SUPPORTING INFORMATION

Elucidating *Escherichia coli* Proteoform Families Using Intact-Mass Proteomics and a Global PTM Discovery Database

Yunxiang Dai[1], Michael R. Shortreed[1], Mark Scalf[1], Brian L. Frey[1], Anthony J. Cesnik[1], Stefan Solntsev[1], Leah V. Schaffer[1], Lloyd M. Smith*[1,2]

[1]Department of Chemistry, University of Wisconsin, Madison, Wisconsin, United States
[2]Genome Center of Wisconsin, University of Wisconsin, Madison, Wisconsin, United States

*Corresponding Author

Table of Contents:

SUPPLEMENTARY EXPERIMENTAL METHODS.

Materials.

- *Escherichia coli* Strain KL334 was obtained from Coli Genetic Stock Center (CGSC #4345) (Yale University, New Haven, CT).
- Sodium butyrate (B5887), urea (U5378), deoxycholic acid (D2510), DL-dithiothreitol solution (DTT) (646563), iodoacetamide (I6125), trifluoroacetic acid (TFA) (T6508), ethyl acetate (270989), acetone (270725), and TWEEN 20 (P7949) were purchased from Sigma-Aldrich Co. (St. Louis, MO).
- 20% sodium dodecyl sulfate (SDS) (161-0418) was purchased from Bio-Rad (Hercules, CA).
- 10X MOPS modified rich buffer (M2101), 10X ACGU solution (M2103), 0.132M dipotassium phosphate solution (M2102), 20% glucose solution (G0520), 5X supplement EZ (M2104), 5X supplement EZ minus lysine (M2118), 4M Tris-HCL pH=7.5 (T5575), phosphate buffered saline (PBS) (P0191), and 0.2 M ammonium bicarbonate buffer (A2012) were purchased from Teknova (Hollister, CA).
- L-lysine:2HCl $^{13}C_6$, $^{15}N_2$ (CNLM-291-H), and L-lysine:2HCl 3,3,4,4,5,5,6,6 D_8 (DLM-2641) were purchased from Cambridge Isotope Laboratory, Inc (Tewksbury, MA).
- 100X HALT™ protease inhibitor cocktail (78430), SeeBlue® plus2 pre-stained protein standard, and formic acid (M1116701000) were purchased from Thermo Fisher Scientific (Waltham, MA).
- Gelfree™ 8100 12% Tris-acetate cartridge (42404), Tris-acetate 5X sample buffer (42302), and HEPES running buffer (42202) were purchased from Expedeon (San Diego, CA).
- NuPAGE® 20X MES SDS running buffer (NP0002), Antioxidant (NP0005), 4X LDS sample buffer (NP0007), and 4-12% Bis-Tris Gel (NP0323) were purchased from Life Technologies (Carlsbad, CA).
- Trypsin (V5111) was purchased from Promega (Madison, WI).
- Acetonitrile (ACN) (AH015-4) was purchased from Honeywell (Morris Plains, NJ).

Cell Lysis.

Cell pellet was resuspended in 1 mL lysis buffer (4% SDS, 100 mM Tris pH=7.5, 10mM DTT, 10 mM sodium butyrate, and 1× HALT™ protease inhibitor) and lysed with heat incubation at 95˚C for 10 min, with vortexing every 2 min. Cell debris was pelleted, and the supernatant was reduced in 10 mM DTT for 30 min. The sample was then alkylated in 20 mM iodoacetamide for 30 min and quenched in 20 mM DTT for 15 min. Proteins were acetone precipitated at -80 ˚C. Protein pellet was collected from centrifugation and air-dried. Lysate proteins were dissolved in 200 µL 1% SDS, and the concentration was measured with BCA (bicinchoninic acid) assay.

Gelfree™ Fractionation.

Based on the protein concentration measured with the BCA assay, sample volume corresponding to 1 mg of total protein was calculated. The volume should not exceed 224 µL. This volume of sample was added in a new 1.7 mL tube. A volume of 60 µL sample buffer, 16 µL 1 M DTT, and water were added into the tube to make the total volume of 300 µL. The tube was incubated at 50 ˚C for 10 min in a water bath and was cooled to room temperature. It was split into 2 halves, each containing 500 µg protein for one Gelfree™ cartridge channel. The 12% cartridges were used in this study. Storage buffer was removed from the channels and was replaced with running buffer. Each of the 150 µL sample (500 µg protein) were loaded into each loading chamber of the 2 channels. A standard running method to separate the sample into 12 fractions based on molecular weight was selected. The running current was ensured to be above 3 mA. Between each step, fractions in the collection chambers from the two channels were combined into a new 2 mL low-retention tube. The collection chambers were rinsed and replenished with new running buffer. These steps were repeated 12 times for each fraction collection. Throughout the run, the running buffer was changed twice according to the standard procedure. The collected fractions were stored at -20 ˚C for subsequent sample preparation. PAGE gel analysis was used to validate separation quality and molecular weight distribution on aliquots of the fractions (SI Figure S-4).

eFASP Procedure.

For samples to be analyzed by bottom-up proteomics, eFASP (*J. Proteome Res.* 2014, 13, 1885-1895) was performed to digest and remove detergent as follows. Filter units and eFASP collection tubes were passivated by soaking in 5% Tween 20 overnight. Filter units and collection tubes were rinsed at least three times using nanopure water. About 40 µL aliquots of fractions 3 to 12 were diluted with in 860 µL eFASP exchange buffer (8 M Urea, 0.10% deoxycholic acid). A volume of 450 µL of each sample was each transferred to a passivated filter unit, which was then inserted into an unpassivated tube, centrifuged at 14 000 g at 15 ˚C for 10 min. The flowthrough was discarded and the rest of the sample was added into its respective filter unit and the centrifugation was repeated. A volume of 200 µL exchange buffer was added in each filter unit, and the samples were centrifuged at 14 000 g at 15 ˚C for 10 min. The flowthrough was discarded. This washing step was repeated three times. Then 200 µL eFASP reducing buffer (8 M Urea, 20 mM DTT) was added to each filter unit, and the samples were incubated at room temperature for 30 min. The samples were then centrifuged at 14,000 g at 15 ˚C for 10 min. Flowthrough was discarded again. 200 µL eFASP alkylation buffer (8 M Urea, 50 mM iodoacetamide, 50mM ammonium bicarbonate) was added in each filter unit, and the samples were incubated at room temperature in the dark for 30 min. 15 µL DTT was added in each sample and incubated for 10 min. The samples were centrifuged again at 14 000 g at 15 ˚C for 10 min, and the flowthrough was discarded. 200 µL eFASP digestion buffer (1 M Urea, 50mM ammonium bicarbonate, 0.1% DCA) was added in each filter unit, and the samples were centrifuged at 14 000 g at 15 ˚C for 10 min. The flowthrough was discarded. This washing step was repeated three times. The filters were transferred to passivated collection tubes. A volume of 100 µL digestion buffer and 1 µg trypsin was added into each filter unit. Tubes and filter units were wrapped in Parafilm®, and incubated at 37 ˚C overnight with no rotation. Following digestion, Parafilm® was removed and the tubes were centrifuged at 14 000 g at 15 ˚C for 10 min. Then, 50 µL of 50 mM ammonium bicarbonate was added in each filter unit and centrifuged at 14 000 g at 15 ˚C for 10 min. This step was repeated twice. The flowthrough was transferred to new 1.7 mL low-retention tubes. A volume of 200 µL ethyl acetate and 200 µL 1% TFA was added to the samples, which then were shaken for 1 min. Samples were centrifuged at 15 800 g at 15 ˚C for 2 min. The top layer was removed from each tube, and another 200 µL ethyl acetate was added, and shaken for 1 min. The samples were centrifuged again at 15 800 g at 15 ˚C for 2 min, and the top layer was removed. Samples were dried in a Savant SpeedVac™ Concentrator for about 150 min. Dry tube contents were dissolved in 180 µL 0.1% TFA and vortexed to mix.

C18 Solid-phase Extraction.

Extraction pipette tips (Agilent Technologies, A57003100K, OMIX C18 tips) were activated by pipetting 180 µL 70% ACN up and down three times, and were washed three times by pipetting and discarding 180 µL of 0.1% TFA. Each sample (digested fractions 3 to 12) was pipetted up and down 3 times using the washed tip. Extraction tips were than washed with 0.1% TFA as above. Peptides were eluted from the tip by repeated (5 ×) aspiration of 150 µL of 70% ACN with 0.1% TFA in a 600 µL low retention tube. Eluted samples were evaporated to dryness in the SpeedVac™ for about 50 min. Tube contents were reconstituted in 200 µL of 95:5 H_2O:ACN with 0.1% formic acid.

Orbitrap Velos Mass Spectrometer Methods.

a. Intact-mass.

Pellets from methanol-chloroform precipitation were redissolved in 32 µL 95:5 H_2O:ACN solution with 0.1% formic acid. Intact protein solutions were gently vortexed and centrifuged on a bench-top centrifuge for 1 min. Solutions were carefully transferred into HPLC sample vials, leaving behind undissolved substances. Samples were analyzed by HPLC-ESI MS (nanoAcquity, Waters, and LTQ Velos, Thermo Fisher Scientific). Two technical replicates were performed for each sample fraction. Each technical replicate consisted of about 6 μL of each sample. The actual volume for each sample's injection was decided by targeting a total ion chromatogram (TIC) value of ~10^9. Generally, smaller volumes of higher fractions were injected, as they had a higher concentration of protein. HPLC separation employed a 100 × 365 μm fused silica capillary microcolumn packed with 20 cm of polymeric reverse phase resin (PLRP-S, 5 μm, 1000 Å, Phenomenex), with an emitter tip pulled to approximately 1 μm using a laser puller (Sutter instruments). The mobile phase contained penta-asparagine synthetic peptide (Asn5) as a lock mass standard at a concentration of 1 μg/mL. The Asn5 synthesized peptide was reconstituted in 95:5 H2O:ACN, 0.2% formic acid, and desalted and purified from the trityl-containing product by loading over a C18 solid phase purification cartridge (Sep-Pak 50mg, Waters) and taking the flow-through. Asn5 produced an $(M+H)^+$ peak at 589.2325 Da in every spectrum, which was used as a lock mass standard. Intact proteins were loaded on-column at a flow rate of 500 nL/min for 30 min, and then eluted over 67 min at the same flow rate with a gradient of 5% to 85% ACN, in 0.1% formic acid. Full-mass scans were performed in the FT orbitrap between 550 and 1600 m/z at 100 000 resolution. A total of 5 μscans were averaged, using an AGC target setting of 10^6, with a maximum fill time of 500 milliseconds. Source fragmentation (SF) was set to 30.

b. Bottom-up.

10 μL of each eFASP sample, after C18 solid-phase extraction, was transferred into HPLC sample vials. They were analyzed by HPLC-ESI MS/MS (nanoAcquity, Waters, and LTQ Velos, Thermo Fisher Scientific). Two technical replicates were performed for each sample fraction. Each technical replicate consisted of 5 μL of each sample. HPLC separation used a 100 × 365 μm fused silica capillary microcolumn packed with 20 cm of 1.7 μm-diameter/130 Å pore size C18 beads (Waters BEH), with an emitter tip pulled to approximately 1 μm using a laser puller (Sutter instruments). Peptides were loaded on-column at a flow rate of 400 nL/min for 30 min and then eluted over 125 min at a flow rate of 300 nL/min with a gradient of 2% to 30% ACN in 0.1% formic acid. Full-mass scans were performed in the FT orbitrap between 300 and 1500 m/z at a resolution of 60 000, followed by 10 MS/MS HCD (higher-energy collisional dissociation) scans of the 10 most intense parent ions at 42% relative collision energy and 7 500 resolution, with a mass range starting at 100 m/z. Dynamic exclusion was enabled with a repeat count of two over the duration of 30 seconds and an exclusion window of 120 seconds.

Spectral Deconvolution.

Intact-mass raw files were deconvoluted into monoisotopic components using Protein Deconvolution 4.0 software (Thermo Scientific). Minimum number of detected charge was set to 2. Fit factor and remainder threshold were set at 70% and 10% respectively. Spectra from the retention time (RT) range of 25 to 105 min were deconvoluted. Sliding window was used with 0.4 min target average spectrum width and an offset value of 34%. Different charge ranges were adjusted for deconvoluting different fractions: +5 to +30 for fractions 3–6; +5 to +40 for fractions 7–9; and +5 to +50 for the highest fractions 10–12. The output spreadsheets were loaded into Proteoform Suite for subsequent proteoform analysis.

Proteoform Suite.

Deconvoluted files (in .xlsx file format) and calibration files (in .tsv file format) were loaded into Proteoform Suite (v0.1.12 available at https://smith-chem-wisc.github.io/ProteoformSuite ; Cesnik, A.J. et al. "Proteoform Suite: software for constructing, quantifying and visualizing proteoform families" manuscript submitted for review). Raw experimental components were extracted from the loaded files. NeuCode pairs were identified. Lysine count and NeuCode pair intensity ratios were plotted in histograms (SI Figure S-1). Acceptable intensity ratio was set to 1.40 minimum and 3.00 maximum according to the ratio histogram, to include variance around the expected mixing ratio of 2:1 (“light”:“heavy”). NeuCode pairs were aggregated to eliminate redundancy by allowing 10 ppm mass tolerance, and 5 min retention time tolerance. A maximum of 3 missed monoisotopic masses and 2 missed lysine counts were allowed. A catalog of theoretical proteoforms was generated using either the *E. coli* database downloaded from UniProt, or constructed from the G-PTM-D strategy (see MetaMorpheus paragraph below). Cleaved N-methionine option was unchecked, allowing full sequence of proteoforms in the catalog. Carbamidomethylation (CAM) of cysteine was included as a fixed modification. A maximum of 3 PTMs (annotated in UniProt and oxidation at methionine residues) per proteoform were allowed, and the minimum peptide length was set to 7. Five decoy databases were generated with these setting to test the ET false discovery rate (FDR) In the process of making ET pairs, the ET peak width was set to 0.05 Da. “No man’s land” was set to between 0.19 Da and 0.94 Da; fractional masses in this range are unlikely to occur naturally from a PTM in the mass range we considered, and any ET pairs with fractional masses in this range were excluded from the final list. The grouped ET peaks with a difference corresponding to an exact match (0 Da mass difference) or any of a defined set of modifications were selected. A ±0.03 Da error between the peak mass difference and theoretical modification set mass difference was allowed when selecting the peaks, based on the observed ET peak position for 0 Da (0.0246 Da in the UniProt analysis; 0.0237 Da in the G-PTM-D analysis). This error parameter was not built into the Proteoform Suite software, but applied manually by comparing a peak’s theoretical mass difference to its actual mass difference. It also marked the mass accuracy of identified experimental proteoforms. The average and median FDR for ET peaks was determined from comparison of experimental proteoforms with each of the five decoy databases. The FDR for each peak was equal to the ratio of decoy pairs within ±0.025 Da of the peak’s mass difference, to the number of ET pairs in the peak. Accepted mass difference peaks (SI Table 5) were determined based on low FDR (average ≤ 20%; median ≤ 13%), and a mass difference within 0.03 Da to common modification sets. The same parameters were used in making EE pairs. Accepted EE peaks (SI Table 5) had FDRs lower than 17% and a mass difference corresponding to a set of common modifications. . A weighted average FDR of the selected peaks were calculated for each ET/EE comparison to show an overall FDR value (∑[average FDR of the peak]·[number of pairs in the peak]/[total number of pairs]) (SI Table S-5). The resulting FDR values were all lower than 3% in this study. Proteoform families were assembled based on the ET and EE pairs of the accepted peak values. A Cytoscape script file (.txt file format), style file (.xml file format), edge and node tables (.tsv file format) were exported to build the proteoform families in Cytoscape. The total computational time in Proteoform Suite is 35 min on a 2-core computer with 8 GB RAM.

Cytoscape.

In Cytoscape (v.3.4.0), the script file exported from Proteoform Suite was loaded under “Tools-Execute Command File”. Plotted networks were subject to circle layout arrangement by name under “Layout”. Scale was adjusted to improve legibility. Additional graphic alterations were added for the families of interest (Figure 4 and Figure 5 in manuscript).

MetaMorpheus.

The reference *E. coli* protein database (in .xml format) was downloaded from UniProt and loaded into MetaMorpheus (v0.0.80 available at https://github.com/smith-chem-wisc/MetaMorpheus/releases). Uncalibrated bottom-up spectra files (.raw) were loaded in MetaMorpheus. Three tasks were added sequentially to carry out the G-PTM-D strategy. First, a “New Calibrate Task” was added, which included an initial search of uncalibrated data using default settings of search parameters: precursor mass tolerance at 10 ppm, maximum missed trypsin cleavages at 2, product mass tolerance at 0.01 Da, and b and y ions searched. Common fixed modification (carbamidomethylation of cysteine) and common variable modification (oxidation of methionine) were selected respectively in the fixed and variable modification columns. The results of this database search were employed by MetaMorpheus to calibrate the spectral data. Theoretical m/z (mass divided by charge) spectral peaks of peptides identified with under 1% FDR (false discovery rate) were compared with experimentally measured m/z values to calculate errors. The errors were plotted against measured m/z and other variables, and were analyzed by linear regression to carry out an initial global calibration. Then a random forest regression method was employed to generate a calibration curve attuned to local patterns. Measured m/z values were adjusted through these models, and calibrated spectra file were written in the mzML format. Second, a "New G-PTM-D Task" was added with search parameters and modification settings similar to those in the calibration task, except the precursor mass tolerance was set at 2 ppm. Under the G-PTM-D modification column, PTM categories of interest were selected ("Mod", "fatty acid", and "metal" in this study) for search and were added to expand the UniProt database. Finally, a "New Search Task" was added with default settings: classic search mode, 2 maximum missed trypsin cleavages, precursor mass difference at 10 ppm around zero, and product mass tolerance at 0.01 Da. Protein aggregation and histogram analysis were allowed for the convenience of search result interpretation. MetaMorpheus would use the newly constructed G-PTM-D database for this search task when this search was put after the G-PTM-D task. The total computational time in MetaMorpheus is 4 hr 50 min on a 24-core computer with 128 GB RAM (calibration task: 4 hr; G-PTM-D database building: 20 min; final search: 30 min).

IHF Subunit Beta Proteoform Family Oxidation Localization in Figure 4A.

To ensure the PTM annotations on the two modified theoretical proteoforms in Figure 4A (right) are in fact oxidation of phenylalanine and tyrosine, but not the common variable oxidation of methionine, we examined the bottom-up proteomics data searched with the G-PTM-D database. We found peptides (residue 43-56 and 47-56) containing oxidized phenylalanine (Phe48, Phe51) and tyrosine (Tyr55). These two peptides do not contain M in the sequence. Moreover, in Proteoform Suite software, all oxidized methionines were included in generating the theoretical proteoforms (see above), therefore, the annotated oxidations of the theoretical proteoforms in the IHF subunit beta family are valid. Although technically intact-mass proteoform analysis using Proteoform Suite does not localize PTMs, one can infer PTM locations from bottom-up data.

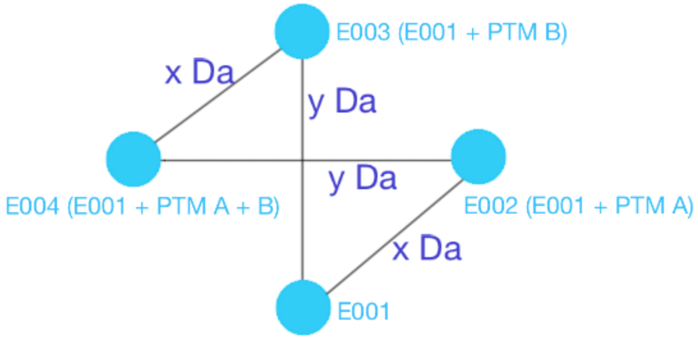
L7/L12 Proteoform Family Alteration in Figure 5.

The following nodes (experimental proteoforms) were eliminated from the Proteoform Suite output (SI Table S-6, S-10) to manually correct analysis errors and better present this family example: E1852 (Figure 5A) was not included in the G-PTM-D family (Figure 5B) because it did not match any entries in the G-PTM-D catalog with the selected PTMs; E2425 in the G-PTM-D family (Figure 5B) and E2499 in both families (Figure 5A and 5B) were not included. These two low abundance (only one observation) experimental proteoforms did not merge with other proteoforms (E2256 and E2145 respectively) during the aggregation step, due to retention time differences larger than 5 min. The following edges (mass differences representing a modification or combination of modifications) were eliminated in the G-PTM-D family because proteoforms connected by these edges were found to have better PTM assignments: unmodified—E300 (58 Da, previously selected as oxidation + acetylation), E15—E300 (58 Da, previously selected as oxidation + acetylation), E70—E300 (44 Da, previously selected as carboxylation), E134—E736 (44 Da, previously selected as carboxylation), E515—E1276 (44 Da, previously selected as carboxylation), E70—E154 (294 Da, previously selected as CHDH fatty acid), E134—E759 (294 Da, previously selected as CHDH fatty acid), and E516—E759 (58 Da, previously selected as oxidation + acetylation). The identity of experimental proteoform E806 remained unclear. It is deduced to be a proteoform of missed monoisotopic mass (+ 1 Da artifact), with trimethylation and loss of ammonia at the N-terminal serine (+ 28 Da).

Butterfly Network.

This term refers to a “butterfly-” or “bowtie-shaped” connection of 4 experimental proteoforms with 4 edges, typically within a larger proteoform family visualized using Cytoscape. This pattern of network, when observed in an assembled proteoform family, validates the PTM annotations of the 4 proteoforms.

Here is how one looks:



The 4 proteoforms in a “butterfly network”, for example, E001, E002, E003, E004 (from the smallest molecular weight to the largest), have 4 edges connecting each other (E001-E002, E001-E003, E002-E004, and E003-E004), with mass difference values reflecting a selected PTM. In these 4 edges, E001-E002 and E003-E004 are equal to x Da (PTM A). E001-E003 and E004-E004 are equal to y Da (PTM B). If the lightest proteoform (E001) is identified, then E002 can be assigned as E001 with an additional PTM A, and E003 can be assigned as E001 with an additional PTM B. Finally, the heaviest proteoform (E004) can be assigned as E001 with both PTM A and B, and this was confirmed by two paths (E001—x Da—E002—y Da—E004, or E001—y Da—E003—x Da—E004).

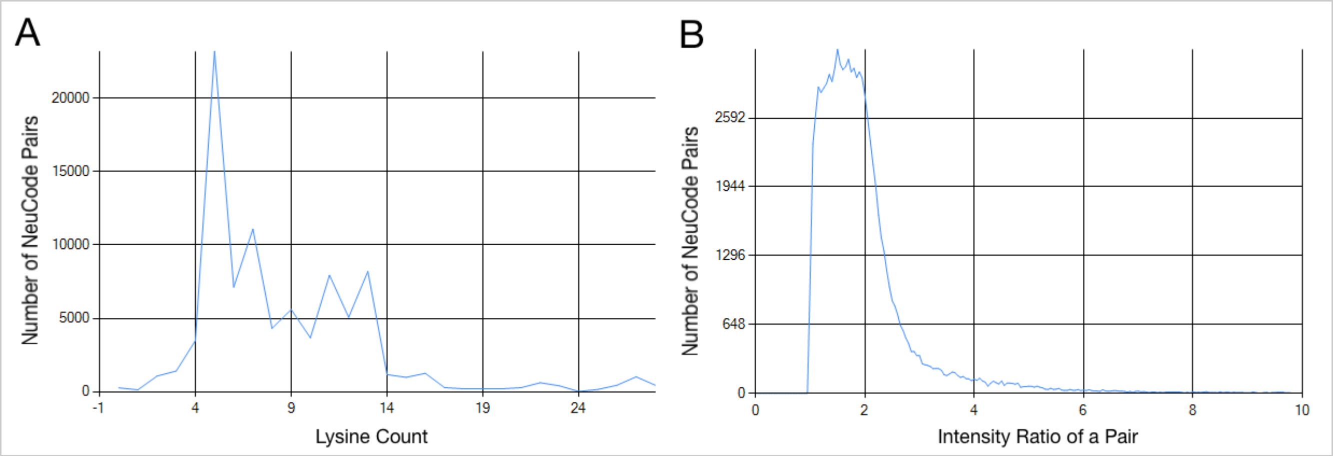
Alternative Parameter Settings in Data Analysis.

We explored the use of different values of the Proteoform Suite parameter settings and their effects upon the number of proteoforms identified and the associated FDRs. Changing the NeuCode pair ratio range from [1.4:1 to 3.0:1] (the range employed for the results presented in this work) to [1.2:1 to 6.0:1], and using a stricter aggregating mass tolerance (±2 ppm instead of ±5 ppm), led to the identification of many additional experimental proteoforms, including counterparts of the LSA-adducted species in the L7/L12 family (Figure 5B). The retention times of these new proteoforms from the L7/L12 family were no more than 5 min different from those of their related proteoforms, indicating they belong to this proteoform family. However, these parameter values doubled the number of proteoforms selected for further analysis (now ~4400), which would substantially increase the number of redundant PTM connections during the final family assembly stage. When choosing parameters for any type of data analysis, there is often a trade-off between reporting additional identifications and having a smaller set of clear and more confident results. In this case, we chose to utilize the original values of the analysis parameters. This trade-off involved in setting the value of the program parameters is a common issue in the analysis of complex data sets. An N-terminal methionine-cleaved catalog was also evaluated for analysis of the aggregated proteoforms. Despite yielding a smaller search space, the resultant FDRs of the selected peaks when using this catalog were found to be larger than those from the methionine-retained catalog. Because a considerable portion of the *E. coli* proteome retains N-terminal methionine, the methionine-retained catalog made more matches with our *E. coli* data. Moreover, the search time was the same for the analyses using either catalog. The steps before making the catalog (extracting raw components and aggregation) take the majority of the analysis time. ET and EE comparison steps usually complete within 1 min.

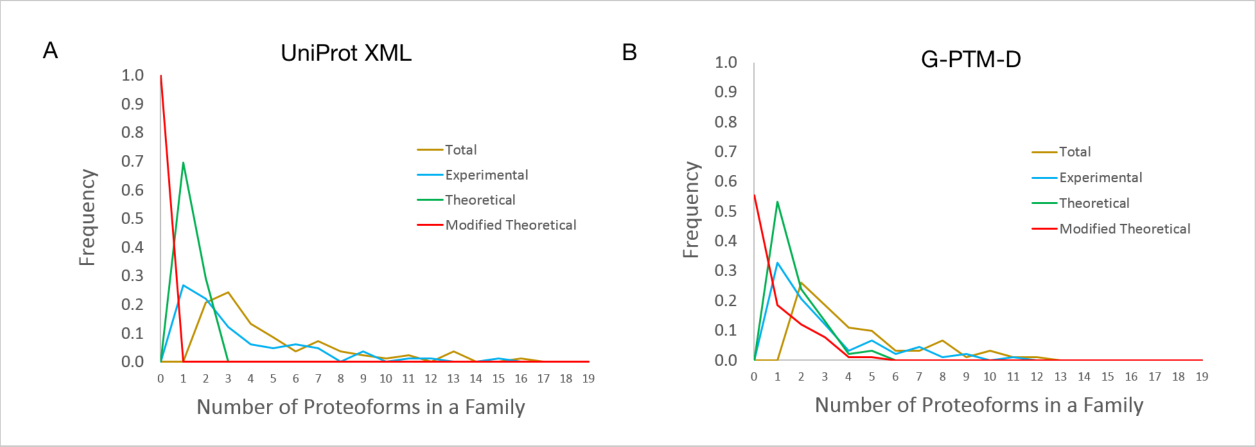
RAW DATA FILES.

All raw data files are available on the MassIVE platform (MSV000081144, ftp://massive.ucsd.edu/MSV000081144). There are 58 files of mass spectra (in .raw format) from intact-mass proteomics (10 Gelfree™ fractions of 3 biological replicates with 2 technical replicates each, but with fraction 7 of the 1st biological replicate missing due to an error). There are 20 files of mass spectra from bottom-up proteomics (10 fractions of one biological replicate with 2 technical replicates each).

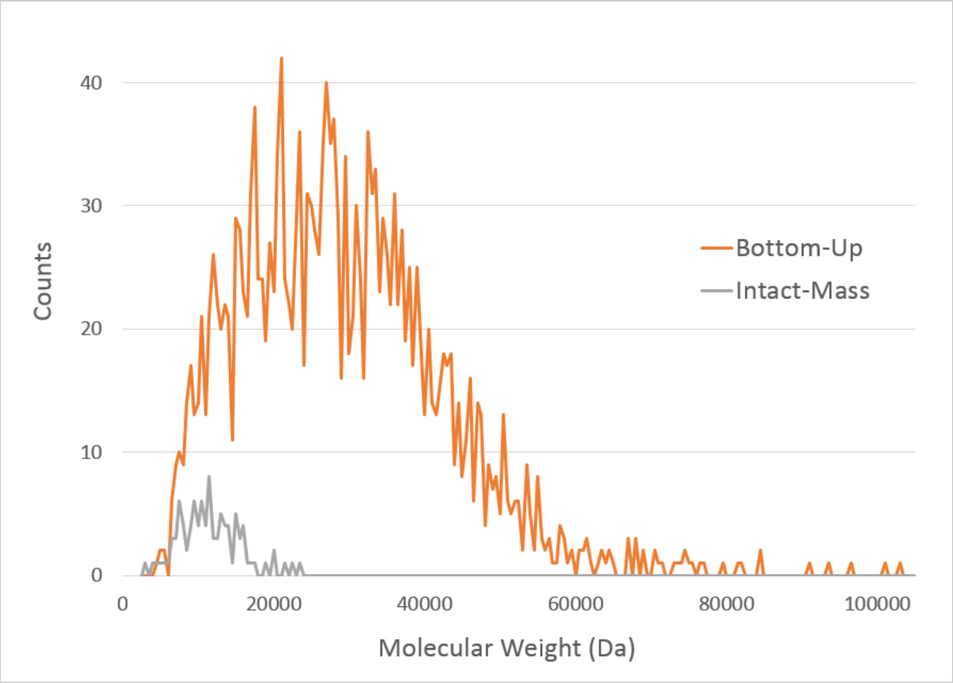
SUPPLEMENTARY FIGURES.



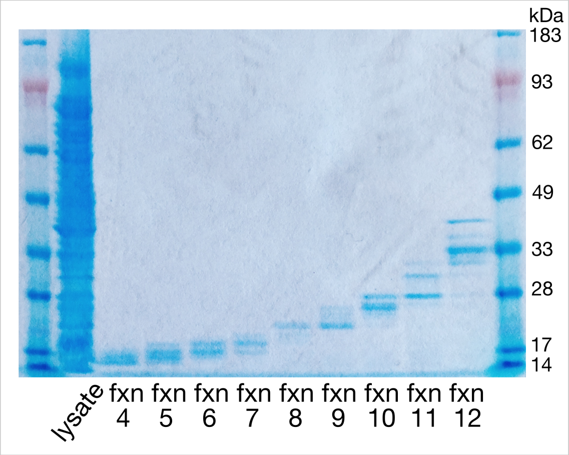
SI Figure S-1. (A) Lysine count and (B) intensity ratio distribution of the 90 259 NeuCode pairs that Proteoform Suite identified in this study. The intensity ratios of "light" to "heavy" concentrated between 1.6:1 to 2.0:1, which was close to the mixing ratio of "light" and "heavy" pellets at 2:1. NeuCode pairs whose intensity ratio was between 1.4:1 to 3.0:1 were accepted to aggregate into experimental proteoforms.



SI Figure S-2. Histogram of the number of proteoform members in identified families. (A) Using a UniProt XML database, most identified families have a single unmodified theoretical proteoform. (B) Using a G-PTM-D database, families include more modified theoretical proteoform members.



SI Figure S-3. Histogram of proteins identified from bottom-up (orange trace) and protein families identified from intact-mass using G-PTM-D database (grey trace). All the mapped proteins from the bottom-up search have lysine counts from 2 to 26. Their molecular weight range matches Gelfree™ fractionation's protein mass limitation (<45 kDa). The molecular weight range of the 97 identified families (including 5 ambiguous families, mapped using the lowest proteoform molecular weight) from intact-mass proteomics matches the Orbitrap limit (<25 kDa). The dynamic range of intact-mass and lysine count identification has yet to be widened.



SI Figure S-4. Gelfree™ size-based fractionation was confirmed by polyacrylamide gel electrophoresis (PAGE).