

In Search of a Universal Method: A Comparative Survey of Bottom-Up Proteomics Sample Preparation Methods

Gina Varnavides, Moritz Madern, Dorothea Anrather, Natascha Hartl, Wolfgang Reiter, and Markus Hartl*



Cite This: *J. Proteome Res.* 2022, 21, 2397–2411



Read Online

ACCESS |

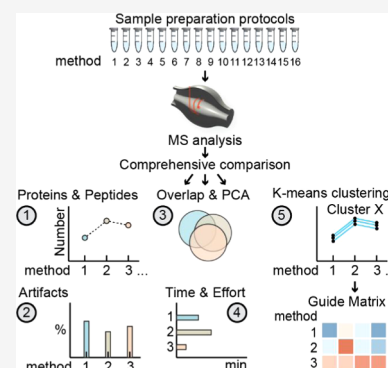
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Robust, efficient, and reproducible protein extraction and sample processing is a key step for bottom-up proteomics analyses. While many sample preparation protocols for mass spectrometry have been described, selecting an appropriate method remains challenging since some protein classes may require specialized solubilization, precipitation, and digestion procedures. Here, we present a comprehensive comparison of the 16 most widely used sample preparation methods, covering in-solution digests, device-based methods, and commercially available kits. We find a remarkably good performance of the majority of the protocols with high reproducibility, little method dependency, and low levels of artifact formation. However, we revealed method-dependent differences in the recovery of specific protein features, which we summarized in a descriptive guide matrix. Our work thereby provides a solid basis for the selection of MS sample preparation strategies for a given proteomics project.

KEYWORDS: proteomics, sample preparation, in-solution digest, SPEED, FASP, iST, S-Trap, SP3, EasyPep, mass spectrometry



INTRODUCTION

State-of-the-art mass-spectrometry-based proteomics workflows are sophisticated multistep processes, combining different methodologies and instrumentation. Clearly, the data quality of an experiment depends on the characteristics and limitations of each step, with errors or biases propagating from the first step throughout the whole experiment. For this reason, the sample preparation protocol is a key determinant in defining what proportion of the proteome is available for the ensuing analysis. Moreover, the robustness and reproducibility of this step will define the degree of data variation and potential systematic bias. Ideally, the universal sample preparation protocol would efficiently and robustly isolate all proteins of any given sample to near completeness. In reality, such comprehensive isolation is very challenging as proteins constitute a heterogeneous group of macromolecules in terms of physicochemical properties and subcellular localization. In addition, other sample characteristics, such as rigid cell walls and tissues that are difficult to lyse or interfering cellular components (e.g., nucleic acids, metabolites, etc.), can greatly affect isolation efficacy and analysis and need to be addressed.

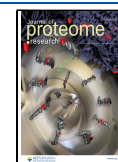
To solve these problems, different sample preparation methods have been developed that can be divided into in-solution digestion methods and methods using additional devices such as filters or beads for protein immobilization or purification, or both. Classical in-solution digestion (ISD) protocols essentially differ in the choice of buffer systems, which are either based on chaotropic denaturants, such as urea

or guanidine hydrochloride (GnHCl), or surfactants, such as ionic detergent sodium-dodecyl-sulfate (SDS) or bile salt sodium deoxycholate (SDC), as they effectively solubilize and denature proteins.^{1–3} Recently, a novel ISD strategy, Sample Preparation by Easy Extraction and Digestion (SPEED), has been published⁴ that uses neither detergents nor chaotropic agents for protein extraction but solely relies on dissolving proteins in trifluoroacetic acid (TFA). Further, ISD protocols often require protein precipitation using either acetone,^{5–7} alcohols such as ethanol,⁸ or chloroform/methanol⁹ to avoid carry-over of nonprotein components that might interfere with downstream processing or analysis.

Device-based approaches (hereafter referred to as “cleanup methods”) aim to remove interfering substances before digestion in “reactors” or on beads. For example, Filter-Aided Sample Preparation (FASP),¹⁰ utilizing molecular weight cutoff (MWCO) membranes, and suspension trapping (S-Trap),¹¹ applying three-dimensional porous quartz filter materials, capture proteins on filters enabling detergent removal, protein digestion, and peptide recovery. Single-pot, solid-phase-enhanced sample preparation (SP3)¹² (and also

Received: May 8, 2022

Published: August 25, 2022



SP4),¹³ uses on-bead-based purification and digestion of proteins in a single tube, exploiting the property of denatured proteins to be nonspecifically immobilized on microparticles by protein aggregation.¹⁴ Finally, the original in-StageTip (iST) method utilizes C18 discs prepared in pipette tips or cartridges to trap proteins for digestion and subsequently to desalt the peptides.¹⁵ Based on these and similar methodical concepts, commercially available MS sample preparation kits in different formats have been developed for iST (PreOmics), S-Trap (ProtiFi), and in-solution digests coupled to peptide cleanup columns (EasyPep, Thermo Scientific).

Overall, this almost overwhelming number of protocols and variants with their apparent advantages and disadvantages make the selection of a suitable method for a given project difficult. Although previous studies compared selected sets of protocols, often focusing on particular aspects or on presenting a new method, a comparison including the most commonly used in-solution, device-based, and commercial methods had yet to be conducted.^{1,3,4,14,16–20} It is also debatable whether there is a truly universal method that exhibits no or negligible extraction bias, as has been proposed for some protocols,^{4,19} and that is applicable to all types of samples. Proving universality is an almost futile task, as it would require the comparison of a set of methods for a virtually endless list of cell types, tissues, body fluids, and organisms. However, Glatter et al.¹ and Doellinger et al.⁴ convincingly demonstrated for a selection of ISD protocols and device-based protocols that there are organism- and buffer-specific differences in extraction efficiency when comparing samples of different bacterial and human cell lines. From these studies, it can be expected that such differences will further increase when comparing even more diverse sets of sample types, e.g., mammalian tissues, plants, or fungi. In contrast, investigating differences in proteome composition for a given set of methods in a defined sample type is more feasible and allows to answer whether and how protocols differ in their extraction properties for the given sample type. In combination with more practical considerations, like processing time, ease of use, and consumable costs, this could help in making a more informed decision for a particular sample preparation strategy and serve as a blueprint for similar studies in other sample types.

Here, we prepared proteomes from HeLa cells applying classical ISD protocols based on urea-, GnHCl-, and SDC-based buffer systems as well as SPEED,⁴ FASP,¹⁰ S-Trap^{11,21} (ProtiFi), and SP3¹² protocols and two commercial kits: iST¹⁵ (PreOmics) and EasyPep (Thermo Scientific). We therefore present a comprehensive quantitative and qualitative comparison of 16 of the most widely used MS sample preparation methods. Our experimental design maximizes reproducibility and comparability and allows for unbiased statistical analyses to extract differences between the methods. The individual methods show a similar proteome extraction efficacy and coverage based on identified proteins and peptides. Method-induced peptide artifacts seem to be negligible. However, an exploratory analysis based on *k*-means clustering revealed qualitative differences in extracted proteomes, which we mapped to features derived from protein databases. The results were summarized into a descriptive guide matrix that highlights specific enrichment of protein features such as structure, abundance, and localization for individual methods. Consequently, our study provides a solid comparison of the currently most widely used sample preparation protocols in

proteomics and can be used as an aid in selecting MS sample preparation strategies.

■ MATERIALS AND METHODS

Human Cell Culture

HeLa cells were cultivated in Dulbecco's modified Eagle's medium (DMEM 4.5 g/L glucose) (Sigma-Aldrich) supplemented with 10% fetal bovine serum (FCS) (Sigma-Aldrich), 1% L-glutamine (Sigma-Aldrich), and 1% penicillin–streptomycin (Sigma-Aldrich) in 15 cm dishes under 5% CO₂ at 37 °C. Cells were harvested at ~80% confluency by 5 min treatment with trypsin (Sigma-Aldrich) at 37 °C, followed by a 1:1 dilution with full media to stop the digest. Cells were pelleted by centrifugation (5 min at 500g, 23 °C) and washed with 1× phosphate-buffered saline (PBS). Aliquots of 2.0E6 cells were subsequently snap-frozen in liquid N₂ and kept at –80 °C until lysis.

In-Solution Protocols

In-Solution Digests. HeLa cells (2.0E6 cells) were dissolved in 100 μL of denaturation buffer, 0.1 M Tris–HCl, pH 8.6, containing either 8 M urea (U), 6 M guanidine HCl (GnHCl), or 1% sodium deoxycholate (SDC), incubated for 10 min at room temperature (U) or at 60 °C (GnHCl, SDC) in a ThermoMixer (Eppendorf), and subsequently disrupted by 2 × 20" high-intensity sonication cycles at 4 °C in a BioRuptor (Diagenode). Protein concentration was determined using the Micro BCA protein assay kit according to the manufacturer's instructions (Thermo Scientific). Each sample was split into two aliquots of 100 μg protein and one additional aliquot of 50 μg. Protein fractions of the two 100 μg aliquots were precipitated using acetone or chloroform–methanol, respectively. Only samples containing GnHCl were precipitated with ethanol instead of acetone since GnHCl is not soluble in the latter. Protein pellets were dissolved in their respective denaturation buffer, and protein concentration was determined as described above. Soluble proteins were reduced using 10 mM dithiothreitol (DTT) for 1 h at 37 °C (U) or 60 °C (SDC, GnHCl) and alkylated for 30 min using 20 mM iodoacetamide (IAA) in the dark. Chaotropic lysis buffers were then diluted to a final concentration of 1 M (urea) and 0.5 M (GnHCl). Proteins were digested overnight at 37 °C using trypsin (Trypsin Gold, Promega) in a 1:30 (w/w) enzyme-to-protein ratio. Digests were stopped by adding 10% trifluoroacetic acid (TFA) to a final concentration of 1%. SDC precipitates were removed by centrifugation (14,000g, 1 min, room temperature (RT)). About 10 μg of resulting peptide samples was desalted on C18 StageTips (triple-plugs)²² and eluted with 80% acetonitrile (ACN) and 0.1% trifluoroacetic acid (TFA). After removal of elution buffer by vacuum centrifugation, samples were resuspended in 0.1% TFA, 2% ACN.

Sample Preparation by Easy Extraction and Digestion (SPEED).⁴ A total of 2.0E6 HeLa cells were resuspended in trifluoroacetic acid (TFA) (Merck) in a sample-to-TFA ratio of 1:4 (v/v), incubated at room temperature for 5 min, and neutralized with 2 M Tris base using 8× volume of TFA used for lysis. Reduction and alkylation of aliquots of 50 μg of protein were achieved by incubation in 10 mM tris(2-carboxyethyl)phosphine (TCEP) and 40 mM 2-chloroacetamide (CAA) at 95 °C for 5 min. Samples were diluted with ddH₂O 1:5, and proteins were digested for 20 h at 37 °C using trypsin (Trypsin Gold, Promega) at an enzyme/protein ratio

of 1:50, as suggested in the original protocol. The digestion was stopped using 2% TFA (final concentration), and peptides were desalted on C18 StageTips and eluted with 80% acetonitrile (ACN) and 0.1% trifluoroacetic acid (TFA). Dried samples were resuspended in 0.1% TFA, 2% ACN.

Device-Based or Cleanup Protocols

Filter-Aided Sample Preparation (FASP).¹⁰ A total of 2.0E6 HeLa cells were resuspended in SDT lysis buffer (4% SDS, 100 mM Tris-HCl, 100 mM DTT, pH 7.6) in a 1:10 (v/v) sample/buffer ratio, incubated at 95 °C for 5 min, and sonicated at 4 °C for two cycles of 20 s at a high-intensity level using a BioRuptor (Diagenode). Samples were clarified by centrifugation at 16,000g for 15 min, at 24 °C. Aliquots of 50 μ g of protein were diluted in urea buffer UA (8 M urea, 0.1 M Tris-HCl, pH 8.5) to a final concentration of 0.5% SDS. Protein extracts were further processed in Microcon 30 kDa Centrifugal Filter Units (Merck) in a tempered centrifuge at 24 °C. Samples were added to the filter unit, washed with UA buffer, centrifuged for 15 min at 14,000g, and incubated with 50 mM IAA for 20 min at room temperature (in the dark). SDS was exchanged by four consecutive washes with UA buffer (centrifugation: 15 min at 14,000g) and a single wash with 50 mM ammonium bicarbonate (ABC) followed by centrifugation for 5–10 min at 14,000g. Proteins were digested using trypsin (Trypsin Gold, Promega) in a 1:50 protein-to-enzyme ratio and incubated for 18 h at 37 °C on a thermoshaker at 600 rpm. The resulting peptides were recovered by centrifugation at 14,000g for 5 min, followed by elution with 50 μ L of 50 mM ABC and repeated centrifugation. Combined eluates were acidified using TFA at a final concentration of 1%.

In-StageTip Sample Preparation (iST). HeLa cell extracts were prepared using the iST 96x sample kit according to the manufacturer's instructions (PreOmics). In short, 2.0E6 cells were lysed by resuspension in a lysis buffer solution at a target protein concentration of 1 mg/mL and heated to 95 °C for 10 min shaking (1000 rpm) followed by two cycles of 20 s of sonication in a BioRuptor (Diagenode). Aliquots containing 50 μ g of protein were transferred into a cartridge and cooled. The digestion solution was added, and proteins were digested for 3 h at 37 °C. Digestion was stopped by adding the "Stop" solution, and peptide purification was achieved by centrifugation for 3 min at 2250g, followed by three rounds of washing and elution into the collection plate using the provided solutions. Peptides were transferred to PCR tubes, dried in a vacuum centrifuge, and resuspended in 0.1% TFA, 2% ACN for MS analysis.

EasyPep. HeLa cell extracts were prepared using the EasyPep Mini MS Sample Prep Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Briefly, 2.0E6 cells were lysed with lysis buffer aiming for a protein concentration of 1 mg/mL, and aliquots containing 50 μ g of protein were treated with Universal nuclease by ten cycles of pipetting up and down until the viscosity was reduced. Reduction and alkylation were achieved by addition of the respective solutions and incubation of samples at 95 °C for 10 min. Once samples were cooled down, the trypsin/Lys-C protease mixture was added and samples were digested for 3 h at 37 °C. Tryptic digestion was stopped using the "Digestion Stop Solution". Peptide Cleanup columns were cleared from the liquid by centrifugation and placed onto 2 mL microcentrifuge tubes. Sample mixtures were transferred into dry Peptide Cleanup columns. Two rounds of consecutive

centrifugation and washing steps were performed. The columns were transferred to 2 mL microcentrifuge tubes, and peptides were eluted by addition of the elution solution and centrifugation at 1500g for 2 min. Samples were dried using a vacuum centrifuge and resuspended in 0.1% TFA, 2% ACN for MS analysis.

Suspension Trapping (S-Trap).¹¹ A total of 2.0E6 HeLa cells were resuspended in lysis buffer LB (10% SDS (w/v), 0.1 M Tris-H₃PO₄, pH 7.55). Cells were disrupted by sonication (two cycles of 20 s at 4 °C) in a BioRuptor (Diagenode), and extracts were cleared by centrifugation at 15,000g for 1 min at 4 °C. Aliquots of 50 μ g of protein were reduced by incubation with 20 mM (final concentration) DTT at 95 °C for 10 min and subsequently alkylated by addition of 40 mM (final concentration) IAA and incubation for 30 min in the dark at room temperature. Samples were acidified with 1.2% (final concentration) phosphoric acid, mixed with a 6 \times volume of S-Trap binding buffer (90% MeOH in 0.1 M Tris-H₃PO₄, pH 7.1), and loaded onto S-Trap columns that were placed in low binding tubes (Axygen). The solvent was removed by centrifugation (4000g), and proteins were washed three times with 150 μ L of S-Trap binding buffer, subsequently digested by addition of digestion buffer (500 mM ABC) containing 1:25 (w/w) trypsin (Trypsin Gold, Promega), and incubated at 37 °C for 3 h. Peptides were eluted in three consecutive steps by addition of 40 μ L of 50 mM ABC, 40 μ L of 0.2% FA, and 35 μ L of 50% ACN, 0.2% FA followed by centrifugation at 4000g, respectively. Eluates were pooled and concentrated in a SpeedVac (Thermo Fisher Scientific). Peptides were resolved in 0.1% TFA, 2% ACN. Aliquots of 10 μ g of peptides were desalted on C18 StageTips (triple-plugs),²² dried in a SpeedVac, and resuspended in 0.1% TFA, 2% ACN.

Single-Pot, Solid-Phase-Enhanced Sample Preparation (SP3).¹² HeLa cells (2.0E6) were resolved in reconstitution buffer (RB)¹² or 1% SDC to a final protein concentration of 1 mg/mL and subsequently lysed, reduced (DTT 5 mM contained in RB), and alkylated using IAA (25 mM final concentration). For protein cleanup and digestion, samples of 50 μ g of protein were first mixed with SP3 beads in a 10:1 (w/w) beads-to-protein ratio. The mixture was then homogenized by adding 1 \times volume of 100% EtOH and incubated for 5 min at 24 °C shaking at 1000 rpm to induce protein binding to the beads. Proteins bound to beads were washed 4 \times with 80% EtOH on a magnetic rack. On-bead digestion was achieved using trypsin (Trypsin Gold, Promega), added in a 1:30 (w/w) enzyme-to-protein ratio, and 20 h incubation at 37 °C in a thermal shaker (1000 rpm). After digestion, beads were pelleted by centrifugation (20,000g, 1 min, 24 °C) and supernatants containing peptides were transferred.

Experimental Design and Quality Control

To enable statistical analysis, we prepared three replicates of equal peptide concentration of each sample preparation method and applied several quality control steps that are summarized in detail below.

Type of Replicates. Starting from a commonly cultured pool of HeLa cells, three independent replicates were prepared for each sample preparation method. These replicates were defined as technical replicates.

Determination of Protein Concentration (of ISD Samples). The protein concentration after cell lysis and

after protein precipitation was determined using the Micro BCA Protein assay kit (Thermo Scientific) according to the manufacturer's guidelines.

Determination of Peptide Concentration. An estimate of 250 ng of peptide per sample was mixed in 0.1% TFA, 2% ACN. Peptide concentrations were determined and adjusted according to UV chromatograms obtained at 214 nm on an UltiMate 3000 RSLC nano-HPLC System (Thermo Scientific), equipped with a monolithic column (PepSwift Monolithic RSLC, Thermo Scientific). To adjust the peptide concentration for MS measurements, peaks were integrated using chromatography software Chromeleon (Thermo Scientific) and peak areas were compared to in-house peptide standards of known concentrations.

Equal Loading of Samples. All samples were adjusted to an estimated concentration of 100 ng/ μ L. The indexed retention time standard (iRT, Biognosys) was added to all samples to a final concentration of 0.1 injection equivalents (IE)/ μ L, allowing continuous monitoring of LC-MS/MS performance. Five microliters of each sample corresponding to 500 ng of peptide with 0.5 IE were subjected to MS analysis. Equal loading of samples was confirmed by checking total summed peptide intensities.

Organization of Batches. Samples were organized into six batches. Batches 1–3 covered ISD protocols (including SPEED), with one replicate of each method per batch. Batches 4–6 were equally organized but with cleanup methods. Samples were measured in a randomized order, and all measurements were separated by wash runs. Before and after each batch, 25 ng of HeLa standard (Pierce) was injected to control system performance. Batches 1–3 and batches 4–6 were run in a single sequence.

Postacquisition QC. The quality of LC-MS runs was continuously monitored by checking the iRT signals in Skyline v20.1.²³ The number of missed cleavages and other metrics of quality control were determined using PTXQC.²⁴

Bridging of Batches. To account for changes in machine performance between batch sequences 1–3 and 4–6, three replicates of each group of batches (SDC-A and EasyPep, respectively) were remeasured in a single sequence of MS measurements. The differences in the number of IDs between these groups were \sim 1%; nevertheless, the number of IDs of all original sample measurements was readjusted by the relative median change factor of the bridge samples. In short, the relative median change factor between the two groups was determined as $[\text{median}(\text{"SDC-A_bridgesamples"}) - \text{median}(\text{"EasyPep_bridgesamples"})] / \text{median}(\text{"EasyPep_bridgesamples"})$. The corrected SDC-A median was calculated as $[\text{group_median}(\text{"EasyPep_samples"}) + \text{group_median}(\text{"EasyPep_samples"}) * \text{relative_median_change_factor}]$. ISD groups were adjusted to the corrected SDC-A group median by their relative change to the original SDC-A group median.

MS Methods

LC-MS/MS analysis was performed on an UltiMate 3000 RSLC nano-HPLC System (Thermo Scientific), containing both a trapping column for peptide concentration (PepMap C18, $5 \times 0.3 \text{ mm}^2$, $5 \mu\text{m}$ particle size) and an analytical column (PepMap C18, $500 \times 0.075 \text{ mm}^2$, $2 \mu\text{m}$ particle size (Thermo Scientific)), coupled to a Q Exactive HF-X Orbitrap (with HCD, higher-energy collisional dissociation mode) mass spectrometer via a Proxeon nanospray flex ion source (all from Thermo Scientific). For peptide chromatography, the concen-

tration of organic solvent (acetonitrile) was increased linearly over 2 h from 1.6 to 28% in 0.1% formic acid at a flow rate of 230 nL/min. For acquisition of MS2 spectra, the instrument was operated in data-dependent mode with dynamic exclusion enabled. The scan sequence began with an Orbitrap MS1 spectrum with the following parameters: resolution, 120,000; scan range, 375–1500 m/z ; automatic gain control (AGC) target, 3×10^6 ; and maximum injection time (IT), 60 ms. The top 20 precursors were selected for MS2 analysis (HCD) with the following parameters: resolution, 15,000; AGC, 1×10^5 ; maximum, IT 54 ms; isolation window, 1.2 m/z ; scan range, 200–2000 m/z ; and normalized collision energy (NCE), 28. The minimum AGC target was set at 1×10^4 , which corresponds to a 1.9×10^5 intensity threshold. Peptide match was set to "preferred". In addition, unassigned, singly, and >6 -charged species and isotopes were excluded from MS2 analysis, and dynamic exclusion was set to 40 s.

MaxQuant Settings

Raw MS data was analyzed using MaxQuant²⁵ software version 1.6.14.0. MS2 spectra were searched against the canonical *Homo sapiens* (human) UniProt database (UP000005640_9606.fasta, release 2020_01, www.uniprot.org) containing 20607 entries, concatenated with the sequences of 397 common laboratory contaminants (extended MaxQuant contaminants database) and the iRT. Enzyme specificity was set to "Trypsin/P", the minimal peptide length was set to 7, and the maximum number of missed cleavages was set to 2. A maximum of five modifications per peptide were allowed. Carbamidomethylation of cysteine was searched as a fixed modification. "Acetyl (Protein N-term)" and "Oxidation (M)" were set as variable modifications. "Match between runs" and LFQ were activated. Results were filtered at a false discovery rate of 1% at the protein and peptide spectrum match level.

FragPipe Analysis

Screening for protein modifications in an unbiased manner was performed using the open search option of MSFragger 3.3 in FragPipe (v16.0).²⁶ All raw files were converted to the mzML format using MSConvert²⁷ with peak picking activated. mzML files were assigned according to sample preparation methods and replicates in the Experiments/Group tab. Default open search parameters were used, with trypsin specificity, -150 to $+500$ Da precursor mass window, oxidation of methionine, and protein N-terminal acetylation as variable modifications and carbamidomethylation of cysteine as fixed modification. PTM-Shepherd was activated at default settings. The observed mass shifts were obtained from the "global.modsummary.tsv" and "global.profile.tsv" tables in the FragPipe output, inspected, and filtered for abundant and relevant modifications.

Computational Methods

Computational analyses were performed using in-house R-scripts (ref 28 and [Supplemental Material Scripts](#)). The data was processed as follows: Proteins only identified by a modified peptide, contaminant proteins as well as protein groups with less than two razor and unique peptides were removed, and LFQ intensities were \log_2 -transformed. Only IDs identified by MS/MS were considered. The data was filtered based on valid values in LFQ intensities with a cutoff of three valid values in at least one group. The remaining missing values were imputed by a constant equal to the minimal \log_2 LFQ intensity across all samples (rounded down to the next

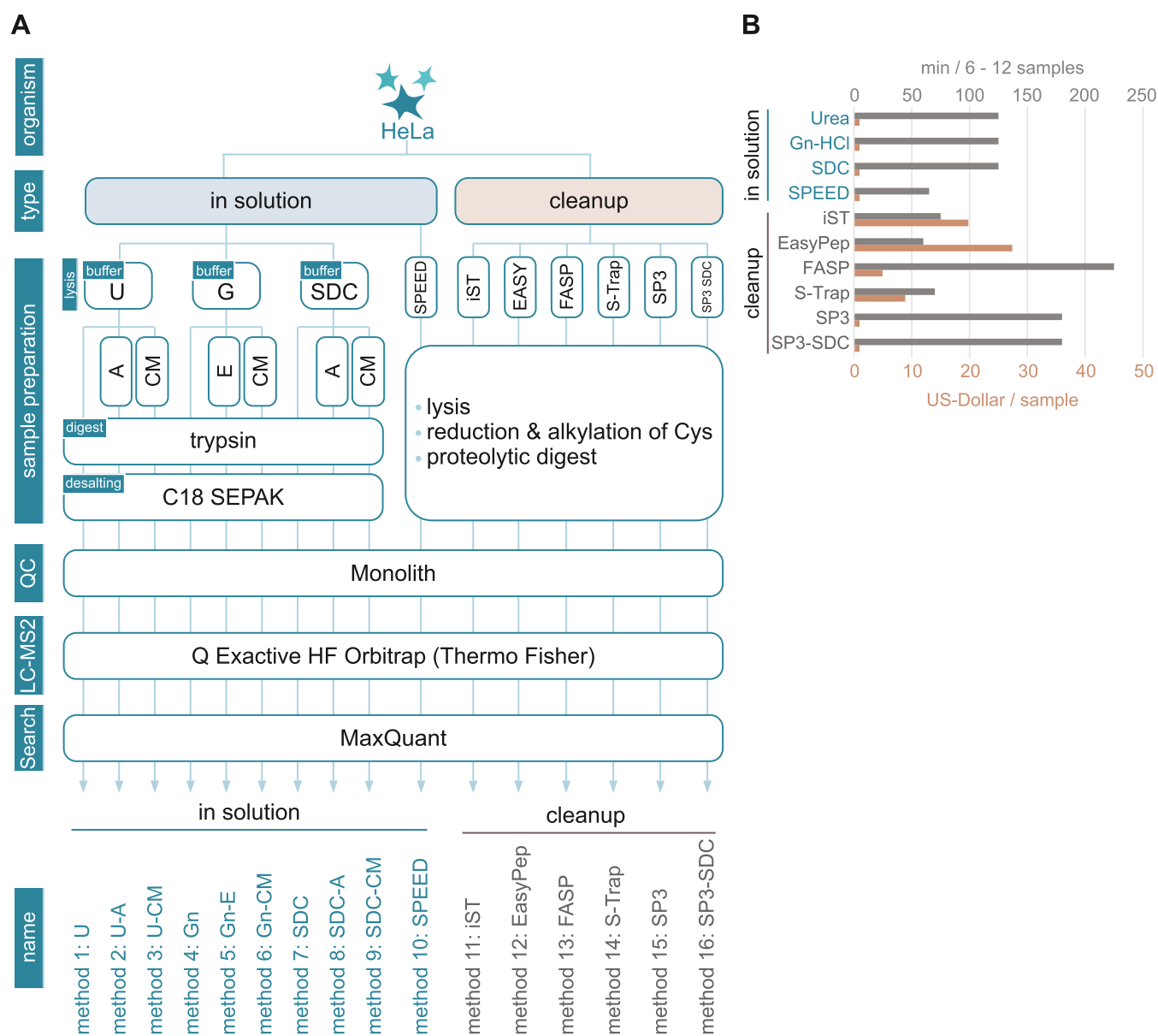


Figure 1. Experimental setup and quality control of applied MS sample preparation methods. (A) Scheme of the experimental design. Proteomes of HeLa cells were prepared according to indicated sample preparation methods. In-solution digest (ISD) protocols (left) covered classical approaches based on urea (U)-, guanidine hydrochloride (G)-, or sodium deoxycholate (SDC)-buffered systems and Sample Preparation by Easy Extraction and Digestion⁴ (SPEED). Lysates prepared with classical ISD protocols were either directly submitted to tryptic digestion or previously mixed with appropriate amounts of acetone (A), ethanol (E), or chloroform/methanol (CM) to precipitate proteins. Cleanup methods covered device-based approaches such as FASP, S-Trap, iST (PreOmics), EasyPep (Thermo Scientific), and SP3-based methods (right). Quality control (QC): peptide concentrations of all samples were determined using UV chromatograms (Monolith) after proteolytic digestion and adjusted for a concentration of 100 ng/ μ l before MS analysis. All samples were analyzed using a quadrupole-orbitrap hybrid MS instrument. MS raw data was analyzed using MaxQuant. (B) Overview of the average cost in US dollars (USD) per sample (brown) and time in minutes needed to process 6–12 samples (gray) excluding digestion and optional C18 peptide cleanup.

integer), in this case, 21. For principal component analysis (PCA) analysis, the `prcomp()` function from the package `stats` (preinstalled in R) was used.

k-Means Clustering. *k*-Means clustering was performed using the function `kmeans()` from the preinstalled R package `stats`. All of the above-described functions are embedded in the in-house script termed *Cassiopeia*.²⁸ Briefly, *Cassiopeia* is an in-house built LaTeX script that runs on R-code and is used for the generation of quality control outputs and statistical outputs and for visualization of information for a given “proteinG-

roups.txt” file as produced by the quantitative proteomics software package MaxQuant.²⁵

Mapping of Protein Features. To map protein features, such as protein abundance level, protein structure, localization in cellular compartments, etc., to the clusters, the results of the *k*-means cluster analysis have been merged with entries of protein databases using an in-house Python script ([Supplemental Material_Scripts](#)). The following databases have been used: Human Protein Atlas ([proteintlas.org](#));²⁹ PhosphoSitePlus;³⁰ PSIPRED;³¹ D2P2;³² Pdbtm;^{33,34} Reactome.org;³⁵ and a database covering the protein expression level,³⁶ information

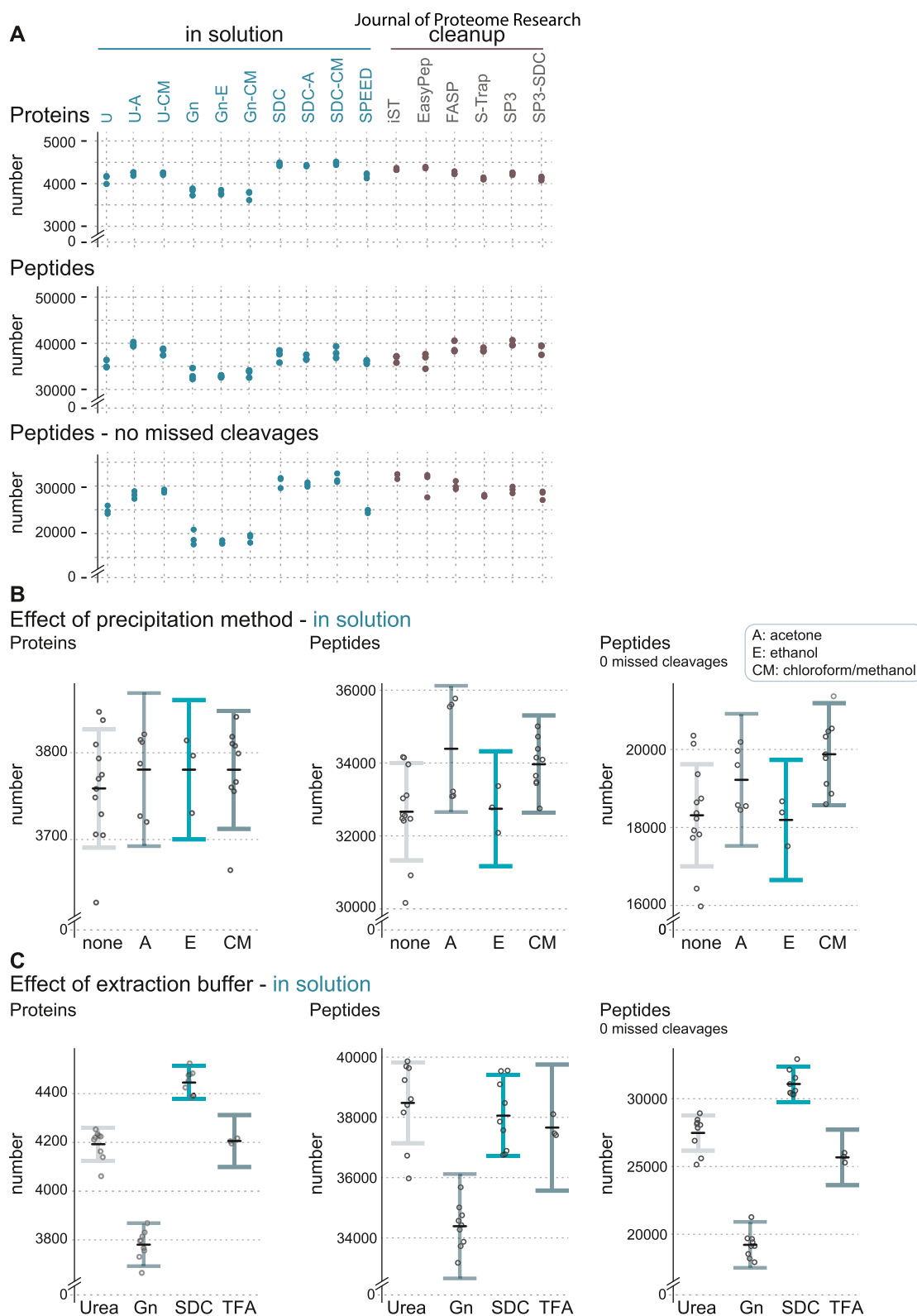


Figure 2. Partial residual plots highlighting effects of applied buffer systems and protein precipitation methods in ISD protocols. (A) Diagram showing a comparison of the total number of identified (by MS/MS) proteins (top), peptides (middle), and peptides with no missed cleavages (bottom). (B) Effects of the applied protein precipitation method on the number of identified (by MS/MS) proteins (left), peptides (middle), and peptides with no missed cleavages (right). A, acetone precipitation; E, ethanol precipitation; CM, chloroform–methanol precipitation. (C) Same as (B) except that the effects of applied buffer systems are shown. Gn, guanidine hydrochloride; SDC, sodium deoxycholate; TFA, SPEED.⁴ Data points represent the predicted number of IDs. Error bars correspond to a 95% confidence interval (CI). Black lines indicate the average predicted number of IDs.

on complexes,³⁷ and aggregator feature.³⁸ Statistical significance for a protein characteristic's enrichment in a cluster was inferred via one-sided Fisher's exact test using the `fisher.test()` function from the R package `stats`. Enrichment factors were calculated as the ratio of observed number/expected number, where the expected number was calculated as the cluster size of cluster k multiplied by the relative frequency of the characteristic n throughout the whole experiment (i.e., enriched compared to the global relative frequency).

Linear Regression Modeling. Within ISD samples, the total number of observed features (proteins, peptides, and peptides with 0 missed cleavages) were analyzed by means of a simple linear model applying the following model formula: $IDs \sim batch + precipitation + buffer$. F/ANOVA tests were applied to test for the significance of the individual variables (batch, precipitation method, and buffer). Linear model predictions were visualized as partial residual plots using the function `effect_plot()` from the R package `jtools`.

Venn Diagrams and UpSetR plots. The MaxQuant ProteinGroups.txt output table was cleared from contaminants, reverse hits, IDs only by site, and iRT (internal retention time standards) hits. Only protein IDs (protein groups) identified by MS/MS in at least two out of three replicates were considered. Area proportional Venn diagrams were created using web application DeepVenn.³⁹ Overlaps of protein IDs (%) were quantified in Python (version 3.9) with pandas data analysis toolkit. Intersecting sets of protein IDs in all methods were further visualized in an UpSet plot, which was generated using a scalable matrix-based visualization script employed by the open-source R package UpSetR.⁴⁰

Cost-Effort Calculations

The financial expenditure for a method was determined by the cost per sample either according to the manufacturer (e.g., 96 samples for iST 96x kit) or calculated by the reagent cost per sample. For SP3, the cost was determined by the usage of magnetic beads solution per sample. The cost of one-time investments such as magnetic racks needed for SP3 protocols is not included in our table. The hands-on times refer to the processing time of 6 (12 samples for SP3 and SP3-SDC) samples in parallel without digestion, and without additional C18 cleanup or vacuum centrifugation times.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE⁴¹ partner repository with the data set identifier PXD030406 and 10.6019/PXD030406.

RESULTS AND DISCUSSION

Experimental Design and Quality Control

To provide a comparative analysis of MS sample preparation methods, we applied 16 widely used protein extraction protocols to isolate whole-cell proteomes from HeLa cells and compared their efficacy on the basis of quantitative and qualitative parameters (Figure 1A). Our experimental setup covered in-solution digest (ISD) protocols and cleanup methods, including commercially available MS sample preparation kits.

For classical ISDs, cells were lysed either at room temperature (urea) or 60 °C (GnHCl, SDC) to optimize cell lysis and protein solubilization. We observed similar protein extraction efficiencies for the three applied buffer systems (Supplemental Figure 1A). Lysates were split into three groups of aliquots of equal protein amounts. The first

group was directly subjected to proteolytic digestion using trypsin. The protein fractions of the remaining aliquots were additionally purified prior to proteolysis by acetone, ethanol (given that GnHCl is not soluble in acetone), or chloroform/methanol protein precipitation, respectively (Figure 1A). Notably, some combinations of buffer systems and precipitation methods, such as urea-based buffer and chloroform/methanol precipitation, resulted in significant sample losses. The highest yields were observed with SDC-based buffers (Supplemental Figure 1A), which correspond to previous observations.^{20,42}

Cleanup samples were prepared as previously described^{10–12,15,21} or according to manufacturer's guidelines, with the exception of SP3, where, additionally to the detergent-heavy buffer system, an easy to prepare buffer consisting of 1% SDC in Tris–HCl (see Materials and Methods for further information) was tested (SP3-SDC). The latter was included since this buffer composition delivered high performance in classical ISDs.^{1,20} Overall, we obtained similar peptide concentrations after tryptic digestion in all cleanup samples (Supplemental Figure 1B), even though proteolysis differed in the reaction mix composition, reaction time, and peptide-to-enzyme ratio (see Materials and Methods).

To achieve equal loading for MS measurements, peptide concentrations of all samples were determined using UV chromatogram peak areas and adjusted accordingly. MS measurements were performed on a quadrupole-orbitrap hybrid MS instrument (Figure 1A). All 16 experimental conditions were analyzed in three technical replicates, resulting in a total of 48 MS runs that were measured in six consecutive batches. The performance of the LC–MS system was monitored by inspecting retention times, intensities, and peak shapes of spike-in standards (iRT) to ensure similar conditions within and between batches. Non-normalized summed protein group intensities indicated that comparable amounts of peptides were submitted to MS measurements (Supplemental Figure 1C).

Cost and Time Effort

Since the expenditure of time and money is important to consider, we determined the average cost in US dollars and hands-on sample processing times for the applied methods (Figure 1B). ISD protocols come at very low consumable costs but are, with the exception of SPEED, considerably more time-demanding than commercial kits. EasyPep, iST, SPEED, and S-Trap protocols were found to have similar hand-on times of around 60 min. FASP, on the other hand, is inherently more time-consuming with long centrifugation steps, taking up to 4 h. Costs ranged from 1\$ (ISD, SPEED, SP3, and SP3-SDC) to 5\$ (FASP), ~10\$ (S-Trap), ~20\$ (iST), or ~30\$ (EasyPep) per sample. From this perspective, SPEED represents a competitive protocol that combines short handling times with low consumable costs.

Global Comparison of Performance

We first compared overall method performance, considering the total numbers of protein groups (protein IDs) and peptides (peptide IDs) identified by LC–MS/MS (Figure 2A, Supplemental Table 1). After filtering data (see Materials and Methods), we retrieved protein IDs ranging from 3500 to 4500 and peptide IDs ranging from 30,000 to 40,000, with SDC-based sample preparations resulting in the highest numbers. ISD protocols based on GnHCl, on the other

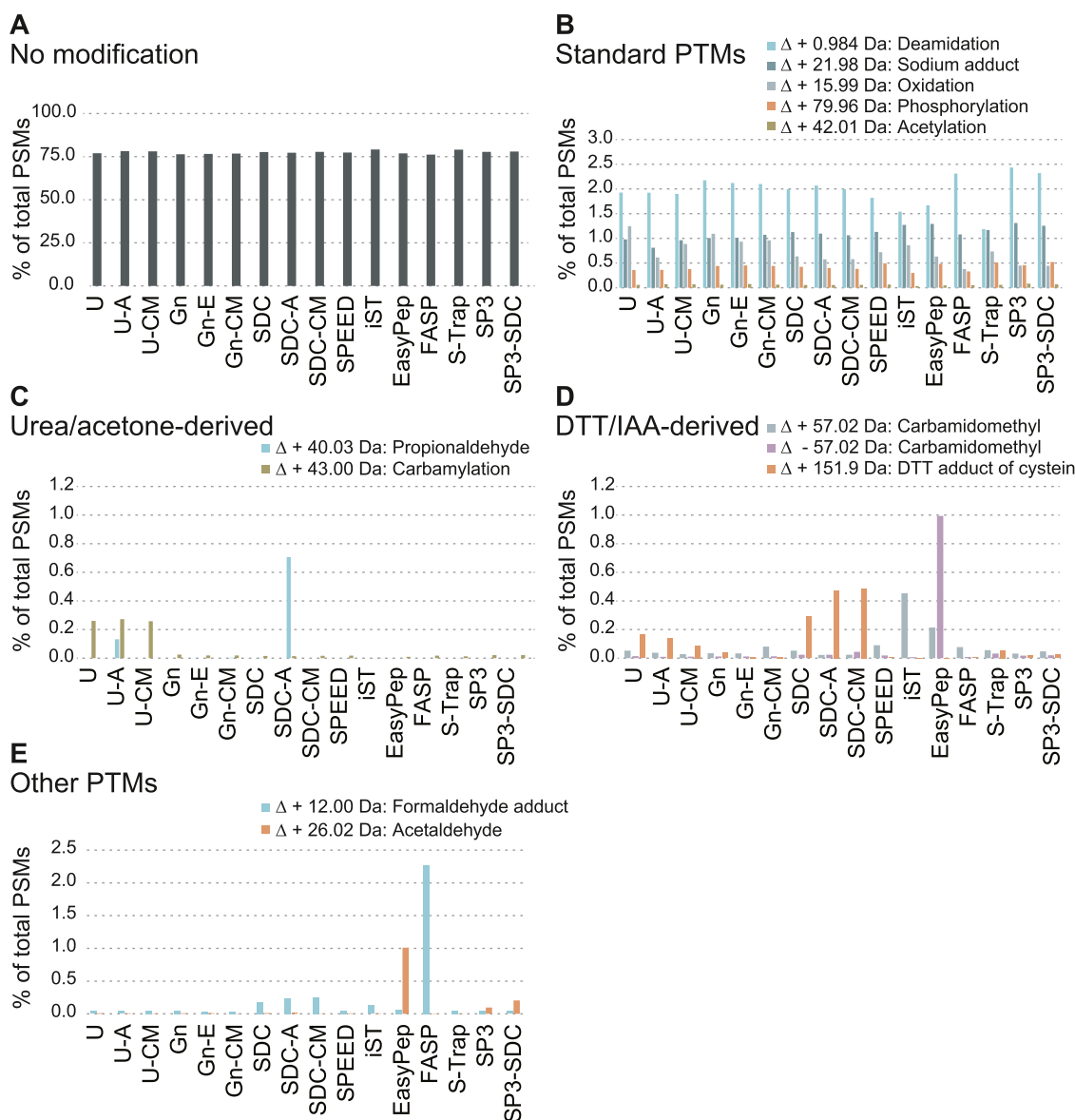


Figure 3. Analysis of covalent peptide modification artifacts created during sample preparation. Open search analysis using MSFragger to identify sample preparation-induced peptide modification artifacts. Bar plots show the sum of PSMs without modification (see also [Supplemental Table 2](#)). (A) Bar plot showing the percentage (y-axis) of PSMs without modification. The x-axis lists the applied sample preparation protocols. (B) Similar to (A) except that exemplary PTMs are shown. Values for oxidation and acetylation represent modifications that were detected in addition to methionine oxidation or protein N-terminal acetylation, which were both specified as variable modifications in the search. (C) Bar plot highlighting previously described artifacts observed in samples prepared using urea buffer (carbamylation) and acetone precipitation (delta mass: +40.03 Da), respectively. y-axis, percentage; x-axis, methods. (D) Similar to (C) except that artifacts derived from reduction (DTT adduct of cysteine) and alkylation (carbamidomethyl) steps are shown. (E) Unknown modifications identified in EasyPep (delta mass: + 26.01 Da) and FASP (delta mass: + 12.00 Da). y-axis, percentage; x-axis, applied methods.

hand, delivered the lowest numbers of identified peptides and proteins.

Extraction buffers containing chaotropes or detergents are known to interfere with the protease activity of trypsin,^{43,44} which results in incomplete protein digestion and consequently in lower proteome coverage due to oversampling of different cleavage forms of abundant peptides. An analysis of missed cleavage frequencies clearly demonstrates strong differences between protocols, with iST and EasyPep showing the highest efficiencies, followed by ISD-SDC protocols ([Supplemental Figure 2A](#)). The high efficiency of iST and EasyPep can be most likely explained by the combined use of trypsin and Lys-C in these kits, in contrast to trypsin alone as in the other

protocols. This suggests that all methods could probably benefit from the use of both enzymes (as previously described in refs 45 and 46), which needs to be considered when comparing results across protocols.

The differences in cleavage efficiency also help to interpret the results of protein and peptide IDs ([Figure 2A](#)). Some methods with high peptide ID numbers show comparably lower protein IDs (e.g., U-A, FASP, SP3). However, when considering peptides with no missed cleavages ([Figure 2A](#), lower panel), it is evident that the lower digestion efficiency in these methods might result in a lower proteome coverage. The excellent performance of ISD-SDC protocols in terms of protein and peptide IDs even without additional use of Lys-C

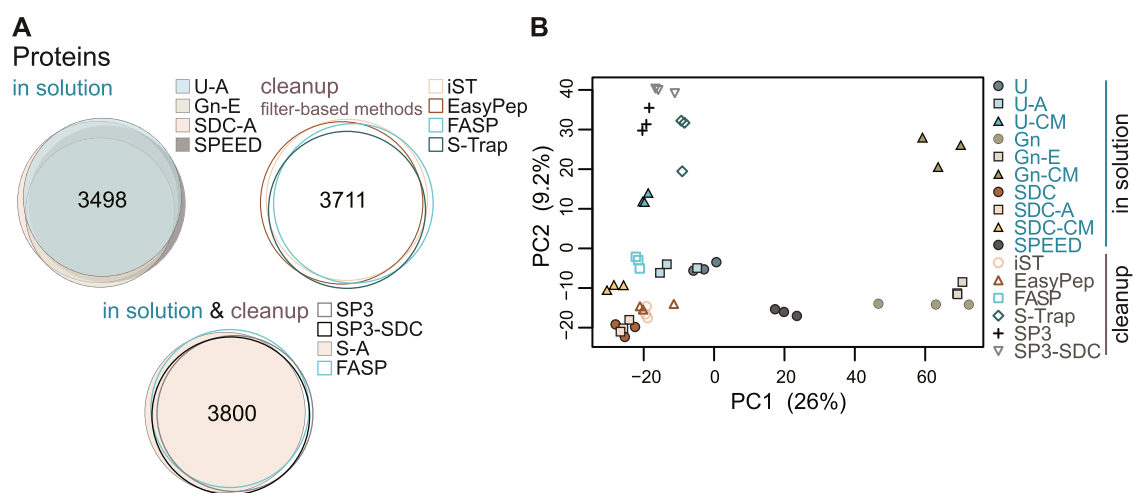


Figure 4. Overlaps of protein identifications and principal component analysis. (A) Venn diagrams depicting the number of overlapping protein IDs (identified by MS/MS) obtained from different sample preparation methods. (B) Principal component analysis (PCA) based on \log_2 -transformed LFQ intensities after normalization and imputation of missing values.

supports the originally reported properties of SDC to enhance trypsin activity and increase digestion efficiency.⁴⁷ Notably, the majority of cleanup protocols and the classical ISD-urea protocols and SPEED showed good performance and rather similar numbers of IDs. Conversely, samples prepared in GnHCl-based buffers displayed the lowest numbers of protein and peptide IDs, suggesting interference of GnHCl with trypsin protease activity even at low concentrations, as reported before.^{3,48}

The values depicted in Figure 2A represent the sum of multiple effects, which hampers an independent evaluation of the impact of single-method parameters, such as protein precipitation. To elucidate the unique impact of variables on the overall performance of ISD protocols, we applied linear regression modeling. In each model, the number of IDs was explained additively by the supposed independent effects of individual precipitation methods and buffer conditions, in addition to batch effects that derive from technical variance during the MS measurements (Supplemental Figure 2B). On the basis of model parameter estimates, we calculated protein and peptide IDs for individual precipitation strategies (Figure 2B) and buffer conditions (Figure 2C) that are corrected for the effects of all other model variables.

In general, protein precipitation only minimally affected the efficiency of protocols, with acetone and chloroform–methanol precipitation being slightly advantageous compared to the other methods (Figure 2B). The strongest impact on method performance is caused by the type of extraction buffer, which confirms that effective protein digestion is a key determinant for proteome coverage. It is possible that there are additional interaction effects between variables. For example, the bimodal data distribution in acetone precipitated samples could hint that acetone precipitation efficiency is influenced by buffer type. However, such potential effects are difficult to resolve statistically with the current study design and with the available number of data points and would require further and more specific experiments. Generally, the SDC-based buffer resulted in the highest numbers of identified proteins and peptides even without precipitation, whereas other methods like urea ISD clearly benefitted from precipitation protocols. Certainly, as mentioned before, these results have been obtained with HeLa cells and might not be directly translatable

to other cells, tissues, or organisms with more challenging properties or specific requirements.

Sample Preparation Artifacts

We next tested whether individual sample preparation methods are prone to protein modification artifacts. We reanalyzed the MS raw data by applying an open search strategy with the FragPipe proteomic software package.²⁶ The open search allows identifying modified peptides from MS data without the need to specify modifications of interest before the analysis.^{49,50} We used the number of PSMs to estimate the abundance of modifications and observed that the majority of PSMs (76–80%) originated from unmodified peptides (Figure 3A). Most of the detected modifications were equally abundant in the different samples (Figure 3B and Supplemental Table 2), suggesting that they are either naturally occurring PTMs or inevitable, method-independent sample preparation artifacts. Nevertheless, we observed method-specific modifications and adducts, some of which have been previously described.^{51–54} Notably, all method-specific modifications were low in abundance ($\leq 1\%$). For example, peptides in ISD-urea samples showed increased levels of carbamylation (Figure 3C), a well-known artifact for this buffer compound.⁵⁴

Peptide artifacts deriving from reduction and alkylation steps could be observed in several methods (Figure 3D). Despite reports on the disadvantages of using dithiothreitol (DTT) and iodoacetamide (IAA), we selected this protocol for the ISD as it is probably the most widely used and because it also allowed comparisons to other standard protocols such as FASP. Interestingly, the alkylation-related artifacts were rather rare and appeared not as problematic as reported in the literature.⁵⁵ Although typical artifacts like off-target alkylation or DTT adducts could be detected, they were found to occur at low levels ($<0.5\%$ or mostly lower), as also reported by Hains and Robinson.⁵⁶ Carbamidomethylated and carboxymethylated methionine or their according neutral losses⁵⁵ as well as potential dialkylation with IAA were not detected or occurred at levels below 0.01%. Among the minor effects, EasyPep and iST showed slightly elevated levels of off-target carbamidomethylation (+57.0215 Da predominantly on lysine and histidine), and in addition EasyPep displayed higher levels of unmodified cysteines (−57.0215 Da), suggesting nonoptimal

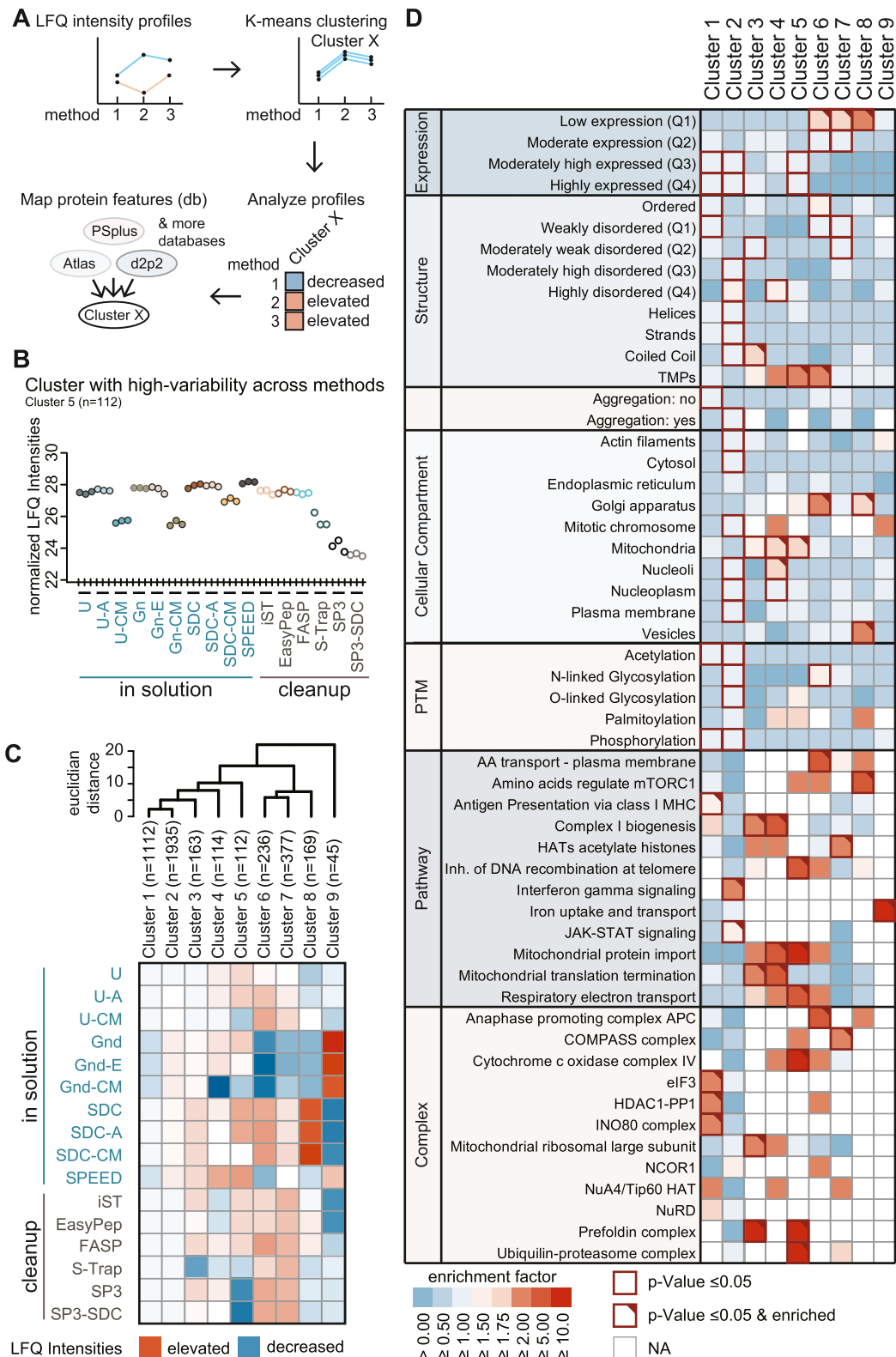


Figure 5. Exploratory *k*-means cluster analysis and descriptive guide matrix. (A) Schematic illustration of *k*-means cluster analysis. (B) Representative cluster-specific ($n = 112$) profile plot of sample preparation methods resulting from an exploratory cluster analysis using *k*-means. Methods (x -axis) are plotted against normalized \log_2 -transformed LFQ intensities (y -axis). Plots depict the coordinates of *k*-means cluster centers. (C) Heatmap showing cluster-specific deviations in the efficiency of sample preparation methods. The color code represents average normalized \log_2 LFQ intensities. Dendrogram (top) depicts hierarchical relationships of clusters based on ultrametric euclidean distances. (D) Matrix depicting the enrichment and significance of protein features (y -axis) in each *k*-means cluster (x -axis). The color code indicates the enrichment factor of protein features. Red frames indicate significance (p -value < 0.05). Red triangles, enrichment factor ≥ 2 .

reaction conditions for alkylation of free thiols.⁵¹ Unfortunately, the type and concentration of chemicals used in these kits are not disclosed; however, based on the “one-pot” reaction conditions and published information,¹⁵ it can be assumed that chemicals other than IAA and DDT are used. Nevertheless, their impact on artifacts and general method performance appears to be rather small when compared to the other protocols in this study. ISD-SDC, and to a minor extent S-Trap, resulted in increased levels of DTT adducts on cysteine (+151.9966 Da). We further recorded a modification seemingly specific to acetone precipitation with a delta mass of +40.0313 Da (propionaldehyde) in ISD-urea and especially ISD-SDC samples (Figure 3C), possibly constituting acetone adducts.⁵³ Finally, we observed enrichment of a delta mass of +26.0157 Da in EasyPep samples, likely corresponding to N-terminal acetaldehyde Schiff base formation,⁵² and a delta mass of +12.00 Da (formaldehyde adduct), previously described to be specific to FASP samples⁵⁷ (Figure 3E).

The open search strategy might not exhibit the sensitivity to reveal all modifications and artifacts occurring in the samples. However, it provided a rather unbiased, broad overview and revealed that only a negligible fraction of peptides was affected by method-induced modifications, indicating that artifacts induced by sample preparation pose only a minor problem for the protocols as they were applied in our study.

Proteome Coverage and Qualitative Differences

Apart from the numbers of proteins and potential artifacts, the most important question is certainly whether methods differ in terms of identity and quantity of the proteins they extract. We investigated whether the individual sample preparation methods covered largely similar or distinct fractions of the HeLa proteome (Figure 4A). Based on this analysis, it appears that overall proteome coverage is rather comparable. We observed a predominant overlap of protein IDs when comparing classical ISD methods and SPEED (3498 proteins, 75.3% overlap). Similar observations were made when comparing the cleanup methods FASP, S-Trap and commercial kits EasyPep, iST (3711 proteins, 78.9% overlap) or when comparing SDC-A, FASP with SP3-based methods (3800 proteins, 81.9% overlap) (Figure 4A). The overlap of all 16 methods (2989 proteins) was 61.6% (Supplemental Figure 3), but this lack of overlap is certainly also driven to a large extent by missing identifications of rather low abundant peptides due to the stochastic nature of data-dependent acquisition. It is clear though that a simple analysis of overlaps in protein IDs does not allow to reveal specific or more subtle differences.

In contrast, a principal component analysis (PCA) of label-free quantified protein intensities separated out distinct clusters for the replicates corresponding to the different sample preparation methods, pointing toward qualitative differences in preparation-dependent variables (Figure 4B). We observed clustering according to buffer and precipitation conditions, with chloroform/methanol precipitation being more distant from other approaches. Distinct grouping of SP3-derived samples was also observed, irrespective of the applied buffer systems, suggesting that the magnetic bead-mediated protein pulldown poses a key variable for method-specific protein extraction. Furthermore, iST and EasyPep clustered close to SDC-ISD protocols, suggesting similarity in their methodology.

To further elucidate method-specific differences systematically, we carried out an explorative *k*-means cluster analysis and

thereby classified variation patterns in protein intensities (Figure 5A). We first defined the optimal number of clusters using the sum of squares within (SSW) distances to the next cluster center. Our approach defined nine *k*-means centers of the cluster ($k = 9$) as the optimal number, each showing a distinct method-dependent signature pattern of center-normalized LFQ intensities (Supplemental Figure 4A). Each cluster therefore consists of an individual set of protein IDs (Figure 5B and Supplemental Figure 4B). A downshift in center-normalized LFQ intensities suggests a method-dependent decrease in protein isolation efficacy in a given cluster. The opposite is true for observed upshifts. For clusters with a large number of elements, such as clusters 1 ($n = 1112$) and 2 ($n = 1935$), we observed similar performance of all sample preparation methods (Supplemental Figure 4B). This suggests that the majority of proteins are effectively extracted independent of the applied protocol, which is also in agreement with the Venn diagrams (Figure 4A). Method-specific up- or downshifts in center-normalized LFQ intensities were prominent in clusters of smaller size, such as cluster 9 ($n = 45$), showing the most profound differences. Shifts in LFQ intensities were generally trending downward. Figure 5C summarizes the relative efficiency of sample preparation methods for each cluster in a heatmap (Figure 5C) and highlights that all methods display distinct profiles with unique features.

Clear clustering suggests that the respective proteins share common properties. We performed enrichment analysis on protein features that we extracted from selected databases, such as the Human Protein Atlas for subcellular localization;²⁹ PhosphoSitePlus for known PTMs;³⁰ PSIPRED for information on the secondary structure;³¹ D2P2 providing a score for disordered regions;³² Pdbtm for transmembrane domains;^{33,34} Reactome.org for cellular pathways;³⁵ and three databases covering protein expression levels,³⁶ complex information,³⁷ and aggregation features³⁸ (Figure 5A). To determine which properties promote effective extraction by a given sample preparation method, we calculated cluster-specific enrichment for individual protein features (Figure 5D and Supplemental Table 3). By combining the information in Figure 5C,D, one can determine which methods can be used to purify specific protein features. In detail, Figure 5C (or Supplemental Figure 4B) illustrates whether a given method works well with a cluster (e.g., all ISD-Gnd methods are well suited for purification of proteins of cluster 9), while Figure 5D (or Supplemental Table 3) shows the cluster properties (the only protein feature enriched in the given example is “ion uptake and transport”).

As stated above, cluster 1 ($n = 1112$) comprises a high number of proteins that become efficiently isolated by all sample preparation methods (Supplemental Figure 4B). We found several features connected to the histone deacetylase (HDAC) 1 complex to be enriched in cluster 1, indicating that the nuclear fraction of proteins can be purified with all tested methods at equal efficacy. Clusters 4 ($n = 114$) and 5 ($n = 112$) showed enrichment of several mitochondrion-associated properties, such as mitochondrial protein import, mitochondrial translation termination, respiratory electron transport, cytochrome c oxidase complex IV, and mitochondrial ribosomal large subunit (Figure 5D). The fact that CM-based precipitation showed lower center-normalized LFQ intensity levels in clusters 4 and 5 (Figure 5C and Supplemental Figure 4B) suggests that these protocols should

be avoided for mitochondrial proteomics. Conversely, ISD (without CM) and SPEED protocols seem to be well suited for mitochondrial protein extraction, as they resulted in the highest intensity levels (Figure 5C,D and Supplemental Figure 4B). Cluster 8 ($n = 169$) showed enrichment of vesicle- and membrane-associated protein properties (Figure 5D), which is consistent with the good performance of ISD-SDC in this group⁵⁸ (Figure 5C). Finally, proteins associated with iron uptake and transport were exclusively found to be enriched in cluster 9 ($n = 45$) (Figure 5D). Successful extraction of this set of proteins seems to be best achieved using ISD protocols based on GnHCl buffers.

Certainly, the efficacy of protein extraction of all applied methods could be further optimized. Here, we provide a basis for doing so, indicating steps in sample preparation protocols that could be further fine-tuned. As suggested in previous reports,^{1,4,13,19,20} different combinations of buffer components and buffer systems, reactor types, proteolytic digestion protocols, and the use of nucleases could be implemented. Changes to protocols should, however, be made with caution since cross-compatibility of reagents is not always guaranteed. For example, we occasionally observed gel-like phases in extracts when we used SDC in conjunction with phosphate buffers (unpublished observation). Our data also suggests that omitting a protein precipitation step during MS sample preparation can still result in sufficient proteome coverage for HeLa cells. Yet, we generally advise including a protein precipitation step to avoid carry-over of nonprotein cellular components such as lipids, nucleic acids, metabolites, etc., which could cause problems during later steps of sample preparation.

In general, different cell types and organisms may require different adaptations. To give an example, we observed that using buffer systems containing urea in combination with chloroform–methanol precipitation resulted in significant losses when proteins were extracted from *Saccharomyces cerevisiae* cells (unpublished observation). Doellinger et al.⁴ have shown that the SPEED protocol outperforms other protocols when processing bacterial samples. Furthermore, it is well known that samples from plants or fungi often require specific protocols due to the high level of interfering metabolites.

Previous comparisons of sample preparation methods across species have shown that extraction biases do exist and that therefore a universal method is rather unlikely.^{1,4} Our study additionally demonstrates that even within the same sample type there is no one-fits-all protocol because all methods have their own peculiarities. For example, even though the SPEED protocol performs well in many aspects it also exhibits an extraction bias toward certain protein groups, e.g., for proteins associated with the Golgi apparatus and transport to the plasma membrane (see cluster 6 and cluster 8, Figure 5C,D). However, despite these clear differences for specific clusters, our data also show that most methods, with the exception of GnHCl, perform overall rather similar in this cell type, which allows choosing methods rather on other parameters like ease of use, processing times, etc.

In summary, despite similar proteome coverage, we could extract qualitative differences between the different protocols that represent varied purification efficacy for certain sets of proteins. The presented matrix, the underlying data set, and the according methodology may serve as a guideline for the

choice of a best-suited sample preparation method for a specific group of proteins of interest.

CONCLUSIONS

The present study provides an in-depth and solid comparison of 16 of the most widely used MS sample preparation protocols in a human cell line. Careful attention has been paid to quality control and experimental design to maximize reproducibility and comparability and to allow for unbiased statistical analyses. We demonstrate that the applied protocols had an overall rather similar performance with a low degree of protein modification artifacts and similar protein extraction efficiencies. Our analysis further revealed method-specific protein clusters, and we summarized their features in a guide matrix to assist in choosing an appropriate method. Urea-acetone, SDC-acetone, and FASP protocols perform well in terms of the number of covered protein/peptide IDs and enrichment of all classes of proteins. In addition, these methods are also comparatively cheap. A similar degree of performance was observed for the commercial kits, with the additional benefit that materials and reagents are provided in a standardized manner and handling is straightforward. SPEED delivered in general a good performance and its simplicity and low price make it an attractive alternative. However, our data also showed that several methods (SPEED, FASP, S-Trap, and SP3) could benefit from further refinements (e.g., trypsin and Lys-C digest). Finally, we also highlighted methods preferable for enrichment for specific protein characteristics. For example, ISD in combination with GnHCl buffer is well suited for the isolation of proteins associated with iron uptake and transport, however, at the cost of reduced efficacy of digestion and an overall lower proteome coverage.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.2c00265>.

Selected quality control steps from our experimental approach (Figure S1), missed cleavages and batch effect (Figure S2), overlap of identified proteins between all methods (Figure S3), and *k*-means cluster centers (Figure S4) (PDF)

Number of proteins, peptides, and missed cleavages (Table S1) (XLSX)

Open search results (Table S2) (XLSX)

Enrichment analysis (Table S3) (XLSX)

Suppl.Scripts, scripts described in Materials and Methods (ZIP)

AUTHOR INFORMATION

Corresponding Author

Markus Hartl – Max Perutz Labs, Mass Spectrometry Facility, Vienna Biocenter Campus (VBC), 1030 Vienna, Austria; Center for Molecular Biology, Department of Biochemistry and Cell Biology, University of Vienna, 1030 Vienna, Austria; orcid.org/0000-0002-4970-7336; Email: markus.hartl@univie.ac.at

Authors

Gina Varnavides – Max Perutz Labs, Mass Spectrometry Facility, Vienna Biocenter Campus (VBC), 1030 Vienna, Austria; orcid.org/0000-0001-8597-1233

Moritz Madern – Max Perutz Labs, Mass Spectrometry Facility, Vienna Biocenter Campus (VBC), 1030 Vienna, Austria; Center for Molecular Biology, Department of Biochemistry and Cell Biology, University of Vienna, 1030 Vienna, Austria

Dorothea Anrather – Max Perutz Labs, Mass Spectrometry Facility, Vienna Biocenter Campus (VBC), 1030 Vienna, Austria

Natascha Hartl – Max Perutz Labs, Mass Spectrometry Facility, Vienna Biocenter Campus (VBC), 1030 Vienna, Austria; Center for Molecular Biology, Department of Biochemistry and Cell Biology, University of Vienna, 1030 Vienna, Austria

Wolfgang Reiter – Max Perutz Labs, Mass Spectrometry Facility, Vienna Biocenter Campus (VBC), 1030 Vienna, Austria; Center for Molecular Biology, Department of Biochemistry and Cell Biology, University of Vienna, 1030 Vienna, Austria; orcid.org/0000-0003-1266-8975

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jproteome.2c00265>

Author Contributions

M.H. conceptualized the study. G.V., D.A., and M.H. designed experiments. G.V., D.A., and N.H. performed experiments. G.V., D.A., M.M., W.R., and M.H. analyzed the data. G.V., W.R., and M.H. wrote the paper. All authors edited the text. All authors read and approved the final manuscript.

Funding

Open Access is funded by the Austrian Science Fund (FWF). G.V. was supported by FEMtech of the Austrian Forschungsförderungsgesellschaft (FFG). M.M., N.H., W.R., and M.H. were supported by the Austrian Science Fund (FWF) Special Research Program F70.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Natalie Romanov for her help with the integration of protein databases used in the cluster analysis and for critical feedback on the manuscript, Dea Slade for kindly providing HeLa cells, Karl Mechtler and his team for the great collaborative spirit in our joint lab space, and VBCF for providing the LC–MS instrument pool.

REFERENCES

- (1) Glatter, T.; Ahn , E.; Schmidt, A. Comparison of Different Sample Preparation Protocols Reveals Lysis Buffer-Specific Extraction Biases in Gram-Negative Bacteria and Human Cells. *J. Proteome Res.* **2015**, *14*, 4472–4485.
- (2) Waas, M.; Bhattacharya, S.; Chuppa, S.; Wu, X.; Jensen, D. R.; Omasits, U.; Wollscheid, B.; Volkman, B. F.; Noon, K. R.; Gundry, R. L. Combine and Conquer: Surfactants, Solvents, and Chaotropes for Robust Mass Spectrometry Based Analyses of Membrane Proteins. *Anal. Chem.* **2014**, *86*, 1551–1559.
- (3) Poulsen, J. W.; Madsen, C. T.; Young, C.; Poulsen, F. M.; Nielsen, M. L. Using Guanidine-Hydrochloride for Fast and Efficient Protein Digestion and Single-Step Affinity-Purification Mass Spectrometry. *J. Proteome Res.* **2013**, *12*, 1020–1030.

- (4) Doellinger, J.; Schneider, A.; Hoeller, M.; Lasch, P. Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-Free Protocol for Proteomics Based on Acid Extraction. *Mol. Cell. Proteomics* **2020**, *19*, 209–222.

- (5) Piettre, V. A. On the Separation of Proteins of the Serum Title. *Comptes rendus l'Acad mie des Sci.* **1920**, *170*, 1466–1468.

- (6) Fic, E.; Kedracka-Krok, S.; Jankowska, U.; Pirog, A.; Dziedzicka-Wasylewska, M. Comparison of Protein Precipitation Methods for Various Rat Brain Structures Prior to Proteomic Analysis. *Electrophoresis* **2010**, *31*, 3573–3579.

- (7) Botelho, D.; Wall, M. J.; Vieira, D. B.; Fitzsimmons, S.; Liu, F.; Doucette, A. Top-down and Bottom-up Proteomics of SDS-Containing Solutions Following Mass-Based Separation. *J. Proteome Res.* **2010**, *9*, 2863–2870.

- (8) Cohn, E. J.; Hughes, W. L.; Weare, J. H. Preparation and Properties of Serum and Plasma Proteins. XIII. Crystallization of Serum Albumins from Ethanol-Water Mixtures 1a,B. *J. Am. Chem. Soc.* **1947**, *69*, 1753–1761.

- (9) Wessel, D.; Fl gge, U. I. A Method for the Quantitative Recovery of Protein in Dilute Solution in the Presence of Detergents and Lipids. *Anal. Biochem.* **1984**, *138*, 141–143.

- (10) Wi niewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat. Methods* **2009**, *6*, 359–362.

- (11) Zougman, A.; Selby, P. J.; Banks, R. E. Suspension Trapping (STrap) Sample Preparation Method for Bottom-up Proteomics Analysis. *Proteomics* **2014**, *14*, 1006–1000.

- (12) Hughes, C. S.; Moggridge, S.; M ller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J. Single-Pot, Solid-Phase-Enhanced Sample Preparation for Proteomics Experiments. *Nat. Protoc.* **2019**, *14*, 68–85.

- (13) Johnston, H. E.; Yadav, K.; Kirkpatrick, J. M.; Biggs, G. S.; Oxley, D.; Kramer, H. B.; Samant, R. S. Solvent Precipitation SP3 (SP4) Enhances Recovery for Proteomics Sample Preparation without Magnetic Beads *bioRxiv* 2021, DOI: [10.1101/2021.09.24.461247](https://doi.org/10.1101/2021.09.24.461247).

- (14) Batth, T. S.; Tollenaere, M. X.; R ther, P.; Gonzalez-Franquesa, A.; Prabhakar, B. S.; Bekker-Jensen, S.; Deshmukh, A. S.; Olsen, J. V. Protein Aggregation Capture on Microparticles Enables Multipurpose Proteomics Sample Preparation. *Mol. Cell. Proteomics* **2019**, *18*, 1027–1035.

- (15) Kulak, N. A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M. Minimal, Encapsulated Proteomic-Sample Processing Applied to Copy-Number Estimation in Eukaryotic Cells. *Nat. Methods* **2014**, *11*, 319–324.

- (16) Elinger, D.; Gabashvili, A.; Levin, Y. Suspension Trapping (S-Trap) Is Compatible with Typical Protein Extraction Buffers and Detergents for Bottom-Up Proteomics. *J. Proteome Res.* **2019**, *18*, 1441–1445.

- (17) Ludwig, K. R.; Schroll, M. M.; Hummon, A. B. Comparison of In-Solution, FASP, and S-Trap Based Digestion Methods for Bottom-Up Proteomic Studies. *J. Proteome Res.* **2018**, *17*, 2480–2490.

- (18) Sielaff, M.; Kuharev, J.; Bohn, T.; Hahlbrock, J.; Bopp, T.; Tenzer, S.; Distler, U. Evaluation of FASP, SP3, and IST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *J. Proteome Res.* **2017**, *16*, 4060–4072.

- (19) Wi niewski, J. R. Quantitative Evaluation of Filter Aided Sample Preparation (FASP) and Multienzyme Digestion FASP Protocols. *Anal. Chem.* **2016**, *88*, 5438–5443.

- (20) Le n, I. R.; Schw mmle, V.; Jensen, O. N.; Sprenger, R. R. Quantitative Assessment of In-Solution Digestion Efficiency Identifies Optimal Protocols for Unbiased Protein Analysis. *Mol. Cell. Proteomics* **2013**, *12*, 2992–3005.

- (21) HaileMariam, M.; Eguez, R. V.; Singh, H.; Bekele, S.; Ameni, G.; Pieper, R.; Yu, Y. S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics. *J. Proteome Res.* **2018**, *17*, 2917–2924.

- (22) Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for Micro-Purification, Enrichment, Pre-Fractionation and Storage of Peptides for Proteomics Using StageTips. *Nat. Protoc.* **2007**, *2*, 1896–1906.

- (23) Schilling, B.; Rardin, M. J.; MacLean, B. X.; Zawadzka, A. M.; Frewen, B. E.; Cusack, M. P.; Sorensen, D. J.; Bereman, M. S.; Jing, E.; Wu, C. C.; Verdin, E.; Kahn, C. R.; Maccoss, M. J.; Gibson, B. W. Platform-Independent and Label-Free Quantitation of Proteomic Data Using MS1 Extracted Ion Chromatograms in Skyline: Application to Protein Acetylation and Phosphorylation. *Mol. Cell. Proteomics* **2012**, *11*, 202–214.
- (24) Bielow, C.; Mastrobuoni, G.; Kempa, S. Proteomics Quality Control: Quality Control Software for MaxQuant Results. *J. Proteome Res.* **2016**, *15*, 777–787.
- (25) Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372.
- (26) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry-Based Proteomics. *Nat. Methods* **2017**, *14*, 513–520.
- (27) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nat. Biotechnol.* **2012**, *30*, 918–920.
- (28) Madern, M. Moritzmadern/Cassiopeia_LFQ: (V1.1) Zenodo, 2021, DOI: 10.5281/zenodo.5758975.
- (29) Uhlén, M.; Fagerberg, L.; Hallström, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szizyarto, C. A.-K.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P.-H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; von Feilitzen, K.; Forsberg, M.; Persson, L.; Johansson, F.; Zvalnen, M.; von Heijne, G.; Nielsen, J.; Pontén, F. Proteomics. Tissue-Based Map of the Human Proteome. *Science* **2015**, *347*, No. 1260419.
- (30) Hornbeck, P. V.; Zhang, B.; Murray, B.; Kornhauser, J. M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and Recalibrations. *Nucleic Acids Res.* **2015**, *43*, D512–D520.
- (31) Buchan, D. W. A.; Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 Years On. *Nucleic Acids Res.* **2019**, *47*, W402–W407.
- (32) Oates, M. E.; Romero, P.; Ishida, T.; Ghalwash, M.; Mizianty, M. J.; Xue, B.; Dosztányi, Z.; Uversky, V. N.; Obradovic, Z.; Kurgan, L.; Dunker, A. K.; Gough, J. D²P²: Database of Disordered Protein Predictions. *Nucleic Acids Res.* **2012**, *41*, D508–D516.
- (33) Tusnády, G. E.; Dosztányi, Z.; Simon, I. Transmembrane Proteins in the Protein Data Bank: Identification and Classification. *Bioinformatics* **2004**, *20*, 2964–2972.
- (34) Kozma, D.; Simon, I.; Tusnády, G. E. PDBTM: Protein Data Bank of Transmembrane Proteins after 8 Years. *Nucleic Acids Res.* **2012**, *41*, D524–D529.
- (35) Gillespie, M.; Jassal, B.; Stephan, R.; Milacic, M.; Rothfels, K.; Senff-Ribeiro, A.; Griss, J.; Sevilla, C.; Matthews, L.; Gong, C.; Deng, C.; Varusai, T.; Ragueneau, E.; Haider, Y.; May, B.; Shamovsky, V.; Weiser, J.; Brunson, T.; Sanati, N.; Beckman, L.; Shao, X.; Fabregat, A.; Sidiropoulos, K.; Murrillo, J.; Viteri, G.; Cook, J.; Shorsler, S.; Bader, G.; Demir, E.; Sander, C.; Haw, R.; Wu, G.; Stein, L.; Hermjakob, H.; D’Eustachio, P. The Reactome Pathway Knowledgebase 2022. *Nucleic Acids Res.* **2022**, *50*, D687–D692.
- (36) Beck, M.; Schmidt, A.; Malmstroem, J.; Claassen, M.; Ori, A.; Szymiorska, A.; Herzog, F.; Rinner, O.; Ellenberg, J.; Aebersold, R. The Quantitative Proteome of a Human Cell Line. *Mol. Syst. Biol.* **2011**, *7*, 549.
- (37) Romanov, N.; Kuhn, M.; Aebersold, R.; Ori, A.; Beck, M.; Bork, P. Disentangling Genetic and Environmental Effects on the Proteotypes of Individuals. *Cell* **2019**, *177*, 1308–1318.e10.
- (38) Määttä, T. A.; Rettel, M.; Sridharan, S.; Helm, D.; Kurzawa, N.; Stein, F.; Savitski, M. M. Aggregation and Disaggregation Features of the Human Proteome. *Mol. Syst. Biol.* **2020**, *16*, No. e9500.
- (39) Hulsen, T.; de Vlieg, J.; Alkema, W. BioVenn - a Web Application for the Comparison and Visualization of Biological Lists Using Area-Proportional Venn Diagrams. *BMC Genomics* **2008**, *9*, 488.
- (40) Conway, J. R.; Lex, A.; Gehlenborg, N. UpSetR: An R Package for the Visualization of Intersecting Sets and Their Properties. *Bioinformatics* **2017**, *33*, 2938–2940.
- (41) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaino, J. A. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47*, D442–D450.
- (42) Shahinuzzaman, A. D. A.; Chakrabarty, J. K.; Fang, Z.; Smith, D.; Kamal, A. H. M.; Chowdhury, S. M. Improved In-Solution Trypsin Digestion Method for Methanol–Chloroform Precipitated Cellular Proteomics Sample. *J. Sep. Sci.* **2020**, *43*, 2125–2132.
- (43) Viswanatha, T.; Pallansch, M. J.; Liener, I. E. The Inhibition of Trypsin. II. The Effect of Synthetic Anionic Detergents. *J. Biol. Chem.* **1955**, *212*, 301–309.
- (44) Gabel, D. The Denaturation by Urea and Guanidinium Chloride of Trypsin and N-Acetylated-Trypsin Derivatives Bound to Sephadex and Agarose. *Eur. J. Biochem.* **1973**, *33*, 348–356.
- (45) Betancourt, L. H.; Sanchez, A.; Pla, I.; Kuras, M.; Zhou, Q.; Andersson, R.; Marko-Varga, G. Quantitative Assessment of Urea In-Solution Lys-C/Trypsin Digestions Reveals Superior Performance at Room Temperature over Traditional Proteolysis at 37 °C. *J. Proteome Res.* **2018**, *17*, 2556–2561.
- (46) Glatter, T.; Ludwig, C.; Ahrné, E.; Aebersold, R.; Heck, A. J. R.; Schmidt, A. Large-Scale Quantitative Assessment of Different in-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/Trypsin Proteolysis over Trypsin Digestion. *J. Proteome Res.* **2012**, *11*, 5145–5156.
- (47) Masuda, T.; Tomita, M.; Ishihama, Y. Phase Transfer Surfactant-Aided Trypsin Digestion for Membrane Proteome Analysis. *J. Proteome Res.* **2008**, *7*, 731–740.
- (48) Ren, D.; Pipes, G. D.; Liu, D.; Shih, L.-Y.; Nichols, A. C.; Treuheit, M. J.; Brems, D. N.; Bondarenko, P. V. An Improved Trypsin Digestion Method Minimizes Digestion-Induced Modifications on Proteins. *Anal. Biochem.* **2009**, *392*, 12–21.
- (49) Yu, F.; Teo, G. C.; Kong, A. T.; Haynes, S. E.; Avtonomov, D. M.; Geiszler, D. J.; Nesvizhskii, A. I. Identification of Modified Peptides Using Localization-Aware Open Search. *Nat. Commun.* **2020**, *11*, No. 4065.
- (50) Geiszler, D. J.; Kong, A. T.; Avtonomov, D. M.; Yu, F.; Leprevost, F.; da, V.; Nesvizhskii, A. I. PTM-Shepherd: Analysis and Summarization of Post-Translational and Chemical Modifications From Open Search Results. *Mol. Cell. Proteomics* **2021**, *20*, No. 100018.
- (51) Boja, E. S.; Fales, H. M. Overalkylation of a Protein Digest with Iodoacetamide. *Anal. Chem.* **2001**, *73*, 3576–3582.
- (52) Chalkley, R. J.; Baker, P. R.; Medzihradsky, K. F.; Lynn, A. J.; Burlingame, A. L. In-Depth Analysis of Tandem Mass Spectrometry Data from Disparate Instrument Types. *Mol. Cell. Proteomics* **2008**, *7*, 2386–2398.
- (53) Güray, M. Z.; Zheng, S.; Doucette, A. A. Mass Spectrometry of Intact Proteins Reveals +98 u Chemical Artifacts Following Precipitation in Acetone. *J. Proteome Res.* **2017**, *16*, 889–897.
- (54) Kollipara, L.; Zahedi, R. P. Protein Carbamylation: In Vivo Modification or in Vitro Artefact? *Proteomics* **2013**, *13*, 941–944.
- (55) Müller, T.; Winter, D. Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by

Iodine-Containing Reagents. *Mol. Cell. Proteomics* **2017**, *16*, 1173–1187.

(56) Hains, P. G.; Robinson, P. J. The Impact of Commonly Used Alkylating Agents on Artifactual Peptide Modification. *J. Proteome Res.* **2017**, *16*, 3443–3447.

(57) Tang, L.; Wu, Z.; Wang, J.; Zhang, X. Formaldehyde Derivatization: An Unexpected Side Reaction During Filter-Aided Sample Preparation. *Anal. Chem.* **2020**, *92*, 12120–12125.

(58) Alfonso-Garrido, J.; Garcia-Calvo, E.; Luque-Garcia, J. L. Sample Preparation Strategies for Improving the Identification of Membrane Proteins by Mass Spectrometry. *Anal. Bioanal. Chem.* **2015**, *407*, 4893–4905.