate the charge suppression, which resulted from mutual electrostatic interactions of the charged groups in the peptides/proteins.[51] The charge suppression becomes more significant in highly charged peptides/proteins because when the total charge of an analyte increases, the effect of any additional charge on its μ_{ef} decreases.[57,58] Histones are highly charged in acidic BGE, so that the charge suppression in the enriched histone sample ought to be strong, leading to the need for a smaller prefactor of Q for accurate prediction of μ_{ef}. Unless stated otherwise, the eq 5 is used for predicting the μ_{ef} of histone proteoforms in the work.

$$\text{predicted}\mu_{ef} = \ln(1 + 0.233 \times Q)/M^{0.411} \quad (5)$$

We then evaluated the influence of PTMs on the charge and μ_{ef} of histone proteoforms. Certain PTMs such as N-terminal acetylation and phosphorylation reduce the charge (Q) of peptides/proteins by roughly one charge unit as demonstrated in early studies,[42,59] thus decreasing their μ_{ef} significantly. To perform quantitative evaluation on the influence of PTMs to μ_{ef} of histone proteoforms, we defined the electrophoretic mobility relative difference (EMRD), a parameter that represents the influence of a PTM to the μ_{ef} of a proteoform. The detailed calculation of EMRD is in the Experimental Section (eq 3 and eq 4). In theory, the EMRDs of N-terminal acetylated histone proteoforms tend to be less than 0 compared to the unmodified proteoforms due to the fact that the PTM reduces the number of positive charges of proteoforms in CZE. A data set containing 89 histone proteoforms, of which 48 were unmodified and 41 had only N-terminal acetylation, was used. As shown in Figure 2C, the EMRDs of unmodified histone proteoforms centers around zero as expected. The median of EMRDs of N-terminal acetylated histone proteoforms is below 0, indicating that N-terminal acetylation indeed has negative effect on the μ_{ef} of histone proteoforms.

We further evaluated the charge reducing effect of N-terminal acetylation on histone proteoforms. A high R^2 value of 0.976 was gained between the observed and the predicted μ_{ef} of all proteoforms in the data set even without any charge adjustment (charge adjustment unit = 0) to the N-terminal acetylated proteoforms, suggesting that the μ_{ef} difference between a N-terminal acetylated histone proteoform and its unmodified counterpart is small due to histones' high charge-to-size ratios, Figure 2D. Because histones are highly charged in the 5% (v/v) AA (pH 2.4) BGE, the changes in charge (Q) caused by the N-terminal acetylation have limited overall impact on their charge-to-size ratios in our CZE conditions, leading to small effect on their μ_{ef}. Interestingly, when we reduced the positive charge (Q) of N-terminal acetylated histone proteoforms stepwise, we observed the best R^2 value when the charge (Q) was reduced by 1 unit, Figure 2D. The data here highlight the challenges of achieving baseline separations of different proteoforms of one histone protein using our CZE condition, as demonstrated in SI Figure S1A. We need to note that the CZE conditions (e.g., BGE composition) can modulate the μ_{ef} of histones substantially as demonstrated by the works from the Zhong group.[60,61] We expect that systematic optimizations of the CZE condition could amplify the effect of the PTMs (e.g., N-terminal acetylation) on histone proteoforms' μ_{ef}, thus leading to better separations of modified and unmodified histone proteoforms.

Histone PTMs. Histones are usually post-translationally modified heavily and the ability of characterizing the PTMs and their combinations on histones is vital for understanding the molecular mechanisms of epigenetic regulations. Here we showed some examples of the identified histone proteoforms with PTMs.

As depicted in Figure 3, the CZE-MS/MS-based platform achieved identification of some modified proteoforms of histone H2A that have low E-values (well below 1×10^{-9}) and high C-Scores (better than 150) of proteoforms. Figure 3A–C show three different proteoforms of histone H2A type 1 and they carry different combinations of PTMs. Proteoform I

https://dx.doi.org/10.1021/acs.analchem.0c04237
Anal. Chem. 2021, 93, 4417–4424

(Figure 3A) has methylated at the Q (Q104me). Proteoform II (Figure 3B) carries the R3me2 and Q104me. Proteoform III (Figure 3C) has one phosphorylation at the N-terminus (S), one R3me2, and one methylation at the position 104 (Q104me). We need to point out that the localizations of the PTMs in the three proteoforms are tentative due to the limited cleavage coverage. More studies are needed to achieve better confidence for PTM localization and to gain a better understanding of the roles played by the combinations of different PTMs on histone proteoforms in epigenetic regulation of gene expression.

We further verified the identifications of the proteoforms in Figure 3A−C by manually checking the annotated tandem mass spectra. The deconvoluted and annotated tandem mass spectrum of the Proteoform I is shown in Figure 3D. Clear and strong signals of b and y ions of the proteoform were observed. Agreeing with the data in Figure 3A, the matched fragment ions in Figure 3D failed to cover the region with the tentative Q104me. SI Figure S3 shows the annotated tandem mass spectra of the proteoforms in Figure 3A−C. Because the three histone H2A proteoforms have an identical C-terminus, the matched high-abundance y ions corresponding to their C-termini are the same (e.g., $y21^+$, $y17^+$, and $y13^+$). By contrast, the high-abundance b ions of the three proteoforms (i.e., $b50^+$ and $b49^+$) have different masses corresponding to the tentative S1ph (+80 Da) and R3me2 (+28 Da). For example, the $b49^+$ of the Proteoform II is 28 Da heavier than that of the Proteoform I (5166.91 vs. 5138.92 Da), SI Figure S3B,C, suggesting a dimethylation or two methylations on the N-terminus of the Proteoform II. Similarly, the $b50^+$ ion of the Proteoform III is 80 Da heavier than that of the Proteoform II (5409.93 vs. 5329.98 Da), SI Figure S3C,D, indicating that besides the dimethylation or two methylations, there is another phosphorylation modification on the N-terminus of the Proteoform III. Two significant neutral loss peaks of $b50^+$ and $b49^+$ (−98 Da) in SI Figure S3D further indicate the existence of a phosphorylation modification on the Proteoform III. We noted that some strong peaks in the SI Figure S3C,D did not match with the Proteoform II and III, most likely due to the cofragmentation of different histone proteoforms with very similar m/z values, SI Figure S3E,F, further highlighting the challenges of delineating histone proteoforms.

One significant limitation of the current data set is the low backbone cleavage coverage of proteoforms from the HCD, impeding the accurate assignments and localization of PTMs on the identified proteoforms. Besides boosting the backbone cleavage coverage, the μ_{ef} of identified proteoforms should help confirm the PTMs on proteoforms since some PTMs (e.g., acetylation and phosphorylation) could decrease their μ_{ef} significantly.[42] The proteoform III in Figure 3C is a perfect example. It is 80-Da heavier and has a lower μ_{ef} compared to Proteoform II (migration time: 25.79 vs 23.86 min). The data suggests that the 80-Da modification is the reason for the reduction of μ_{ef} for Proteoform III. The combination of mass shift (80 Da) and μ_{ef} change provides higher confidence for the phosphorylation assignment on the Proteoform III. We need to point out that more studies are certainly required to demonstrate the usefulness of μ_{ef} of histone proteoforms for improving the assignment of PTMs. These studies most likely will need large and high-quality histone proteoform data sets.

CONCLUSIONS

We documented the use of CZE-MS/MS for large-scale TDP of histones for the first time and reported the identification of nearly 400 histone proteoforms from less than 300-ng protein material. The results demonstrate that CZE-MS/MS is a useful alternative with high sensitivity to RPLC-MS/MS for delineation of the histone code at the proteoform level. *The MS raw data have been deposited to the ProteomeXchange Consortium via the PRIDE*[62] *partner repository with the data set identifier PXD021889.*

We need to point out that the backbone cleavage coverage of identified histone proteoforms is limited in this work due to the unsatisfying performance of HCD for histone fragmentation. We will integrate electron or photon-based fragmentation methods (e.g., ECD/ETD and UVPD) with CZE-MS/MS for TDP of histones in our future studies for better fragmentation coverage. We will also improve the CZE separation of histone proteoforms via optimizing the compositions of BGEs.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.0c04237.

The experimental details; summary of intact proteoforms identified by CZE-MS/MS and RPLC-MS/MS; electropherogram of CZE-MS analysis of histones and overlap of proteoforms of Histone H3.1 with PTMs identified by CZE-MS/MS and RPLC-MS/MS; average mass of identified proteoforms as a function of SEC fraction number; annotated tandem mass spectra of some histone proteoforms (PDF)

Identified histone proteoforms using CZE-MS/MS and SEC-CZE-MS/MS (XLSX)

AUTHOR INFORMATION

Corresponding Author

Liangliang Sun − *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0001-8939-5042; Phone: 1-517-353-0498; Email: lsun@chemistry.msu.edu

Authors

Daoyang Chen − *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States*

Zhichang Yang − *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States*

Xiaojing Shen − *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0003-2079-9115

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.0c04237

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the support from the National Institute of General Medical Sciences (NIGMS) through Grant R01GM125991 and the National Science Foundation through Grant DBI1846913 (CAREER Award).

REFERENCES

(1) Su, Z.; Denu, J. M. *ACS Chem. Biol.* **2016**, *11*, 564−74.
(2) Moradian, A.; Kalli, A.; Sweredoski, M. J.; Hess, S. *Proteomics* **2014**, *14*, 489−97.
(3) Zheng, Y.; Huang, X.; Kelleher, N. L. *Curr. Opin. Chem. Biol.* **2016**, *33*, 142−50.
(4) Kouzarides, T. *Cell* **2007**, *128*, 693−705.
(5) Allis, C. D.; Jenuwein, T. *Nat. Rev. Genet.* **2016**, *17*, 487−500.
(6) Zhao, Y.; Garcia, B. A. *Cold Spring Harbor Perspect. Biol.* **2015**, *7*, a025064.
(7) Karch, K. R.; Denizio, J. E.; Black, B. E.; Garcia, B. A. *Front. Genet.* **2013**, *4*, 264.
(8) Yuan, Z. F.; Arnaudo, A. M.; Garcia, B. A. *Annu. Rev. Anal. Chem.* **2014**, *7*, 113−28.
(9) El, K. S.; Crespo, M.; Govin, J.; Pflieger, D. *Proteomes* **2018**, *6*, E29.
(10) Sidoli, S.; Garcia, B. A. *Expert Rev. Proteomics* **2017**, *14*, 617−626.
(11) Garcia, B. A.; Mollah, S.; Ueberheide, B. M.; Busby, S. A.; Muratore, T. L.; Shabanowitz, J.; Hunt, D. F. *Nat. Protoc.* **2007**, *2*, 933−8.
(12) Chen, Y.; Sprung, R.; Tang, Y.; Ball, H.; Sangras, B.; Kim, S. C.; Falck, J. R.; Peng, J.; Gu, W.; Zhao, Y. *Mol. Cell. Proteomics* **2007**, *6*, 812−9.
(13) Tan, M.; Luo, H.; Lee, S.; Jin, F.; Yang, J. S.; Montellier, E.; Buchou, T.; Cheng, Z.; Rousseaux, S.; Rajagopal, N.; Lu, Z.; Ye, Z.; Zhu, Q.; Wysocka, J.; Ye, Y.; Khochbin, S.; Ren, B.; Zhao, Y. *Cell* **2011**, *146*, 1016−28.
(14) Dai, L.; Peng, C.; Montellier, E.; Lu, Z.; Chen, Y.; Ishii, H.; Debernardi, A.; Buchou, T.; Rousseaux, S.; Jin, F.; Sabari, B. R.; Deng, Z.; Allis, C. D.; Ren, B.; Khochbin, S.; Zhao, Y. *Nat. Chem. Biol.* **2014**, *10*, 365−70.
(15) Zheng, Y.; Sweet, S. M.; Popovic, R.; Martinez-Garcia, E.; Tipton, J. D.; Thomas, P. M.; Licht, J. D.; Kelleher, N. L. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 13549−54.
(16) Jaffe, J. D.; Wang, Y.; Chan, H. M.; Zhang, J.; Huether, R.; Kryukov, G. V.; Bhang, H. E.; Taylor, J. E.; Hu, M.; Englund, N. P.; Yan, F.; Wang, Z.; Robert McDonald, E., 3rd; Wei, L.; Ma, J.; Easton, J.; Yu, Z.; deBeaumount, R.; Gibaja, V.; Venkatesan, K.; Schlegel, R.; Sellers, W. R.; Keen, N.; Liu, J.; Caponigro, G.; Barretina, J.; Cooke, V. G.; Mullighan, C.; Carr, S. A.; Downing, J. R.; Garraway, L. A.; Stegmeier, F. *Nat. Genet.* **2013**, *45*, 1386−91.
(17) Leroy, G.; Dimaggio, P. A.; Chan, E. Y.; Zee, B. M.; Blanco, M. A.; Bryant, B.; Flaniken, I. Z.; Liu, S.; Kang, Y.; Trojer, P.; Garcia, B. A. *Epigenet. Chromatin* **2013**, *6*, 20.
(18) Garcia, B. A.; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. *Nat. Methods* **2007**, *4*, 487−9.
(19) Phanstiel, D.; Brumbaugh, J.; Berggren, W. T.; Conard, K.; Feng, X.; Levenstein, M. E.; McAlister, G. C.; Thomson, J. A.; Coon, J. J. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 4093−8.
(20) Jiang, T.; Hoover, M. E.; Holt, M. V.; Freitas, M. A.; Marshall, A. G.; Young, N. L. *Proteomics* **2018**, *18*, e1700442.
(21) Holt, M. V.; Wang, T.; Young, N. L. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2514−2525.
(22) Maile, T. M.; Izrael-Tomasevic, A.; Cheung, T.; Guler, G. D.; Tindell, C.; Masselot, A.; Liang, J.; Zhao, F.; Trojer, P.; Classon, M.; Arnott, D. *Mol. Cell. Proteomics* **2015**, *14*, 1148−58.
(23) Garcia, B. A.; Mollah, S.; Ueberheide, B. M.; Busby, S. A.; Muratore, T. L.; Shabanowitz, J.; Hunt, D. F. *Nat. Protoc.* **2007**, *2*, 933−8.
(24) Liao, R.; Wu, H.; Deng, H.; Yu, Y.; Hu, M.; Zhai, H.; Yang, P.; Zhou, S.; Yi, W. *Anal. Chem.* **2013**, *85*, 2253−9.
(25) Pesavento, J. J.; Kim, Y. B.; Taylor, G. K.; Kelleher, N. L. *J. Am. Chem. Soc.* **2004**, *126*, 3386−7.
(26) Boyne, M. T., 2nd; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. *J. Proteome Res.* **2006**, *5*, 248−53.
(27) Thomas, C. E.; Kelleher, N. L.; Mizzen, C. A. *J. Proteome Res.* **2006**, *5*, 240−7.
(28) Siuti, N.; Roth, M. J.; Mizzen, C. A.; Kelleher, N. L.; Pesavento, J. J. *J. Proteome Res.* **2006**, *5*, 233−9.
(29) Parks, B. A.; Jiang, L.; Thomas, P. M.; Wenger, C. D.; Roth, M. J.; Boyne, M. T., 2nd; Burke, P. V.; Kwast, K. E.; Kelleher, N. L. *Anal. Chem.* **2007**, *79*, 7984−91.
(30) Pesavento, J. J.; Bullock, C. R.; LeDuc, R. D.; Mizzen, C. A.; Kelleher, N. L. *J. Biol. Chem.* **2008**, *283*, 14927−37.
(31) Pesavento, J. J.; Yang, H.; Kelleher, N. L.; Mizzen, C. A. *Mol. Cell. Biol.* **2008**, *28*, 468−86.
(32) Tian, Z.; Tolić, N.; Zhao, R.; Moore, R. J.; Hengel, S. M.; Robinson, E. W.; Stenoien, D. L.; Wu, S.; Smith, R. D.; Paša-Tolić, L. *Genome Bio.* **2012**, *13*, R86.
(33) Dang, X.; Singh, A.; Spetman, B. D.; Nolan, K. D.; Isaacs, J. S.; Dennis, J. H.; Dalton, S.; Marshall, A. G.; Young, N. L. *J. Proteome Res.* **2016**, *15*, 3196−203.
(34) Holt, M. V.; Wang, T.; Young, N. L. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2548−2560.
(35) Greer, S. M.; Brodbelt, J. S. *J. Proteome Res.* **2018**, *17*, 1138−1145.
(36) Aguilar, C.; Hofte, A. J.; Tjaden, U. R.; van der Greef, J. *J. Chromatogr. A* **2000**, *926*, 57−67.
(37) Sarg, B.; Faserl, K.; Kremser, L.; Halfinger, B.; Sebastiano, R.; Lindner, H. H. *Mol. Cell. Proteomics* **2013**, *12*, 2640−56.
(38) Lubeckyj, R. A.; Basharat, A. R.; Shen, X.; Liu, X.; Sun, L. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 1435−1445.
(39) McCool, E. N.; Lubeckyj, R. A.; Shen, X.; Chen, D.; Kou, Q.; Liu, X.; Sun, L. *Anal. Chem.* **2018**, *90*, 5529−5533.
(40) Valaskovic, G. A.; Kelleher, N. L.; McLafferty, F. W. *Science* **1996**, *273*, 1199−202.
(41) Shen, X.; Yang, Z.; McCool, E. N.; Lubeckyj, R. A.; Chen, D.; Sun, L. *TrAC, Trends Anal. Chem.* **2019**, *120*, 115644.
(42) Chen, D.; Lubeckyj, R. A.; Yang, Z.; McCool, E. N.; Shen, X.; Wang, Q.; Xu, T.; Sun, L. *Anal. Chem.* **2020**, *92*, 3503−3507.
(43) Zhu, G.; Sun, L.; Dovichi, N. J. *Talanta* **2016**, *146*, 839−43.
(44) McCool, E. N.; Lubeckyj, R.; Shen, X.; Kou, Q.; Liu, X.; Sun, L. *J. Visualized Exp.* **2018**, *140*, 58644.
(45) Chen, D.; Shen, X.; Sun, L. *Analyst* **2017**, *142*, 2118−2127.
(46) Sun, L.; Zhu, G.; Zhang, Z.; Mou, S.; Dovichi, N. J. *J. Proteome Res.* **2015**, *14*, 2312−21.
(47) Wojcik, R.; Dada, O. O.; Sadilek, M.; Dovichi, N. J. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 2554−60.
(48) LeDuc, R. D.; Taylor, G. K.; Kim, Y. B.; Januszyk, T. E.; Bynum, L. H.; Sola, J. V.; Garavelli, J. S.; Kelleher, N. L. *Nucleic Acids Res.* **2004**, *32*, W340−5.
(49) Krokhin, O. V.; Anderson, G.; Spicer, V.; Sun, L.; Dovichi, N. J. *Anal. Chem.* **2017**, *89*, 2000−2008.
(50) Cifuentes, A.; Poppe, H. *J. Chromatogr. A* **1994**, *680*, 321−40.
(51) Mittermayr, S.; Olajos, M.; Chovan, T.; Bonn, G. K.; Guttman, A. *TrAC, Trends Anal. Chem.* **2008**, *27*, 407−417.
(52) Han, X.; Wang, Y.; Aslanian, A.; Fonslow, B.; Graczyk, B.; Davis, T. N.; Yates, J. R., 3rd *J. Proteome Res.* **2014**, *13*, 6078−86.
(53) McCool, E. N.; Sun, L. *Sepu* **2019**, *37*, 878−886.
(54) Bonenfant, D.; Coulot, M.; Towbin, H.; Schindler, P.; van Oostrum, J. *Mol. Cell. Proteomics* **2006**, *5*, 541−52.
(55) Geis-Asteggiante, L.; Ostrand-Rosenberg, S.; Fenselau, C.; Edwards, N. J. *Anal. Chem.* **2016**, *88*, 10900−10907.
(56) Riley, N. M.; Westphall, M. S.; Coon, J. J. *J. Proteome Res.* **2017**, *16*, 2653−2659.
(57) Grossman, P. D.; Colburn, J. C.; Lauer, H. H. *Anal. Biochem.* **1989**, *179*, 28−33.
(58) Cifuentes, A.; Poppe, H. *Electrophoresis* **1997**, *18*, 2362−76.
(59) Chen, D.; Ludwig, K. R.; Krokhin, O. V.; Spicer, V.; Yang, Z.; Shen, X.; Hummon, A. B.; Sun, L. *Anal. Chem.* **2019**, *91*, 2201−2208.
(60) Lee, J.; Perez, L.; Liu, Y.; Wang, H.; Hooley, R. J.; Zhong, W. *Anal. Chem.* **2018**, *90*, 1881−1888.
(61) Lee, J.; Chen, J.; Sarkar, P.; Xue, M.; Hooley, R. J.; Zhong, W. *Anal. Bioanal. Chem.* **2020**, *412*, 6189−6198.
(62) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.;

https://dx.doi.org/10.1021/acs.analchem.0c04237
Anal. Chem. 2021, 93, 4417−4424

Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. *Nucleic Acids Res.* **2019**, *47*, D442−D450.

https://dx.doi.org/10.1021/acs.analchem.0c04237
Anal. Chem. 2021, 93, 4417−4424