Supporting Information Part I

Robustness and Ruggedness of Isoelectric Focusing and Superficially Porous Liquid Chromatography with Fourier transform Mass Spectrometry

John R. Corbett[2,3], Dana E. Robinson[2], Steven M. Patrie[1,2,*]

[1,*]Department of Chemistry, Northwestern University, 2145 Sheridan Rd, Evanston, IL, 60208
Email: steven.patrie@northwestern.edu. Telephone: 847-491-3731
[2]Department of Pathology, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390
[3]Department of Bioengineering, UT Dallas, 800 W. Campbell Rd., Richardson, TX 75080

Correspondence: Dr. Steve M. Patrie

[1,*]Department of Chemistry, Northwestern University, 2145 Sheridan Rd, Evanston, IL, 60208
Email: steven.patrie@northwestern.edu. Telephone: 847-491-3731

This word file contains the first part of the Supplemental Information, including additional experimental details and 6 supporting figures. An excel file contains the second part of the Supporting Information, comprised of 5 supporting tables.

Multi-dimensional Binning

Datasets associated with individual IEF fractions were automatically binned to avoid duplicate reporting of redundant masses observed in adjacent IEF fractions and for the calculation of summed intensities and weighted p*I* values for each proteoform observed. Initial binning criteria of ± 3 min, ± 3 p*I* units, > 1,000 intensity, and 30 ppm and 2 Da for high and low resolution respectively were used. Weighted p*I* ($_{w}$p*I*) estimates were calculated by $\sum_{i=1}^{i=n}(pI_i) * (relative\ intensity_i)$, where n was the number of putative masses binned. Propagation of error for $_{w}$p*I* was calculated by $E_T = \sqrt{(s_{pI})^2 * (e_{pI})^2}$ where $e_{pI} = 0.15$. Limits of detection (LOD) were calculated from the slope of the calibration curve and standard deviation of the slope for 2D IEF-SPLC-MS. LOD = (((3.3 * 8 standard deviation of the curve) / (slope of the curve)) / x) where x = 7.5 since only 20 µL of each IEF fraction was analyzed by SPLC-FTMS [1]. The $_{w}$p*I* change associated with incorporation of each modification class for the standard proteins was determined by inclusion of phosphorylation (pKa 1.5) and sialylation (pKa 3.0) into the pKa sets within the program Isoelectric Point Calculator (IPC)[2]. For fragment assignment, *E. coli* (strain K12) proteome was downloaded from UniProt and a database shotgun annotated in ProSightPC 3.0 (Thermo) with only six concurrent PTMs considered. The theoretical proteoforms were then processed with ProPAS to create theoretical hydrophobicity and p*I*[3]. Protein/proteoform identification was accomplished with a previously reported custom search engine run on a Windows 2012 Server[4]. The tool supports absolute mass searches and reports of Poisson-based P-scores of identified proteins/proteoforms[5]. Absolute mass test used a 1.2 Dalton (Da) mass tolerance and a 10 ppm fragment mass tolerance. Alignment of *E. coli* protein hydrophobicity (LC retention time vs. GRAVY) and weighted p*I* was performed by calibration curves comparing observed vs. theoretical physiochemical property values, respectively, for the 214 unique proteoforms identified by NSD experiments (**Table S1**). These calibration curves were subsequently applied to the triplicate *E. coli* proteome runs to create post-calibrated hydrophobicity and weighted p*I* values.

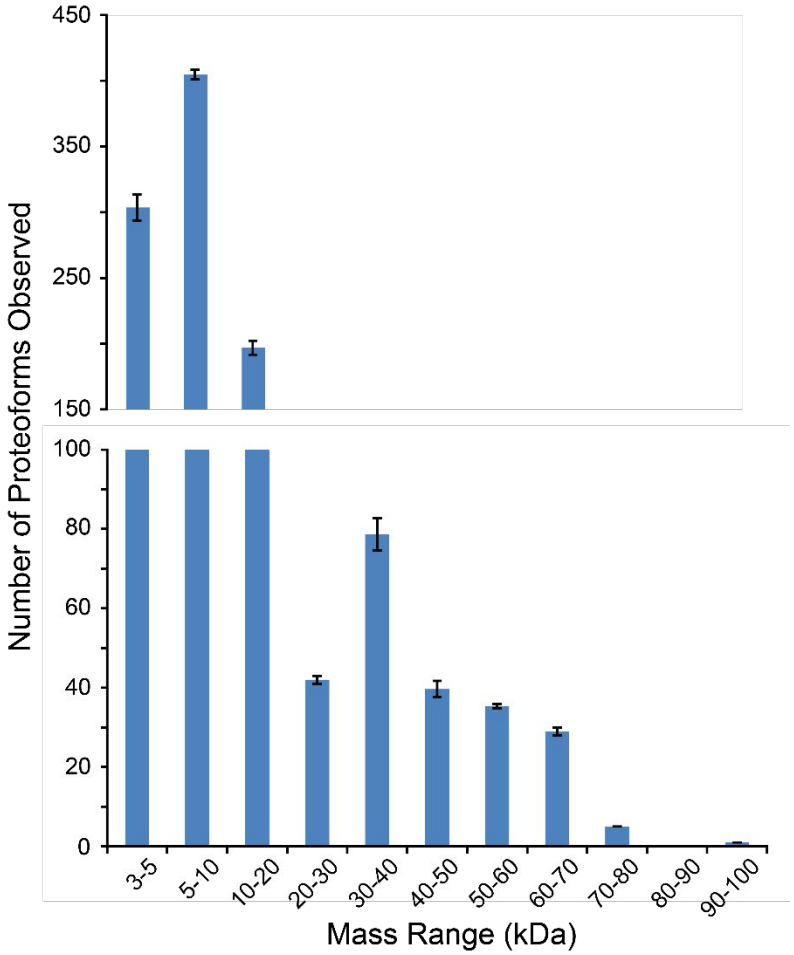


Figure S1: Average number of observed proteoforms and their respective mass ranges (kDa) for the triplicate high and low resolution *E. coli* proteome runs.

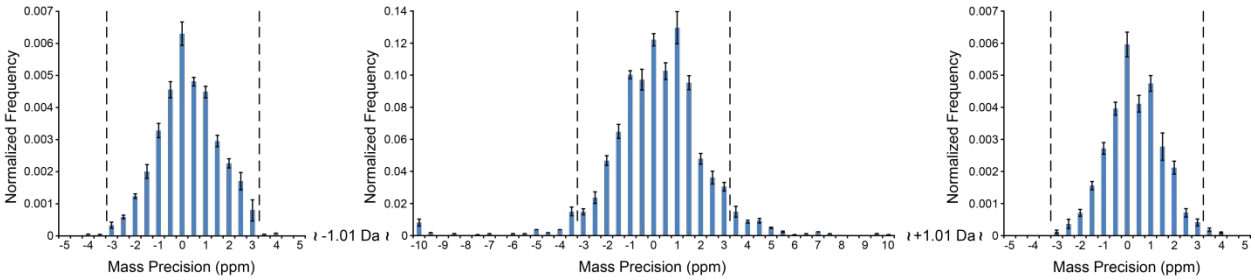


Figure S2: Mass precision histograms for the high resolving power extracted proteoforms. Bar lines represent 95% confidence interval.

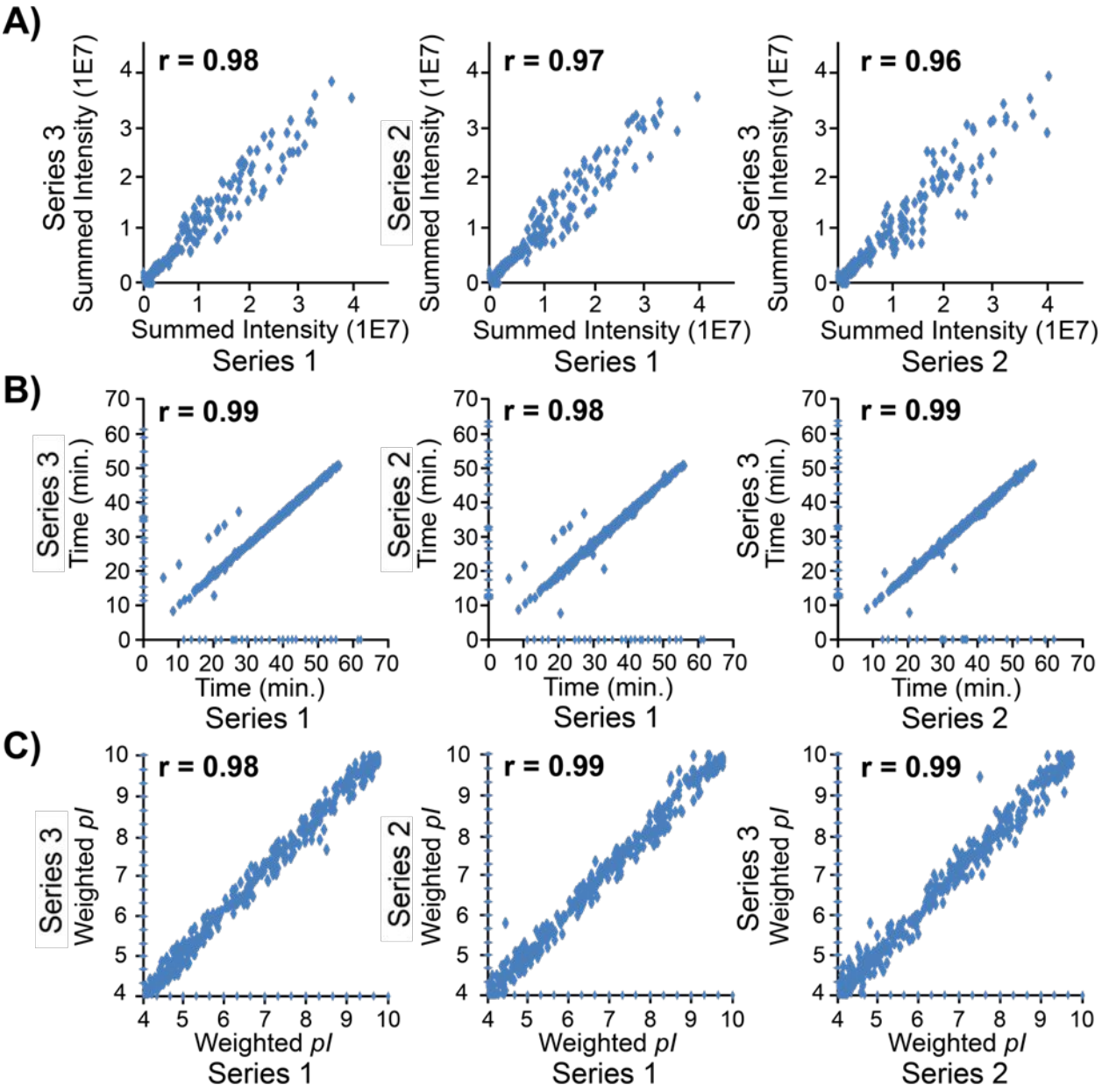


Figure S3: A-C) Pearson correlation studies conducted on proteoform summed intensities, retention time, and weighted p*I*s for the high resolution *E. coli* datasets.

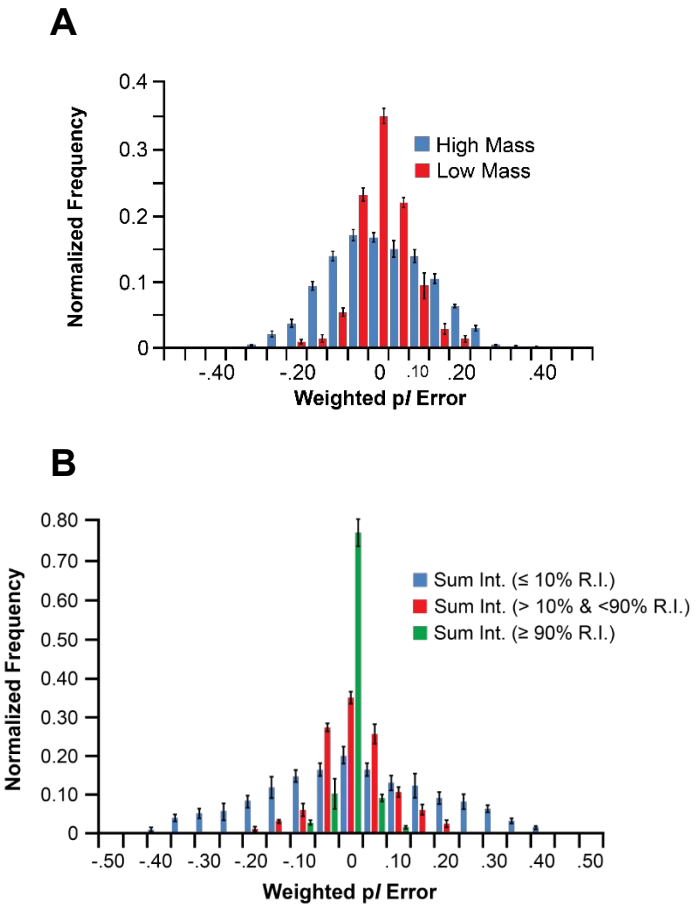


Figure S4: Weighted p*I* histograms for the high, medium, and low summed intensity ranges. R.I. stands for relative intensity.

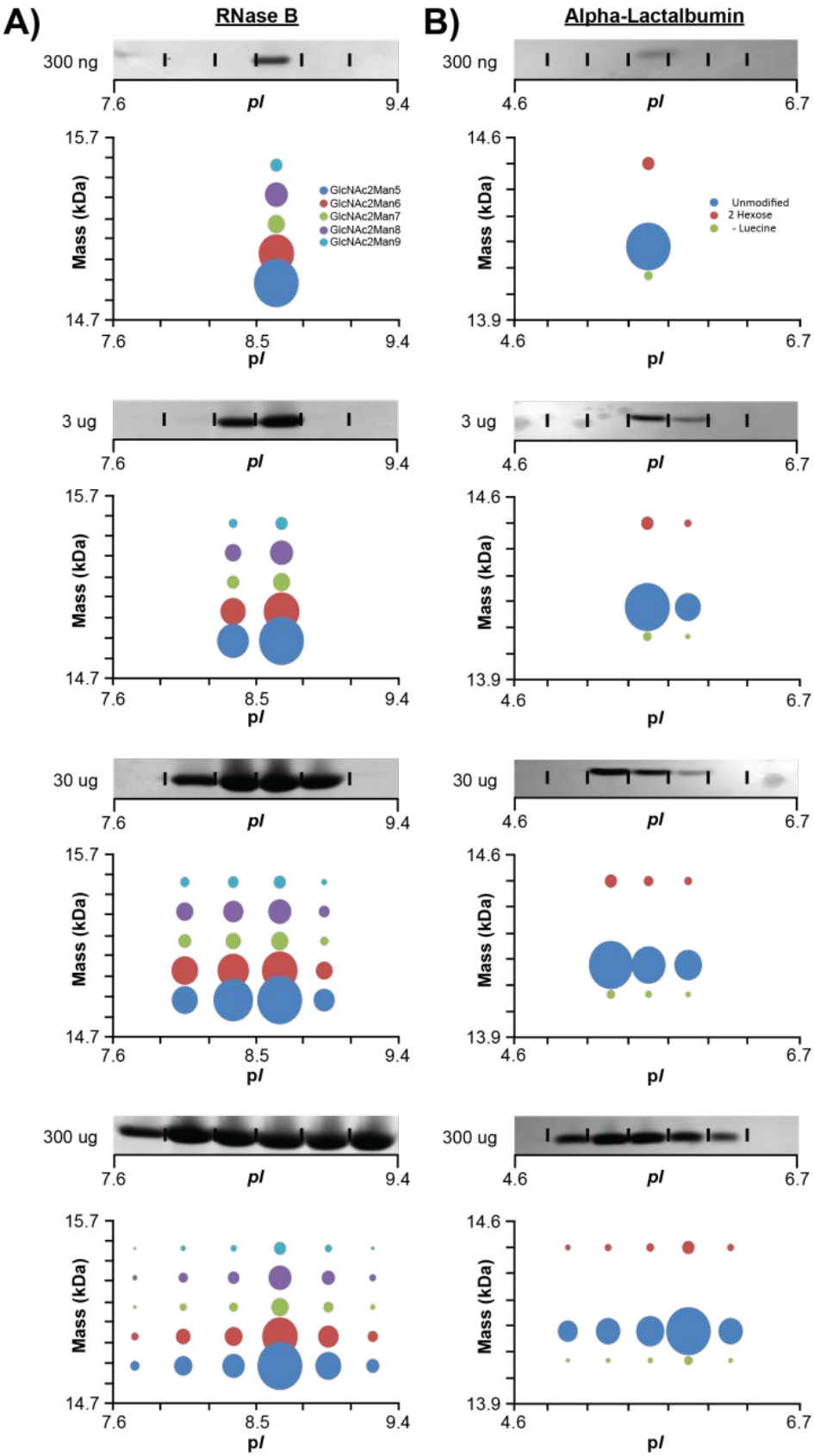


Figure S5: Evaluation of sample broadening and pre-binned signal intensity for the standard proteins at four IEF sample loads (0.3, 3, 30 and 300 µgrams). Representative silver stain gel for the standard proteins (A) RNase B, (B) α-lactalbumin, (C) BSA, and (D) transferrin at each sample load (C & D*, see next page*). Each respective Mass vs. p*I* bubble chart highlights an average signal intensity (bubble size) for various proteoforms observed for each protein. The average signal intensity was determined from three technical replicates at each sample load (n=3).

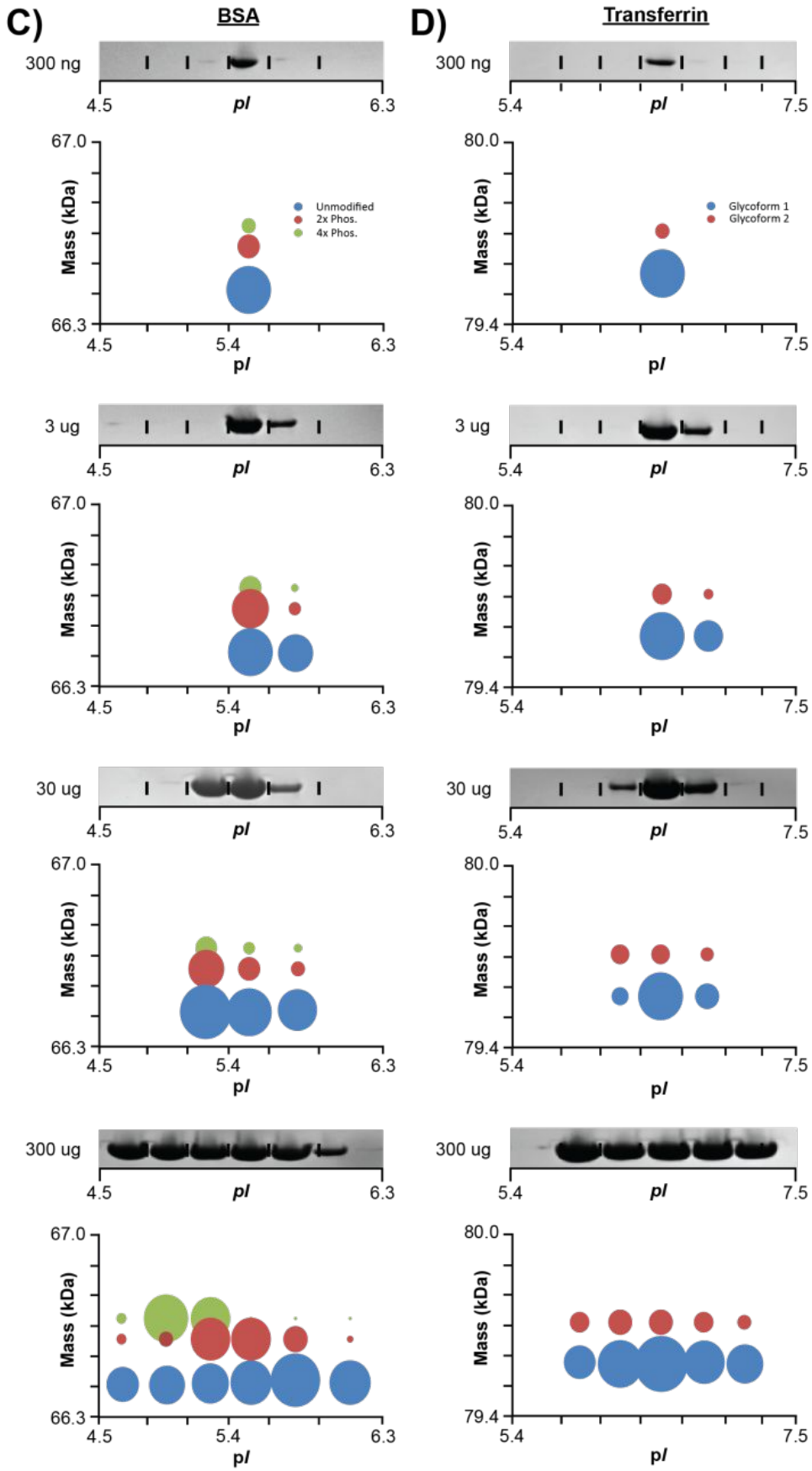


Figure S5: Continued.

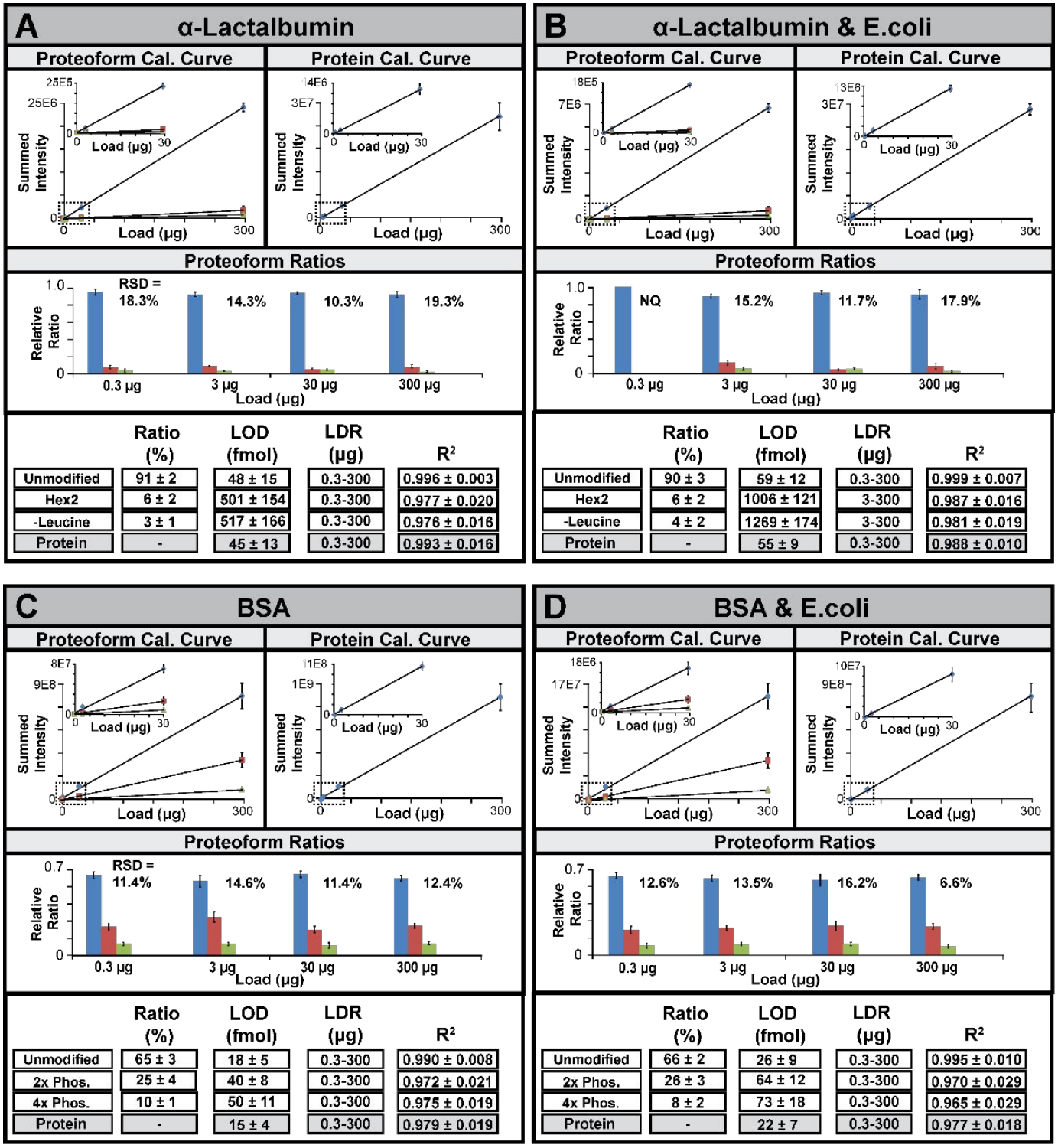
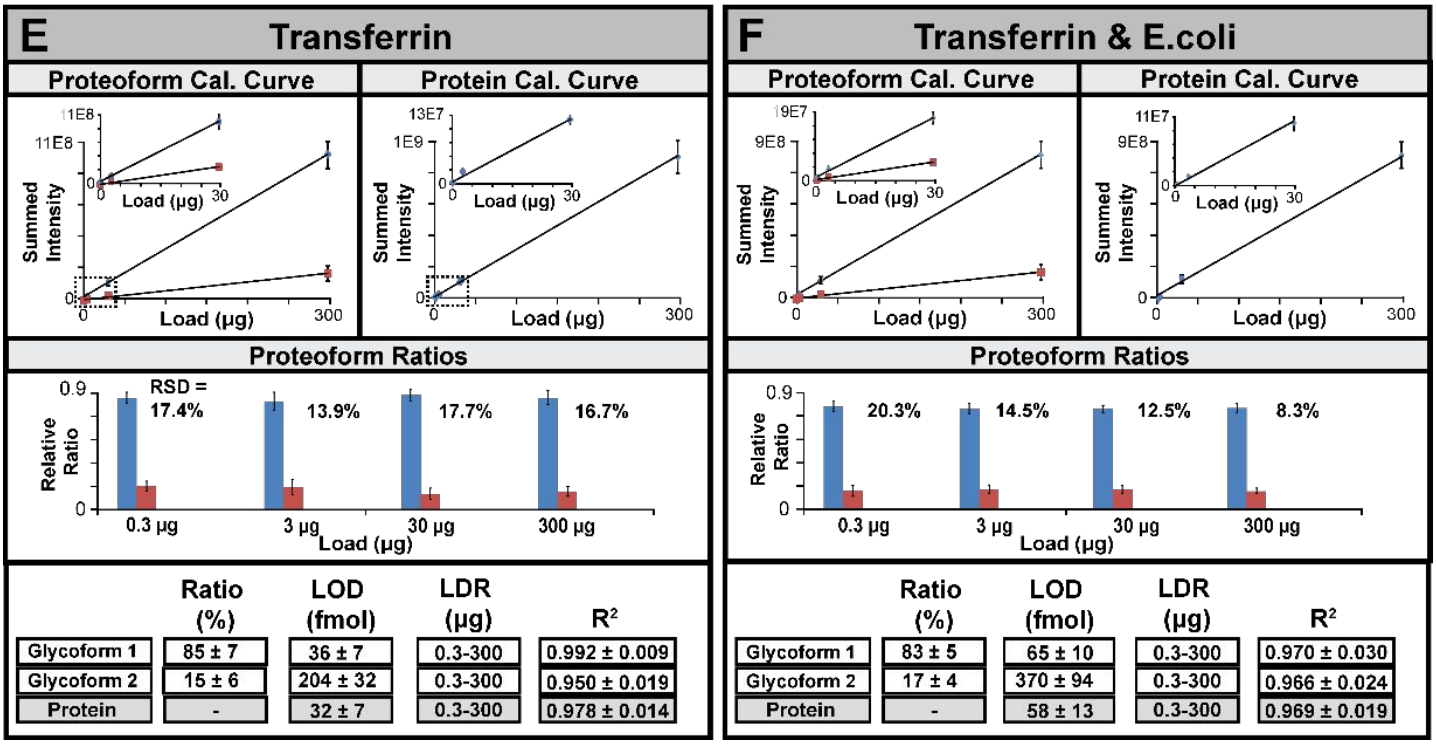


Figure S6: A-F) Averaged calibration curves (n=3) for the proteoforms (upper left panels), estimated theoretical total protein (upper right panels), proteoform ratios (middle panels), and tabulated FOM (lower panels) for non-spiked and spiked α-lactalbumin (A, B), BSA (C, D), and transferrin (E, F, *see next page*) across the four loading amounts. List of slopes and standard deviations are presented in **Table S5**.



	Ratio (%)	LOD (fmol)	LDR (µg)	R^2
Glycoform 1	85 ± 7	36 ± 7	0.3-300	0.992 ± 0.009
Glycoform 2	15 ± 6	204 ± 32	0.3-300	0.950 ± 0.019
Protein	-	32 ± 7	0.3-300	0.978 ± 0.014

	Ratio (%)	LOD (fmol)	LDR (µg)	R^2
Glycoform 1	83 ± 5	65 ± 10	0.3-300	0.970 ± 0.030
Glycoform 2	17 ± 4	370 ± 94	0.3-300	0.966 ± 0.024
Protein	-	58 ± 13	0.3-300	0.969 ± 0.019

Figure S6: Continued.

References

1. Zhang, J.; Roth, M. J.; Chang, A. N.; Plymire, D. A.; Corbett, J. R.; Greenberg, B. M.; Patrie, S. M., Top-down mass spectrometry on tissue extracts and biofluids with isoelectric focusing and superficially porous silica liquid chromatography. *Anal Chem* **2013,** *85* (21), 10377-84.
2. Kozlowski, L. P., IPC - Isoelectric Point Calculator. *Biol Direct* **2016,** *11* (1), 55, pmcid:PMC5075173.
3. Wu, S.; Zhu, Y., ProPAS: standalone software to analyze protein properties. *Bioinformation* **2012,** *8* (3), 167-9, pmcid:PMC3283891.
4. Plymire, D. A.; Wing, C. E.; Robinson, D. E.; Patrie, S. M., Continuous Elution Proteoform Identification of Myelin Basic Protein by Superficially Porous Reversed-Phase Liquid Chromatography and Fourier Transform Mass Spectrometry. *Anal Chem* **2017,** *89* (22), 12030-12038, pmcid:PMC5735417.
5. Meng, F.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L., Informatics and multiplexing of intact protein identification in bacteria and the archaea. *Nat. Biotechnol.* **2001,** *19* (10), 952-7.