

RESEARCH ARTICLE OPEN ACCESS

Comparative Analysis of Proteoform Clean-Up Methods for In-Depth Top-Down Proteomics

Alyssa A. Williams¹ | Kendalyn N. Firenza² | Amy K. Carfagno¹ | Jake T. Kline¹ | Luca Fornelli^{1,2} ¹School of Biological Sciences, University of Oklahoma, Norman, Oklahoma, USA | ²Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma, USACorrespondence: Luca Fornelli (luca.fornelli@ou.edu)

Received: 16 May 2025 | Revised: 29 September 2025 | Accepted: 9 October 2025

Funding: This work was supported by NIH R35GM147397 awarded to Luca Fornelli, a University of Oklahoma Honors College Undergraduate Research Opportunities Program (UROP) grant awarded to Kendalyn Firenza, and the University of Oklahoma Libraries' Open Access Fund.**Keywords:** detergent removal | PEPPi-MS | proteomics | sample preparation | top-down mass spectrometry

ABSTRACT

Proteoforms, the endogenous forms of proteins that take into account all sources of variation (genetic as well as co- and post-translational), may be more relevant for understanding complex biological mechanisms, including disease phenotypes, than generically defined “protein families”. Mass spectrometry-based top-down proteomics often utilizes electrophoretic techniques that rely on sodium dodecyl sulfate (SDS) to fractionate proteoforms prior to liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis. Methanol-chloroform-water (MCW) precipitation is commonly utilized for SDS removal due to its efficacy and low cost, yet it may lead to poor recovery of smaller proteoforms. As a technical contribution to top-down proteomics, four commercial SDS clean-up alternatives were benchmarked against MCW. Results indicate that MCW yields fewer proteoform identifications, particularly among small and acidic proteoforms. The analysis of post-translational modifications identified using the different clean-up methods indicates increased prevalence of methylation modifications post-MCW clean-up. Among the commercial kits, DetergentOUT and HiPPR achieved SDS removal comparable to MCW but at a higher cost. For studies sensitive to loss of low molecular weight or acidic proteoforms, these kits may offer an advantage. Alternatively, at a lower cost, Minute SDS provides sufficient SDS removal and broader proteome coverage.

1 | Introduction

The completion of the Human Genome Project provided an invaluable source of information to identify the genetic causes of a number of diseases [1–3]. However, it became rapidly apparent that solely investigating the genetic roots of disease is insufficient to fully explain the observed variations in onset and progression. It is necessary to shift the focus to the molecular effectors of

cellular functions, particularly proteins, to elucidate the mechanisms underlying complex aberrant phenotypes. A now widely accepted notion proposes that the definition of proteins as a mere sequence of amino acids representing the information encoded in a gene is largely inadequate, as sources of variations introduced at the nucleic acid level (e.g., single-nucleotide polymorphisms, alternative splicing) and directly to the amino acids (e.g., post-translational and co-translational modifications)

Abbreviations: AnExSP, anion-exchange disk-assisted sequential sample preparation; AGC, automatic gain control; BCA, bicinchoninic acid assay; CV, compensation voltage; DDA, data-dependent acquisition; FA, formic acid; FAIMS, high-field asymmetric waveform ion mobility spectrometry; GELFrEE, gel-eluted liquid fraction entrapment electrophoresis; GLOB, globular; HiPPR, high protein and peptide recovery; MCW, methanol/chloroform/water; NCE, normalized collision energy; PEPPi, passively eluting proteins from polyacrylamide gels as intact species; SP, signal peptide; TDP, top-down proteomics; TM, transmembrane.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDeriv](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Proteomics* published by Wiley-VCH GmbH.

Summary

This study aims to support the growing top-down proteomics (TDP) community by addressing a practical challenge in sample preparation: efficient SDS removal following molecular weight-based fractionation. As TDP gains traction for its ability to directly characterize proteoforms and their post-translational modifications (PTMs), accessible and reliable workflows become increasingly important. Detergent removal is a key step, as SDS, though essential for protein solubilization and separation, interferes with mass spectrometry. Traditional methods like methanol-chloroform-water (MCW) precipitation can be labor-intensive and may lead to loss of small proteoforms. In this work, we compare four commercially available SDS clean-up kits to MCW based on SDS removal efficiency, proteoform recovery, PTM retention, and ease of use. Our findings provide practical guidance regarding each method's trade-offs, aiming to inform more streamlined, reproducible TDP workflows. By benchmarking these kits, we contribute to ongoing efforts to democratize TDP and enhance its utility in biological and clinical research.

- The golden standard for SDS removal is methanol-chloroform-water (MCW) precipitation, which is labor-intensive and, as demonstrated by both earlier reports and the present work, may lead to loss of small proteoforms.
- This study benchmarks four commercially available SDS clean-up kits against MCW, considering SDS removal efficiency, proteoform recovery, retention of post-translationally modified proteoforms, and ease of use.
- Resin-based clean-up methods, DetergentOUT and HiPPR, were highly effective in SDS removal, and comparable to MCW clean-up.
- Minute SDS and SurfactAway kits, on the other hand, allowed the identification of a slightly higher number of proteoforms.

may substantially affect protein function and longevity [4–6]. These protein forms, or proteoforms [7], encompass all aspects of variation in expressed gene products; therefore, their analysis may more accurately model differences in disease phenotypes compared to the consideration of generic protein families [8]. This underscores the importance of accurately mapping the proteoform landscape.

Mass spectrometry (MS)-based top-down proteomics (TDP) bypasses the protein digestion step central to bottom-up proteomics, allowing for the direct analysis of proteoforms without incurring the protein inference problem afflicting peptide-centric proteomic approaches [9, 10]. Within the small molecular weight range (e.g., 0–30 kDa), TDP has been successfully applied in a number of studies of the human proteome, including untargeted *in vitro* experiments leading to the identification of thousands of proteoforms [11–15], the targeted characterization of individual proteoforms linked to specific diseases such as cancer and cardiac diseases [16, 17], and even the analysis of human tissues [18–22]. Despite recent efforts to simplify the higher-throughput analysis of proteoforms up to 60 kDa [23, 24], TDP is typically performed

on proteoforms <30 kDa due to signal dilution effects that complicate the detection of larger proteoforms [25]. Achieving comprehensive backbone fragmentation of larger species with a single dissociation method is another factor limiting sequence coverage and thus adequate characterization of larger proteoforms [26]. For this reason, the field remains reliant on molecular weight-based protein fractionation methods prior to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. While size exclusion chromatography has been successfully employed for off-line fractionation [27], electrophoretic methods are more commonly applied. These include gel-eluted liquid fraction entrapment electrophoresis (GELFrEE) [28] and passively eluting proteins from polyacrylamide gels as intact species (PEPPI) [29]. Molecular weight-based fractionation substantially reduces sample complexity, allowing greater coverage of the proteoform landscape [30]. Though the use of sodium dodecyl sulfate (SDS) is necessary for proper protein solubilization, unfolding, and homogeneous charging during gel-based separation, this detergent suppresses electrospray ionization [31]. Thus, its removal is necessary prior to LC-MS/MS analysis [14].

Methanol/Chloroform/Water (MCW) precipitation has long been the gold standard for SDS removal from GELFrEE and PEPPI fractions, since this method is effective and inexpensive [32]. Another inexpensive method, cold acetone precipitation, provides insufficient SDS removal for LC-MS/MS analysis and may introduce artificial proteoform modifications [33]. However, recent reports showed that MCW may yield lower recovery of small proteoforms, particularly <15 kDa [30, 34]. Introduced by Takemori and co-workers, the anion-exchange disk-assisted sequential sample preparation (AnExSP) method is an alternative SDS removal approach based on a stage tip [35] prepared using a solid phase extraction disc [36]. The use of AnExSP provides improved recovery of smaller proteoforms compared to MCW [30], and the method has been further refined by replacing the use of membrane filtration with single-pot, solid-phase-enhanced sample preparation (SP3), resulting in increased recovery of proteins carrying post-translational modifications (PTMs) [34]. However, both versions of the AnExSP protocol are relatively labor-intensive and require the manual preparation of stage tips, potentially introducing reproducibility issues if not performed by expert users. Multiple attempts at democratizing TDP have recently been coordinated by the Consortium for Top-Down Proteomics [37–39], but these have not included improvements in detergent clean-up that would streamline the sample preparation process and benefit new TDP practitioners. Toward this end, we compared four commercially available detergent clean-up kits in terms of intact proteome coverage and SDS removal efficiency (Figure S1). In addition to determining the number of proteoforms identified, clean-up methods were evaluated for their biases in proteoform retention based on physico-chemical properties and the types of PTMs present. Specifically, we subjected 0–30 kDa proteoforms from an *Escherichia coli* whole cell lysate to clean-up using High Protein and Peptide Recovery (HiPPR) Detergent Removal Spin Columns (Thermo Scientific, Waltham, MA, USA), DetergentOUT (G-Biosciences, St. Louis, MO, USA), Minute SDS-Remover (Invent Biotechnologies, Plymouth, MN, USA), and SurfactAway (Biotech Support Group, Monmouth Junction, NJ, USA). To assess possible differences in effects between prokaryotic and eukaryotic samples, we also

evaluated 0–30 kDa proteoforms from the Jurkat human cell line to assess the consistency of trends in clean-up method performance. Results from the commercial solutions were benchmarked against MCW in terms of proteoform identifications and efficiency of SDS removal, with each clean-up approach also evaluated in terms of ease of use and cost.

2 | Materials and Methods

2.1 | Proteoform Extraction

E. coli (strain K12 25922) was grown in LB media at 37°C with shaking at 200 rpm. Mature cultures (OD₆₀₀ of ~0.7 for LB) were harvested by centrifugation. Cell pellets were lysed by sonication on ice (10 cycles of 10 s on, 20 s off, 50% amplitude for a total of 500 J) in a detergent lysis buffer (20 mM Tris pH 7.5, 100 mM NaCl, 1% Nonidet P-40 in water) supplemented with HALT protease inhibitor cocktail (Thermo Scientific, Waltham, MA, USA) [19]. Lysates were clarified by centrifugation, and the supernatant was retained and quantified by BCA Gold assay (Thermo Scientific, Waltham, MA, USA).

Jurkat cells (ATCC, Manassas, VA, USA) were seeded at 1e5 and grown to a density of roughly 1e6 in DMEM media supplemented with 10% FBS and 1% PenStrep at 37°C and 5% CO₂ over a period of 72 h. Cell pellets were lysed in detergent lysis buffer supplemented with HALT, and lysates were clarified, as previously described for *E. coli* lysis. Lysates were quantified using the Pierce BCA Protein assay (Thermo Scientific, Waltham, MA, USA).

2.2 | PEPPI Fractionation

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed using an in-house prepared gel (1.5 mm thick, 10 wells) composed of a 10% resolving gel and Tris/Glycine/SDS running buffer (Bio-Rad, Hercules, CA, USA). Samples (300 µg *E. coli* or Jurkat whole cell protein extract per gel lane) were mixed with 5× loading buffer (10% SDS, 500 mM dithiothreitol [DTT], 50% glycerol, 250 mM Tris pH 6.8, 0.5% Bromophenol Blue) and subsequently separated at a constant amperage of 25 mA. Following Kline et al. [19], PEPPI was used to isolate the 0–30 kDa protein fraction from each gel lane. Proteoforms extracted from each lane were subsequently combined and split into equal volume aliquots to allow for three clean-up technical replicates for each of the five clean-up methods. Successful recovery of the 0–30 kDa fraction was confirmed by analyzing a small aliquot (~1/15 of the total volume) of each sample via SDS-PAGE using a 12% resolving gel.

2.3 | Detergent Removal

Immediately before MS analysis, triplicate samples were cleaned by either MCW, DetergentOUT, HiPPR, Minute SDS-Remover, or SurfactAway and reconstituted in mobile phase A (vide infra). The MCW clean-up was performed with a solution of 600 µL methanol, 150 µL chloroform, and 450 µL ultrapure water added to the sample, mixed thoroughly, and centrifuged at 13,000 rpm for 10 min at room temperature. This step yielded a biphasic solution, and the upper aqueous layer was carefully removed.

450 µL of methanol was added to the remaining lower phase, followed by another centrifugation at 13,000 rpm for 10 min at room temperature. The resulting protein pellet was washed with 450 µL of methanol, followed by another centrifugation at 13,000 rpm for 10 min at room temperature, subjected to five additional methanol washes, and then air-dried for 20 min. To remove residual cellular debris, samples were resuspended in mobile phase A and then centrifuged at 13,000 rpm for 10 min at 4°C to collect the supernatant to be placed in a passivated vial (MicroSolv Reduced Surface Activity glass vials coated with bovine serum albumin and rinsed thoroughly prior to sample addition). Protocols for commercial kits were carried out according to the manufacturer's instructions except for the following changes. Zeba columns were used after Minute SDS-Remover and SurfactAway clean-ups due to remaining salts and dye post-clean-up. The HiPPR kit used either a 1:2 sample-to-settled resin ratio or a 2:1 ratio, with an additional desalting step in the latter case. Samples cleaned with all protocols, except for MCW, were placed in a Vacufuge Plus (Eppendorf, Framingham, MA) to concentrate samples before dilution in mobile phase A. BCA analysis (Thermo Scientific, Waltham, MA) was performed to determine protein recovery after each of the five clean-up methods. Jurkat PEPPI sample clean-up was performed using MCW, HiPPR, and Minute SDS methods to evaluate whether similar trends in clean-up method performance were observed for samples from both eukaryotic and prokaryotic organisms.

2.4 | SDS Concentration Assay

The method of Cao et al. [40] was applied to determine residual SDS concentrations remaining after cleaning. SDS concentration curves were acquired at 450 nm for SDS concentrations within a low-weight range (0.31, 0.63, 1.25, 2.5, and 5 ppm) in triplicate using an Agilent BioTek 800 TS Absorbance Reader (Fisher Scientific, Waltham, MA, USA) with a 96-well plate. Approximately 30 µg of *E. coli* whole-cell lysate was resuspended in 0.1% SDS solution, for consistency with protein and SDS content expected in typical PEPPI fractions [29], to be cleaned using MCW and each of the commercial kits. Residual SDS concentration post clean-up was measured at 450 nm in triplicate. Zeba desalting columns were used for Minute SDS, SurfactAway, and HiPPR (sample to settled resin of 2:1 only).

2.5 | Liquid Chromatography—Tandem Mass Spectrometry Analysis

An Ultimate 3000 HPLC system (Thermo Fisher Scientific, Sunnyvale, CA, USA) was used for chromatographic separation. Mobile phase A was water with 5% ACN and 0.2% formic acid (FA) (v/v); mobile phase B was acetonitrile with 5% water and 0.2% FA (v/v). All samples were resuspended to a final volume of 25 µL with mobile phase A. Samples were separated using an EASY-Spray MAbPac RP column (150 µm i.d. x 150 mm length, particle size 4 µm, Thermo Scientific), heated to 55°C. The flow rate was maintained at 1.5 µL/min throughout the 65-min total run time. The 45-min separation gradient increased from 5%–14% mobile phase B in 1.5 min, then 14%–42% B in 43.5 min. This was followed by two consecutive column washes at 85% B before re-equilibration at 5% B for 6 min.

MS experiments were performed using an Orbitrap Eclipse Tribrid mass spectrometer equipped with a FAIMS Pro unit and an EASY-Spray nanoelectrospray source (Thermo Scientific, San Jose, CA, USA). The ionization voltage was set between 2.1–2.3 kV. All measurements were performed using “intact protein” mode (3 mTorr in the ion routing multiple). Data were acquired using a data-dependent acquisition (DDA) strategy. Broadband mass spectra (MS¹) were recorded at 120,000 (at 200 *m/z*) resolving power (r.p.); higher-energy collisional dissociation (HCD) MS/MS data were collected at 60,000 r.p. (at 200 *m/z*). Normalized collision energy (NCE) for HCD was set at 35%, with the most abundant ions with charge between +4 and +50 isolated in the quadrupole with a 3 *m/z* isolation window. Dynamic exclusion was enabled (30 s duration). AGC targets were set at 4.8e5 and 2.5e5 for MS¹ and MS², respectively, both using a single microscan. To maximize the number of identified proteoforms, each clean-up technical replicate was analyzed in two separate LC-MS/MS analyses applying different sets of compensation voltages (CVs) in each run (i.e., internal CV stepping [14, 19, 41, 42]). The CVs used were as follows: −40 V, −20 V, and 0 V for the first injection of each sample, and −45 V, −30 V, and −5 V for the second. All MS data files have been uploaded to MassIVE (repository number MSV000097808).

2.6 | Data Analysis

Mass spectrometry results files were searched with ProSight PD v 4.3 (Proteinaceous, Inc., Evanston, Illinois, USA) run as a node within Proteome Discoverer 3.1 (Thermo Fisher Scientific, San Jose, CA, USA) against a proteoform database for *E. coli* or human. Each cleaning method data set was searched separately. The FAIMS CV setting was left unspecified in the Spectrum Selector node, given the use of multiple CVs per file. Deconvolution was performed using the High/High cRAWler node with the Xtract algorithm and signal-to-noise threshold of 3 for both precursor and fragment ions. Two database searches were performed using (1) the Annotated Proteoform node with 2.2 Da precursor mass tolerance, 10 ppm fragment mass tolerance, and acetylation as a possible N-terminal modification and (2) the Subsequence Search node with 10 ppm for both subsequence precursor and fragment mass tolerance and acetylation as a possible N-terminal modification. The maximum number of proteoforms per precursor was set to 1 for both searches. In the consensus workflow, the ProSightPD FDR node was used for FDR calculations at the proteoform, isoform, and protein levels, with proteoform identifications filtered at 1% FDR. Proteoform identifications were analyzed and compared among methods based on the calculated Grand Average of Hydropathy (GRAVY) scores and isoelectric points using the Pyteomics software package [43–46] using the identified amino acid sequence for each proteoform. Topology prediction was performed using DeepTMHMM [47].

3 | Results and Discussion

3.1 | Residual SDS Concentration and Protein Recovery Analysis

While all methods removed over 99% of the initial SDS input, the commercial kits varied in their effectiveness when removing

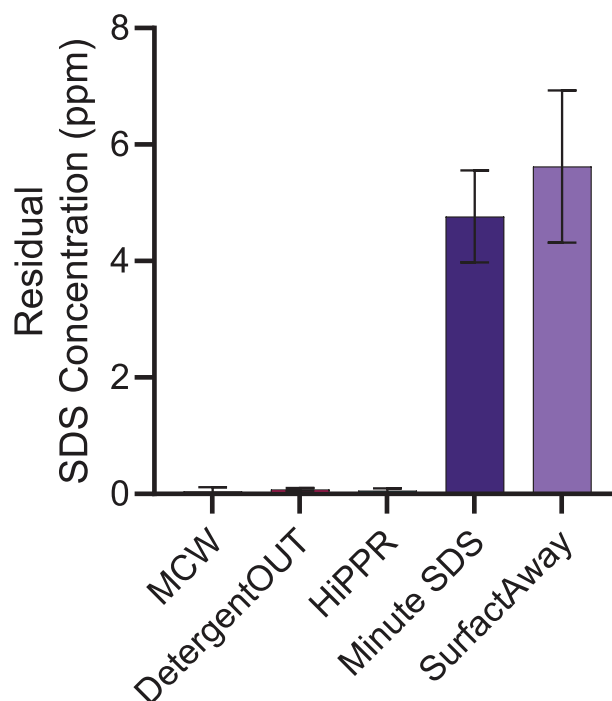


FIGURE 1 | Residual SDS concentration analysis. Approximately 30 μ g of *E. coli* whole-cell lysate resuspended in solutions of 0.1% (1000 ppm) SDS were cleaned in triplicate with each method and residual concentration measured post-clean-up via a colorimetric assay. Error bars represent the standard deviation of residual SDS (ppm) measured for technical triplicates for each clean-up protocol.

SDS from protein in 0.1% solutions (Figure 1). Residual SDS measurements of solutions cleaned with Minute SDS averaged 4.77 ppm, while SurfactAway averaged 5.62 ppm. MCW easily outperformed these precipitation-based kits (0.027 ppm). The only commercial kits competitive with MCW were the column-based kits DetergentOUT and HiPPR, which had residual SDS concentrations of 0.079 and 0.056 ppm, respectively.

To determine if the different detergent removal methods had differences in protein recovery, the amount of protein was measured by BCA (Figure S2A). In general, MCW-cleaned samples retained less protein than those cleaned with commercial kits. When normalized to the input (Figure S2B), the MCW-cleaned samples retained approximately 25% of the 34 μ g input. The commercial kits yielded a higher percent recovery of total protein input, with DetergentOUT, HiPPR, and Minute SDS each recovering ~50% and SurfactAway recovering 34% of the input.

3.2 | Proteoform Identifications across Clean-Up Methods

MCW was used to benchmark the efficiency of commercially available clean-up kits in terms of their ability to identify expressed genes (hereinafter indicated as UniProt accessions) and their proteoforms from 0–30 kDa PEPPI fractions (with protein extraction efficiency confirmed by SDS-PAGE; Figure S3A). MCW was associated with the lowest number of identifications, with 717 proteoforms derived from 242 unique UniProt accessions (Figure 2A). In terms of the total number of identifications,

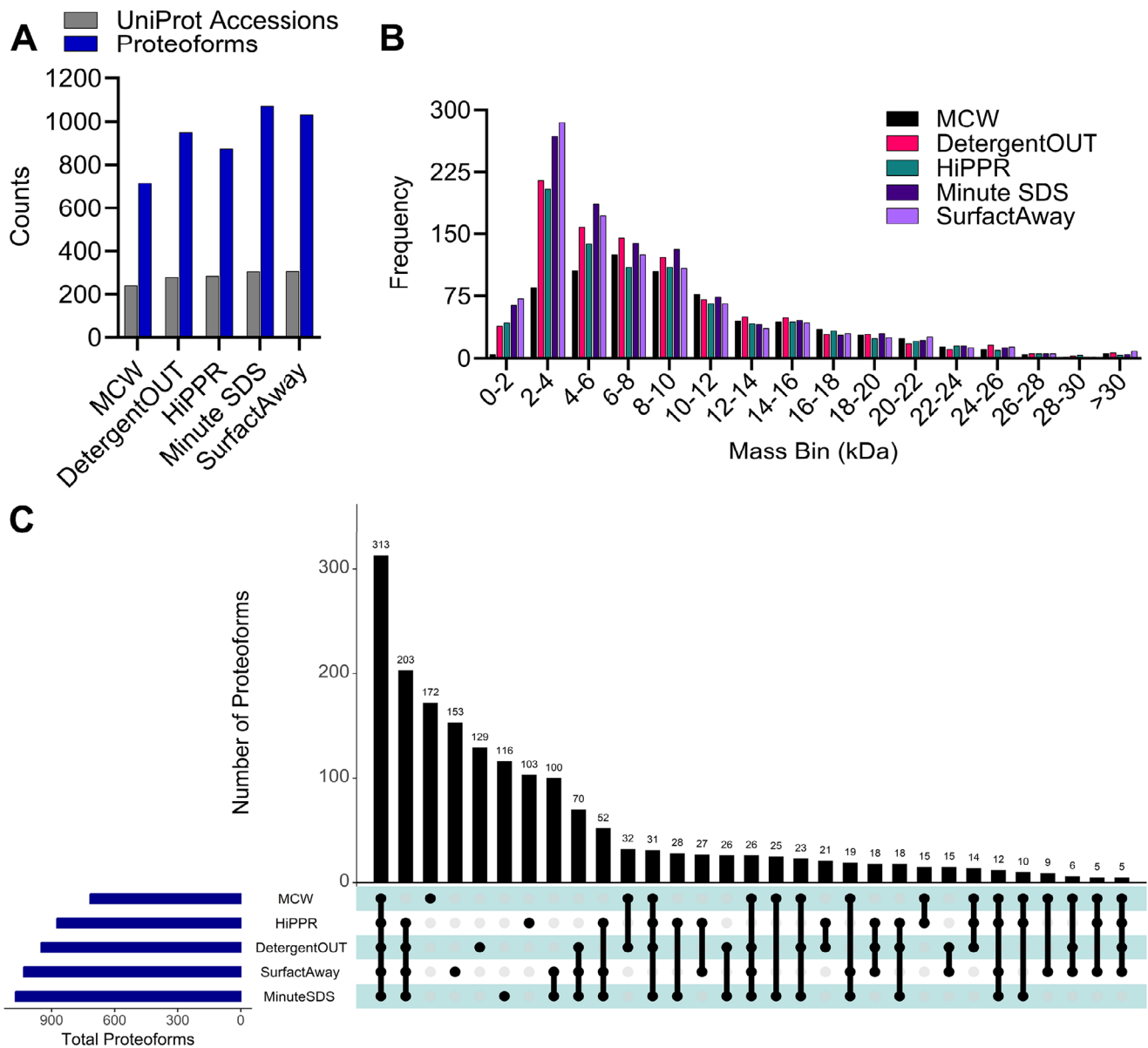


FIGURE 2 | UniProt accession and proteoform identifications from *E. coli* after performing the five SDS clean-up methods. (A) Global count of identified UniProt accessions and unique proteoforms. (B) Mass distributions of unique proteoforms. (C) Comparative distribution of shared proteoforms identified across clean-up methods. The UpSet plot depicts the total number of proteoforms identified for each clean-up method (lower left bar chart) as well as indicating proteoforms in common for each possible subset of sample groups (here, clean-up methods) in the dataset. For example, 313 proteoforms were identified for all five clean-up methods; 203 were identified for all four commercial kits, but not MCW; 172 were identified for MCW only, but for none of the commercial kits.

Minute SDS-Remover and SurfactAway performed similarly, in each case supporting the identification of more than 300 UniProt accessions and more than 1000 proteoforms. The other two commercial kits, DetergentOUT and HiPPR, yielded fewer identifications (for DetergentOUT, 280 UniProt accessions and 950 proteoforms; for HiPPR, 286 UniProt accessions and 875 proteoforms), yet still outperformed MCW. A more detailed assessment of the mass distribution of identified proteoforms indicated that the relatively lower number of identifications observed for MCW was mainly restricted to the 0–6 kDa mass range (Figure 2B). The lower recovery of smaller proteoforms observed for MCW is consistent with previous reports [30, 34]. However, for proteoforms larger than 6 kDa, MCW performed similarly to other methods

in terms of the number of identifications (Figure 2B). The mass distribution of proteoform identifications also indicates that, in the case of the Minute SDS-Remover and SurfactAway kits, the increased identifications may be more concentrated in the very low molecular weight range, particularly in the 2–4 kDa range.

To evaluate the reproducibility of intact proteome coverage among clean-up methods, overlap in proteoform identifications was determined for each set of technical triplicates for each clean-up method (Figure S4). Most methods yielded a reasonable degree of overlap in proteoform identifications among replicates. For MCW, HiPPR, and Minute SDS, at least 70% of total proteoform identifications were observed in at least two technical replicates,

whereas for DetergentOUT, only 62.2% of total proteoform identifications were observed in at least two replicates. Further, in the case of one of the DetergentOUT replicates, 214 proteoforms were identified that were not observed in the other two replicates. Considering the simplicity of the DetergentOUT protocol, this is unlikely to be indicative of introduced human error and may be due to the column resin performance of individual columns. Overall, this analysis indicated that HiPPR and MCW provided the most reproducible performance in terms of the highest percentage of total proteoform identifications observed in at least two replicates.

Finally, we summarized the similarities and differences of proteoform (Figure 2C) and UniProt accession (Figure S5) identifications across the five clean-up methods under analysis using UpSet plots. At the proteoform level, the largest intersecting group consisted of 313 proteoforms identified across all methods. Interestingly, the two next largest groups included 203 proteoforms identified across the four commercial methods but not found in MCW, and 172 proteoforms identified uniquely via MCW, respectively. A detailed investigation of these two groups revealed that proteoforms common to all four commercial methods, but not observed for MCW, were predominantly <10 kDa (Figure S6A). Though there was not a noticeable difference in the hydrophobicity of the unique proteoforms (Figure S6B, Tables S2–S3, p value = 0.5234), proteoforms exclusively identified in the commercial kits were generally more acidic compared to those unique to MCW (Figure S6C), with median pI values of 6.91 and 10.42 for commercial kits and MCW, respectively (Tables S4–S5, p value < 0.0001). Consistent with trends in the global dataset described below, PTM analysis indicated a higher prevalence of methylation and N-terminal acetylation modifications for proteoforms unique to MCW (Table S6).

For a more in-depth comparison of the complete set of proteoforms identified by each method, we examined the physicochemical properties of the proteoforms. Property prediction was performed using the Pyteomics Python package and DeepTHMM web tool based on proteoform primary sequences. A limitation of this approach is that effects of PTMs, for example, in shifting proteoform pI, are not accounted for. Despite this limitation, global patterns in the results are still expected to be informative regarding trends in proteoform identification for the different clean-up methods. Analysis of hydrophobicity was based on the distribution of GRAVY scores of identified proteoforms. All methods showed a similar distribution of hydropathy with a slight preference for less hydrophobic proteoforms (Figure 3A; descriptive statistics of GRAVY scores are available in Table S7, while full results of statistical testing are available in Table S8). Reflecting differences observed between proteoforms unique to MCW and those identified only from the commercial kits (vide supra), analysis of the global set of proteoform identifications also indicates that commercial methods identified proteoforms with lower isoelectric points than those identified by MCW (Figure 3B; descriptive statistics of pI are available in Table S9, while full results of statistical testing is available in Table S10). As noted, this indicates a slight bias against the recovery of acidic proteoforms in the case of MCW.

Proteoform topology prediction performed using DeepTMHMM indicated an additional difference between proteoforms identi-

fied by commercial kits compared to MCW (Figure 3C,D). In general, a greater number of proteoforms identified after clean-up with commercial kits were predicted to contain signal peptides or correspond to globular proteins, the latter of which may be more hydrophilic in nature (Figure 3D). However, while MCW has previously been observed to successfully extract membrane proteins [20], a similar number of proteoforms with predicted transmembrane topology were observed for MCW, SurfactAway, and DetergentOUT in the present data set (Figure 3D). It should also be noted that our lysate preparation is biased toward the recovery of soluble proteins. The low number of membrane proteoform identifications is unsurprising in this regard, with the small pool of membrane proteoforms present in the lysate samples potentially obscuring biases introduced by different clean-up methods in terms of membrane proteoform recovery. Taken together, these results support the idea that there is greater potential for loss of small and acidic proteoforms in the aqueous layer during MCW clean-up, compared to commercial kits that rely on either resin (DetergentOUT and HiPPR) or proprietary formulas to induce precipitation of SDS (Minute SDS-Remover and SurfactAway). To evaluate potential protein loss during the MCW process, we collected the aqueous layer along with two subsequent methanol washes and analyzed them by SDS-PAGE (Figure S7). Minimal protein is present in the aqueous phase, with no visual loss to the methanol washes observed. It is also worth noting that the protein content lost to the aqueous phase appears to be limited to the low-weight region (<15 kDa).

A recent study focusing on non-commercial SDS removal methods highlighted differences in the prevalence of observed PTMs across the tested protocols [34]. In light of those results, we assessed the potential biases in PTM identification across the four commercial methods compared to MCW (Figure 4; raw counts of PTMs identified by each method are available in Table S11). In general, the commercial kits exhibited a slightly higher prevalence of N-terminal truncation products and proteoforms that arise from simultaneous truncations of both termini (Figure 4A). This result aligns with expectations, as commercial kits tend to perform better in retaining very small proteoforms (<6 kDa). We also observed pyruvic acid modifications of Ser residues (Figure S8). This is consistent with the results of the isoelectric point analysis, which indicated that proteoforms identified for the MCW-cleaned samples generally exhibited higher isoelectric points than the commercial kits; that is, a reduced detection of acidic modifications was foreseeable. An increased prevalence of lysine methylation (mono-, di-, and tri-) was observed for proteoforms identified post-MCW clean-up compared to commercial methods, as well as a slightly higher prevalence of N-terminal acetylation (Figure 4B). Both such modifications render proteins slightly less hydrophilic, which, combined with the pI and molecular weight distribution results, provides support for the idea that MCW favors recovery of larger proteoforms yet risks the loss of smaller, acidic proteoforms in the discarded aqueous layer. In general, trends in post-translational modification prevalence were similar for Minute SDS and SurfactAway, both of which remove SDS via similar mechanisms (precipitation and pelleting). Consistent with previous work comparing clean-up of a human cell protein extract with methods including MCW and filter-aided sample preparation, reduced prevalence of N-terminal acetylation and serine phosphorylation was observed for the filter-based methods in the present study (i.e., DetergentOUT and HiPPR)

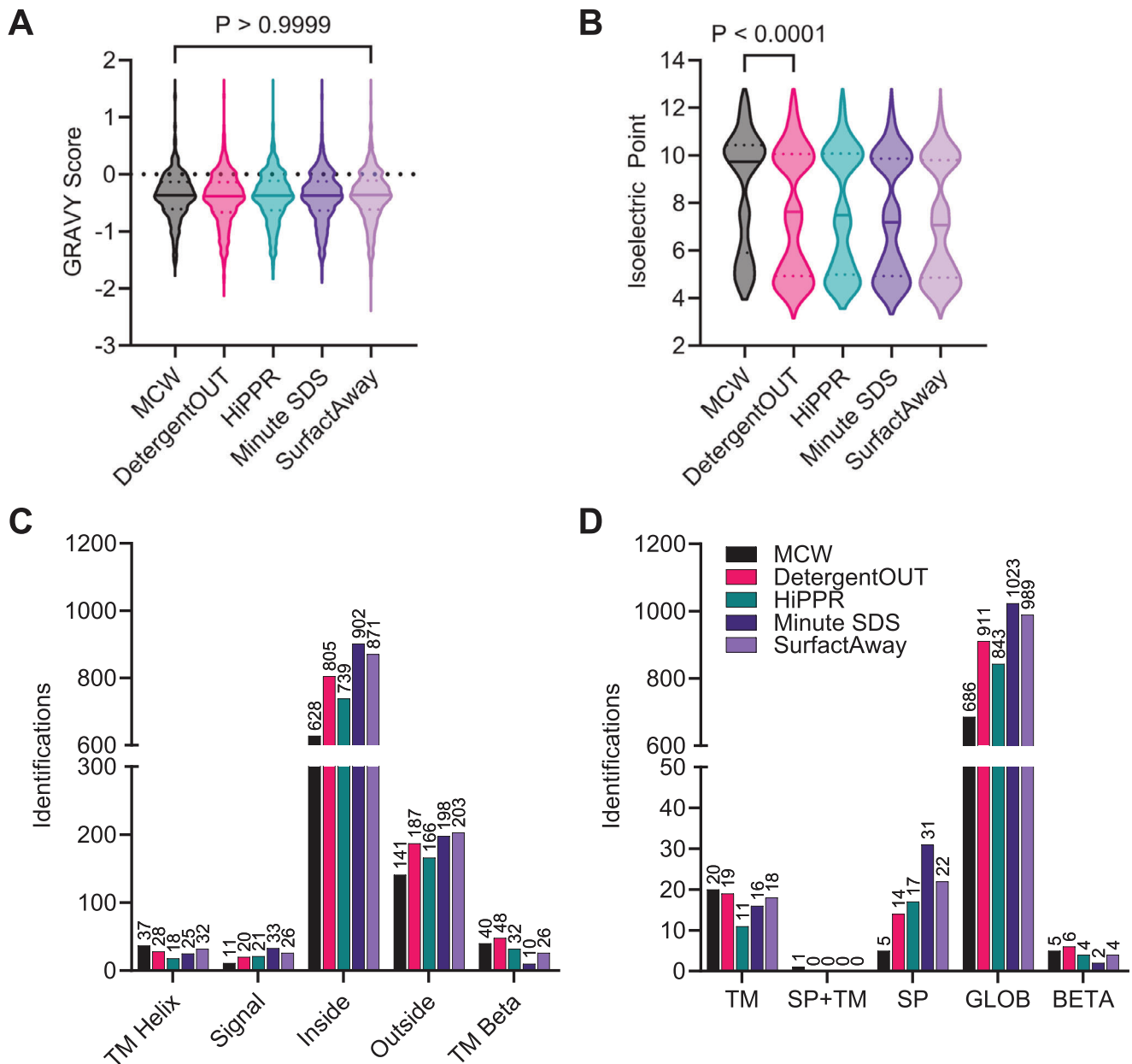


FIGURE 3 | Comparison of hydrophobicity, isoelectric, and topology characteristics of identified proteoforms from *E. coli* after five different detergent removal methods. (A) Comparison of Grand Average of Hydropathy (GRAVY) score distributions for proteoforms identified. (B) Comparison of isoelectric point (pI) distributions for proteoforms identified. In the truncated violin plots shown in (A) and (B), median is indicated by solid line, whereas quartiles are indicated by dotted lines. (C) Prediction of proteoform topology using DeepTMHMM: predicted domains identified in proteoforms from samples after detergent removal. TM Helix: α -helical transmembrane, Signal: signal peptide, Inside: inside cell/cytosolic, Outside: outside of cell/lumen of ER/Golgi/lysosome, and TM Beta: β -barrel transmembrane. (D) Prediction of proteoform topology using DeepTMHMM: topology types for proteoforms identified. BETA, beta-barrel transmembrane proteins; GLOB, globular; SP, signal peptide; SP+TM, transmembrane with signal peptide; TM, transmembrane.

(Figure 4B, Figure S8) [34]. However, in general, instances of lower PTM recovery relative to MCW for the filter-based methods were also observed for the precipitation methods; further, it should be noted that PTM profiles may not be generalizable across samples from different organisms. The commercial kits' reduced identification rates of biologically common PTMs imply that associated chemistries are inadvertently adsorbing some modified proteoforms during detergent removal.

To evaluate whether differences in clean-up protocols observed for the *E. coli* model also applied to human samples, we performed further experiments on the human Jurkat T-cell line. Here, again using the MCW method as a reference, we compared the performance of two commercial kits representing both column-based (HiPPR) and precipitation-based (Minute SDS) clean-up methods. Protein recovery for Jurkat 0–30 kDa fractions isolated via PEPPI was confirmed by SDS-PAGE (Figure S3B). As observed

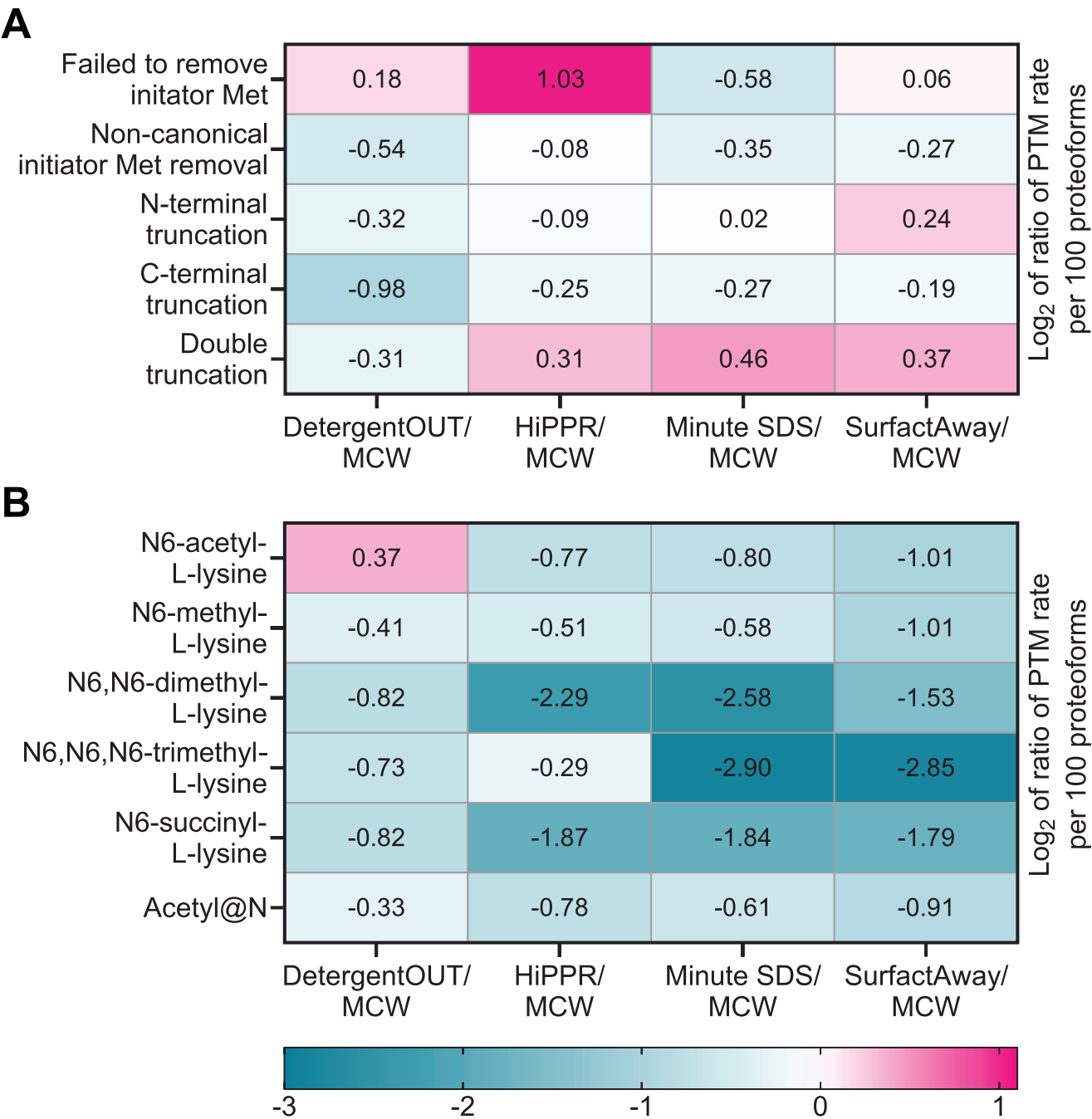


FIGURE 4 | Selected results from PTM analysis of identified *E. coli* proteoforms. (A) PTMs related to sequence changes. (B) Lysine modifications and N-terminal acetylation.

for the *E. coli* results, the number of identified UniProt accessions was relatively consistent for each clean-up method, with Minute SDS affording a higher number of identifications at both the UniProt Accession and proteoform level (Figure S9A). Similar to the *E. coli* experiments, recovery of small proteoforms (<6 kDa) from the human cell lysate samples was relatively less effective when using MCW (Figure S9B). Overlap in UniProt Accession and proteoform identifications observed for the three clean-up methods applied to Jurkat cell lysate are shown in Figures S9C and S9D, respectively.

As shown in Figure S10A, human cell lysate proteoforms identified for each of the three clean-up methods tested exhibit relatively similar GRAVY score distributions as observed for the *E. coli* results. However, for the human cell lysate experiments, the commercial methods exhibited a stronger bias toward recovery of more hydrophilic proteoforms compared to MCW (*p* value < 0.0001, Figure S10A; full descriptive statistics and statistical analysis available in Tables S12–S13). Consistent with the *E. coli* results, for the human cell lysate, commercial kits also exhibited a stronger bias toward recovery of proteoforms

with lower isoelectric points (p value < 0.0001, Figure S10B; full descriptive statistics and statistical analysis available in Tables S14–S15).

Considering predicted proteoform topology, as previously mentioned, the number of predicted transmembrane *E. coli* proteoforms was unexpectedly low for MCW. For the Jurkat experiments, however, MCW yielded a considerable increase in the number of predicted transmembrane proteoform identifications compared to the HiPPR and Minute SDS commercial methods (Figure S10C,D). A possible rationale is that the use of sonication to supplement the detergent lysis buffer during preparation of the *E. coli* lysate limited recovery of membrane proteoforms, which, along with the reduced complexity of the *E. coli* proteome, obscured the potential advantage of MCW in improving membrane proteoform recovery. For both *E. coli* and Jurkat cell lysates, results indicate that commercial kits afforded a higher number of globular proteoform identifications compared to MCW. Overall, general trends observed when comparing proteoforms identified using different clean-up methods were relatively consistent for both the prokaryotic and eukaryotic models evaluated.

3.3 | Practical Considerations

While MCW is the most labor-intensive protocol relative to the commercial kits, all clean-up methods have the potential to require approximately the same amount of time. MCW requires a drying time post clean-up, leaving a dry protein pellet that can be resuspended in the necessary volume of mobile phase A. MCW clean-up takes ~100 min, accounting for 5 washes and not including actual handling time. The commercial kits return samples that are, for the purposes of TDP, too dilute and require the use of a speed vacuum to concentrate to a smaller volume prior to dilution in mobile phase A. Speed vacuum drying time is variable based on humidity, starting volume, and the performance of the unit and may require from 45 min up to 1 h and 30 min to reduce the sample volume to 5–10 μ L, though the time required to clean samples with these kits is only 8 to 16 min.

However, MCW is still the most affordable option. The price per sample for the commercial kits ranges from \$6.42 to \$31.40 (Table 1), compared to \$1.34 for MCW when considering the PEPPi sample volumes used in this study (~90 μ L). Typical volumes for MCW reagents allow for processing of a large number of samples, whereas the DetergentOUT kit is the most limited in terms of the number of samples per kit, based on the number of supplied columns (\$8.40 per sample). Furthermore, the number of samples per kit for HiPPR is limited by the volume of resin slurry required based on PEPPi sample volume (cost up to \$31.40 per sample if omitting desalting for ~90 μ L PEPPi samples). However, typical PEPPi volumes are approximately 250 μ L, which increases the cost of reagents considerably for all commercial kits except for Minute SDS and SurfactAway (Table S1). The cost per sample of DetergentOUT increased to \$22.20. Assuming inclusion of a desalting step, the cost of which increases to \$12.60 for a 250 μ L sample, the cost of HiPPR, Minute SDS, and SurfactAway increased to \$44.00, \$13.10, and \$13.70 per sample, respectively. Cost may be reduced for larger sample volumes by using 4 mL HiPPR columns, as reported by Kline et al., yet the cost is still relatively high (\$33.00 per sample) [19].

Minute SDS and SurfactAway would be less expensive if the desalting step were omitted, but the methods cannot completely clean PEPPi samples without including such a step, such as using Zeba columns as in this study. Due to the excessive residual salt, PEPPi samples cleaned with Minute SDS resulted in lane broadening in the polyacrylamide gels run to confirm protein extraction (Figure S11). SurfactAway-cleaned samples retained the blue color from the Bromophenol Blue introduced during SDS-PAGE, thus also requiring the use of Zeba columns. Cost per 90 μ L sample thus increased by \$6.24 for Minute SDS and Surfact Away based on the cost of the Zeba column clean-up. It should also be noted that the additional step of desalting appears to reduce the amount of protein retained by Minute SDS (Figure S12).

As a supplementary data set, LC-MS/MS analysis of samples prepared with 2:1 sample to HiPPR resin ratio was performed to evaluate whether the cost of HiPPR per sample could be reduced. A desalting column was also used due to some retention of Bromophenol Blue post-clean-up, although to a lesser degree than SurfactAway and Minute SDS. Average residual SDS concentrations increased from 0.30 to 2.01 ppm (Figure S13A) but still remained lower than Minute SDS and SurfactAway. As with Minute SDS, the additional desalting step reduced the total amount of protein recovered post-clean-up (Figure S13B). Notably, total proteoform identifications increased to 1041 (Figure S13C), which appears to reflect increased recovery of proteoforms in the 2–10 kDa range (Figure S13D). Overlap in UniProt accession and proteoform identifications is shown in Figure S14, indicating 380 proteoforms were uniquely identified with inclusion of the additional desalting step (Figure S14B). This increase in the number of identified proteoforms indicates that the higher ratio of sample to settled resin is a cheaper alternative (\$18.80 per sample), even with the added cost of desalting. Interestingly, the 2:1 sample to settled resin HiPPR columns identified a greater number of proteoforms <10 kDa with more terminal truncations than the HiPPR columns with more resin (Table S16). Comparison of hydrophobicity (Figure S15A) and isoelectric point (Figure S15B) distributions of proteoforms unique to each ratio supports the idea that resin may retain certain lower molecular weight, acidic proteoforms. Though no significant difference in hydrophobicity was observed (Tables S17–S18, p value = 0.6893), proteoforms uniquely identified using the 1:2 ratio of sample to settled resin had a significantly higher median pI (9.04) compared to those uniquely identified for the 2:1 sample to resin ratio (7.04) (Tables S19–S20, p value < 0.0001).

4 | Conclusions

Results from the systematic comparison of four commercial SDS detergent removal kits with the standard MCW method indicate that commercial kits may be more effective in retaining lower molecular weight, more acidic proteoforms, thereby contributing to an overall higher number of unique UniProt accession and proteoform identifications, at least in the case of the 0–30 kDa *E. coli* fraction analyzed here. Further, results indicate that the resin-based clean-up methods, DetergentOUT and HiPPR, were similarly effective in SDS removal and comparable to MCW clean-up. However, to the extent residual SDS was present for samples cleaned with precipitation-based commercial kits, LC-MS/MS

TABLE 1 | Comparative cost of each clean-up method per sample based on sample PEPPI volumes used in the study (~90 µL).

Clean-up Method	MCW	DetergentOUT	HiPPR	Minute SDS	SurfactAway	Zeba Spin Desalting
Total price, reagents/kit	\$495.85	\$84.00	\$314.00	\$100.00	\$530.00	\$312.00
Number of samples per set of reagents/kit	370	10	10	555	1333	50
Price per Sample	\$1.34	\$8.40	\$31.40	\$0.18	\$0.40	\$6.24
Price per Sample with Desalting	—	—	\$18.80 ^{a)}	\$6.42	\$6.64	—
Sample Processing Time (min) ^{b)}	100	53-68	61-82	56-71	62-77	—
SDS Removal Efficiency	>99.999%	>99.99%	>99.99%	>99%	>99%	—
% Total Proteoform Identifications Observed in ≥2 Replicates	73.9%	62.2%	74.4%	70.3%	67.1%	—
Total Number of Proteoforms Identified	717	950	875	1072	1033	—
Ranked Ease of Use ^{c)}	5	2	4	1	3	—

^{a)}When reducing the sample to settled resin ratio to 2:1, the number of samples per kit increases to 25, lowering the cost even with the added desalting Zeba column.

^{b)}Variability in sample processing time is due to the drying step, as day-to-day humidity may greatly increase the time required for the speed vacuum. Zeba use, when needed, has been included in processing time.

^{c)}Ranked 1 to 5, 1 the easiest and 5 the most difficult based on active handling time and the number of steps involved.

results did not appear compromised in terms of the number of proteoform identifications. In terms of reproducibility, a higher percentage of total proteoform identifications was observed in at least two technical replicates for the HiPPR and MCW protocols. Practical considerations, including limitations in the number of samples that may be processed with each kit, handling time, and the expense of commercial kits, may make the use of MCW more favorable. The presented analysis indicates the standard MCW approach is competitive with commercial kits for the identification of proteoforms 6–30 kDa, while also yielding an increased prevalence of certain methylation and acetylation modifications in the set of identified proteoforms.

Acknowledgments

This work was supported by NIH R35GM147397 awarded to Luca Fornelli, a University of Oklahoma Honors College Undergraduate Research Opportunities Program (UROP) grant awarded to Kendalyn Firenza, and the University of Oklahoma Libraries' Open Access Fund.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All mass spectrometry data files have been uploaded to MassIVE (repository number MSV000097808).

References

1. bib id="bib1" type="Periodical">[1] D. Dersh, Y. Iwamoto, and Y. Argon, "Tay-Sachs Disease Mutations in HEXA Target the α Chain of

Hexosaminidase A to Endoplasmic Reticulum-Associated Degradation," *Molecular Biology of the Cell* 27 (2016): 3813–3827.

2. bib id="bib2" type="Periodical">[2] I. A. Prior, P. D. Lewis, and C. Mattos, "A Comprehensive Survey of Ras Mutations in Cancer," *Cancer Research* 72 (2012): 2457–2467.

3. J. R. Riordan, J. M. Rommens, B.-S. Kerem, et al., "Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA," *Science* 245 (1989): 1066–1073, <https://doi.org/10.1126/science.2475911>.

4. J. Zecha, W. Gabriel, R. Spallek, et al., "Linking Post-Translational Modifications and Protein Turnover by Site-Resolved Protein Turnover Profiling," *Nature Communications* 13 (2022): 165.

5. Q. Zhong, X. Xiao, Y. Qiu, et al., "Protein Posttranslational Modifications in Health and Diseases: Functions, Regulatory Mechanisms, and Therapeutic Implications," *MedComm* 4 (2023): e261.

6. H. Ryšlavá, V. Doubnerová, D. Kavan, and O. Vaněk, "Effect of Posttranslational Modifications on Enzyme Function and Assembly," *Journal of Proteomics* 92 (2013): 80–109.

7. L. M. Smith and N. L. Kelleher, "Proteoform: A Single Term Describing Protein Complexity," *Nature Methods* 10 (2013): 186–187, <https://doi.org/10.1038/nmeth.2369>.

8. L. M. Smith, J. N. Agar, J. Chamot-Rooke, et al., "The Human Proteoform Project: Defining the Human Proteome," *Science Advances* 7 (2021): eabk0734.

9. T. K. Toby, L. Fornelli, and N. L. Kelleher, "Progress in Top-Down Proteomics and the Analysis of Proteoforms," *Annual Review of Analytical Chemistry* 9 (2016): 499–519, <https://doi.org/10.1146/annurev-anchem-071015-041550>.

10. D. S. Roberts, J. A. Loo, Y. O. Tsybin, et al., "Top-Down Proteomics," *Nature Reviews Methods Primers* 4 (2024): 38.

11. L. C. Anderson, C. J. Dehart, N. K. Kaiser, et al., "Identification and Characterization of Human Proteoforms by Top-Down LC-21 Tesla FT-

- ICR Mass Spectrometry," *Journal of Proteome Research* 16 (2017): 1087–1096, <https://doi.org/10.1021/acs.jproteome.6b00696>.
12. K. R. Durbin, L. Fornelli, R. T. Fellers, P. F. Doubleday, M. Narita, and N. L. Kelleher, "Quantitation and Identification of Thousands of Human Proteoforms Below 30 kDa," *Journal of Proteome Research* 15 (2016): 976–982, <https://doi.org/10.1021/acs.jproteome.5b00997>.
13. L. Fornelli, K. R. Durbin, R. T. Fellers, et al., "Advancing Top-Down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer," *Journal of Proteome Research* 16 (2017): 609–618, <https://doi.org/10.1021/acs.jproteome.6b00698>.
14. P. T. Kaulich, L. Cassidy, K. Winkels, and A. Tholey, "Improved Identification of Proteoforms in Top-Down Proteomics Using FAIMS With Internal CV Stepping," *Analytical Chemistry* 94 (2022): 3600–3607, <https://doi.org/10.1021/acs.analchem.1c05123>.
15. E. N. Mccool, T. Xu, W. Chen, et al., "Deep Top-Down Proteomics Revealed Significant Proteoform-Level Differences Between Metastatic and Nonmetastatic Colorectal Cancer Cells," *Science Advances* 8 (2022): abq6348, <https://doi.org/10.1126/sciadv.abq6348>.
16. E. F. Bayne, K. J. Rossler, Z. R. Gregorich, et al., "Top-Down Proteomics of Myosin Light Chain Isoforms Define Chamber-Specific Expression in the human Heart," *Journal of Molecular and Cellular Cardiology* 181 (2023): 89–97, <https://doi.org/10.1016/j.yjmcc.2023.06.003>.
17. I. Ntai, L. Fornelli, C. J. Dehart, et al., "Precise Characterization of KRAS4b Proteoforms in Human Colorectal Cells and Tumors Reveals Mutation/Modification Cross-Talk," *Proceedings of the National Academy of Sciences* 115 (2018): 4140–4145, <https://doi.org/10.1073/pnas.1716122115>.
18. B. S. Drown, K. Jooß, R. D. Melani, C. Lloyd-Jones, J. M. Camarillo, and N. L. Kelleher, "Mapping the Proteoform Landscape of Five Human Tissues," *Journal of Proteome Research* 21 (2022): 1299–1310, <https://doi.org/10.1021/acs.jproteome.2c00034>.
19. J. T. Kline, M. W. Belford, C. L. Boeser, et al., "Orbitrap Mass Spectrometry and High-Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) Enable the In-Depth Analysis of Human Serum Proteoforms," *Journal of Proteome Research* 22 (2023): 3418–3426, <https://doi.org/10.1021/acs.jproteome.3c00488>.
20. R. D. Melani, V. R. Gerbasi, L. C. Anderson, et al., "The Blood Proteoform Atlas: A Reference Map of Proteoforms in human Hematopoietic Cells," *Science* 375 (2022): 411–418, <https://doi.org/10.1126/science.aaz5284>.
21. R. Tabatabaieian Nimavard, S. A. Sadeghi, M. Mahmoudi, G. Zhu, and L. Sun, "Top-Down Proteomic Profiling of Protein Corona by High-Throughput Capillary Isoelectric Focusing-Mass Spectrometry," *Journal of the American Society for Mass Spectrometry* 36 (2025): 778–786, <https://doi.org/10.1021/jasms.4c00463>.
22. C.-F. Huang, M. A. Hollas, A. Sanchez, et al., "Deep Profiling of Plasma Proteoforms With Engineered Nanoparticles for Top-Down Proteomics," *Journal of Proteome Research* 23 (2024): 4694–4703, <https://doi.org/10.1021/acs.jproteome.4c00621>.
23. L. Fornelli and T. K. Toby, "Characterization of Large Intact Protein Ions by Mass Spectrometry: What Directions Should We Follow?," *Biochimica Et Biophysica Acta (BBA)—Proteins and Proteomics* 1870 (2022): 140758, <https://doi.org/10.1016/j.bbapap.2022.140758>.
24. R. Huguet, C. Mullen, K. Szrentic, et al., "Proton Transfer Charge Reduction Enables High-Throughput Top-Down Analysis of Large Proteoforms," *Analytical Chemistry* 91 (2019): 15732–15739, <https://doi.org/10.1021/acs.analchem.9b03925>.
25. P. D. Compton, L. Zamdborg, P. M. Thomas, and N. L. Kelleher, "On the Scalability and Requirements of Whole Protein Mass Spectrometry," *Analytical Chemistry* 83 (2011): 6868–6874, <https://doi.org/10.1021/ac2010795>.
26. L. Fornelli, K. Szrentic, R. Huguet, et al., "Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques," *Analytical Chemistry* 90 (2018): 8421–8429, <https://doi.org/10.1021/acs.analchem.8b00984>.
27. T. Tucholski, S. J. Knott, B. Chen, P. Pistono, Z. Lin, and Y. Ge, "A Top-Down Proteomics Platform Coupling Serial Size Exclusion Chromatography and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry," *Analytical Chemistry* 91 (2019): 3835–3844, <https://doi.org/10.1021/acs.analchem.8b04082>.
28. J. C. Tran and A. A. Doucette, "Gel-Eluted Liquid Fraction Entrapment Electrophoresis: An Electrophoretic Method for Broad Molecular Weight Range Proteome Separation," *Analytical Chemistry* 80 (2008): 1568–1573, <https://doi.org/10.1021/ac702197w>.
29. A. Takemori, D. S. Butcher, V. M. Harman, et al., "PEPPI-MS: Polyacrylamide-Gel-Based Prefractionation for Analysis of Intact Proteoforms and Protein Complexes by Mass Spectrometry," *Journal of Proteome Research* 19 (2020): 3779–3791, <https://doi.org/10.1021/acs.jproteome.0c00303>.
30. A. Takemori, P. T. Kaulich, L. Cassidy, N. Takemori, and A. Tholey, "Size-Based Proteome Fractionation Through Polyacrylamide Gel Electrophoresis Combined With LC-FAIMS-MS for in-Depth Top-Down Proteomics," *Analytical Chemistry* 94 (2022): 12815–12821, <https://doi.org/10.1021/acs.analchem.2c02777>.
31. R. R. O. Loo, N. Dales, and P. C. Andrews, "Surfactant Effects on Protein Structure Examined by Electrospray Ionization Mass Spectrometry," *Protein Science* 3 (1994): 1975–1983, <https://doi.org/10.1002/pro.5560031109>.
32. D. Wessel and U. I. Flügge, "A Method for the Quantitative Recovery of Protein in Dilute Solution in the Presence of Detergents and Lipids," *Analytical Biochemistry* 138 (1984): 141–143, [https://doi.org/10.1016/0003-2697\(84\)90782-6](https://doi.org/10.1016/0003-2697(84)90782-6).
33. M. Z. Güray, S. Zheng, and A. A. Doucette, "Mass Spectrometry of Intact Proteins Reveals +98 u Chemical Artifacts Following Precipitation in Acetone," *Journal of Proteome Research* 16 (2017): 889–897.
34. A. Takemori, N. Sugiyama, J. T. Kline, L. Fornelli, and N. Takemori, "Gel-Based Sample Fractionation With SP3-Purification for Top-Down Proteomics," *Journal of Proteome Research* 24 (2025): 850–860, <https://doi.org/10.1021/acs.jproteome.4c00941>.
35. J. Rappsilber, M. Mann, and Y. Ishihama, "Protocol for Micro-Purification, Enrichment, Pre-Fractionation and Storage of Peptides for Proteomics Using StageTips," *Nature Protocols* 2 (2007): 1896–1906, <https://doi.org/10.1038/nprot.2007.261>.
36. A. Takemori, Y. Kawashima, and N. Takemori, "Bottom-up/Cross-Linking Mass Spectrometry via Simplified Sample Processing on Anion-Exchange Solid-Phase Extraction Spin Column," *Chemical Communications* 58 (2022): 775–778, <https://doi.org/10.1039/D1CC05529A>.
37. D. P. Donnelly, C. M. Rawlins, C. J. Dehart, et al., "Best Practices and Benchmarks for Intact Protein Analysis for Top-Down Mass Spectrometry," *Nature Methods* 16 (2019): 587–594, <https://doi.org/10.1038/s41592-019-0457-0>.
38. T. Habeck, K. A. Brown, B. Des Soye, et al., "Top-Down Mass Spectrometry of Native Proteoforms and Their Complexes: A Community Study," *Nature Methods* 21 (2024): 2388–2396, <https://doi.org/10.1038/s41592-024-02279-6>.
39. K. Szrentić, L. Fornelli, Y. O. Tsybin, et al., "Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry," *Journal of the American Society for Mass Spectrometry* 31 (2020): 1783–1802.
40. W. Cao, P. Y. Chu, M. L. Bruening, R. L. Shi, H. Hernandez-Barry, and J. C. Tran, "Efficient, Low-Cost, and High-Throughput Sodium Dodecyl Sulfate (SDS) Removal From Protein Digests Using Weak-Anion Exchange," *Journal of the American Society for Mass Spectrometry* 36 (2025): 680–687, <https://doi.org/10.1021/jasms.4c00304>.
41. J. T. Kline, M. W. Belford, J. Huang, et al., "Improved Label-Free Quantification of Intact Proteoforms Using Field Asymmetric Ion Mobil-

ity Spectrometry,” *Analytical Chemistry* 95 (2023): 9090–9096, <https://doi.org/10.1021/acs.analchem.3c01534>.

42. J. M. Fulcher, A. Makaju, R. J. Moore, et al., “Enhancing Top-Down Proteomics of Brain Tissue With FAIMS,” *Journal of Proteome Research* 20 (2021): 2780–2795, <https://doi.org/10.1021/acs.jproteome.1c00049>.

43. A. A. Goloborodko, L. I. Levitsky, M. V. Ivanov, and M. V. Gorshkov, “Pyteomics—A Python Framework for Exploratory Data Analysis and Rapid Software Prototyping in Proteomics,” *Journal of the American Society for Mass Spectrometry* 24 (2013): 301–304, <https://doi.org/10.1007/s13361-012-0516-6>.

44. L. I. Levitsky, J. A. Klein, M. V. Ivanov, and M. V. Gorshkov, “Pyteomics 4.0: Five Years of Development of a Python Proteomics Framework,” *Journal of Proteome Research* 18 (2019): 709–714, <https://doi.org/10.1021/acs.jproteome.8b00717>.

45. J. Kyte and R. F. Doolittle, “A Simple Method for Displaying the Hydropathic Character of a Protein,” *Journal of Molecular Biology* 157 (1982): 105–132, [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0).

46. P. T. Kaulich, K. Jeong, O. Kohlbacher, and A. Tholey, “Influence of Different Sample Preparation Approaches on Proteoform Identification by Top-Down Proteomics,” *Nature Methods* 21 (2024): 2397–2407, <https://doi.org/10.1038/s41592-024-02481-6>.

47. J. Hallgren, K. D. Tsigirgos, M. D. Pedersen, et al., “DeepTMHMM Predicts Alpha and Beta Transmembrane Proteins Using Deep Neural Networks,” *BioRxiv* (2022).

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File 1: pmic70067-sup-0001-Tables.xlsx.

Supporting File 2: pmic70067-sup-0002-SuppMat.pdf.