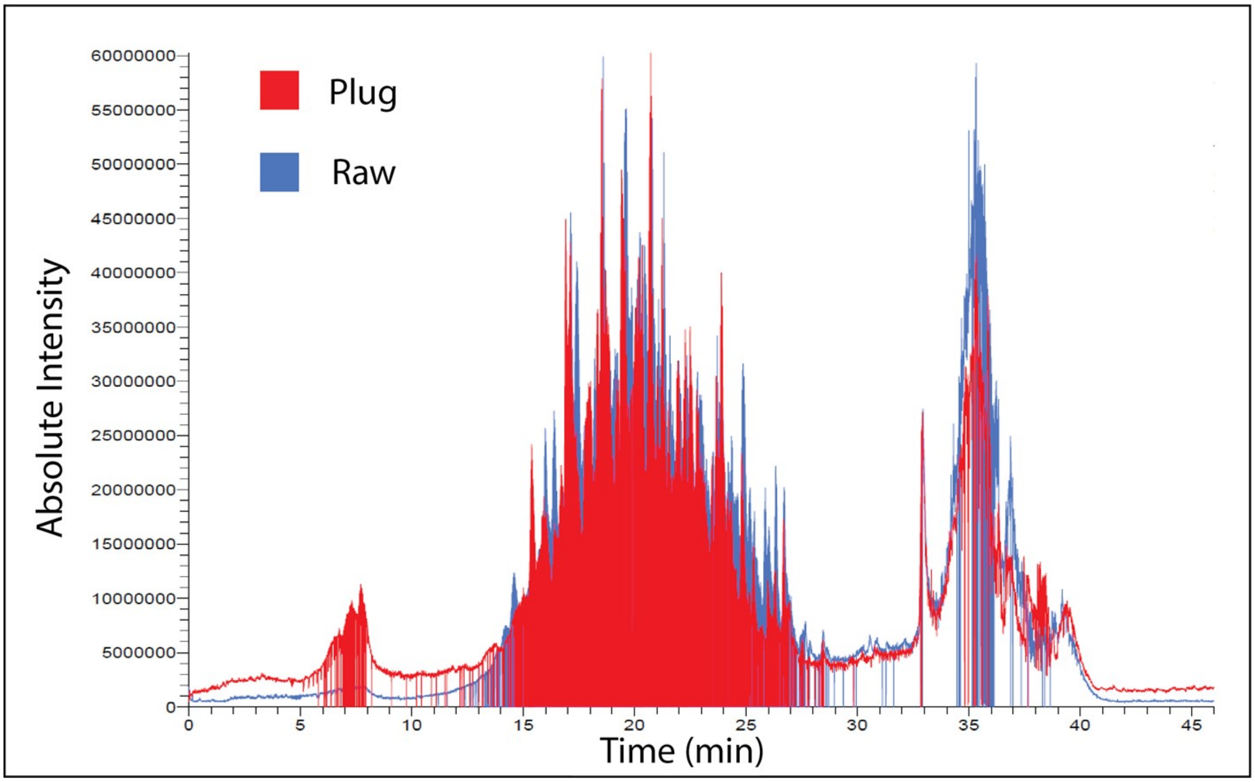
Supporting information for:

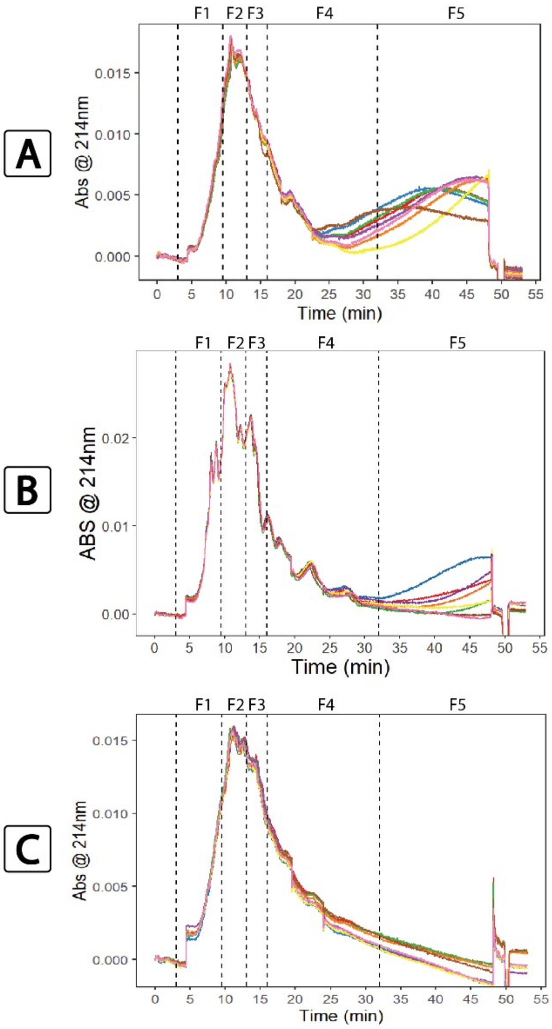
Preparative capillary electrophoresis (CE) fractionation of protein digests improves protein and peptide identification in bottom-up proteomics

Simon D. Weaver, Naviya Schuster-Little, and Rebecca J. Whelan

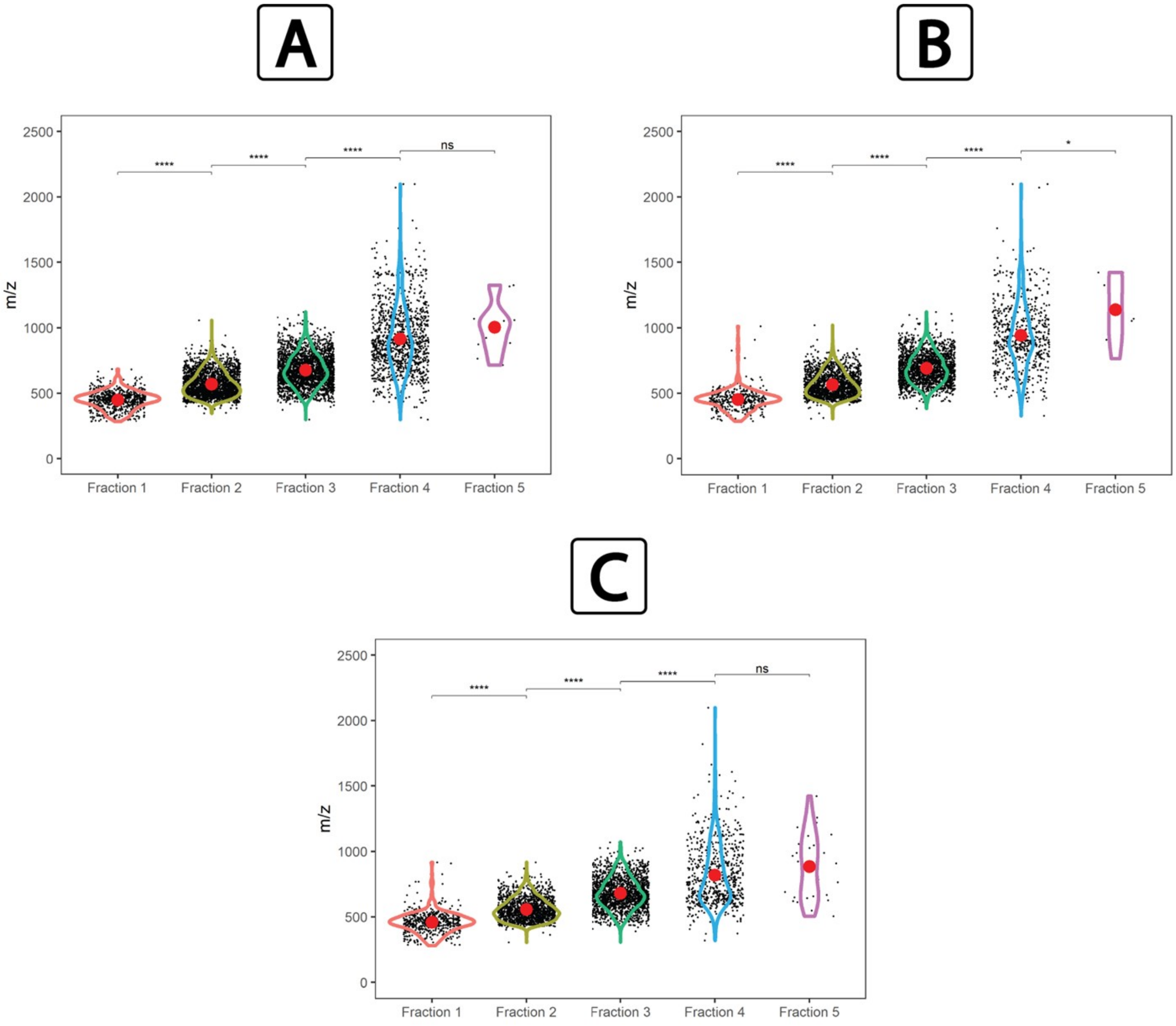
SI Figure 1: Comparison of total ion intensity (TIC) of raw (blue) and plug (red) controls from the Q-Exactive mass spectrometer. Representative injections shown.



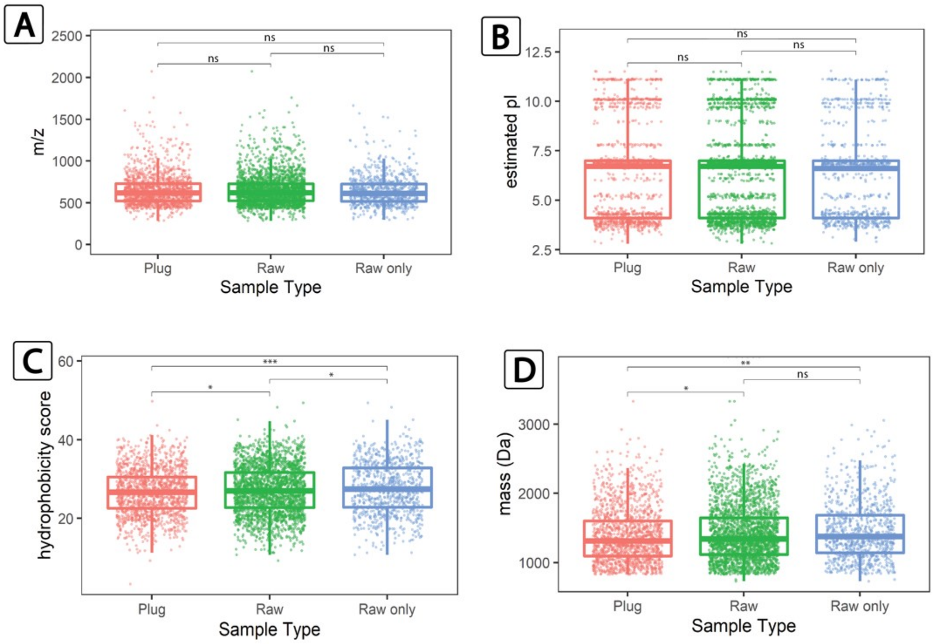
SI Figure 2: Representative overlaid electropherograms of fraction collection steps for (A) *E. coli* cell lysate, (B) Human Serum, and (C) Conditioned Cell Media (CCM) digests. Collected fractions are shown with dotted lines.



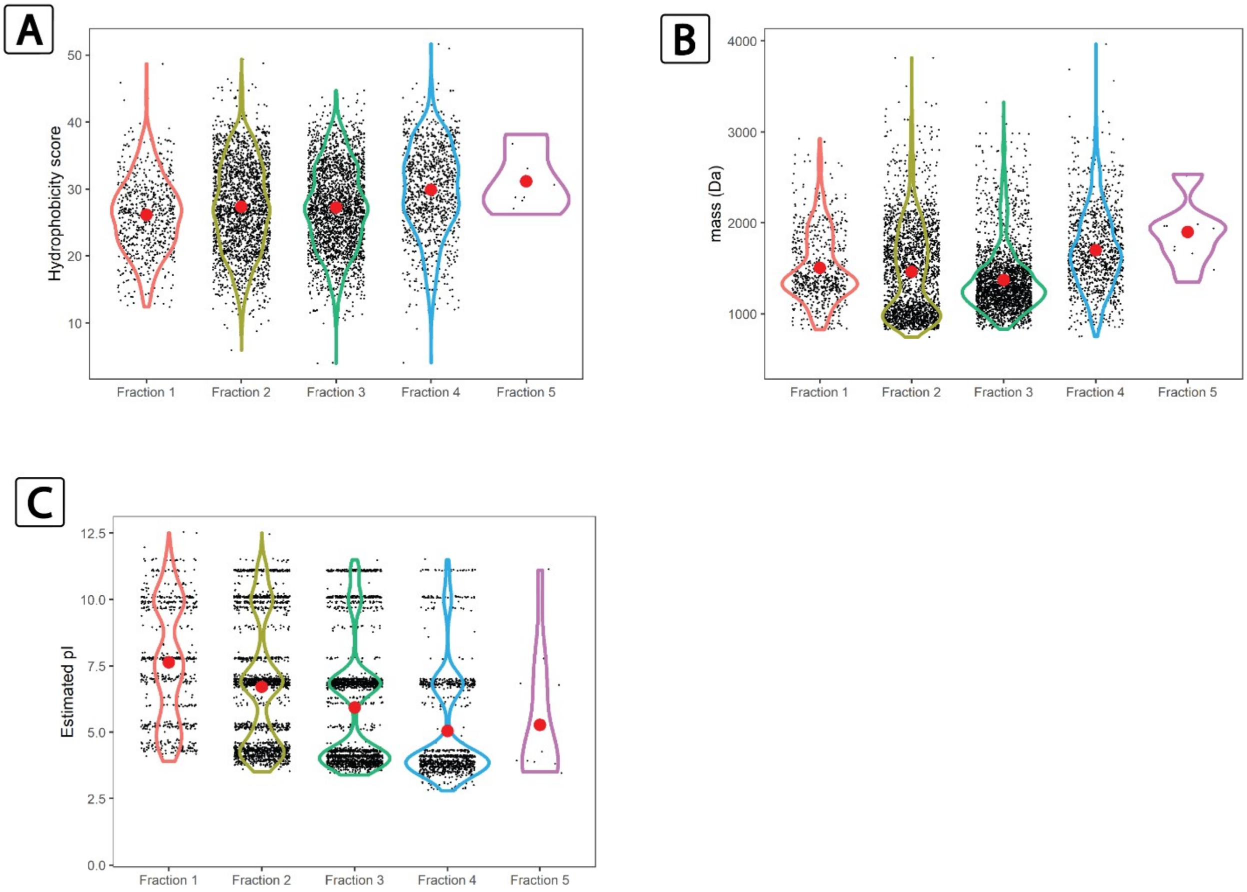
SI Figure 3: Identified peptides from *E. coli* cell lysates displayed as theoretical m/z at pH 2.38 vs. collected fraction number. (A, B, C) Results from three biological replicates. Note that (A) is identical to figure 2B in the main text and is included here for completeness. Red point is the mean value for each fraction. **** signifies a p-value of ≤0.0001 by the Wilcoxon rank-sum test. ns = not significant.



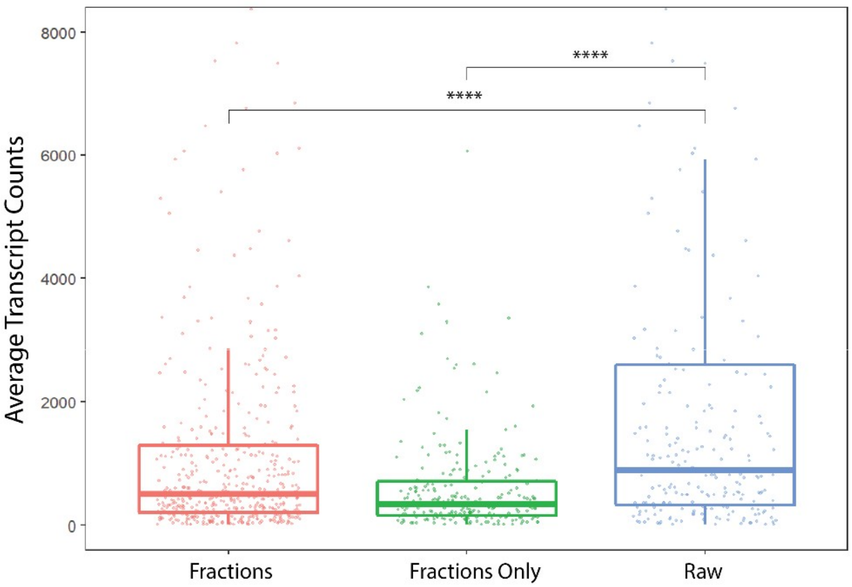
SI Figure 4: Identified peptides from *E. coli* cell lysates displayed as various peptide attributes comparing the raw control and plug control. ‘Raw only’ refers to peptides found uniquely in the raw control, and not in the plug control. Attributes analyzed are: (A) m/z, (B) isoelectric point (pI), (C) hydrophobicity, (D) mass. *** signifies p-value of ≤0.0005, ** signifies p-value of ≤0.005, * signifies p-value of ≤0.05 by the Wilcoxon rank-sum test. ns = not significant. Peptide attributes calculated by Thermo Fisher peptide analysis tool (https://www.thermofisher.com/us/en/home/life-science/protein-biology/peptides-proteins/custom-peptide-synthesis-services/peptide-analyzing-tool.html, accessed 1/24/2022).



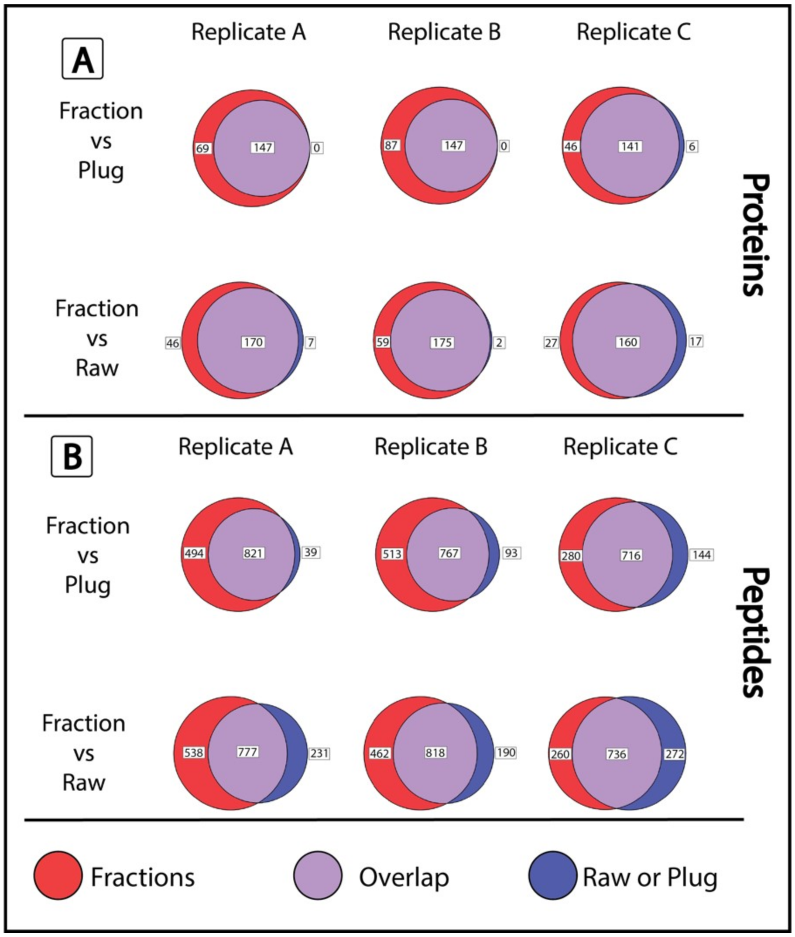
SI Figure 5: Identified peptides from *E. coli* cell lysates displayed as various peptide attributes vs collected fraction number from replicate A. (A) hydrophobicity score by fraction. (B) mass by fraction. (C) isoelectric point (pI) by fraction. Red point is the mean value for each fraction. Peptide attributes calculated by Thermo Fisher peptide analysis tool (https://www.thermofisher.com/us/en/home/life-science/protein-biology/peptides-proteins/custom-peptide-synthesis-services/peptide-analyzing-tool.html, accessed 1/24/2022).



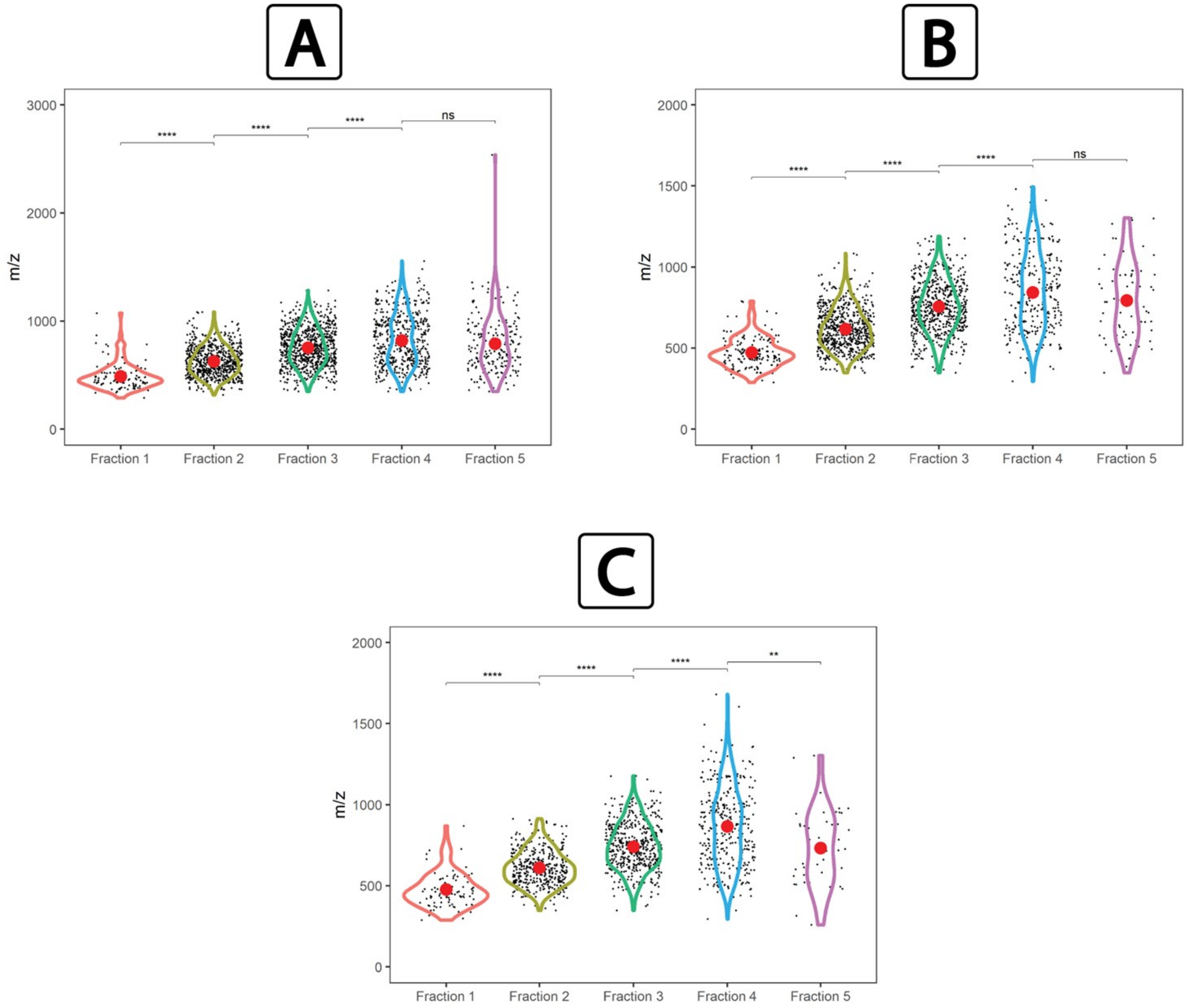
SI Figure 6: Expression analysis of identified proteins in *E. coli* cell lysate by sample type. Fractionated sample compared to raw control, with 'Fractions Only' referring to proteins identified uniquely in the fractionated sample and not in the raw control. **** signifies p-value of ≤0.00005 by Student's t-test.



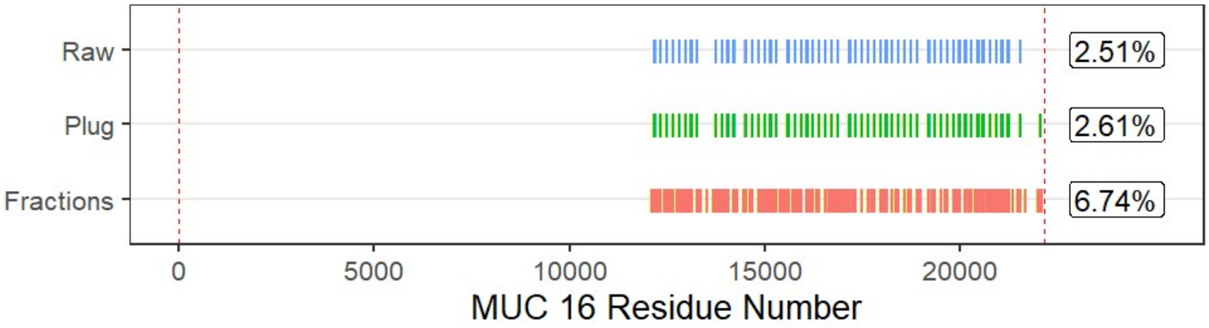
SI Figure 7: Protein and peptide identifications in CE fractionated human serum sample compared to CE plug and raw digest. (A) Overlap in protein identifications between fraction (red) and plug or raw (blue) for three biological replicates. (B) Overlap in peptide identifications between fraction (red) and plug or raw (blue) for three biological replicates. Files from the five fractions (15 injections total) were combined and searched as one sample in PEAKS, to enable direct comparison with the Raw and Plug controls, for which equivalent material by mass was analyzed.



SI Figure 8: Identified peptides from human serum lysates displayed as theoretical m/z at pH 2.38 vs. collected fraction number. (A, B, C) Results from three biological replicates. Red point is the mean value for each fraction. **** signifies a p-value of ≤0.0001 by the Wilcoxon rank-sum test. ns = not significant.



SI Figure 9: Conditioned cell media from OVCAR-3 cells shows the presence of ovarian cancer biomarker MUC16. N-terminus on the left and C-terminus on the right, designated by red dashed lines. Repeat region starts at amino acid 12,071. Identified peptides are shown with vertical colored lines. Percent coverage for each sample listed to the right of the C-terminus. Data shown from replicate B.



Supporting Files:

SI_Fractions_repA_peptides_and_proteins_Ecoli.csv
Database search output for the CE fractionated *E. coli* lysate sample, replicate A.

SI_Fractions_repB_peptides_and_proteins_Ecoli.csv
Database search output for the CE fractionated *E. coli* lysate sample, replicate B.

SI_Fractions_repC_peptides_and_proteins_Ecoli.csv
Database search output for the CE fractionated *E. coli* lysate sample, replicate C.

SI_Plug_control_peptides_and_proteins_Ecoli.csv
Database search output for the CE plug *E. coli* lysate sample.

SI_Raw_control_peptides_and_proteins_Ecoli.csv
Database search output for the raw *E. coli* lysate sample.

SI_Fractions_repA_peptides_and_proteins_HS.csv
Database search output for the CE fractionated human serum sample, replicate A.

SI_Fractions_repB_peptides_and_proteins_HS.csv
Database search output for the CE fractionated human serum sample, replicate B.

SI_Fractions_repC_peptides_and_proteins_HS.csv
Database search output for the CE fractionated human serum sample, replicate C.

SI_Plug_control_peptides_and_proteins_HS.csv
Database search output for the CE plug human serum sample.

SI_Raw_control_peptides_and_proteins_HS.csv
Database search output for the raw human serum sample.

SI_CCM_peptides.csv
Summary of MUC16 specific peptides found in each sample type in the conditioned cell media (CCM) replicates.