Title

Automated “Cells-To-Peptides” Sample Preparation Workflow for High-Throughput, Quantitative Proteomic Assays of Microbes

Journal

Journal of Proteome Research, 18(10)

ISSN

1535-3893

Authors

Chen, Yan
Guenther, Joel M
Gin, Jennifer W
et al.

Publication Date

2019-10-04

DOI

10.1021/acs.jproteome.9b00455

Peer reviewed

An automated ‘cells-to-peptides’ sample preparation workflow for high-throughput, quantitative proteomic assays of microbes

Yan Chen,[1,2,3] *Joel M. Guenther,*[1,4,‡] *Jennifer W. Gin,*[1,2,3] *Leanne Jade G. Chan,*[1,3,||] *Zak Costello,*[1,2,3] *Tadeusz L. Ogorzalek,*[1,2,3] *Huu M. Tran,*[1,4] *Jacquelyn M. Blake-Hedges,*[1,3,5] *Jay D. Keasling,*[1,3,6,7,8,9] *Paul D. Adams,*[1,6,10] *Héctor García Martín,*[1,2,3,11] *Nathan J. Hillson,*[1,2,3] *and Christopher J. Petzold*[1,2,3,*]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, CA, USA

[2]Agile BioFoundry, Lawrence Berkeley National Laboratory, Emeryville, CA, USA

[3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[4]Sandia National Laboratories (NTESS), Livermore, CA, USA

[5]Department of Chemistry, University of California Berkeley, Berkeley, CA, USA

[6]Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA

[7]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA, USA

[8]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Denmark

[9]Center for Synthetic Biochemistry, Synthetic Biology Institute, Shenzhen Institutes for Advanced Technologies, Shenzhen, China

[10]Molecular Biophysics and Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[11]BCAM, Basque Center for Applied Mathematics, Bilbao, Bizkaia, Spain

Corresponding author:

Christopher J. Petzold (cjpetzold@lbl.gov)

ABSTRACT

Mass spectrometry–based quantitative proteomic analysis has proven valuable for clinical and biotechnology related research and development. Driving this value have been improvements in the sensitivity, resolution, and robustness of mass analyzers. However, manual sample preparation protocols are often a bottleneck for sample throughput and can lead to poor reproducibility, especially for applications where thousands of samples per month must be analyzed. To alleviate these issues, we developed a 'cells-to-peptides' automated workflow for Gram-negative bacteria and fungi that includes cell lysis, protein precipitation, resuspension, quantification, normalization, and tryptic digestion. The workflow takes two hours to process 96 samples from cell pellets to the initiation of the tryptic digestion step and can process 384 samples in parallel. We measured the efficiency of protein extraction from various amounts of cell biomass and optimized the process for standard liquid chromatography-mass spectrometry systems. The automated workflow was tested by preparing 96 *Escherichia coli* samples and quantifying over 600 peptides that resulted in a median coefficient of variation of 15.8%. Similar technical variance was observed for three other organisms as measured by highly-multiplexed LC-MRM-MS acquisition methods. These results show that this automated sample preparation workflow provides robust, reproducible proteomic samples for high-throughput applications.

KEYWORDS: Automation, Sample preparation, Proteomics, Bacteria, Fungi, Microbes, Biotechnology, High-throughput

INTRODUCTION

Quantitative proteomic assays are a key component of numerous clinical and biotechnology research studies. They inform biomarker discovery research,[1–3] guide optimization of metabolically engineered organisms,[4] and help characterize protein structure and function.[5] As such, improvements in the sensitivity of mass analyzers, speed of data processing, and data quality have been the primary focus of research across the proteomics field. Yet, hundreds, or even thousands, of samples must be analyzed to establish statistically significant and actionable information. This puts an enormous burden on proteomic sample preparation efforts to increase throughput and minimize variation despite protocols that involve many time-consuming, labor intensive steps. Typical sample preparation protocols include over 20 liquid transfer events, reagent mixing, various centrifugation and incubation steps, and an extended tryptic digestion step.

Recently, several automated sample preparation protocols have been developed for blood plasma, urine, or other soluble protein samples.[6–10] However, protocols that incorporate automated cell lysis are much more difficult to implement because traditional cell lysis protocols such as sonication or shear stress via a French press are not easily applied in plate-based formats compatible with liquid handling systems. Consequently, multiple efforts have been made to develop plate-based sample preparation protocols based on chemical lysis procedures such as filter-aided sample preparation (FASP)[8,11,12] and chaotropes like urea.[13] These types of protocols, such as the Quick 96FASP[8] method, require multiple buffer-exchange steps to remove detergents and chaotropes prior to tryptic digestion. This is a significant limitation for these protocols because molecular weight cut off (MWCO) filter plates compatible with liquid handling systems cannot be

centrifuged as fast as the analogous spin columns that are used for the manual preparation protocols.[14,15] Consequently, these protocols are time consuming and susceptible to variable sample loss. Efforts to circumvent the MWCO filter plate limitations by using polyvinylidene fluoride (PVDF) membranes have been shown to overcome the time limitation, but produce missed cleavages and poor quantitative reproducibility.[15] Similarly, tip-based peptide cleanup steps have been implemented in 96-well plate formats,[16] however, at significant per-sample added expense.

In an effort to reduce variability and improve throughput we developed an automated sample preparation workflow that consists of cell lysis, protein precipitation, protein resuspension, protein quantification and normalization of protein concentration followed by standard bottom-up proteomic procedures of reducing and blocking cysteine residues and tryptic digestion. This comprehensive workflow can process four 96-well plates (384 samples) in parallel in under six hours for subsequent analysis via liquid chromatography-mass spectrometry systems. We evaluated the effectiveness of workflow in a 96-well plate format against manual sample preparation protocols and demonstrated its effectiveness on Gram-negative bacteria (*Escherichia coli*, *Pseudomonas putida*) and fungi (*Saccharomyces cerevisiae*, *Rhodosporidium toruloides*). We optimized protein extraction conditions and measured the variation of the workflow with high-throughput UHPLC-MRM mass spectrometry assays for this workflow.

EXPERIMENTAL SECTION

Strains, media, and growth conditions

Escherichia coli strain BW25113 (JPUB_001327), *Saccharomyces cerevisiae* S288C (JPUB_013514), *Rhodosporidium toruloides* NP11 (JPUB_012600), and *Pseudomonas putida*

KT2440 (JPUB_012605) were cultured in house under the following conditions. For shake flask culturing at 200 RPM, *E. coli* and *P. putida strains* were grown for 24 hours in Luria broth (LB) medium at 37 °C and 30 °C respectively. *S. cerevisiae* and *R. toruloides* strains were grown overnight in YPD medium at 30 ºC. Cells were distributed to 96-well plate format and harvested by centrifugation. Cell pellets were frozen at -80 °C until further processing. For *cultivation* in 96-deep-well and BioLector plate formats, an overnight *E. coli* pre-culture was prepared in LB medium, and was inoculated to either 96-deep well plate, or a BioLector 48-well flat-bottom plate at 0.1 starting OD600 of a total 800 µL culture volume in LB medium. The 96-deep well plate was cultured for 24 hours at 37 °C, shaking at 800 RPM. The BioLector plate experiment was run on a BioLector microbioreactor (m2p-labs, Hauppauge, NY) at 800 rpm at 37 °C and ambient chamber pressure for 24 hours with 85% humidity. For both experiments, cell biomass were harvested after cultivation by centrifugation (3,000 x g for 5 minutes at 4 °C) and then stored at -80 °C until prepared via the automated workflow.

Manual proteomic sample preparation

Manual cell lysis and protein extraction was accomplished using a chloroform/methanol precipitation method as previously described.[17] Briefly, for extracting proteins from samples in 96 well plates, the same chloroform/methanol cell lysis and protein extraction procedure is applied and the volume of reagents is reduced by a factor of 5. Vortexing is substituted by pipette mixing. Cell pellets were resuspended in 80 µL of methanol and 20 µL of chloroform and thoroughly mixed by pipetting. 60 µL of water was subsequently added to samples and mixed by pipetting. Phase separation was induced by centrifugation at maximum speed in a table centrifuge for 10 minutes. The methanol and water layers were removed, then 80 µL of methanol was added to each

well. The plate was centrifuged for another 10 minutes at maximum speed, then the supernatant chloroform and methanol layers were decanted. All resulted protein pellet was resuspended in 100 mM ammonium bicarbonate (AMBIC) buffer supplemented with 20% methanol, and protein concentration was determined by the DC assay (BioRad, Hercules, CA). Prior to protein trypsin digestion at a concentration of 1 mg/mL, protein reduction was accomplished using 5 mM tris 2-(carboxyethyl)phosphine (TCEP) for 30 min at room temperature, and alkylation was performed with 10 mM iodoacetamide (IAA) for 30 min at room temperature in the dark. Overnight digestion with trypsin was accomplished with a 1:50 (w:w) trypsin:total protein ratio.

Automated proteomic sample preparation

The automated sample preparation workflow was executed on a Biomek FXp system (Beckman-Coulter, Indianapolis, IN) with integrated SpectraMax Paradigm Multi-mode microplate reader (Molecular Devices), Sigma centrifuge (Model 6-16K) (Sigma-Aldrich, St. Louis, MO), and cytomat incubator and plate hotel units (ThermoFisher Scientific, Waltham, MA). The workflow consists of ten steps: cell biomass preparation and lysis, protein precipitation, remove metabolites, remove lipids, protein resuspension, protein quantification, protein normalization, and digestion with trypsin (Supplemental table 1). The cell biomass was prepared by resuspending cells in water, transferring the cells into a 96-well protein extraction plate, pelleting the cells by centrifugation for 10 minutes, and subsequent removal of the water. The protein was extracted by transferring methanol and chloroform in a 4:1 (v:v) ratio into the protein extraction plate with the cells with mixing to promote cell lysis. Next, water was added to the solution with mixing followed by centrifugation (5 minutes, 1000 x g) to promote phase separation. The top layer (methanol-water) was then discarded or filtered through a 0.45 µm filter plate and stored for potential metabolite

analysis at a later time. An additional aliquot of methanol was added to the protein extract followed by centrifugation (1 minutes, 100 x g) to pellet the protein at the bottom of the well. At this point the remaining methanol and chloroform mixture was either discarded or filtered through a 0.45 µm filter plate and stored for lipid analysis if so desired. The protein pellet was resuspended by addition of 100 mM AMBIC supplemented with 20% methanol. Initial efforts to resuspend the precipitated protein resulted in poor protein quantification due to weaker aspiration with the automated liquid handler than what is possible via manual aspiration. To achieve better protein resuspension, we performed 80 cycles of pipetting mixing at the maximum allowable aspiration and dispensing speed. The protein concentration was measured using a Bio-Rad DC protein assay kit and a spectrophotometer operated in automation mode. The concentration of each sample was communicated from the plate reader to the liquid handler software and the volume necessary to transfer 50 µg of protein was calculated and transferred to a new plate on the deck, the concentration was normalized to 1 µg/µL by adding 100 mM AMBIC for subsequent trypsin digestion. A one-pot trypsin digestion mixture of TCEP, IAA and trypsin was subsequently added into normalized protein solution to final concentrations of 5 mM, 10 mM, and 1:100 (w:w) ratio, respectively. Finally, the deck of the liquid handling system was cleaned up and all used tip boxes and solvents were removed. Once trypsin was added the digestion was carried out either by a programmed thermocycler (not integrated with the automation platform), or sealed and transported to the integrated cytomat incubator at 37 °C for four hours or overnight.

MRM methods and UHPLC-MS data acquisition

Targeted MRM methods were developed with the assistance of in-house spectral libraries of microbes.[18] MRM selection criteria excluded peptides with Met/Cys residues, tryptic peptides

followed by additional cut sites (KK/RR), and peptides with proline adjacent to K/R cut sites. All possible doubly charged peptides were screened for y-series ions to establish the peptide identity and the most sensitive transitions. The MRM targeted proteomic assays were performed on an Agilent 6460QQQ mass spectrometer system coupled with an Agilent 1290 UHPLC system (Agilent Technologies, Santa Clara, CA). Mobile phase A consisted of 0.1% FA (Thermo Scientific) in LC-MS grade water (Burdick & Jackson, Muskegon, MI), and mobile phase B consisted of 0.1% FA in LC-MS grade acetonitrile (Burdick & Jackson). Twenty (20) µg peptides were separated on an Ascentis Express Peptide C18 column [2.7 mm particle size, 160 Å pore size, 5 cm length × 2.1 mm inside diameter (ID), coupled to a 5 mm × 2.1 mm ID guard column with the same particle and pore size, operating at 60 °C; Sigma-Aldrich, St. Louis, MO] operating at a flow rate of 0.4 mL/min via gradients depending on the organism of our investigations. To comprehensively evaluate the variance of automated sample preparation platform for microorganism we investigated, highly multiplexed assays that target hundreds of peptides were developed in either a single, scheduled 25 min gradient UHPLC-MRM-MS analysis for *E. coli* or a single, scheduled 30 min gradient UHPLC-MRM-MS analysis for the other three microorganisms. For the 25-min LC run, peptides were loaded to the column equilibrated with 5% B and hold for 0.6 min, followed by a linear gradient elution to 35% B over 20.4 min. The column was washed at 80% B for two minutes, and then equilibrated to 5% B for 1.5 min before loading next sample. For the 30-min LC run, peptides were loaded to the column equilibrated with 4% B and hold for 1 min, followed by a linear gradient elution to 40% B over 20 min. The column was washed at 90% B for 2 min, and then equilibrated to 4% B for 4 min before loading next sample. The eluted peptides were ionized via an Agilent Jet Stream ESI source operating in positive ion mode with the following source parameters: gas temperature = 250 °C, gas flow = 13 L/min,

nebulizer pressure = 35 psi, sheath gas temperature = 250 °C, sheath gas flow = 11 L/min, capillary voltage = 3500 V, nozzle voltage = 0 V. The transitions are acquired in dynamic MRM mode with scanning retention time window of 0.6 minutes and a total cycle time of one second. All MRM transitions *m/z* values are included in Supplemental tables 2-5. The data were acquired using Agilent MassHunter version B.08.02. MRM methods were generated with and all acquired data were analyzed by Skyline software version 4.20 (MacCoss Lab Software) and are available along with additional method details on Panoramaweb (https://panoramaweb.org/automated-sample-prep.url). The MRM peptide quantitation results were combined into total peak area per protein. The data were analyzed by using a custom python script to quantify and visualize error. For each dataset, mean, variance, and coefficient of variation were computed for replicates. The data were visualized with violin plots for each condition showing the distribution of the coefficients of variation across all proteins and peptides in the data set. Scatter plots were created that show the relationship between mean and coefficient of variation for all proteins and peptides in the data set. The primary dataset was visualized by using Plotly for violin and box-and-whisker plots.

RESULTS & DISCUSSION

Automated bottom-up proteomic protocol for microbes

As the number of proteomic samples increases, traditional manual proteomic sample preparation procedures become a bottleneck in terms of resources and data quality. Thus, efforts to automate liquid transfer steps of cell lysis, protein extraction, and digestion are attractive to both minimize variation and improve resource utilization. While several automated sample preparation protocols have been developed previously, they utilize detergents or chaotropes that must be removed prior to LC-MS analysis. Consequently, one goal of this work was to minimize the steps needed to

remove compounds that are used for cell lysis but are detrimental to LC-MS analysis. The chloroform-methanol (C:M) protocol[19] effectively lyses and precipitates protein from a wide variety of microbial organisms and it is routinely used for manual sample preparation of proteomic samples from Gram-negative bacteria.[4,17,20] For this automated workflow we established the utility and reproducibility of the C:M protocol for a plate-based proteomic sample preparation workflow (Figure 1A) applied to both Gram-negative bacteria and, by pretreatment with zymolyase to digest the cell wall glucan, fungal species such as *S. cerevisiae* and *R. toruloides*. One of the primary challenges of applying the C:M protocol in a plate-based format is isolation of the protein that precipitates at the interface between the chloroform and methanol layers. However, by removing the upper methanol-water layer with the automated liquid handling system (Figure 1B), followed by the addition of a second aliquot of methanol the protein pellet settles to the bottom of the plate. This allowed facile removal of the methanol-chloroform mixture via liquid handling systems, thereby simplifying subsequent processing steps. Here, the precipitated protein was resuspended and all 96 wells were quantified by using a Lowry-based protein quantification method via a plate reader integrated into the liquid handler system (Figure 1A). After the concentration of protein in each well was measured, fifty micrograms of protein was transferred automatically to a new 96-well plate for reduction with TCEP, blocking of cysteine moieties by using IAA, and digestion with trypsin. Sample preparation of cells subjected to the C:M protocol described above take two hours from the beginning of the lysis step to the addition of trypsin (Figure 1C) for one 96-well plate. The entire workflow takes 377 minutes (6.3 hours) to prepare 384 samples (four 96-well plates) from cell biomass to beginning of tryptic digestion.

Figure 1

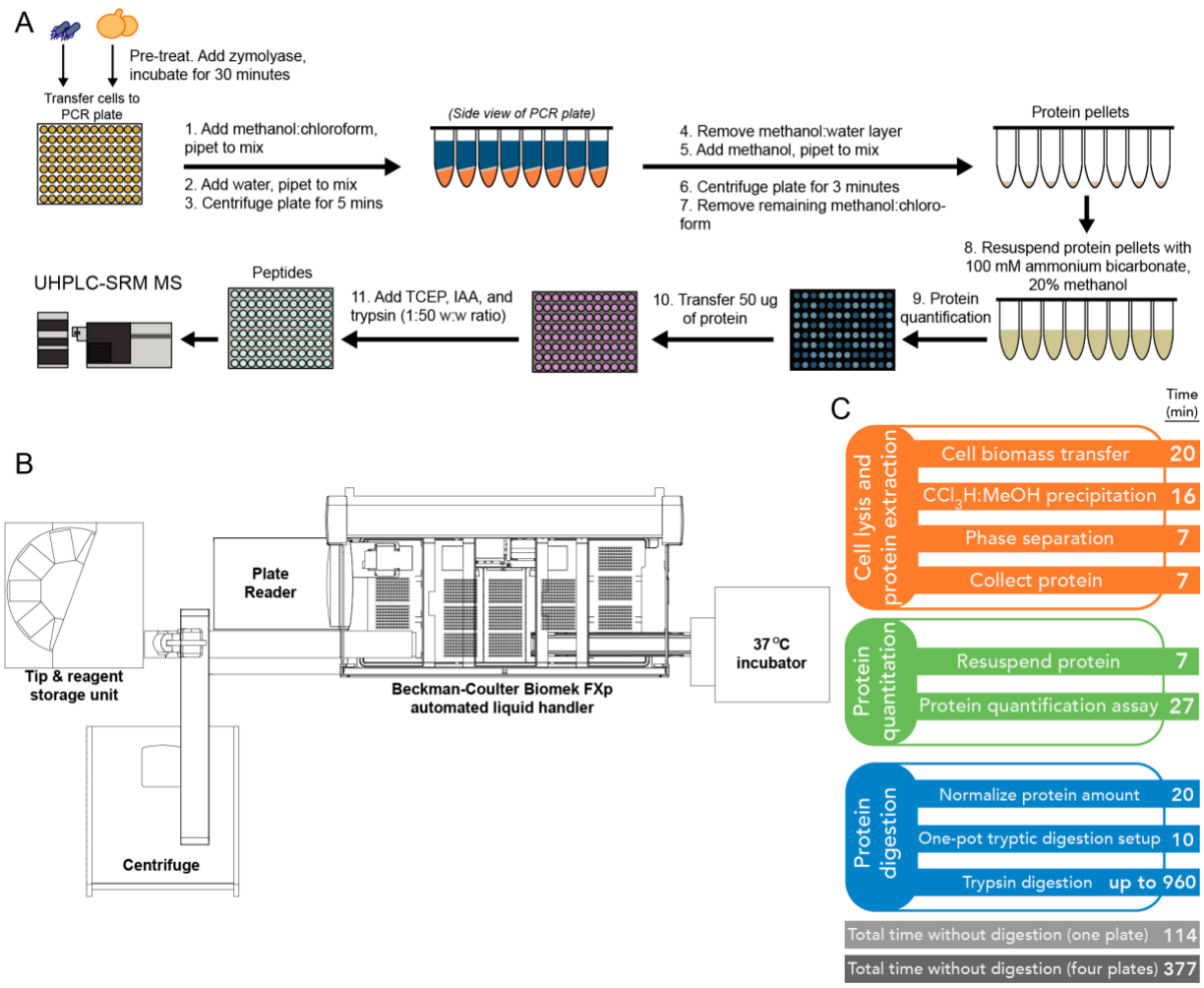


Figure 1. Automated sample preparation protocol. (A) Automated sample preparation steps; (B) Automated liquid handling system layout; (C) Timeline of the full workflow. Detailed times and reagents for each step can be found in Supplemental table 1.

Optimization of the automated 'cells-to-peptides' protein extraction process

To evaluate the effectiveness of C:M precipitation on variable amounts of cellular biomass in plate format, a single *E. coli* culture was grown overnight at 37 °C in LB media, diluted to different cell

densities, and distributed across a 96-well PCR plate. Estimates based on optical density can be highly variable while accurate cell count procedures are low throughput, consequently, for this comparison, cell biomass corresponding to 0.5, 1.0, 2.0, and 3.0 OD*mL (1 OD = ~ 5 x 10_8 *E. coli* cells) of culture were aliquoted into 96-well deep plates (n = 96 for each amount of cells). The cells were centrifuged and the supernatant was decanted by inverting the plate, after which the cells were subjected to automated C:M cell lysis, resolubilization with a mixture of 100 mM AMBIC/methanol (80%/20%), followed by protein quantification. The amount of protein extracted followed a near linear increase from the different amounts of cells with the median protein amounts of 18, 56, and 112 µg/well from 0.5, 1.0, and 2.0 OD*mL, respectively (Figure 2A). The protein recovery efficiency of the automated protocol was assessed by comparison to the total extractable protein by manual sample preparation (Figure 2B) of 0.5, 1.0, and 2.0 OD*mL of cells prepared in plate format with multi-channel pipettes. Automated cell lysis and protein precipitation experiments (n = 24) indicated that more protein was lost for 0.5 and 2.0 OD*mL and above of cells likely due to poor pellet formation or due to protein remaining in the methanol-water solution and removal via pipetting. Given the plate format and the volumes of the reagents selected, the most efficient process was observed for 1.0 OD*mL of cells. However, to ensure a sufficient amount of peptide digest for data acquisition and minimize liquid transfer variation, we selected 2.0 OD*mL of *E. coli* cell biomass for subsequent experiments. Likewise, resolubilization of *R. toruloides* and *P. putida* proteins showed similar amounts and variability to *E. coli* indicating that this protocol is applicable to a wide variety of microbial organisms. This shows that performing the C:M protocol and resuspending the protein under these conditions yields highly reproducible amounts of protein that is sufficient for common nano-flow and standard-flow LC-MS methods. Optimization of the workflow for smaller or larger initial biomass amounts could be

achieved by increasing centrifugation force and time, or by adjusting the volume of methanol and chloroform used for the protein extraction step as well as adjusting the height of the pipetting step to remove the methanol-water layer.

Figure 2

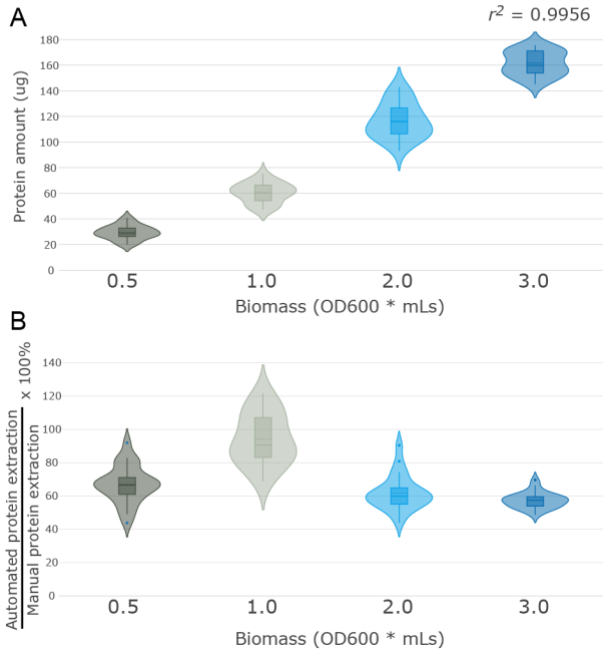


Figure 2. Comparison of automated sample preparation workflow protein extraction (A) and protein extraction efficiency (automated/manual process; B) from different amounts of *E. coli* cell biomass. The cell biomass amounts and their average protein amount yields fit a simple linear regression model of $r_2 = 0.9956$.

Reproducibility of the automated workflow measured by targeted proteomic assays

Once the optimized automated C:M method was established, we assessed the reproducibility of the automated workflow against expert and non-expert manual sample preparation protocols via targeted proteomic assays. For comparison, the same amount *E. coli* biomass (2 OD*mL) was distributed across a plate (96 replicate samples) and prepared either via the automated workflow

or manually by using multi-channel pipettes by an expert in proteomics sample preparation that had experience preparing over 2000 samples with the C:M protocol and by a non-expert who had experience preparing over 100 samples with the C:M protocol (Figure 3A). Fifty micrograms of protein from the automated or manual C:M extraction protocols were digested at a final concentration of 1 µg/µL and 20 µL were loaded onto a reverse phase C18 chromatography column connected to an UHPLC-QQQ system operating in dynamic MRM mode. The MRM method (available at: https://panoramaweb.org/automated-sample-prep.url) targeted 600 peptides from 367 *E. coli* proteins known to be expressed under these culture conditions. Among the three methods, we observed that the automated workflow produced similar variance for peptide measurements to that of the expert whereas the highest variance data was produced by the nonexpert. The median coefficient of variation for the automated workflow was 15.8% and 13.8% for peptide and protein measurements, respectively (Figure 3B). Importantly, 75% of the total peptide CVs were below 20%. As expected, the majority of peptides and proteins with high variance had the lowest intensity and an overall trend of decreasing variance with increasing signal intensity was observed (Figures 3C, 3D), suggesting that low signal intensity was the main factor in high technical variability. Even though the automated sample preparation workflow showed slightly higher variance overall it achieved similar performance as experts in the sample preparation field. The inter-day reproducibility of the automated sample preparation platform was evaluated by assaying plate samples that were prepared on three separate days. The median CVs of 600 peptides from these separate assessments ranged from 15.8 to 19.2%, suggesting that use of automation for sample processing provides good day-to-day consistency. (Supplemental figure 1).

Figure 3

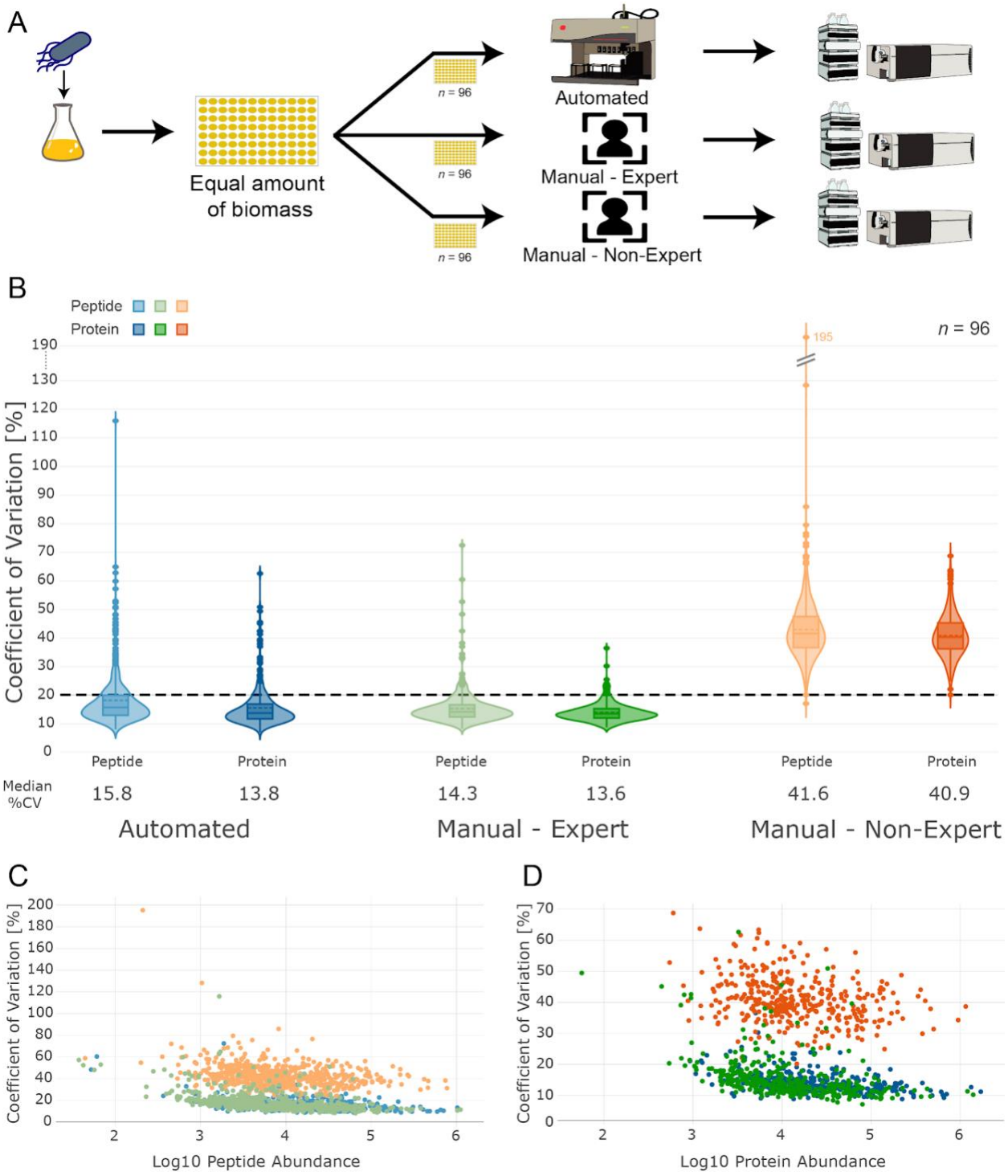


Figure 3. Reproducibility of the automated and manual sample preparation workflows as measured by UHPLC-MRM-MS data acquisition. (A) Experiment design and (B) violin and scatter plots of the percent coefficient of variation (% CV) for 600 peptides (C) from 367 *E. coli*

proteins (D). The violin plots illustrate the kernel density estimation of the CV of the automated, manual-expert, and manual-non-expert workflows. Inside each violin plot is a box plot summarising ranges (IQR, whiskers) and individual medians (solid lines), averages (dashed lines) and the black dotted line indicates the 20% CV value. The scatter plots plot the % CV for each peptide and protein (*y*-axis) versus the log10 of the peptide intensity (*x*-axis) (E). The total amount of protein loaded on the UHPLC-MS system were normalized during sample preparation but the peptide counts were not normalized post-acquisition.

Automated protein extraction from industrially favored bacteria *P. putida* demonstrated another successful application of the automated platform on Gram-negative bacteria. The extension of its application on other organisms, such as *S. cerevisiae* and *R. toruloides* in our tests, requires additional procedure of weakening their cell wall structure and more starting biomass. Similarly, we distributed cell biomass from single cultures of these microorganisms, and investigated the variances presented in the peptide samples that were prepared through the automated workflow. Highly multiplexed MRM assays were developed for targeting 340, 305, and 401 peptides in *P. putida*, *S. cerevisiae*, and *R. toruloides*, respectively. Our results showed similar variance distributions on these microorganisms in comparison to what we have observed in the tests of *E. coli* cell culture (Figure 4). As observed for *E. coli*, low peptide signal intensity was the main factor in high technical variability. The peptide median CV achieved at 11.6%, 14.7%, and 15.8% for *P. putida*, *S. cerevisiae* and *R. toruloides*, respectively. The similar CV distributions observed for these organisms suggests that the automated sample preparation workflow is applicable for a wide variety of applications.

Figure 4

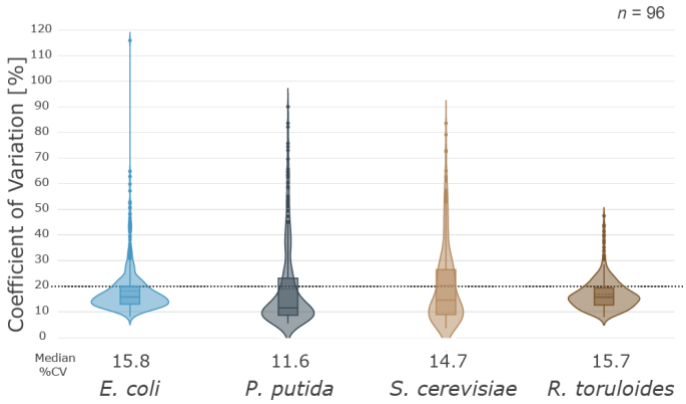


Figure 4. Reproducibility of the automated sample preparation workflow as measured by UHPLC-MRM-MS data acquisition for four different organisms. (A) Violin plots of the %CV for 600, 340, 305, 401 peptides from *E. coli*, *P. putida*, *S. cerevisiae*, *R. toruloides* proteins, respectively (n = 96). The total amount of protein loaded on the UHPLC-MS system were normalized but the peptide counts were not normalized post-acquisition.

Identification of sources of variance in the automated sample preparation workflow

To identify the sources of variance across the workflow we isolated several steps in the process and assessed the error of each one. As shown in Figure 2 above, we observed protein extraction variations from distributed biomass samples in plates, however, this variation should have no contribution to the overall variance since the amount of protein that was digested for LC-MS acquisition was normalized after determining the concentration. Consequently, the variance is likely due to the protein quantification, normalization, and the tryptic digestion steps. We tested the variability of the protein quantification step by distributing two concentrations (1.9 µg/µL and 4.8 µg/µL), that were determined manually via the DC protein quantification assay of a whole-cell lysate protein mixture produced from a single culture, across two 96-well plates and subjected them to the automated protein quantitation process. The same protein samples were quantified

manually by using the same Bio-Rad DC assay (n = 36 each). We observed CVs of 2.0% and 3.5% with the manual assay and CVs of 5% and 7% with the automated protocol for the 1.9 and 4.8 µg/µL samples, respectively. When performing the protein resuspension step manually, it is easy to meticulously resuspend the protein with extra pipetting force and time, thus, leading to more accurate protein quantification and normalization steps. This factor likely contributes to the small CV distribution difference observed between the samples prepared by human expert and the automated workflow above. Future workflow optimization efforts will focus on improving this step in the automated workflow.

To investigate the possible sources of the observed variation in the tryptic digestion step, we distributed 50 µg of protein, that was extracted from a single *E. coli* cell culture, across a 96-well plate and initiated the automated workflow from the trypsin digestion step (Figure 5A). By using the same MRM method that targeted 600 peptides to assess the total variance from tryptic digest to LC-MS acquisition procedures the median peptide CV for the trypsin digestion components was 8.7% (Figure 5B) with over 89% of peptides CV under 20%. This error along with approximately 7% of the error observed in the complete automated workflow coming from the protein quantification and normalization steps accounts for essentially all of the variability (16.1% for peptides) in the process.

Figure 5

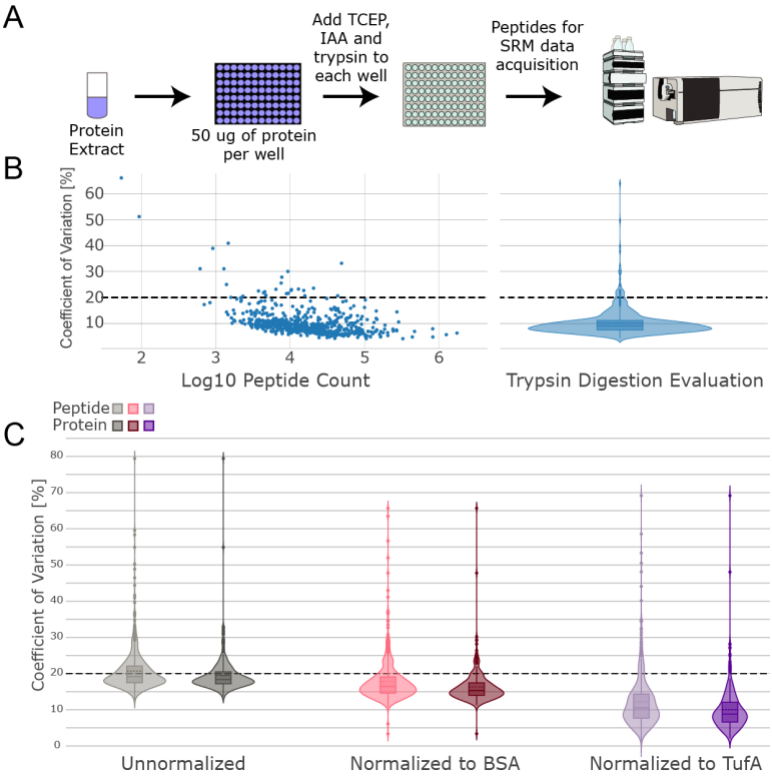


Figure 5. Identification of the sources of variation in the automated sample preparation process. (A) Tryptic digest experimental design and (B) scatter and violin plots of the % CV for 600 peptides from 367 *E. coli* proteins (n = 96). (C) Comparison of un-normalized, BSA-normalized, and TufA-normalized % CVs from the automated sample preparation workflow for 600 peptides from 367 *E. coli* proteins (n = 96). The total amount of protein loaded on the UHPLC-MS system were normalized and the peptide counts were normalized post-acquisition to internal standards as indicated.

A common practice to improve the data acquisition accuracy and reduce variance is to introduce internal standards into the samples at different steps of the proteomic sample preparation process.[21] These standards could be stable-isotope labeled proteins, stable-isotope labeled synthetic peptides, or an endogenous protein that is considered to be constant in the tested samples. We tested the

normalization effect to internal standards for the automated workflow at two points in the process. First, by spiking a constant amount of peptides from a heterologous protein, tryptically digested peptides of bovine serum albumin (BSA), into each sample at the end of the workflow prior to data acquisition to determine the contribution of the LC-MS analysis to the observed variation. Second, by normalizing the data to an endogenous *E. coli* protein, Elongation factor Tu-1 (TufA), to account for variance across the entire automated workflow. The peak areas of 600 peptides, as measured from 96 replicate samples prepared by the automated workflow, ranged from 10^3 to $>10^6$ counts with median CVs of 19% (59% below 20% CV), 16% (80%), and 11% (89%) for un-normalized, BSA-normalized and TufA-normalized, respectively (Figure 5C). This improvement of data quality could be observed in all the tests we have presented above suggesting that introducing internal standards from the beginning of the automation process greatly improves the data acquisition accuracy and reduce technical variance. This workflow is well-suited for Qualis-SIS[22] or single-point external reference material[23] data acquisition methods that have been developed for multiplexed, quantitative proteomic assays. Furthermore, the precision (<20% CV) of this workflow is suitable for targeted peptide measurements across Exploratory studies (Tier 3) and, with appropriate isotopically-labeled internal standards, for Clinical Bioanalysis/Diagnostics (Tier 1) and quantitative Research-use (Tier 2) studies as proposed by Carr *et al*.[24]

Reproducibility of protein levels from cells cultured in different formats

Cultivation in multi-well plate formats are a useful way to facilitate large experimental designs and increase sample throughput. While growth has been shown to be reproducible in various plate-based culture conditions, the variance in the proteomes of cells from different formats has not been comprehensively characterized. We tested whether culturing cells in low volume growth format,

such as 96-deep well plate or BioLector[25,26] flower plate, would increase the variance among replicates in comparison to culturing in 5 mL test tube culture conditions. To assess this question, we grew *E. coli* BW25113 cells in LB medium for 24 h in three culture platforms and prepared peptides from them with the automated workflow (Figure 6A). The CVs of 357 *E. coli* proteins in 18 biological replicates after normalization to TufA showed overall low technical variance but slightly different at three cultivation platforms (Figure 6B). Test tube platform achieved the best median CV at 5.4%, followed by 96-deep well plate at 5.5%, and BioLector plate at 11.4% sequentially. Notably, over 95% proteins measured in test tube and 96-deep well plate platforms achieved CV below 20%. The three cultivation platforms were well separated in PCA based on their peptides quantity differences (Figure 6C), suggesting there are differences between the proteomes of the three cultivation platforms. However, there was a relatively high correlation between them with no correlation dropping below 0.85 (Figure 6D) across cultivation condition and extremely high (>99%) correlation within replicates of a given cultivation type.

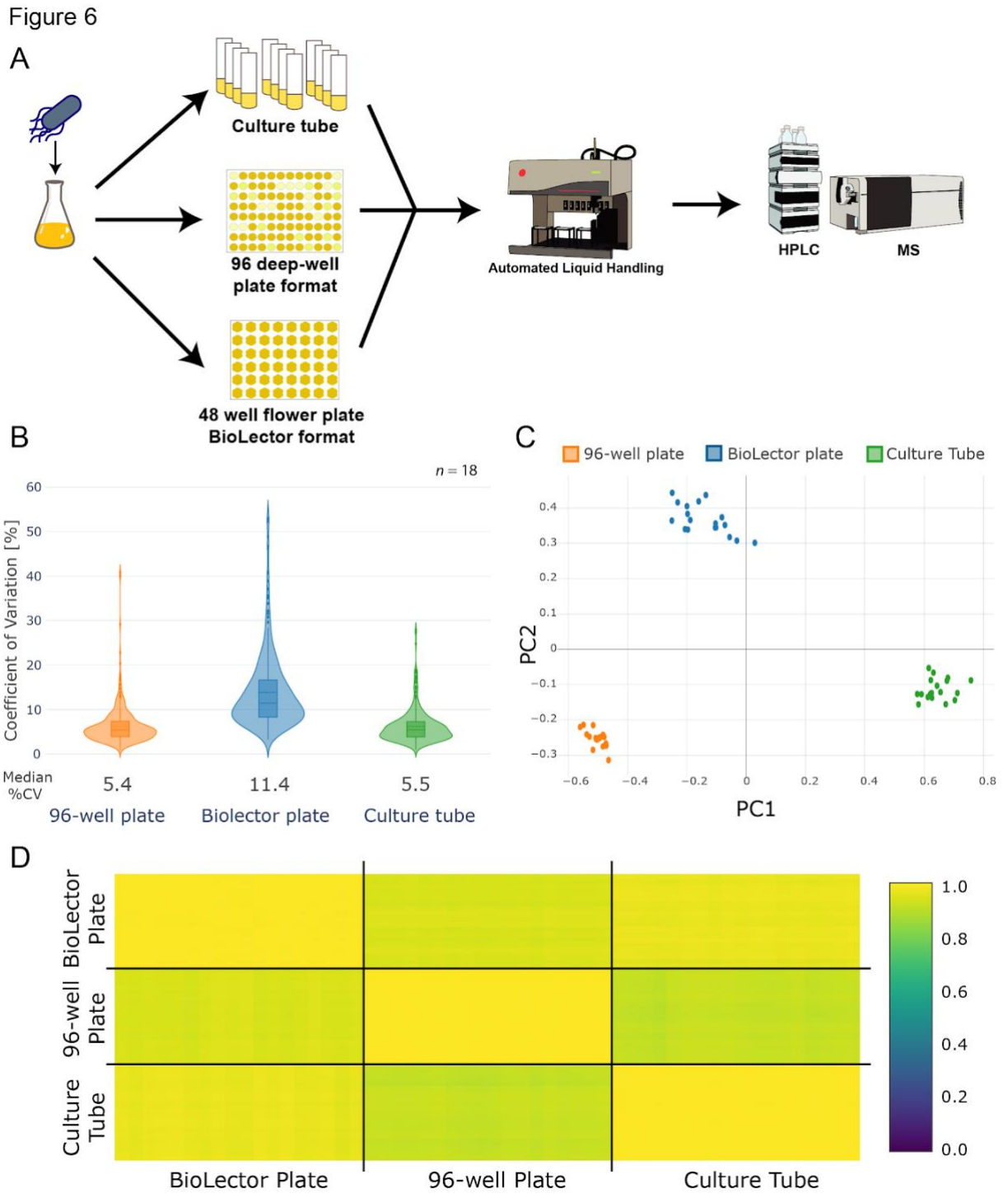


Figure 6. Variation analysis of culture condition samples (A) experiment design and (B) violin plots of the % CV for 367 *E. coli* proteins (n = 18). (C) Principal-component analysis (PCA) demonstrates that protein expression for replicates in the same culture condition cluster tightly

together. (D) A pairwise correlation analysis on replicates was performed to further determine how replicates relate to one another within and between formats. In all cases the determined correlation coefficient was above 0.85 and showed good agreement between formats. The total amount of protein loaded on the UHPLC-MS system were normalized to 20 µg and the peptide counts were normalized to TufA abundance post-acquisition.

The correlation analysis showed that the BioLector samples agree slightly more with culture tubes than does the 96-well plate format. The lower correlation may result from physiological differences during biological growth, or micro-cultivation condition differences in the incubator.[27] Pairwise comparison between plate formats and test tube identified 127 genes as differentially expressed by more than 2-fold with statistical significance (*p*-value = 0.05; Supplemental table 2). Interestingly, the majority of these genes (121) are down-regulated and are mostly identified in BioLector plate samples. Functional enrichment analysis (DAVID[28] version 6.8) of these down-regulated genes against the background of 367 measured proteins indicated that proteins involved in 'protein synthesis and refolding' were highly enriched, including FKBP-type 22 kDa peptidyl-prolyl cis-trans isomerase (FklB), FKBP-type peptidyl-prolyl cis-trans isomerases (FkpA, SlyD), and trigger factor (Tig) as the highest scored enriched group. Additionally, the proteins from central carbon metabolism, particularly pentose phosphate pathway genes, including glucokinase (Glk), transketolase 1 (TktA), transketolase (TktB), transaldolase A (TalA), transaldolase B (TalB), fructose-bisphosphate aldolase class 1 (FbaB), and ATP-dependent 6-phosphofructokinase isozyme 2 (PfkB) were enriched. Overall, our data suggest that the high-throughput cultivation platforms contribute negligible variance to proteome measurements, thus are as suitable to conduct experiments as using traditional test tube.

CONCLUSIONS

Here, we present an automated proteomics workflow for cell lysis, protein precipitation, robust tryptic digestion (cells-to-peptides) and data acquisition by LC-MRM mass spectrometry. Cell lysis and protein precipitation by using the chloroform-methanol protocol is amenable to parallel processing via automated liquid handling systems for a variety of cell types and efficiently isolates proteins from cellular lipids and metabolites. We demonstrated that the workflow is effective for quantitative proteomic studies of Gram-negative bacteria (*E. coli*, *P. putida*) and fungi such as *S. cerevisiae* and *R. toruloides* in plate format. The automated workflow was tested by preparing 96 *E. coli* samples and quantifying over 600 peptides that resulted in a median CVs of 15.8%. Further testing identified the protein resuspension and quantification steps as the biggest contributors to the observed variability. Importantly, the consistent performance of the automation platform would alleviate manual labor, risk of repetitive stress injuries, and the overall stress level of researchers that come from preparing thousands of proteomic samples.

The time required for the workflow can be optimized for specific experimental needs or resource availability. By using data acquisition methods of ten minutes total duration[18], the workflow supports a throughput of 144 samples per day per instrument. While we applied this workflow to targeted proteomic methods, it is also useful for comparative shotgun proteomic experiments. This automated workflow can be modified to include recovery of the culture supernatant prior to C:M mediated cell lysis permits analysis of secretome samples for protein production applications or integrated into fully-automated cell culturing, sampling, and sample preparation experiments. Even though this workflow utilizes a highly-specialized liquid handling system, the basic pipetting

steps can be implemented on many off-the-shelf automation platforms while the centrifugation and protein quantification steps can be done manually. Future work will include extending the workflow to Gram-positive bacteria and mammalian cells, reducing variation in the protein quantification step, and optimizing the metabolomic and lipidomic extractions so they can be paired with proteomics to provide a systems-level view of the microbe. Overall, this automated sample preparation workflow provides robust, reproducible proteomic samples for high-throughput applications.

ASSOCIATED CONTENT

The Supporting Information is available free of charge on the ACS Publications website. All generated skyline files are available in the Panorama Public repository at this link: https://panoramaweb.org/automated-sample-prep.url and data are available via ProteomeXchange with identifier PXD014182.

SUPPORTING INFORMATION

The following supporting information is available free of charge at ACS website

http://pubs.acs.org

Supplemental figure 1 (PDF)

Six supplemental tables (MS excel)

Figure S1. Violin plots of the inter-day variability of the automated workflow

Table S1. Detailed times and reagents for each step of the automated workflow

Table S2. MRM transition parameters for *E. coli*

Table S3. MRM transition parameters for *P. putida*

Table S4. MRM transition parameters for *S. cerevisiae*

Table 52. MRM transition parameters for *R. toruloides*

Table S6. Up-regulated and down-regulated proteins in cultivation platforms

AUTHOR INFORMATION

Corresponding author

E-mail: cjpetzold@lbl.gov

Present Addresses

‖Calico Biosciences, South San Francisco, CA, USA

ǂLabcyte Inc., San Jose, CA, USA

Author Contributions

Conceptualization, Y.C. and C.J.P.; Investigation, Y.C., J.M.G., L.J.G.C., T.O., H.M.T.., J.M.B.-H., and J.W.G.; Formal Analysis, Z.C., and Y.C.; Visualization, Y.C., Z.C., and C.J.P.; Supervision, J.D.K., H.G.M., N.J.H., and P.D.A.; Writing - Original Draft, Y.C. and C.J.P;

Writing - Review & Editing, all authors. All authors have given approval to the final version of the manuscript.

Notes

J.D.K. has financial interests in Amyris, Lygos, Constructive Biology, Demetrix, Napigen and Maple Bio.

ACKNOWLEDGMENTS

The authors thank Mark Forrer, Mark Kulawik, and William Morrell for help with informatics and data sharing. The proof-of-concept work and resources were part of the Joint BioEnergy Institute (JBEI; http://www.jbei.org) and extension of the procedure for *Rhodospiridium toruloides* and identification of the sources of error were part of the Agile BioFoundry (ABF; http://agilebiofoundry.org) supported through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy,

completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. HGM was also supported by the Basque Government through the BERC 2018-2021 program and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2017-0718.

ABBREVIATIONS

REFERENCES

(1) Zhu, W.; Smith, J. W.; Huang, C.-M. Mass spectrometry-based label-free quantitative proteomics. *J. Biomed. Biotechnol.* **2010**, *2010*, 840518.

(2) Shi, T.; Fillmore, T. L.; Sun, X.; Zhao, R.; Schepmoes, A. A.; Hossain, M.; Xie, F.; Wu, S.; Kim, J.-S.; Jones, N.; et al. Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 15395–15400.

(3) Cominetti, O.; Núñez Galindo, A.; Corthésy, J.; Oller Moreno, S.; Irincheeva, I.; Valsesia, A.; Astrup, A.; Saris, W. H. M.; Hager, J.; Kussmann, M.; et al. Proteomic Biomarker Discovery in 1000 Human Plasma Samples with Mass Spectrometry. *J. Proteome Res.* **2016**, *15*, 389–399.

(4) Brunk, E.; George, K. W.; Alonso-Gutierrez, J.; Thompson, M.; Baidoo, E.; Wang, G.; Petzold, C. J.; McCloskey, D.; Monk, J.; Yang, L.; et al. Characterizing Strain Variation in Engineered E. coli Using a Multi-Omics-Based Workflow. *Cell Syst.* **2016**, *2*, 335–346.

(5) Aebersold, R.; Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **2016**, *537*, 347–355.

(6) Fu, Q.; Kowalski, M. P.; Mastali, M.; Parker, S. J.; Sobhani, K.; van den Broek, I.; Hunter, C. L.; Van Eyk, J. E. Highly reproducible automated proteomics sample preparation workflow for quantitative mass spectrometry. *J. Proteome Res.* **2018**, *17*, 420–428.

(7) Switzar, L.; van Angeren, J.; Pinkse, M.; Kool, J.; Niessen, W. M. A. A high-throughput sample preparation method for cellular proteomics using 96-well filter plates. *Proteomics* **2013**, *13*, 2980–2983.

(8) Yu, Y.; Bekele, S.; Pieper, R. Quick 96FASP for high throughput quantitative proteome analysis. *J. Proteomics* **2017**, *166*, 1–7.

(9) Dayon, L.; Núñez Galindo, A.; Corthésy, J.; Cominetti, O.; Kussmann, M. Comprehensive and Scalable Highly Automated MS-Based Proteomic Workflow for Clinical Biomarker Discovery in Human Plasma. *J. Proteome Res.* **2014**, *13*, 3837–3845.

(10) Lee, J.; Kim, H.; Sohn, A.; Yeo, I.; Kim, Y. Cost-Effective Automated Preparation of Serum Samples for Reproducible Quantitative Clinical Proteomics. *J. Proteome Res.* **2019**, *18*, 2337–2345.

(11) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–362.

(12) Arul, A.-B.; Byambadorj, M.; Han, N.-Y.; Park, J. M.; Lee, H. Development of an Automated, High-throughput Sample Preparation Protocol for Proteomics Analysis. *Bull. Korean Chem. Soc.* **2015**, *36*, 1791–1798.

(13) Specht, H.; Harmange, G.; Perlman, D. H.; Emmott, E.; Niziolek, Z.; Budnik, B.; Slavov, N. Automated sample preparation for high-throughput single-cell proteomics. *BioRxiv* **2018**.

(14) Potriquet, J.; Laohaviroj, M.; Bethony, J. M.; Mulvenna, J. A modified FASP protocol for high-throughput preparation of protein samples for mass spectrometry. *PLoS One* **2017**, *12*, e0175967.

(15) Berger, S. T.; Ahmed, S.; Muntel, J.; Cuevas Polo, N.; Bachur, R.; Kentsis, A.; Steen, J.; Steen, H. MStern Blotting-High Throughput Polyvinylidene Fluoride (PVDF) Membrane-Based Proteomic Sample Preparation for 96-Well Plates. *Mol. Cell Proteomics* **2015**, *14*, 2814–2823.

(16) 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.

(17) Batth, T. S.; Singh, P.; Ramakrishnan, V. R.; Sousa, M. M. L.; Chan, L. J. G.; Tran, H. M.; Luning, E. G.; Pan, E. H. Y.; Vuu, K. M.; Keasling, J. D.; et al. A targeted proteomics toolkit for high-throughput absolute quantification of Escherichia coli proteins. *Metab. Eng.* **2014**, *26*, 48–56.

(18) Chen, Y.; Vu, J.; Thompson, M. G.; Sharpless, W. A.; Chan, L. J. G.; Gin, J. W.; Keasling, J. D.; Adams, P. D.; Petzold, C. J. A rapid methods development workflow for high-throughput quantitative proteomic applications. *PLoS One* **2019**, *14*, e0211582.

(19) Wessel, D.; Flügge, 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.

(20) Weaver, L. J.; Sousa, M. M. L.; Wang, G.; Baidoo, E.; Petzold, C. J.; Keasling, J. D. A kinetic-based approach to understanding heterologous mevalonate pathway function in E. coli. *Biotechnol. Bioeng.* **2015**, *112*, 111–119.

(21) Bourmaud, A.; Gallien, S.; Domon, B. A quality control of proteomic experiments based on multiple isotopologous internal standards. *EuPA Open Proteom.* **2015**, *8*, 16–21.

(22) Mohammed, Y.; Percy, A. J.; Chambers, A. G.; Borchers, C. H. Qualis-SIS: automated standard curve generation and quality assessment for multiplexed targeted quantitative proteomic experiments with labeled standards. *J. Proteome Res.* **2015**, *14*, 1137–1146.

(23) Pino, L. K.; Searle, B. C.; Huang, E. L.; Noble, W. S.; Hoofnagle, A. N.; MacCoss, M. J. Calibration Using a Single-Point External Reference Material Harmonizes Quantitative Mass Spectrometry Proteomics Data between Platforms and Laboratories. *Anal. Chem.* **2018**, *90*, 13112–13117.

(24) Carr, S. A.; Abbatiello, S. E.; Ackermann, B. L.; Borchers, C.; Domon, B.; Deutsch, E. W.; Grant, R. P.; Hoofnagle, A. N.; Hüttenhain, R.; Koomen, J. M.; et al. Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol. Cell Proteomics* **2014**, *13*, 907–917.

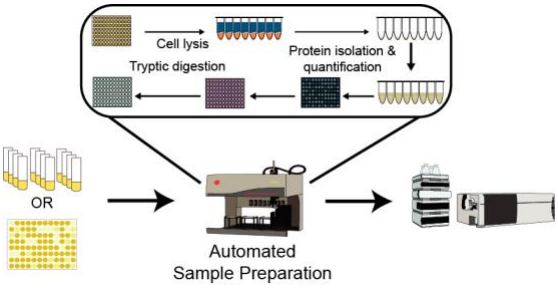
(25) Kensy, F.; Engelbrecht, C.; Büchs, J. Scale-up from microtiter plate to laboratory fermenter: evaluation by online monitoring techniques of growth and protein expression in Escherichia coli and Hansenula polymorpha fermentations. *Microb. Cell Fact.* **2009**, *8*, 68.

(26) Funke, M.; Buchenauer, A.; Mokwa, W.; Kluge, S.; Hein, L.; Müller, C.; Kensy, F.; Büchs, J. Bioprocess control in microscale: scalable fermentations in disposable and user-friendly microfluidic systems. *Microb. Cell Fact.* **2010**, *9*, 86.

(27) Kensy, F.; Zang, E.; Faulhammer, C.; Tan, R.-K.; Büchs, J. Validation of a high-throughput fermentation system based on online monitoring of biomass and fluorescence in continuously shaken microtiter plates. *Microb. Cell Fact.* **2009**, *8*, 31.

(28) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57.

For TOC Only


Cell lysis
Protein isolation &
quantification
Tryptic digestion
OR
Automated
Sample Preparation