Membrane ultrafiltration-based sample preparation method and sheath-flow CZE-MS/MS for top-down proteomics

Zhichang Yang and Liangliang Sun*

Department of Chemistry, Michigan State University, 578 S Shaw Ln, East Lansing, Michigan 48824 United States

* Corresponding author.

Email: lsun@chemistry.msu.edu; Phone: 517-353-0498

Abstract

Mass spectrometry (MS)-based denaturing top-down proteomics (dTDP) identify proteoforms without pretreatment of enzyme proteolysis. A universal sample preparation method that can efficiently extract protein, reduce sample loss, maintain protein solubility and be compatible with following up liquid-phase separation, MS and tandem MS (MS/MS) is vital for large-scale proteoform characterization. Membrane ultrafiltration (MU) was employed here for buffer exchange to efficiently remove the sodium dodecyl sulfate (SDS) detergent in protein samples used for protein extraction and solubilization, followed by capillary zone electrophoresis (CZE)-MS/MS analysis. The MU method showed good protein recovery, minimum protein bias, and nice compatibility with CZE-MS/MS. Single-shot CZE-MS/MS analysis of an *E. coli* sample prepared by the MU method identified over 800 proteoforms.

Keywords: Denaturing top-down proteomics, Mass spectrometry, Membrane ultrafiltration, capillary zone electrophoresis-mass spectrometry, proteoform, sodium dodecyl sulfate

1. Introduction

Denaturing top-down proteomics (dTDP) aims to characterize proteoforms in cells with high throughput.[1-3] It is becoming an important tool for better understanding of protein structure, PTMs and function in biological system. Mass spectrometry (MS)-based dTDP has achieved great advance due to tremendous efforts in development of proteoform liquid-phase separation,[4-14] MS instrumentation,[7, 15, 16] and new bioinformatics tools for proteoform identifications (IDs) through database search,[17-19] leading to thousands of proteoform IDs from a complex proteome.

A high throughput and comprehensive proteoform characterization can't do without an efficient and comprehensive extraction of proteins with high recovery, good reproducibility, minimum bias and free of MS incompatible salts, chaotropes and detergents.[20] Protein extraction is normally implemented with assistance of additive such as detergents and chaotropic reagent for a thorough protein extraction and denaturing. However, these detergents and chaotropic reagents need to be removed before MS analysis since they can cause significant ion suppression. Multiple strategies were developed for protein clean-up including membrane ultrafiltration (MU),[21] chloroform-methanol precipitation (CMP),[22] and single-spot solid-phase sample preparation using magnetic beads (SP3).[23, 24] Membrane ultrafiltration (MU) has been widely used by the bottom-up proteomics community for the filter-aided sample preparation (FASP) method to remove sodium dodecyl sulfate (SDS) before enzymatic digestion of proteins.[21] Basically, a protein sample in 1-5% (w/v) SDS solution is loaded onto a commercialized membrane filter unit with a 10-30-kDa molecular weight cut off (MWCO), followed by washing with a 8 M urea solution to remove SDS, which is based on the fact that 8 M urea can destroy the hydrophobic interaction between SDS and proteins. The cleaned protein can then be recovered with designated buffer for dTDP analysis or be subjected to enzyme digestion for Bottom-up analysis. We systematically compared MU, SP3 and CMP method for preparing protein ready for dTDP analysis through capillary zone electrophoresis mass spectrometry (CZE-MS) and concluded that MU method can be a universal strategy for dTDP sample preparation, due to its nature of high efficiency, high protein recovery, comprehensiveness and compatibility with downstream MS analysis. A workflow of MU

sample processing for dTDP is shown in **Figure 1**.[25] With sample prepared from MU strategy, we applied dynamic pH junction-based CZE-MS/MS on *E. coli* proteoform analysis and achieved over 800 proteoform IDs in a single shot. Here we provided a detail description on the MU strategy and highlighted some critical steps for high protein recovery.

2. Materials

All reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless stated otherwise. All solvents are prepared with water and reagents at LC-MS grade. All processing buffer/solution needs to be prepared fresh.

2.1. Prepare protein for MU clean-up

1. LB (Luria-Bertani) medium for E. coli culture.

2. Sterile phosphate buffered saline (PBS) without calcium and magnesium.

3. 50 mL Falcon tube.

3. 500 mL conical flask.

4. E. coli (strain K-12 substrain MG1655).

5. Cell lysis buffer: 1%(w/w) SDS, 100 mM NH_4HCO_3 (pH 8.0), 5 mg/mL protease inhibitor, 5 mg/mL phosphatase inhibitor. Store in 4 °C.

6. Centrifuge (compatible with 1.5 mL Eppendorf tube with maximum speed of 14,800 rpm, e.g., Thermo Scientific Legend Micro 21)

7. Ultrasonication (capable of inducing cell disruption, homogenization, and emulsification through cavitation, e.g., Branson Sonifier 250)

2.2. Membrane Ultrafiltration

1. Urea buffer: 8M urea, 100 mM NH_4HCO_3 (pH 8.0).

2. NH_4HCO_3 buffer: 100 mM NH_4HCO_3 (pH 8.0).

3. 30 kDa MWCO membrane unit (Millipore: MRCF0R030).

2.3. Prepare LPA (Linear polyacrylamide)-coated capillary

1. CE separation capillary: 360/50 (O.D./I.D.) fused silica capillary.

2. 1 M NaOH.

3. 1 M HCl.

4. γ-MAPS solution: 50%(v/v) 3-(trimethoxysilyl)propyl-methacrylate (in methanol (MeOH)).

5. Acrylamide solution: 40mg/mL acrylamide solution.

6. APS buffer: 5% w/v ammonium persulfate buffer (APS).

7. Coating solution: mix 3.5 µL APS and 500 µL acrylamide solution.

8. Hydrofluoric acid: 40% hydrofluoric acid.

9. 10 M NaOH solution.

10. 600 µL Eppendorf tubes.

2.4. CZE-MS

1. Background Electrolyte (BGE): 20% Acetic Acid.

2. Sheath buffer: 10% MeOH, 0.2% formic acid.

3. CE autosampler: CMP Scientific autosampler.

4. CE interface: electrokinetically pumped sheath flow CE-MS interface (EMASS II, CMP scientific, Brooklyn, NY).

5. Mass Spectrometer: Q-Exactive HF mass spectrometer (Thermo Fisher Scientific).

3. Method

3.1. Prepare protein for MU clean-up

1. Culture the *E. coli* (strain K-12 substrain MG1655) in the LB (Luria-Bertani) medium at 37 °C until OD600 reached 0.7.

2. Harvest the *E. coli* cells by centrifugation at 4000 rpm for 10 min. Wash the cell pellet with PBS buffer three times to remove the leftover culture medium.

3. Add 400 µL cell lysis buffer into the cell pellet. Pipette up and down a few times to lysis the cell. For thorough lysis, immerse sample tube into ultrasonication equipment filled with ice water and sonicate for 10 min (*see* **Note 1**).

4. Spin down the sample tube with speed of 14,000 g for 5 min. The proteins present in the supernatant. Take the supernatant out carefully and measure the protein concentration of the supernatant by BCA protein assay. Aliquot the protein into 600 µL Eppendorf tube with 100 µg in each tube and store the protein in -80 °C before use.

3.2. Ultrafiltration buffer exchange

1. Rinse the ultrafiltration membrane with 200 µL 100 mM NH_4HCO_3 buffer for membrane pretreatment.

2. Dilute the protein solution with 8 M urea buffer to concentration up to 1 µg/µL if the protein original concentration is too high (*see* **Note 2**).

3. Load the protein solution onto membrane and spin down the ultrafiltration unit with high speed until minimum solution remnants above the membrane (*see* **Note 3**).

4. Wash the membrane with 100 µL 8 M urea buffer at least two times and 100 µL 100 mM NH_4HCO_3 buffer 3 times through high speed (< 14,000 g) spin-down (*see* **Note 4**).

5. Add 50 to 100 µL of 100 mM NH_4HCO_3 onto the membrane, pipette up and down a couple times to resuspend the protein left on the membrane. Additional resuspension can be performed through 5 min of vortex.

6. Flip the membrane unit onto a new collection tube and collect the protein solution through a quick spin down.

7. Measure the protein concentration with BCA protein assay to estimate protein recovery and to adjust the amount of protein to be loaded for follow up MS identification (*see* **Note 5**).

3.3 Prepare capillary for CE-MS

1. Capillary pretreatment: Flush the capillary with following buffer in order: 1 M NaOH for more than 30 min; water for 30 min; 1M HCl for 30 min; water for 30 min; MeOH for 30 min.

2. Dry the capillary with Nitrogen gas flushing for 10 min.

3. Flush the capillary with 50% 3-(trimethoxysilyl)propyl-methacrylate for 10 min and seal both end with rubber. Incubate the capillary at room temperature for at least 24 h.

4. Rinse the capillary with MeOH and dry the capillary with nitrogen gas.

5. Capillary coating: Degas the LPA coating solution. Introduce LPA coating solution into the capillary and seal both end with rubber. Incubate the capillary in a 50 °C water bath for 50 min. *See* **Note 6.**

6. Flush the capillary with water and store the capillary at room temperature if not use right away. *See* **Note 7.**

7. Etch capillary: to accommodate the capillary into interface emitter with minimum sample dilution. One end of capillary needs to be etched with hydrofluoric (HF) acid to reduce the outer diameter. Use lighter fire to burn out the outer coating of the capillary, in the middle part that is about 1 cm away from the end, to create a 1-2 cm etching segment. Wipe out the coating residue with wet paper towel. Install the capillary into a 600 µL Eppendorf tube through a hole pierced at bottom of the tube so that the etching segment is in the Eppendorf tube and the pierced hole is clogged tightly by the capillary (as shown in **Figure 2**). Add 60-80 µL HF acid into the tube so the HF acid can cover part the etching segment. Leave the capillary in HF acid for 90 min. Carefully remove the HF acid in the tube and deposit the HF acid into 10 M NaOH buffer to neutralize the HF acid. Rinse the etched end with water. Cutoff the etching segment and leave a desired length of etched capillary (~5 mm) (*see* **Note 8**).

3.4 CE-MS

1. Install the etched tip of the capillary into the glass emitter of the interface. The glass emitter (orifice controlled at 20-40 µm) needs to be filled with sheath buffer. Control the

distance of the etched capillary tip to the emitter orifice less than 500 μm and the distance of the emitter orifice to the MS entrance around 2 mm.

2. Perform the CE separation with the autosampler. Set the ESI voltage 2 kV. Flush the capillary with 15 psi for 10 min with BGE. For 500 nL sample injection, immerse the distal end of the capillary into sample solution, apply 5 psi for 90 second.(*see* **Note 9**) For separation, immerse the distal end of the capillary back into BGE buffer and apply 30 kV voltage for 115 min. Flush the capillary with 15 psi for 10 min for capillary clean-up and equilibrate after separation is finished. The example electropherograms of the prepared *E. coli* sample are shown in **Figure 3**. *See* **Note 10**.

3. During the separation, MS instrument is operated in data-dependent mode. Acquire MS1 data with full scan range of 600-2000 m/z. Set the MS1 resolution 120,000 (at 200 m/z), set the AGC 1E6 and set maximum injection time 50 ms. Allow up to 3 precursor ions for HCD fragmentation in tandem mass spectra. Acquire MS2 spectra at 60000 resolution (at 200 m/z), AGC target value of 1E5 and max injection time of 200 ms. Set precursor ion isolation width to 4 m/z. Set the dynamic exclusion 30 s.

4. Data analysis: Perform database search with TopPIC Suite.[17] Convert RAW files into mzML file using msconvert tool.[36] Process the mzML file with TopFD software for spectral deconvolution. Process the msalign files resulted from spectral deconvolution with TopPIC software for database search. Specify the designated database, uniport *E. coli* in this case. Specify the maximum shift value, 1 in this case. Specify the approach for proteome FDR evaluation, target-decoy in this case. Set the PrSM-level FDR to 1% and proteoform-level to 5%. *See* **Note 11**.

4. Notes

1. For thorough cell lysis, sonication was set at power output of 6. To avoid accumulated heating during long session of sonication, sonication session can be divided into multiple sessions with break of 30 seconds.

2. Total protein amount loaded onto the ultrafiltration membrane should not exceed 100 μg to avoid protein precipitation. It is recommended to dilute the protein with urea first to disturb the interaction of protein and SDS and facilitate protein dissolution.

3. During ultrafiltration through centrifugation, try not to exceed 14,000 g for centrifugation speed as too high of speed might cause membrane clogging with protein molecules and protein loss. A spin-down on the protein solution to remove any precipitate is also recommended before ultrafiltration.

4. Make sure minimum solution left above the membrane after each cycle of wash.

5. An SDS-PAGE analysis can be performed to check the comprehensiveness of sample preparation (whether specific MW proteins are missing compared to pre-processed sample).

6. When preparing LPA coating solution, the degas step is crucial as oxygen will influent the polymerization process. A continuous bubbling from the degassing tube that is immersed in the coating solution should be maintained for an efficient degas. Avoid disturb the coating solution after the degas step and introduce the coat solution into the capillary with vacuum pump. The incubation time should be at least 50 min. Insufficient incubation could lead incomplete reaction and unstable coating. Incubating for too long (for example 1.5 hours) could lead to capillary clogging.

7. When the reaction time is too long, it is hard to push the polymer in the capillary out. HPLC pump can be used to flush the capillary in this case.

8. Hydrofluoric acid is very dangerous and needs to be processed with care in the hood. Follow the appropriate safety procedure. All the tubes and tips involved in the etching steps need to be placed in proper trash container.

9. The total loading amount of protein should be adjusted. The loading amount in Figure 3 was 400 ng. Too high of loading amount can cause protein precipitation during CE separation and odd current flow chart. Dilute the protein with 100 mM NH_4HCO_3 if the concentration is too high. A loading amount of 100 to 500 ng should be applicable.

10. It is proved that proteins extracted with SDS as additive tend to be more hydrophobic than extracted with urea as additive.[25] Hydrophobic protein is more insoluble in aqueous sample buffer such as NH_4HCO_3 and can precipitate out during CE

separation. To facilitate the dissolution, we applied 20% acetic acid as BGE as opposed to 5% that was regularly used in our previous studies.

11. 1% proteoform FDR can also be used for more strict setting. 800 proteoforms can be identified typically using the CZE-MS/MS system from the E.coli sample in a single run with 5% FDR at proteoform level.

Acknowledgements

We thank Prof. Heedeok Hong's group at the Department of Chemistry of Michigan State University for kindly providing the *E. coli* cells for this project. 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. Smith LM, Kelleher NL (2013) Consortium for Top Down, P., Proteoform: a single term describing protein complexity. Nature methods 10: 186-187.
2. Toby TK, Fornelli L, Kelleher NL (2016) Progress in Top-Down Proteomics and the Analysis of Proteoforms. Annual review of analytical chemistry 9: 499-519.
3. Smith LM, Kelleher NL (2018) Proteoforms as the next proteomics currency. Science 359: 1106-1107.
4. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL (2011) Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature 480: 254-258.
5. Catherman AD, Durbin KR, Ahlf DR, Early BP, Fellers RT, Tran JC, Thomas PM, Kelleher NL (2013) Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. Molecular & cellular proteomics MCP 12: 3465-3473.
6. Ahrne E, Martinez-Segura A, Syed AP, Vina-Vilaseca A, Gruber AJ, Marguerat S, Schmidt A (2015) Exploiting the multiplexing capabilities of tandem mass tags for high-throughput estimation of cellular protein abundances by mass spectrometry. Methods 85: 100-107.
7. Anderson LC, DeHart CJ, Kaiser NK, Fellers RT, Smith DF, Greer JB, LeDuc RD, Blakney GT, Thomas PM, Kelleher NL, Hendrickson CL (2017) Identification and Characterization of Human Proteoforms by Top-Down LC-21 Tesla FT-ICR Mass Spectrometry. Journal of proteome research 16: 1087-1096.
8. Cai W, Tucholski T, Chen B, Alpert AJ, McIlwain S, Kohmoto T, Jin S, Ge Y (2017) Top-Down Proteomics of Large Proteins up to 223 kDa Enabled by Serial Size Exclusion Chromatography Strategy. Analytical chemistry 89: 5467-5475.
9. Liang Y, Jin Y, Wu Z, Tucholski T, Brown KA, Zhang L, Zhang Y, Ge Y (2019) Bridged Hybrid Monolithic Column Coupled to High-Resolution Mass Spectrometry for Top-Down Proteomics. Analytical chemistry 91: 1743-1747.
10. McCool EN, Lubeckyj RA, Shen X, Chen D, Kou Q, Liu X, Sun L (2018) Deep Top-Down Proteomics Using Capillary Zone Electrophoresis-Tandem Mass

Spectrometry: Identification of 5700 Proteoforms from the Escherichia coli Proteome. Analytical chemistry 90: 5529-5533.
11. Yu D, Wang Z, Cupp-Sutton KA, Liu X, Wu S (2019) Deep Intact Proteoform Characterization in Human Cell Lysate Using High-pH and Low-pH Reversed-Phase Liquid Chromatography. Journal of the American Society for Mass Spectrometry 30: 2502-2513.
12. Ansong C, Wu S, Meng D, Liu X, Brewer HM, Deatherage Kaiser BL, Nakayasu ES, Cort JR, Pevzner P, Smith RD, Heffron F, Adkins JN, Pasa-Tolic L (2013) Top-down proteomics reveals a unique protein S-thiolation switch in Salmonella Typhimurium in response to infection-like conditions. Proceedings of the National Academy of Sciences of the United States of America 110: 10153-10158.
13. Lubeckyj RA, Basharat AR, Shen X, Liu X, Sun L (2019) Large-Scale Qualitative and Quantitative Top-Down Proteomics Using Capillary Zone Electrophoresis-Electrospray Ionization-Tandem Mass Spectrometry with Nanograms of Proteome Samples. Journal of the American Society for Mass Spectrometry 30: 1435-1445.
14. Liu Z, Wang R, Liu J, Sun R, Wang F (2019) Global Quantification of Intact Proteins via Chemical Isotope Labeling and Mass Spectrometry. Journal of proteome research 18: 2185-2194.
15. Riley NM, Westphall MS, Coon JJ (2017) Activated Ion-Electron Transfer Dissociation Enables Comprehensive Top-Down Protein Fragmentation. Journal of proteome research 16: 2653-2659.
16. Shaw JB, Li W, Holden DD, Zhang Y, Griep-Raming J, Fellers RT, Early BP, Thomas PM, Kelleher NL, Brodbelt JS (2013) Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. Journal of the American Chemical Society 135: 12646-12651.
17. Kou Q, Xun L, Liu X (2016) TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. Bioinformatics 32: 3495-3497.
18. Cai W, Guner H, Gregorich ZR, Chen AJ, Ayaz-Guner S, Peng Y, Valeja SG, Liu X, Ge Y (2016) MASH Suite Pro: A Comprehensive Software Tool for Top-Down Proteomics. Molecular & cellular proteomics : MCP 15: 703-714.

19. Sun RX, Luo L, Wu L, Wang RM, Zeng WF, Chi H, Liu C, He SM (2016) pTop 1.0: A High-Accuracy and High-Efficiency Search Engine for Intact Protein Identification. Analytical chemistry 88: 3082-3090.
20. Donnelly DP, Rawlins CM, DeHart CJ, Fornelli L, Schachner LF, Lin Z, Lippens JL, Aluri KC, Sarin R, Chen B, Lantz C, Jung W, Johnson KR, Koller A, Wolff JJ, Campuzano IDG, Auclair JR, Ivanov AR, Whitelegge JP, Pasa-Tolic L, Chamot-Rooke J, Danis PO, Smith LM, Tsybin YO, Loo JA, Ge Y, Kelleher NL, Agar JN (2019) Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. Nature methods 16: 587-594.
21. Wisniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. Nature methods 6: 359-362.
22. Wessel D, Flugge UI (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Analytical biochemistry 138: 141-143.
23. Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM, Krijgsveld J (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. Molecular systems biology 10: 757.
24. Hughes CS, Moggridge S, Muller T, Sorensen PH, Morin GB, Krijgsveld J (2019) Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nature protocols 14: 68-85.
25. Yang Z, Shen X, Chen D, Sun L (2020) Toward a Universal Sample Preparation Method for Denaturing Top-Down Proteomics of Complex Proteomes. Journal of proteome research 19: 3315-3325.
26. Kessner D, Chambers M, Burke R, Agus D, Mallick P (2008) ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24: 2534-2536.

Figure Captions

Figure 1. Workflow of the dTDP sample preparation using membrane ultrafiltration. Reproduced from ref. 25 with permission from American Chemical Society, copyright (2020)

Figure 2. Set-up of capillary etching

Figure 3. Electropherograms of dTDP CE-MS analysis on two replicates of *E. coli* proteome processed by membrane ultrafiltration. Reproduced from ref. 25 with permission from American Chemical Society, copyright (2020)

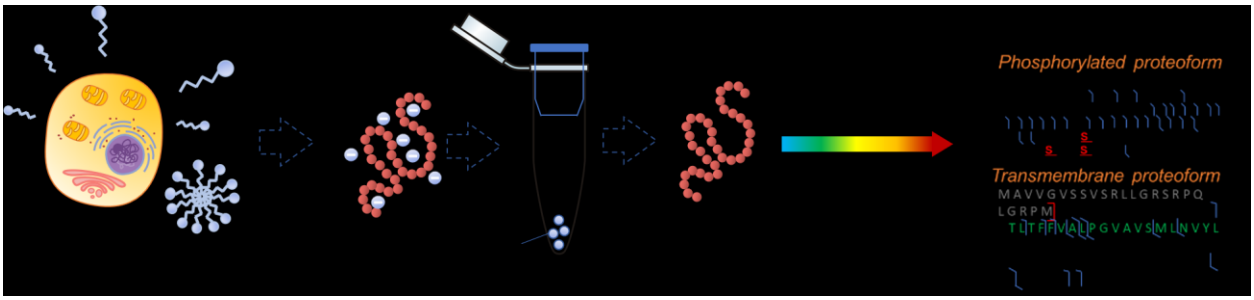
Phosphorylated proteoform
Transmembrane proteoform
MAVVGVSSVSRLLGRSRPQ
LGRPM

Figure 1

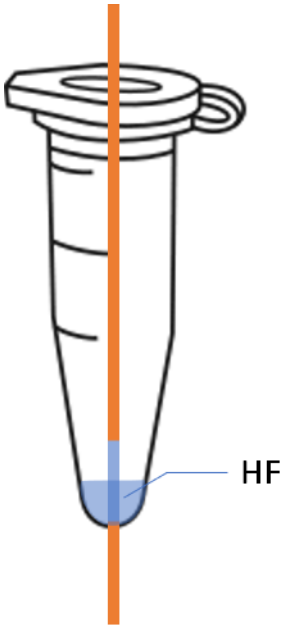

Figure 2

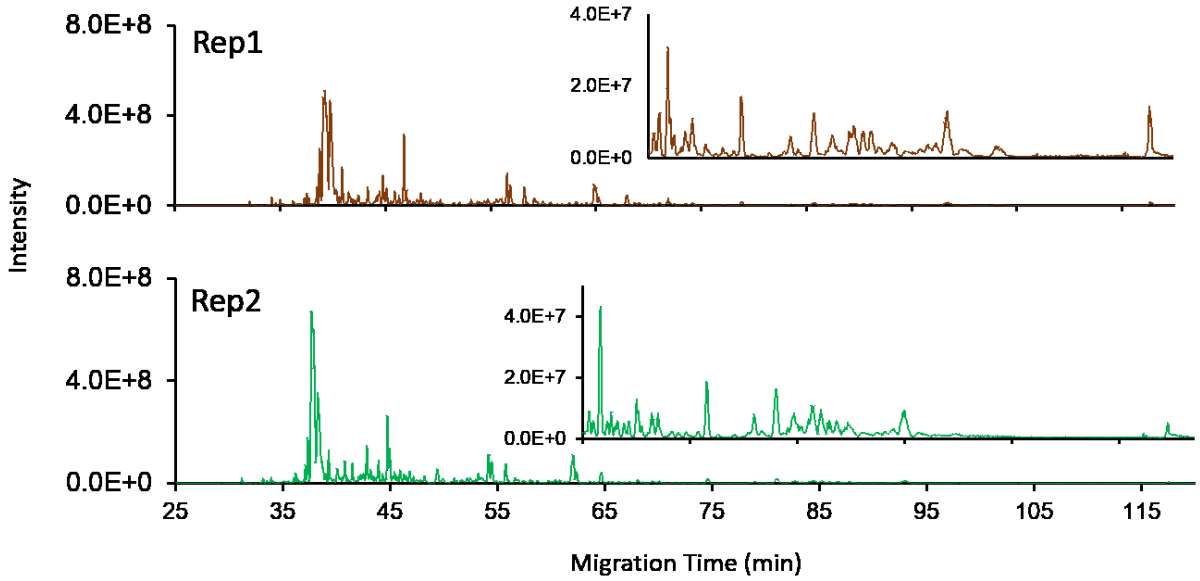
Rep1
Rep2
8.0E+8
4.0E+8
0.0E+0
4.0E+7
2.0E+7
0.0E+0
Intensity
25
35
45
55
65
75
85
95
105
115
Migration Time (min)

Figure 3