

# Use of Hybrid Data-Dependent and -Independent Acquisition Spectral Libraries Empowers Dual-Proteome Profiling

Patrick Willems, Ursula Fels, An Staes, Kris Gevaert, and Petra Van Damme\*

Cite This: *J. Proteome Res.* 2021, 20, 1165–1177

Read Online

ACCESS |



Metrics &amp; More



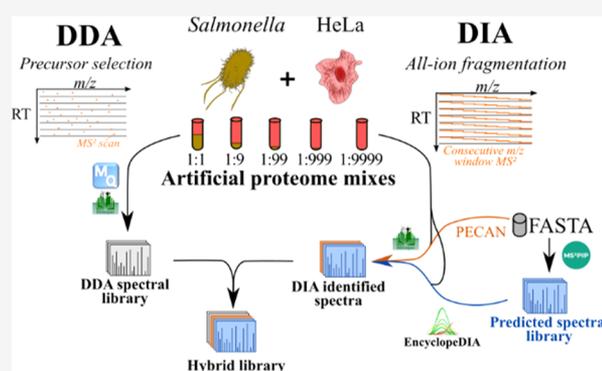
Article Recommendations



Supporting Information

**ABSTRACT:** In the context of bacterial infections, it is imperative that physiological responses can be studied in an integrated manner, meaning a simultaneous analysis of both the host and the pathogen responses. To improve the sensitivity of detection, data-independent acquisition (DIA)-based proteomics was found to outperform data-dependent acquisition (DDA) workflows in identifying and quantifying low-abundant proteins. Here, by making use of representative bacterial pathogen/host proteome samples, we report an optimized hybrid library generation workflow for DIA mass spectrometry relying on the use of data-dependent and *in silico*-predicted spectral libraries. When compared to searching DDA experiment-specific libraries only, the use of hybrid libraries significantly improved peptide detection to an extent suggesting that infection-relevant host-pathogen conditions could be profiled in sufficient depth without the need of a priori bacterial pathogen enrichment when studying the bacterial proteome. Proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD017904 and PXD017945.

**KEYWORDS:** bacterial pathogen/host interaction, data-dependent acquisition (DDA), data-independent acquisition (DIA), *Salmonella*, spectral library



## INTRODUCTION

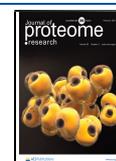
Among others, proteomics aims to identify and quantify changes in protein levels. Bearing in mind that relatively small genomes of microorganisms such as that of the bacterium *Salmonella Typhimurium* encoding approximately 4500 protein-coding genes, bacterial proteomics can be complex, as already a significant number of proteins need to be profiled. In fact, these numbers get even much higher when considering proteoforms, that is, multiple protein products arising from a single gene.<sup>1,2</sup>

The still preferred method to analyze proteomes is by means of liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS). Specifically, in bottom-up proteomic approaches, proteins are digested to peptides by proteases (e.g., trypsin) which are subsequently analyzed by LC–MS/MS. Mass spectrometers are then mostly operated in the so-called data-dependent acquisition (DDA) mode. Here, following an MS1 scan, a number of peptide ions of sufficiently high intensity are selected for further fragmentation to deliver tandem MS (MS2) spectra for sequence identification following database searching.<sup>3</sup> Peptide quantification in DDA is routinely based on the MS1 intensity of the precursor ion, which can be hampered by interference with chemical noise, resulting in a decreased dynamic range due to difficulties when trying to quantify low-abundant peptides.<sup>4,5</sup> On the other hand, in the so-called targeted proteomic approaches, for example, selected reaction

monitoring (SRM) and parallel reaction monitoring (PRM), quantifications are based on MS2 scans that are composed of multiple fragments ions, therefore being more robust and reproducible.<sup>6</sup> However, in this way, only a limited number of peptides can be quantified and thus do not permit proteome-wide discovery. In contrast, mass spectrometers operating in data-independent acquisition (DIA) mode aim to combine the broad identification power of DDA with accurate quantification available by targeted approaches.<sup>7</sup> In DIA mode, all precursors in a predetermined *m/z* isolation window are fragmented together irrespective of their abundances. This overcomes the stochastic and thus irreproducible nature of DDA which biases DDA toward the identification of (highly) abundant peptides.<sup>4</sup> Furthermore, in DIA mode, peptide quantification is performed at the MS2 level, resulting in more robust quantification versus the interference-prone MS1 quantification in DDA mode. As the direct relation between the precursor and fragment ions is lost,

Received: May 24, 2020

Published: January 19, 2021



analysis of DIA data requires however more sophisticated algorithms compared to the analysis of DDA data.<sup>8</sup>

Although spectrum-centric DDA search algorithms are not suited for the analysis of the complex MS2 spectra generated by DIA, annotated MS2 spectra and retention times (RTs) from DDA searches are widely used for constructing spectral libraries to query DIA data.<sup>9–11</sup> The sensitivity of these approaches is thus inherently limited to the stochastic limitations of DDA. Spectral libraries can be created in several ways, leading to more or less extensive libraries, with the library size influencing the accuracy and specificity of identifications and quantifications.<sup>12</sup> Besides the use of DDA-based spectral libraries, an alternative way of creating spectral libraries is by using algorithms that accurately predict MS2 fragmentation spectra such as MS<sup>2</sup>PIP<sup>13</sup> and Prosit.<sup>14</sup> MS<sup>2</sup>PIP (MS2 peak intensity prediction) is a data-driven tool<sup>13</sup> that is trained on different types of public DDA data to predict MS2 peak intensities.<sup>15</sup> Prosit<sup>14</sup> on the other hand is trained on MS2 spectra of synthetic peptides and mass spectrometry data generated in the context of the Proteome-Tools project, which has the overall aim to provide a high-quality reference MS2 data of synthetic peptides.<sup>16</sup> It is noteworthy that whole-proteome spectral libraries predicted by MS<sup>2</sup>PIP and Prosit have been used to search narrow-window DIA data and generate chromatogram libraries using EncyclopeDIA.<sup>17–19</sup> This DIA-only approach allows to empirically correct peptide predictions and limit the spectral library search to peptides identified in the high-sensitive narrow-window detection runs. Using DIA-only workflows with chromatogram libraries shows even better performance than DDA-based experimental libraries, thus bypassing the need for sample-specific DDA analysis.<sup>17,18</sup> Next to predicted spectral libraries, the so-called library-free approaches also overcome the stochastic limitation of DDA by querying DIA data for the best supporting evidence of peptide detection. The peptide-centric algorithm Peptide Centric Analysis or PECAN<sup>20</sup> incorporates a sequence-based RT predictor to improve the sensitivity of detection by filtering DIA data based on expected RTs. In a DIA-only workflow, PECAN queries a peptide list derived from a background proteome on DIA data, and provides output auxiliary scores of experimental and expected RTs, which are further processed by Percolator<sup>21</sup> to report confident peptide and protein identifications. Another library-free approach includes algorithms such as DIA-Umpire that perform deconvolution of DIA data to DDA-like pseudo MS2 spectra, which can then be identified with DDA-based database-searching approaches.<sup>22</sup>

In bacterial infection biology, a comprehensive mapping of the proteome profiles of both host and pathogen is needed.<sup>23</sup> Such dual-proteome profiles are currently lacking likely because of the scarce amount of protein material originating from the bacterial pathogen in most commonly used infection models. Since most intracellular bacterial replicate inside the host or bacterial viability is affected inside the host, bacteria enumeration varies during the time course of an infection. More specifically, in *in vitro* infection models using *Salmonella*-infected human epithelial HeLa cells, bacterial enumeration estimates <10 bacteria per HeLa cell at early times and up to 100 bacteria at late times in infection.<sup>24</sup> Considering these numbers, bacteria-to-host protein content varies from ~1:1000 to 1:100 during the infection process. Despite the improved sensitivity and speed of contemporary mass spectrometers, the current technologies still do not permit us to study host and pathogen proteomes simultaneously with sufficient depth, requiring prior enrichment of bacteria which can typically be attained by

selective host lysis and differential centrifugation.<sup>25,26</sup> Moreover, from the host point of view, bystander host cells can foremost be removed in order to exclusively profile the proteomes of infected host cells. This is commonly done using fluorescent bacterial reporter strains and fluorescence-activated cell sorting.<sup>27,28</sup> Of note, the elimination of host bystander contributions concomitantly enriches for bacterial content as typically only a fraction of cells get infected.<sup>29</sup>

In the present work, we aimed to establish a DIA-MS workflow to improve the overall sensitivity of protein identification and quantification of complex *Salmonella*–host mixtures containing only a fairly low amount of peptide material derived from *Salmonella* without bacterial pre-enrichment. We compared the performance of MS<sup>2</sup>PIP-predicted and DDA-based spectral libraries, and, to improve on DDA sensitivity, we extended the spectral library<sup>30</sup> by including MS2 spectra obtained from pre-fractionated LC–MS/MS DDA data of *Salmonella* grown under different infection-relevant conditions. Finally, by integrating (predicted) DDA libraries and library-independent approaches, a hybrid spectral library was created,<sup>31</sup> which could achieve an up to 2- to 3-fold increase in consistently quantified human or spiked-in *Salmonella* proteins and peptides.

## ■ EXPERIMENTAL PROCEDURES

### HeLa Cell Culture

HeLa cells (epithelial cervix adenocarcinoma, American Type Culture Collection, Manassas, VA, USA; ATCC CCL-2) were cultured in GlutaMAX containing Dulbecco's modified Eagle medium (Gibco, cat no. 31966047) supplemented with 10% fetal bovine serum (Gibco, cat no. 10270-106) and 50 units/mL penicillin and 50 µg/mL streptomycin (Gibco; cat no. 5070-063). Cells were cultured at 37 °C in a humidified atmosphere with 5% CO<sub>2</sub> and passaged at a 1:8 ratio every 4 days.

### Bacterial Strain and *Salmonella* Cultivation Conditions

The *Salmonella enterica* serovar Typhimurium wild-type strain SL1344<sup>32</sup> (Genotype: hisG46, Phenotype: His(-); biotype 26i), herein referred to as *Salmonella*, was obtained from the *Salmonella* Genetic Stock Center (SGSC, Calgary, Canada; cat no. 438). Bacterial growth was performed in liquid Lennox (L) growth medium (10 g/L Bacto tryptone, 5 g/L Bacto yeast extract, 5 g/L NaCl), Luria Beltrami (LB)-Miller broth (10 g/L Bacto tryptone, 5 g/L Bacto yeast extract, 10 g/L NaCl) or variants of phosphate carbon nitrogen (PCN) medium<sup>33</sup> (lnSPI2; pH 5.8, 0.4 mM Pi), SPI2-inducing PCN (pH 5.8, 0.4 mM inorganic phosphate) containing low levels (10 µM) of magnesium sulfate (PCN medium was stored at 4 °C and brought at room temperature for cultivation). Viewing the auxotrophic nature of the SL1344 strain used, all PCN media were supplemented with histidine to a final concentration of 5 mM. For bacterial cultivation, single colonies were picked from LB plates, inoculated in 8 mL of liquid Lennox (L) growth medium (L-broth), and grown overnight at 37 °C with agitation (180 rpm). Subsequently, the overnight cultures with an optical density measured at 600 nm (OD<sub>600</sub>) of ~4.8 were diluted 1:200 (~OD<sub>600</sub> 0.02) in T175 flasks in 50 mL of L-medium without antibiotics and grown under ten different (infection-relevant) growth conditions as reported in ref 34.<sup>34</sup> More specifically, bacteria were grown to early-exponential growth phase (EEP; OD<sub>600</sub> 0.1), mid-exponential growth phase (MEP; OD<sub>600</sub> 0.3), late-exponential growth phase (LEP; OD<sub>600</sub> 1.0), early-stationary phase (ESP; OD<sub>600</sub> 2.0), and late-stationary phase (LSP; OD<sub>600</sub> 2.0 + 6 h of extra growth). Besides, environmental

shocks in LB were performed on MEP-grown bacteria by the addition of NaCl to a final concentration of 0.3 M and continued growth for 10 min or, in the case of anaerobic shock, growth for an additional 30 min without agitation in a filled and tightly screwed 50 mL Falcon tube. For growth in variants of PCN minimal medium,<sup>33</sup> overnight-grown LB cultures were washed twice in PCN medium before resuspension at O.D.<sub>600</sub> 0.02. Cells were grown in SPI2-inducing PCN or low-magnesium SPI2-inducing PCN. The nitric oxide shock conducted in PCN (InSPI2) was performed at OD<sub>600</sub> 0.3 by the addition of the nitric oxide donor spermine NONOate to a final concentration of 250  $\mu$ M for 20 min (nitric oxide shock (InSPI2)).<sup>35</sup> Bacterial cells were collected by centrifugation (2600g, 10 min) at 4 °C and the supernatant was discarded. Samples were flash-frozen in liquid nitrogen and stored at -80 °C until further processing.

### Proteome Extractions and Sample Preparation

Cell pellets were resuspended in guanidinium chloride (Gu.HCl)-containing lysis buffer (4 M Gu.HCl, 50 mM ammonium bicarbonate (pH 7.9)) at  $5 \times 10^9$  *Salmonella* and  $1 \times 10^7$  HeLa cells per 500  $\mu$ L of lysis buffer and mechanically lysed by three rounds of freeze–thaw cycles in liquid nitrogen. The lysates were sonicated (Branson probe sonifier output 4, 50% duty cycle,  $2 \times 30$  s, 1 s pulses) followed by centrifugation (16,100g, 10 min) at 4 °C, to remove cellular debris. The protein concentration of the supernatant was determined by Bradford measurement according to the manufacturer's instructions (Bio-Rad, cat no. 5000006).

Samples mixtures were made from trypsin-digested total *Salmonella* and/or HeLa protein lysate(s). More specifically, for the infection-relevant complex *Salmonella* sample, an equimolar mix of *Salmonella* protein samples originating from *Salmonella* grown in the 10 infection-relevant conditions was made. Alternatively, in the case of complex *Salmonella*–host mixtures, *Salmonella* protein lysate (S) dilution series in protein lysates of human HeLa cells (H) (hereafter referred to as artificial mixtures) were made by mixing proteome samples prior to digestion in a 1:9 ratio and making dilutions series thereof to obtain complex S/H proteome mixtures with the corresponding *Salmonella*/HeLa protein ratios of 1:99, 1:999, and 1:9999. In addition, a sample containing equal amounts of *Salmonella* and HeLa proteins (1:1 ratio) was prepared. All spiked-in samples were prepared in triplicate. For all protein mixtures, an aliquot equivalent to 400  $\mu$ g of total protein was transferred to a 1.5 mL Eppendorf tube, twice diluted with liquid chromatography (LC)-grade water and precipitated overnight with 4 volumes of -20 °C acetone. The precipitated protein material was recovered by centrifugation (3500g, 15 min) at 4 °C, and pellets were washed twice with -20 °C 80% acetone and air-dried upside down at room temperature until no residual acetone odor remained. Pellets were resuspended in 200  $\mu$ L of TFE (2,2,2-trifluoroethanol) digestion buffer (10% TFE, 100 mM ammonium bicarbonate, pH 7.9) with sonication at 4 °C (Branson probe output 20; 1 s pulses) until a homogeneous suspension was reached. All samples were digested overnight at 37 °C using a Trypsin/Lys-C Mix (mass spec grade, Promega, Madison, WI) (enzyme/substrate of 1:100 w/w) while mixing (550 rpm). Samples were acidified with TFA to a final concentration of 0.5% and cleared from insoluble particulates by centrifugation (16,100g, 15 min) at 4 °C and the supernatant was transferred to new Eppendorf tubes. Methionine oxidation was performed by the addition of hydrogen peroxide to a final concentration of 0.5% for 30 min at 30 °C. Solid-phase

extraction of peptides was performed using a C18 reversed-phase sorbent containing 100  $\mu$ L pipette tips according to the manufacturer's instructions (Agilent, Santa Clara, CA, USA, cat no. A57003100K). The pipette tip was conditioned by aspirating the maximum pipette tip volume of water/acetonitrile (ACN), 50:50 (v/v) and the solvent was discarded. After equilibration of the tip by washing three times with the maximum pipette tip volume in 0.1% TFA in water, 100  $\mu$ L of the acidified peptide mixtures (~200  $\mu$ g) was dispensed and aspirated for 10 cycles for maximum binding efficiency. The tip was washed three times with the maximum pipette tip volume of 0.1% TFA in water/ACN, 98:2 (v/v) and the bound peptides were eluted in LC–MS/MS vials with the maximum pipette tip volume of 0.1% TFA in water/ACN, 30:70 (v/v). The samples were vacuum-dried in a SpeedVac concentrator and redissolved in 100  $\mu$ L (infection-relevant *Salmonella* peptide mixture) for subsequent reversed-phase (RP-HPLC) fractionation (see below) or, for LC–MS/MS analysis, in 50  $\mu$ L (artificial mixtures) or 20  $\mu$ L (fractionated RP-HPLC samples obtained from the infection-relevant *Salmonella* peptide mixture) of 2 mM tris(2-carboxyethyl)-phosphine (TCEP) in 2% ACN spiked with an indexed RT (iRT) peptide mix (Biognosys, Schlieren, Switzerland) according to the manufacturer's instructions<sup>36</sup> for RT prediction (see below). Samples were stored at -20 °C until further analysis.

### RP-HPLC Fractionation of the Complex *Salmonella* Peptide Mixture

The infection-relevant, complex *Salmonella* peptide mixture (100  $\mu$ L) was acidified by the addition of 5  $\mu$ L of glacial acetic acid and the peptide mixture (corresponding to an equivalent of 400  $\mu$ g of digested protein) fractionated at pH 5 using an HPLC Agilent series 1100 instrument. The sample was trapped for 16 min on a reversed-phase trapping column (35 mm  $\times$  300  $\mu$ m I.D., 5  $\mu$ m beads C18 material (Dr. Maisch, Ammerbuch, Germany), fritted and packed in-house). Next, the sample was separated on an analytical column (150 mm  $\times$  250  $\mu$ m I.D., 3  $\mu$ m beads C18 material fritted and packed in-house) using a 100 min gradient from solvent A (10 mM ammonium acetate, pH 5.5) to solvent B (10 mM ammonium acetate, 70% ACN, pH 5.5) at a constant flow rate of 3  $\mu$ L/min. The constant flow rate is achieved with an Agilent's 1100 series capillary pump in the microflow mode with the flow controller at 20  $\mu$ L/min. After the gradient, the column was run at solvent B for 5 min, switched to solvent A, and re-equilibrated for 20 min. One-minute fractions were collected in MS vials over a time interval of 65 min and automatically pooled, this being the restarting of the fraction collection cycle every 10 min, resulting in a total of 10 (pooled) fractions. Peptides were detected by absorbance at 214 and 280 nm. The fractions were then vacuum-dried in a SpeedVac concentrator and re-dissolved in 20  $\mu$ L of 2 mM TCEP in 2% ACN spiked with the iRT peptide mix as described above. Samples, referred to as *Salmonella* pre-fractionated samples, were stored at -20 °C until further analysis.

### LC–MS/MS Data Acquisition

From each artificial mixture and *Salmonella* pre-fractionated sample, 10 and 2  $\mu$ L were injected onto the column, corresponding to 2 and 4  $\mu$ g peptide material, respectively, for LC–MS/MS analysis on an Ultimate 3000 RSLC nano system in-line connected to a Q Exactive HF BioPharma mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Trapping was performed at 10  $\mu$ L/min for 4 min in loading solvent A [0.1% TFA in water/ACN (98:2, v/v)] on a 20 mm trapping column [made in-house, 100  $\mu$ m internal diameter

(I.D.), 5  $\mu\text{m}$  beads, C18 Reprosil-HD, Dr. Maisch, Ammerbuch, Germany]. After flushing from the trapping column, the peptides were loaded and separated on an analytical 200  $\mu\text{m}$  PAC column with C18-encapped functionality (Pharma-Fluidics, Belgium) kept at a constant temperature of 50  $^{\circ}\text{C}$ . Peptides were eluted using a non-linear gradient reaching 9% MS solvent B [0.1% formic acid (FA) in water/ACN (2:8, v/v)] in 15 min, 33% MS solvent B in 105 min, 55% MS solvent B in 125 min, and 99% MS solvent B in 135 min at a constant flow rate of 300 nL/min, followed by a 5 min wash with 99% MS solvent B and re-equilibration with MS solvent A (0.1% FA in water). For both analyses, a pneu-Nimbus dual-column ionization source was used (Phoenix S&T), at a spray voltage of 2.6 kV and a capillary temperature of 275  $^{\circ}\text{C}$ . For the first analysis (DDA mode), the mass spectrometer automatically switched between MS and MS2 acquisition for the 16 most abundant ion peaks per MS spectrum. Full-scan MS spectra (375–1500  $m/z$ ) were acquired at a precursor resolution of 60,000 at 200  $m/z$  in the Orbitrap analyzer after accumulation to a target value of 3,000,000. The 16 most intense ions above a threshold value of 13,000 were isolated for higher-energy collisional dissociation (HCD) fragmentation at a normalized collision energy of 28% after filling the trap at a target value of 100,000 for maximum 80 ms injection time using a dynamic exclusion of 12 s. MS2 spectra (200–2000  $m/z$ ) were acquired at a resolution of 15,000 at 200  $m/z$  in the Orbitrap analyzer.

Another 10  $\mu\text{L}$  aliquot from each artificial mixture was analyzed using the same mass spectrometer in the DIA mode. Nano LC conditions and gradients were the same as used for DDA. Full-scan MS spectra ranging from 375 to 1500  $m/z$  with a target value of  $5 \times 10^6$  were followed by 30 quadrupole isolations with a precursor isolation width of 10  $m/z$  for HCD fragmentation at a normalized collision energy of 30% after filling the trap at a target value of  $3 \times 10^6$  for a maximum injection time of 45 ms. MS2 spectra were acquired at a resolution of 15,000 at 200  $m/z$  in the Orbitrap analyzer without multiplexing. The isolation intervals ranged from 400 to 900  $m/z$  with an overlap of 5  $m/z$ .

### Processing of DDA Data

Raw data files corresponding to 10 fractions of the complex, infection-relevant *Salmonella* peptide mixture and 5 artificial dual-proteome mixtures (1:1, 1:9, 1:99, 1:999, and 1:9999 *Salmonella*/HeLa samples, in triplicates), were searched in parallel using MaxQuant<sup>37</sup> (version 1.6.10.43). Protein databases for searching the obtained spectra were either the UniProt knowledgebase (UniProtKB) proteomes for *Salmonella* pre-fractionated samples (proteome UP000008962, 4657 proteins) or the *Salmonella* proteome database concatenated to the human UniProtKB database for artificial mixtures (proteomes UP000008962 [4657 proteins] and UP000005640 [74,449 proteins]). In addition, MaxQuant built-in contaminant proteins and the 11 iRT peptide sequences (Biognosys-11) were included in the search.<sup>36</sup> Methionine oxidation to methionine-sulfoxide was set as a fixed modification, and in the case of artificial mixtures, protein N-terminal acetylation was set as variable modification. To augment peptide quantification, we performed matching-between-runs with a match time window of 1.2 min and an alignment time window of 20 min and performed label-free quantitation with the LFQ algorithm using default settings in MaxQuant. We used the enzymatic rule of trypsin/P with a maximum of two missed cleavages. The peptide-to-spectrum match level was set at 1% FDR. Protein

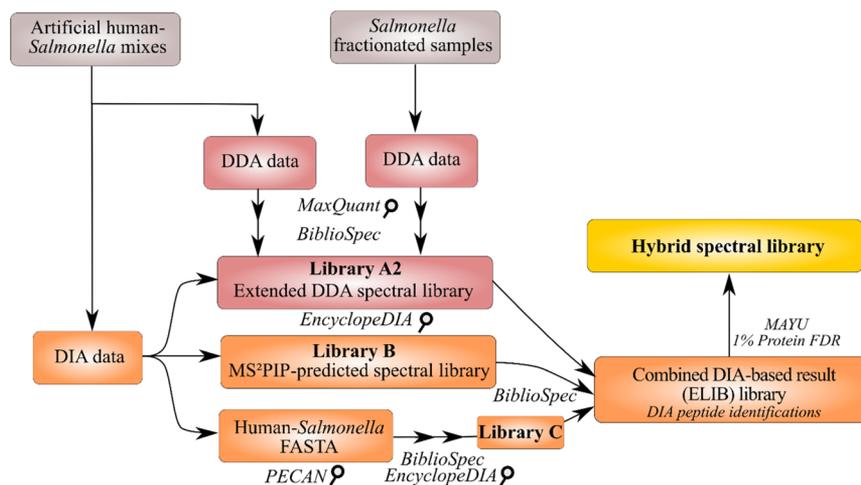
FDR—calculated by employing a reverse database strategy—was set at 1%. For protein quantification in the proteinGroups.txt file, only unique peptides were considered and all modifications were allowed. For other search parameters not specified here, default MaxQuant settings were used.

### Processing of DIA Data

**DDA-Based Spectral Library Construction.** The “msms.txt” files outputted using MaxQuant were used as input for the creation of redundant BLIB spectral libraries (artificial mixtures and *Salmonella* pre-fractionated samples) using BiblioSpec (version 2.1).<sup>38</sup> Redundant spectra were subsequently filtered using the “BlibFilter” function, requiring entries to have at least 5 peaks (“-n 5”). The DDA-based spectral library created from the MaxQuant results of the artificial mixtures is referred to as Library A1. Peptide sequences uniquely identified in the *Salmonella* pre-fractionated samples were appended to Library A1 to augment the detection capacity of *Salmonella* peptides. This extended DDA-based spectral library is referred to as Library A2 and a detailed overview is given in Table S1. We transformed the peptide RTs present in the BLIB library to iRTs using the spiked-in iRT peptides (Biognosys-11). To this end, empirical RTs of the top-scoring iRT peptide identifications (lowest posterior error probability, “msms.txt”) in the DDA samples were used to fit a linear trend line and scale the RTs. For the artificial mixtures and *Salmonella* pre-fractionation analyses, the corresponding trend lines were  $\text{iRT} = 1.220 \text{ RT} - 74.566$  and  $\text{iRT} = 1.189 \text{ RT} - 75.039$ , respectively. The updated BLIB files were then converted to DLIB format using EncyclopeDIA, using the combined human–*Salmonella* UniProtKB proteome FASTA as background.<sup>19</sup>

**Library-free Searching of Proteome FASTA by PECAN/Walnut.** DIA raw data files were converted to mzML by MSConvert using vendor peakPicking and enabling the “SIM as spectra” option. Pre-processed DIA samples were searched against a compilation of the *Salmonella* UniProtKB proteome (UP000008962, 4657 proteins) and human Swiss-Prot proteome (UP000005640, 20,367 proteins) using the EncyclopeDIA built-in PECAN algorithm.<sup>19,20</sup> We opted to solely search the human Swiss-Prot protein database, which resulted in a  $\sim 3$ -fold reduction in the protein database search space (25,024 vs 74,449 proteins when combining *Salmonella* [UP000008962] and human [UP000005640] UniProtKB references proteomes), in order to minimize the theoretical peptide search space. This was desired to limit the size of a predicted spectral library for all possible tryptic peptides (see below) and overall runtime and memory usage. Since 36,494 out of 36,668 (*i.e.*, 99.53%) of all human peptides identified using MaxQuant matched a Swiss-Prot protein entry, no drastic loss in identifications is anticipated. Default settings were used, except for methionine oxidation (to methionine-sulfoxide) being set as fixed modification and considering a maximum length of 25 amino acids and HCD as the fragmentation type.

**Construction of an MS<sup>2</sup>PIP-Based Spectral Library.** MS2 spectra were predicted by MS<sup>2</sup>PIP (version 20190312)<sup>13</sup> for tryptic peptides derived from an *in silico* digest of the *Salmonella* UniProtKB proteome and human Swiss-Prot proteome (trypsin/P, peptide length 7–25 AA, mass 500–5000 Da, one missed cleavage, N-terminal initiator methionine removal considered) in the case of 2+ and or 3+ peptide precursor fit within 400–900  $m/z$  (scanned range DIA). This yields a total of 1,586,777 predicted MS2 spectra for 1,151,386 peptides solely matching human proteins, 197,782 spectra for



**Figure 1.** Proteome data analysis workflow of hybrid spectral library construction. DDA data from artificial dual-proteome mixtures and complex *Salmonella* pre-fractionated samples were searched using MaxQuant,<sup>37</sup> and an extended DDA spectral library (Library A2) of both datasets was constructed using BiblioSpec.<sup>38</sup> In parallel, entire proteomes of human and *Salmonella* were searched either by searching an MS<sup>2</sup>PIP-predicted spectral library (Library B) with EncyclopeDIA<sup>19</sup> or searching the combined FASTA using PECAN<sup>20</sup> for which an EncyclopeDIA ELIB library was constructed (Library C). A combined hybrid spectral library was constructed from EncyclopeDIA ELIB libraries when searching Library A2, B, and C searches. The hybrid spectral library only contained non-redundant peptide entries matching proteins with MAYU protein FDR  $\leq 1\%$  (see [Experimental Procedures](#)).

144,156 peptides matching *Salmonella* proteins, and 117 spectra for 110 peptides matching both species. We set methionine oxidation (to methionine-sulfoxide) as a fixed modification for MS2 prediction by MS<sup>2</sup>PIP. Predicted spectra were supplemented with DeepLC-predicted RTs using a model trained on RTs of 35,206 non-redundant peptides identified in DIA PECAN searches (peptide Q-value  $< 0.01$ ) (as described in the above section). The MS<sup>2</sup>PIP-based spectral library is referred to as Library B.

**Hybrid Library Construction.** DIA-based result (ELIB) libraries generated by EncyclopeDIA after PECAN processing were combined into a single redundant library. After conversion to BLIB format and adding the “redundant” tag in library info (sqlite3), the “BlibFilter” function was used to create a non-redundant DIA spectral library as described above. All raw DIA data were searched against this PECAN result library with EncyclopeDIA, and the outputted ELIB library is referred to as Library C. Aiming at integrating the results for the three searched libraries, we ran Percolator (v3.5) independently on the combined EncyclopeDIA scoring features of the three searches (Library A2, B, and C) for all samples with similar options as in EncyclopeDIA internal Percolator processing ( $-y -N 200000 -no-terminate$ ). Subsequently, we used the obtained Percolator score as the discriminant score for MAYU FDR estimation for peptides matching non-ambiguously to human Swiss-Prot or *Salmonella* proteins.<sup>39</sup> Afterward, spectral libraries were filtered for entries matching proteins with a MAYU protein-FDR  $\leq 1\%$ . For the 7192 unique protein entries that passed this criterium, all spectra corresponding to peptides with an EncyclopeDIA Percolator peptide-level FDR  $\leq 1\%$  were retained for hybrid spectral library construction. To avoid redundant entries in the hybrid spectral library, we appended the EncyclopeDIA ELIB library of the Library A2 search (39,120 peptides) with additional peptide entries found in search results of Library B and/or Library C. Initially, 8795 additional peptide entries were appended from the filtered results of Library C and afterward, another 11,541 additional peptide entries from

Library B results. A detailed overview of the hybrid spectral library is given in [Table S2](#).

**EncyclopeDIA Spectral Library Searching and Peptide Quantification.** The resulting mzML files were searched against Library A2, B, or hybrid spectral (DLIB) libraries using EncyclopeDIA software (version 0.90)<sup>19</sup> with default settings. Sample-specific Percolator output files and EncyclopeDIA result (ELIB) libraries were stored. Per setup, a combined EncyclopeDIA result library was created consisting of the three replicates. This performs a Percolator re-running of the combined results and provides peptide and protein quantifications at a 1% peptide and protein Q-value, respectively. For quantification, the number of minimum required and quantifiable ions was set at 5 with aligning between samples enabled.

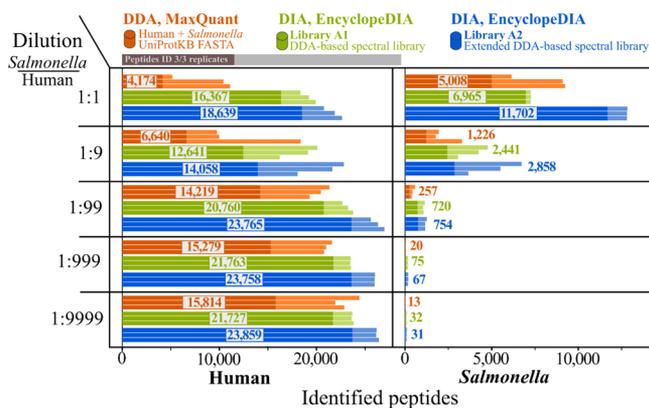
**SpectraST Library Construction and Searches.** We used SpectraST<sup>40</sup> (version 5.0) for spectral library searches of the DDA data. Library A2 and Library B were used to generate a SpectraST spectral library, appending MS<sup>2</sup>PIP peptide ions to Library A2 (option  $-cJA$ ). Afterward, a concatenated target-decoy library was generated, using the precursor swap method<sup>41</sup> (options  $-cAD -cc -cy1 -c\_DPS$ ). The default search parameters of the SpectraST were used except for specifying a precursor isolation window of 0.01 Th and outputting the top 3 ranked hits. The FDR was estimated by  $FDR = n_d/n_t$ , where  $n_t$  and  $n_d$  are the number of PSMs of target and decoy peptides, respectively.

## RESULTS

### DDA-Based Spectral Library Searching of DIA Data Improves Detection of Low-Abundant Peptides

To mimic the proteome complexity of bacterial host cell infections, we generated five artificial *Salmonella*–human proteome mixtures (1:1, 1:9, 1:99, 1:999, and 1:9999) in triplicate, referred to as artificial dual-proteome mixtures. Following trypsin digestion, each sample was analyzed by LC–MS/MS in both DDA and DIA modes. First, DDA data were analyzed using MaxQuant against a composite database containing *Salmonella* and human UniProtKB protein entries (see [Experimental Procedures](#)). As anticipated, with decreasing

*Salmonella* protein content, the number of identified *Salmonella* peptides decreases, whereas the number of identified human peptides increases (Figure 2, orange bars, top to bottom).



**Figure 2.** Peptide identifications in artificial dual-proteome mixtures acquired in DDA and DIA modes. The number of identified non-redundant human (left panel) or *Salmonella* (right panel) peptide sequences (x-axis, peptide Q-value  $\leq 0.01$ ) is shown per replicate across artificial mixtures (y-axis, 1:1 to 1:9999). Bars showing triplicate samples are used to display data acquired in the DDA mode and searched with MaxQuant (orange) and data acquired in the DIA mode and searched with EncyclopeDIA against Library A1 (the non-extended DDA-based spectral library, green) and Library A2 (extended DDA-based spectral library, blue). The dark-colored portion of the bars and corresponding numbers within indicate the number of peptide sequences consistently identified in all three replicate samples.

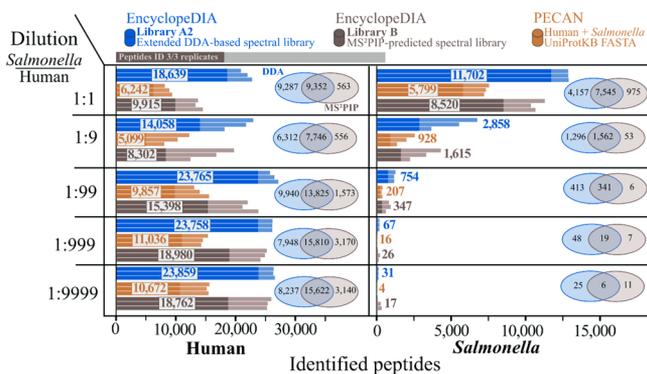
Indeed, although approximately 5008 non-redundant *Salmonella* peptides are consistently (*i.e.*, in all three replicates) identified in the 1:1 dilution, this number decreases to 1226 (24.5%) and only 257 (5.1%) in the 1:9 and 1:99 dilutions, respectively.

In a next phase, MaxQuant results were used to create a DDA-based spectral library designated as “Library A1” (see [Experimental Procedures](#)), encompassing a total of 57,056 MS2 spectra corresponding to 44,932 non-redundant peptide sequences of which 34,475 (76.7%) uniquely matched to human proteins and 10,422 (23.2%) to *Salmonella* proteins. However, when analyzing related shotgun samples exclusively consisting of *Salmonella* peptides on the same MS instrument, we previously routinely identified between 15,000 and 25,000 *Salmonella* peptides with MaxQuant<sup>42</sup> and when loading equal amounts of total peptides as determined by microfluidic spectroscopy,<sup>43</sup> suggesting that mixing of human and *Salmonella* proteomes limits *Salmonella* peptide identification and thus DDA-based spectral library construction. To extend the number of *Salmonella* peptides in this library, we performed an offline RP-HPLC pre-fractionation of a digest of a complex proteome mixture obtained from mixing equal proteome amounts from *Salmonella* grown *in vitro* under 10 different (infection relevant) conditions, as reported in ref 34 (MaxQuant results, see Dataset PXD017904). This way, additional 12,945 non-redundant *Salmonella* peptide sequences (14,709 spectra) were appended to Library A1, referring to the extended DDA-based library as “Library A2” (for entries see [Table S1](#)). DIA data from artificial mixtures were searched against both spectral libraries using EncyclopeDIA<sup>19</sup> (Figure 2, green and blue bars, respectively). In the case of human peptides, DIA analysis approximately doubles the number of peptide identifications in the 1:1 and 1:9 dilutions

compared to DDA analysis and identifies up to ~6000 to 9000 additional human peptides compared to DDA when decreasing *Salmonella* peptide input further. Overall, this suggests that with increasing abundance, *Salmonella* peptide ions obstruct selection and fragmentation of human peptide ions in the DDA mode, while clearly much less interference is observed when samples are analyzed in the DIA mode.<sup>44</sup> Furthermore, when looking at artificial mixtures across all dilutions, DIA identifies most peptides consistently (*i.e.*, in all three replicates). Notably, extending Library A1 with additional *Salmonella* peptides (Figure 2, left panel, blue vs green bars) increases human peptide identifications, likely due to increased *Salmonella* peptide identifications included in peptide-level FDR scoring. Logically, a similar trend holds true when inspecting *Salmonella* peptide identifications (Figure 2, right panel). For instance, in all dilutions, the DIA data queried with Library A2 more than doubles the number of consistently identified *Salmonella* peptides compared to MaxQuant DDA identifications. Taken together, offline fractionation methods can provide a useful asset to improve DIA identification rates, especially in the case of complex proteome mixes with components of low abundance (*e.g.*, dual-proteome mixes such as bacterial pathogen infection of human hosts). Most likely, grasping the whole multi-species peptide mixture complexity is impossible in the given LC-MS/MS setting—as illustrated here by the merely ~10,000 *Salmonella* peptides identified, whereas on average, more than double the amount of *Salmonella* peptides are identified in pure proteome samples.

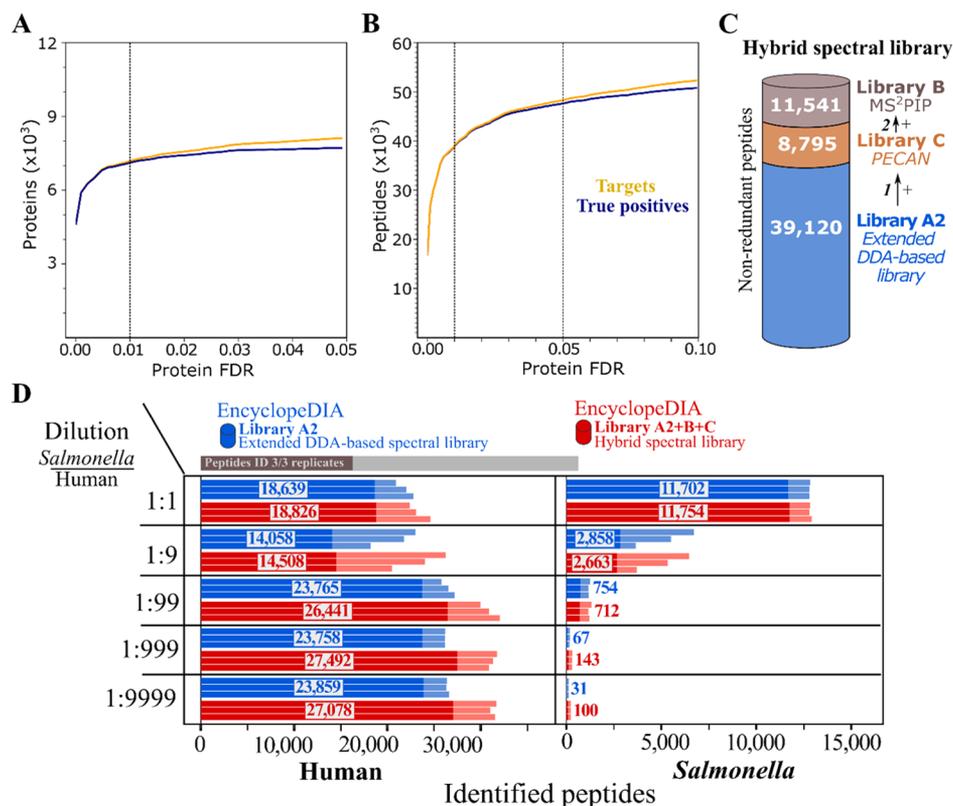
#### DIA-Only Workflows and Predicted Spectral Libraries Can Join Forces with DDA Spectral Libraries

Although our DIA spectral library searches clearly outperformed MaxQuant DDA analysis in terms of peptide identification, an important limitation is that only those peptides originally identified in DDA analysis can be detected. To further tap into the peptide discovery potential of DIA data, we implemented two workflows independent of DDA data. First, we used the DIA library-free PECAN software to search the human and *Salmonella* proteome. When using PECAN, the number of peptide identifications is in-line with or just below DDA data-based identifications but lower as compared to when our DIA data was queried with Library A2, in line with previous reports.<sup>19</sup> Second, we generated a spectral library of an *in silico* tryptic proteome digest of the human and *Salmonella* proteomes using MS<sup>2</sup>PIP.<sup>13</sup> In total, 1,784,677 MS<sup>2</sup> spectra were predicted for 2+/3+ peptide precursors (395 to 905 *m/z*) for 1,295,652 peptides with enzymatic settings similar to PECAN search settings (Trypsin/P, 7 to 25 amino acid long peptides with a maximum of one missed cleavage). DIA-based peptide RTs obtained from the PECAN search results were used to train and predict RTs for all peptides using DeepLC.<sup>45</sup> We then used the MS<sup>2</sup>PIP-predicted spectral library, referred to as “Library B”, to query our artificial mixtures with EncyclopeDIA (Figure 3, left panel, grey bars). Overall, the searches identified a reduced number of consistently identified human peptides per dilution series, while nonetheless identifying a similar number of peptides per run in 1:999 and 1:9999 dilution series. This relatively lower consistency is likely due to the increased number of peptide sequences, ~35-fold, in Library B compared to Library A2. The use of large database sizes already showed to increase variation in peptide and protein quantification.<sup>12</sup> Importantly, and as illustrated in the respective Venn diagrams (insets Figure 3, left panel), thousands of peptides not present

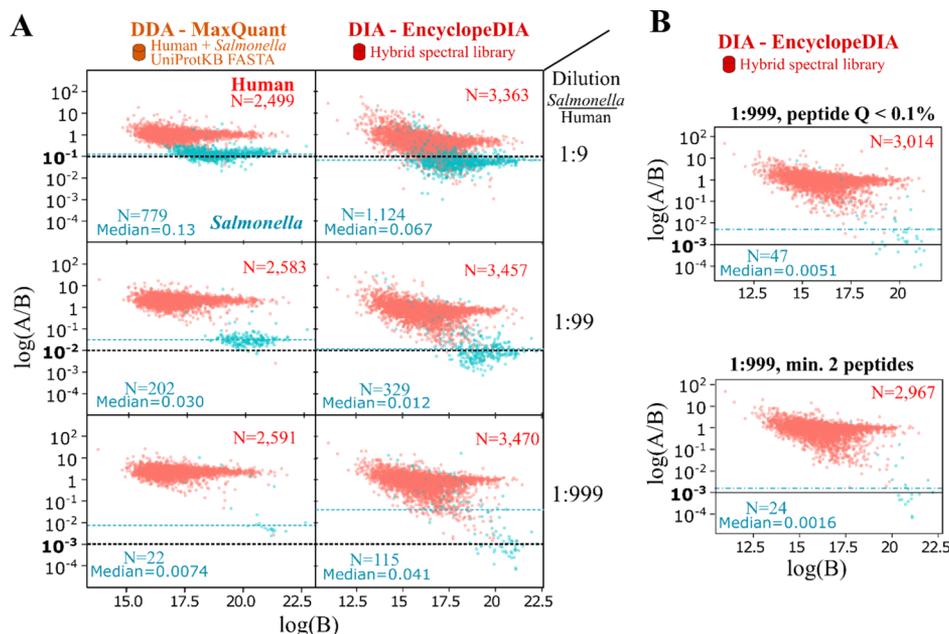


**Figure 3.** DIA-based peptide identifications using different search strategies. The number of identified non-redundant human (left panel) or *Salmonella* (right panel) peptide sequences ( $x$ -axis, peptide  $Q$ -value  $\leq 0.01$ ) are shown per replicate across artificial mixtures ( $y$ -axis, 1:1 to 1:9999). Bars showing triplicate samples are used to display data acquired in the DIA mode and searched with the Library A2 (blue), Library B (ochre), and EncyclopeDIA built-in PECAN algorithm (Walnut), using as input the human–*Salmonella* FASTA (brown). The dark-colored portion of the bars and corresponding numbers within indicate the number of peptide sequences consistently identified in all three samples. Venn diagrams indicate the overlap of consistently identified peptide sequences between Library A2 and Library B.

either in the DDA results nor Library A2 were identified when searching library B, demonstrating the potential of this approach to identify peptides in DIA not found in DDA. Turning our attention to *Salmonella* identified peptides, searching Library B shows relatively lower peptide identification rates ( $\sim 60$ – $80\%$ ) than Library A2. Nonetheless, a significant agreement among results is observed, as for instance, in the 1:1 dilutions, 7545 out of 8520 peptide sequences (88.6%) identified in the Library B search were also identified when querying the artificial mixtures with Library A2. In a next phase, we inspected the properties of peptides identified by searching Library B that were not found in the DDA searches (and thus not part of Library A2). For instance, 1814 peptides were discovered using Library B in the 1:1 sample, of which 933 human and 881 *Salmonella* peptides (Figure S1). In less diluted human samples 1:999 and 1:9999, more than 5000 peptides are uniquely identified using Library B. When comparing the intensity of peptides included in Library A2 and those discovered by searching Library B, the novel MS<sup>2</sup>PIP-identified peptide distributions are slightly lower, and yet more outspoken in the 1:999 and 1:9999 samples (Figure S1A). Hence, these observations suggest that Library B, thus searching of MS<sup>2</sup>PIP-predicted spectra, enable the detection of relatively lower abundant peptide species missed in DDA. Next, we checked whether the novel MS<sup>2</sup>PIP-based peptides matched



**Figure 4.** Hybrid spectral library construction. (A) Number of target and true positive protein identifications (orange and blue, respectively) in function of MAYU estimated protein FDR. (B) Number of target and true positive peptide identifications (orange and blue, respectively) in function of MAYU-estimated protein FDR. (C) Hybrid spectral library consisting of 59,475 non-redundant peptide entries matching 7194 proteins (MAYU protein FDR  $\leq 1\%$ ). The filtered EncyclopeDIA ELIB library from the Library A2 search was appended stepwise with additional peptide entries from Library B (MS<sup>2</sup>PIP-predicted spectral library, step 2) and subsequently, with additional peptide entries from Library C (PECAN results, step 1) and subsequently, with additional peptide entries from Library B (MS<sup>2</sup>PIP-predicted spectral library, step 2). For more details on the hybrid spectral library construction, see [Experimental Procedures](#). (D) Number of identified non-redundant human (left panel) or *Salmonella* (right panel) peptide sequences ( $x$ -axis, peptide  $Q$ -value  $\leq 0.01$ ) are shown per replicate across artificial mixtures ( $y$ -axis, 1:1 to 1:9999). Bars showing triplicate samples are used to display data acquired in the DIA mode and searched with Library A2 (blue) or the hybrid spectral library (red). The dark-colored portion of the bars and corresponding numbers within indicate the number of peptide sequences consistently identified in all three replicate samples analyzed.



**Figure 5.** *Salmonella* protein quantification throughout *Salmonella*–human proteome dilutions. (A) Log-transformed ratios ( $\log(A/B)$ , y-axis) of human (red) and *Salmonella* (blue) proteins with  $B$  being the protein intensity in the 1:1 dilution series mixture and  $A$ , the protein intensity in 1:9, 1:99, and 1:999 dilution series (from top to bottom). MaxQuant DDA quantification (left) was compared to EncyclopeDIA quantification searching the hybrid spectral library (right). The number of plotted human and *Salmonella* proteins were indicated. Anticipated and median *Salmonella* protein ratios were plotted (black and blue dotted lines, respectively). Protein quantification was performed by the average of the three most abundant proteotypic peptides per protein (see Experimental Procedures). (B) EncyclopeDIA protein quantification results when applying a 0.1% Percolator peptide FDR filter (top) or when requiring a minimum of two proteotypic peptide quantifications per protein.

protein entries that were also matched by other peptides in the DDA searches. In all dilutions, approximately 85% of MS<sup>2</sup>PIP-based peptides matched a protein identified in the MaxQuant searches of the DDA data (Figure S1B). As such, the majority of MS<sup>2</sup>PIP-based peptides increase the protein sequence coverage for proteins present in the DDA spectral library in turn increasing identification and quantification confidence.

Next to a conventional MaxQuant database search (Figures 2 and S2, orange bars), we combined Library A2 and Library B to search DDA data with SpectraST<sup>40</sup> (Figure S2, ochre bars). Spectral library searching gained popularity due to their high sensitivity and fast running time, which therefore might improve the detection of *Salmonella* peptides. At 1:1 and 1:9 dilutions, MaxQuant identifies slightly more consistently identified *Salmonella* peptides which largely overlaps with SpectraST identified peptides (~80%, Figure S2A). However, SpectraST identifies more peptides at a 1:99 dilution (270 vs 257 peptides) and 1:999 dilution (50 vs 20 peptides), suggesting spectral library searching as a more sensitive method to identify low-abundant *Salmonella* peptides. For instance, peptides IVIRPLPLPVIR and TNVPHIFAIGDIVGQPMLAHK were identified by both SpectraST and MaxQuant in 1:1, 1:9, and 1:99 dilution series but solely using SpectraST in the 1:999 dilution (Figure S2B,C).

#### Hybrid Spectral Library Searching Combining DIA-Only and DDA-Based Results Improves Peptide Detection

In a next phase, we generated a hybrid spectral library to attest its potential to further increase proteome coverage (Experimental Procedures, Figure 1). Hybrid spectral libraries are essentially merged libraries comprising results of different DDA or DIA analysis workflows, for example, combining predicted spectral libraries with experimental libraries,<sup>46</sup> or DDA-based and DIA-only spectral libraries.<sup>31</sup> Here, we merged the EncyclopeDIA

ELIB libraries when querying library A2, Library B, and a non-redundant spectral library of PECAN-identified peptides (designated “Library C”) to generate a single hybrid library (Library A2 + B + C, see Experimental Procedures). One of the key challenges is to maintain comparable FDR control to merged spectral libraries.<sup>47</sup> To this end, we performed integrated Percolator processing for all EncyclopeDIA scoring features of the three spectral library searches combined. Afterward, the Percolator score was used as the discriminant score for peptide- and protein-level error rate control by MAYU.<sup>39</sup> It can be observed that both curves did not yet reach saturation, and a total of 7194 proteins (5086 human and 2108 *Salmonella* proteins) were below a stringent protein FDR of 1% (Figure 4A,B). For hybrid library construction, we first filtered the EncyclopeDIA ELIB library (*i.e.*, DIA-calibrated results) result when searching Library A, including spectra of 39,120 peptides matching these 7194 proteins with an MAYU 1% protein FDR (Table S2), instead of the initial 57,877 peptides (Table S1). After similar protein FDR filtering, DIA-only identified peptides were appended stepwise with 8795 additional peptides of Library C (*i.e.*, PECAN result library) and yet additional 11,541 peptides identified solely by Library B searches (Figure 4C, Table S2). EncyclopeDIA searches with the obtained hybrid spectral library (Library A2 + B + C) resulted in an increased number of human peptide identifications, most outspoken in 1:99 to 1:999 dilution series with an average of 3000–4000 additional peptides consistently identified (Figure 4D). In the case of *Salmonella*, an increased peptide identification rate is only observed in the 1:999 and 1:9999 dilution series. This beneficial effect might arise due to the inclusion of more low-abundant peptides species solely identified in DIA. In addition, stricter protein FDR control used for hybrid library construction might also be a plausible

reason as in these low-input *Salmonella* samples as relatively high proportions of “false *Salmonella* targets” (also referred to as  $\pi_0$ <sup>48,49</sup>) can be anticipated, in our case due to their (extreme) low abundance and inclusion of an heterogeneous *Salmonella* pre-fractionation experiment. Such larger  $\pi_0$  necessitates stricter error rate control,<sup>48</sup> as implemented here in our hybrid spectral library construction (Figure 4A,B).

### DIA Improves Quantification of Low-Abundant Proteins

Besides peptide identification rates, we evaluated and compared protein quantifications across dilutions between both acquisition modes. Aiming at assessing the performance per dilution series independently, that is, as a proxy for biological infection conditions where the bacterial proteome content is limiting and thus 1:1 and 1:9 dilution conditions are thus typically not representative, we ran the three replicates of each dilution in MaxQuant enabling matching-between-runs and used the alignment-between-runs function in EncyclopeDIA. In order to handle identical protein interference of MaxQuant and EncyclopeDIA results, we used the average intensity of the three most intense proteotypic peptides (unambiguously matched and considering Leu and Ile as indistinguishable) for determining the corresponding protein intensity, that is, the TOP3 quantification method<sup>30</sup> also used in the label-free quantification benchmark tool LFQbench.<sup>51</sup> Importantly, we did include protein quantifications based on single peptides (given the low abundance of *Salmonella* proteins) and, unlike DIA, considered missing values for DDA peptide quantifications by averaging over the non-zero quantifications across replicates. DIA hybrid spectral library searches quantified 1124 *Salmonella* proteins with a median protein ratio of 0.067 in the 1:9 dilution series, or an additional 345 proteins compared to MaxQuant (Figure 5A, top). In the more representative infection-relevant host-pathogen 1:99 condition, the hybrid spectral search quantified 329 proteins with a median protein ratio of 0.012, while MaxQuant only quantified 202 proteins with a 0.03 median ratio (Figure 5A, middle). In the extremely challenging 1:999 dilution, more than 5-fold the number of MaxQuant-quantified proteins were identified—although with an increased number of outlying ratios and a median ratio of 0.041 (Figure 5A, bottom). In general, in both 1:99 and 1:999 dilution series, it can be observed that quantified *Salmonella* proteins are typically of higher abundance in the 1:1 dilution series ( $x$ -axis,  $\log(B)$ ), and the few protein quantification outliers not in-line with the anticipated dilution frequency are relatively low abundant in the 1:1 dilution series, thus likely representing false targets. Tackling this issue, we tested two more stringent filtering criteria for 1:999 dilution series. First, requiring a 10-fold stricter Percolator peptide  $Q$ -value of 0.1%, which deliver 47 quantified proteins at a median ratio of 0.051 (Figure 5B, top). Alternatively, requiring at least two peptide quantifications per peptide narrows down protein quantifications to 24 proteins with a median ratio of 0.0016. Thus, stricter filtering of DIA results here enables higher-confidence *Salmonella* protein quantifications matching the anticipated dilution.

We also manually inspected the more extreme cases of quantifiable *Salmonella* peptides in the 1:999+ dilutions in our DIA workflow. As a representative example, the peptide ILADIAVFDK (doubly charged) maintains similar  $MS^2$  peak shapes throughout 1:1, 1:9, and 1:99 and to a lesser extent in 1:999 and 1:9999 dilutions, while the intensity decreases according to the dilution factor (Figure S3). However, at increased dilutions, flanking noise peaks are evidenced and

become dominant over genuine peptide peaks—similar as observable for SpectraST-identified peptides at a 1:999 dilution in DDA (Figure S2B,C). Undoubtedly, low peptide abundance represents a clear challenge for achieving correct identification and quantification in the 1:999 and 1:9999 setups irrespective of the acquisition mode used.

### DISCUSSION

MS acquisition of artificial human–*Salmonella* dual-proteome mixtures in the DIA mode significantly improved identification and quantification of low-abundant *Salmonella* proteins compared to the DDA mode, increasing the discriminative power of LC–MS/MS. Among the artificial mixtures (1:1 to 1:9999) used in this study to effectuate comparative analysis between both acquisition modes, the 1:99 dilution closely reflects the host/pathogen ratio under actual infection conditions in the case of *Salmonella*,<sup>23</sup> where a scarce amount of protein material comes from the bacterial pathogen, making the simultaneous study of host and pathogen proteomes challenging. Overall, the DIA mode improved detection and quantification of *Salmonella* peptides and protein quantifications at lower 1:99 and 1:999 dilutions (Figures 3 and 5). Also, in 1:1 *Salmonella*–human proteome mixtures, human and *Salmonella* peptide/protein identification rates were significantly boosted when searching DDA-based and/or hybrid spectral libraries compared to a conventional DDA database search (Figure 3). Taken together, the increased identification of peptides with lower abundance in DIA points to the stochastic nature of DDA. Notably, to correct for DDA stochastic sampling, more advanced DDA precursor selection algorithms have been developed such as MaxQuant.Live.<sup>52</sup> In addition, next to a conventional database search, we performed a spectral search with SpectraST<sup>40</sup> on a library composed of Library A2 and Library B spectra. Although not outperforming MaxQuant for human peptide identification, it did improve detection of peptide identification at 1:999 dilution (Figure S2). Another alternative is to combine spectral library-based scoring features with conventional scoring features as demonstrated by Prosit and  $MS^2$ PIP-based spectral libraries.<sup>14,53</sup> In contrast to a recent report,<sup>11</sup> offline fractionation of a *Salmonella* sample proved to increase *Salmonella* peptide identification, nearly doubling *Salmonella* peptide identification in the 1:1 *Salmonella*–human proteome mixture. This effect is likely attributable to the dual-proteome complexity, rendering DDA unable to fully grasp *Salmonella* proteome complexity. Hence, in such multi-species proteome analyses, comprehensive spectral libraries and pre-fractionation of a species of interest can be advisable. However, this brings along spectral library heterogeneity which may increase the risk of “false targets” within the library,<sup>48</sup> especially at lower dilutions where stricter filtering can be advisable (Figure 5B).

Besides assessing the performance of DDA-based spectral libraries, we tested alternative DIA workflows that enable us to search both *Salmonella* and human proteomes. The library-free EncyclopeDIA built-in PECAN algorithm (Walnut) delivered lower peptide identifications, in-line with earlier observations.<sup>19</sup> In addition, we made use of  $MS^2$ PIP-predicted spectral libraries (Library B), predicting RT with DeepLC trained on DIA identifications made by PECAN. Similar to recent studies reported on the use of Prosit and Prism-predicted spectral libraries,<sup>14,54</sup> we achieved a similar performance compared to searching DDA-based spectral libraries (Figure 3, Library A2). Interestingly, both alternative approaches allowed us to identify

18,736 novel peptides not found in DDA data. Further inspection of non-DDA peptides identified by searching Library B pointed to relatively lower-abundant peptides that might have been missed in DDA precursor selection. Notably, the majority of these peptides (85%) matched proteins identified by other peptides in DDA (Figure S1). Combining the strengths of both approaches, we generated a hybrid spectral library merging EncyclopeDIA ELIB libraries from Library A2, B, and C searches, filtering for peptides matching 7192 proteins with a protein FDR  $\leq 1\%$  as assessed by MAYU<sup>39</sup> (Figure 4A–C). The resulting hybrid spectral library identified  $\sim 27,000$  peptides in all 3 replicates of concentrated (1:99+) human samples (Figure 4D), which is 1.35-fold higher than searching Library A2 alone ( $\sim 20,000$  peptides, Figure 3), and 1.8-fold higher than a DDA analysis ( $\sim 15,000$  peptides, Figure 2). As such, integrating DIA-based peptide identifications from DDA-based spectral library searches with DIA-only approaches resulted in a drastic increase in peptide identification. Hybrid spectral libraries were very recently reported to improve proteome coverage.<sup>19,31,46</sup> Here, we showed that this greatly facilitates dual-proteome profiling, as complex proteome mixtures pose an enormous challenge to DDA, making DDA-only libraries far from comprehensive. Taken together, making use of DIA-only approaches, predicted spectral libraries, and perhaps publicly available DDA data or DDA-based spectral libraries, DIA analyses no longer seem to depend on sample-specific complementary DDA runs, although these can further improve the performance, as also shown to be the case in our study. Notably, other DIA-only analyses such as spectral deconvolution algorithms, for example, DIA-Umpire,<sup>22</sup> could further strengthen hybrid libraries. Besides EncyclopeDIA, other peptide-centric DIA analysis software algorithms could be of value, such as the recently developed DIA-NN algorithm that also supports library-free searches.<sup>55</sup> In this regard, the provided human–*Salmonella* datasets could serve as a valuable benchmarking tool for wide-window DIA analysis. Another promising avenue would be to run narrow-window detection DIA runs to generate a dual-proteome chromatogram library as described previously by Searle et al.<sup>19</sup> Using predicted libraries to search narrow-window DIA data greatly improves peptide detection and the empirical correction for chromatogram library delivers highly performant libraries.<sup>17,18</sup> Hence, although DDA and library-free PECAN (Library A2 and C, respectively) searches yield additional identifications and motivates the hybrid library construction in our wide-window DIA analysis, these contributions are expected to weaken and might possibly be neglectable when making use of narrow-window chromatogram search strategies. Nevertheless, for DIA analysis not supported by additional detection-only chromatogram libraries, hybrid library creation can serve as an alternative workflow to augment DIA detection and quantification.

When judging protein quantification, both MaxQuant (DDA) and EncyclopeDIA (DIA) delivered *Salmonella* protein quantifications in-line with dilution series up to 1:999 dilutions (Figure 5A). In the infection-relevant 1:99 sample, 329 proteins were quantified with a relative median intensity ratio of 0.012 compared to 1:1 equal mixed samples. Notably, MaxQuant also delivers 202 (–39%) quantified proteins and has overall better accuracy, which is clearer at a 1:999 ratio (Figure 5A). However, applying stricter FDR threshold or requiring minimum 2 peptide quantifications per proteins did improve accuracy of DIA protein quantifications (Figure 5B). Note that we did not include 1:9999 protein quantifications as the few that were found had inconsistent protein ratios. When judging, for

instance, the ILADIAVFDK/2+ precursor that was found across all dilutions (Figure S3), it is clear how peptide peaks at a 1:9999 dilution become nearly impossible to distinguish from random noise. However, artificial mixtures do suggest that infection-relevant host pathogen conditions with an approximate  $\sim 1:100$  dilution could be profiled in sufficient proteome depth without prior bacteria enrichment, especially so when increasing LC–MS/MS run time by, for instance, offline pre-fractionations or the use of parallel DIA runs with narrow  $m/z$  windows in consecutive  $m/z$  ranges as demonstrated in the EncyclopeDIA workflow.<sup>19</sup>

In our study, we compared DDA and DIA analysis workflows, which is not straightforward given the evident differences between spectrum-centric and peptide-centric searches. Alternatively, the provided data provide an interesting and challenging benchmark case for label-free quantification algorithms. For instance, artificial proteome mixes of human, yeast, and *Escherichia coli* have been used before for quantification by DIA algorithms.<sup>51,56</sup> In addition, a *Plasmodium falciparum* proteome was diluted to a 1:99 ratio in an uninfected human red blood cell lysate for DIA quantification.<sup>57</sup> Here, we sampled more extreme dilutions, providing an interesting challenge for state-of-the-art quantification algorithms. Moreover, the spectral libraries created and provided as the Supporting Information online will assist future DIA-based research on *Salmonella*-infected epithelial host cells.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00350>.

Peptide intensity distributions and characterization of MS<sup>2</sup>PIP-only identified peptides; peptide identifications in artificial dual-proteome mixtures including SpectraST DDA searches; and quantification of a doubly charged peptide among artificial mixtures in the DIA mode (PDF)  
Overview of peptide precursors in Library A2 (XLSX)  
Hybrid spectral library overview (XLSX)

## ■ AUTHOR INFORMATION

### Corresponding Author

Petra Van Damme – Department of Biochemistry and Microbiology, Ghent University, Ghent 9000, Belgium; [orcid.org/0000-0001-9090-027X](https://orcid.org/0000-0001-9090-027X); Phone: +32 09 264 51 29; Email: [Petra.VanDamme@ugent.be](mailto:Petra.VanDamme@ugent.be)

### Authors

Patrick Willems – Department of Biochemistry and Microbiology, Ghent University, Ghent 9000, Belgium; Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent 9000, Belgium; VIB-UGent Center for Plant Systems Biology, Ghent 9052, Belgium; [orcid.org/0000-0003-4667-2294](https://orcid.org/0000-0003-4667-2294)

Ursula Fels – Department of Biochemistry and Microbiology, Ghent University, Ghent 9000, Belgium; VIB-UGent Center for Medical Biotechnology, Ghent 9052, Belgium

An Staes – VIB-UGent Center for Medical Biotechnology, Ghent 9052, Belgium; Department of Biomolecular Medicine, Ghent University, Ghent 9000, Belgium

Kris Gevaert – VIB-UGent Center for Medical Biotechnology, Ghent 9052, Belgium; Department of Biomolecular Medicine,

Ghent University, Ghent 9000, Belgium;  [orcid.org/0000-0002-4237-0283](https://orcid.org/0000-0002-4237-0283)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jproteome.0c00350>

### Author Contributions

P.W., U.F., A.S., K.G., and P.V.D. designed research. P.W., U.F., and P.V.D. performed research. P.W. analyzed data. P.W., U.F., and P.V.D. wrote the paper, and all authors contributed to finalizing the manuscript text and gave approval to the final version of the manuscript.

### Funding

P.V.D. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (PROPHECY grant agreement no. 803972) and support from the Research Foundation—Flanders (FWO-Vlaanderen), project number G.0511.20N. K.G. acknowledges support from a Ghent University Concerted Research Actions (grant BOF14/GOA/013).

### Notes

The authors declare no competing financial interest.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE<sup>58</sup> partner repository with the dataset identifier PXD017904 for DDA artificial mixtures and *Salmonella* pre-fractionated samples and with the dataset identifier PXD017945 for the DIA artificial mixtures. In addition, the generated spectral libraries, FASTA databases, and search results are organized in the Open Science Framework (OSF) project page<sup>59</sup> and are available at <https://osf.io/u96bk/>.

### ACKNOWLEDGMENTS

We would like to acknowledge Brian Searle for his valuable feedback during the implementation of EncyclopeDIA software.

### ABBREVIATIONS

LC-MS/MS, liquid chromatography coupled to tandem mass spectrometry; RT, retention time; iRT, indexed retention time; MS<sup>2</sup>PIP, MS<sup>2</sup> peak intensity prediction

### REFERENCES

- (1) Smith, L. M.; Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nat. Methods* **2013**, *10*, 186–187.
- (2) Fijalkowska, D.; Fijalkowski, I.; Willems, P.; Van Damme, P. Bacterial riboproteogenomics: the era of N-terminal proteoform existence revealed. *FEMS Microbiol. Rev.* **2020**, *44*, 418–431.
- (3) Kapp, E.; Schutz, F. Overview of tandem mass spectrometry (MS/MS) database search algorithms. *Current Protocols in Protein Science*; John Wiley & Sons, 2007, Chapter 25, Unit25.2.
- (4) Michalski, A.; Cox, J.; Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **2011**, *10*, 1785–1793.
- (5) Panchaud, A.; Scherl, A.; Shaffer, S. A.; von Haller, P. D.; Kulasekara, H. D.; Miller, S. L.; Goodlett, D. R. Precursor Acquisition Independent From Ion Count: How to Dive Deeper into the Proteomics Ocean. *Anal. Chem.* **2009**, *81*, 6481–6488.
- (6) Burgess, M. W.; Keshishian, H.; Mani, D. R.; Gillette, M. A.; Carr, S. A. Simplified and efficient quantification of low-abundance proteins at very high multiplex via targeted mass spectrometry. *Mol. Cell. Proteomics* **2014**, *13*, 1137–1149.

(7) Venable, J. D.; Dong, M.-Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **2004**, *1*, 39–45.

(8) Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **2018**, *14*, e8126.

(9) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Miladinović, S. M.; Cheng, L.-Y.; Messner, S.; Ehrenberger, T.; Zanotelli, V.; Butscheid, Y.; Escher, C.; Vitek, O.; Rinner, O.; Reiter, L. Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. *Mol. Cell. Proteomics* **2015**, *14*, 1400–1410.

(10) Weisbrod, C. R.; Eng, J. K.; Hoopmann, M. R.; Baker, T.; Bruce, J. E. Accurate Peptide Fragment Mass Analysis: Multiplexed Peptide Identification and Quantification. *J. Proteome Res.* **2012**, *11*, 1621–1632.

(11) Michalik, S.; Depke, M.; Murr, A.; Gesell Salazar, M.; Kusebauch, U.; Sun, Z.; Meyer, T. C.; Surmann, K.; Pfortner, H.; Hildebrandt, P.; Weiss, S.; Palma Medina, L. M.; Gutjahr, M.; Hammer, E.; Becher, D.; Pribyl, T.; Hammerschmidt, S.; Deutsch, E. W.; Bader, S. L.; Hecker, M.; Moritz, R. L.; Mader, U.; Volker, U.; Schmidt, F. A global *Staphylococcus aureus* proteome resource applied to the in vivo characterization of host-pathogen interactions. *Sci. Rep.* **2017**, *7*, 9718.

(12) Barkovits, K.; Pacharra, S.; Pfeiffer, K.; Steinbach, S.; Eisenacher, M.; Marcus, K.; Uszkoreit, J. Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition. *Mol. Cell. Proteomics* **2020**, *19*, 181–197.

(13) Degroeve, S.; Martens, L. MS2PIP: a tool for MS/MS peak intensity prediction. *Bioinformatics* **2013**, *29*, 3199–3203.

(14) Gessulat, S.; Schmidt, T.; Zolg, D. P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; Reimer, U.; Ehrlich, H.-C.; Aiche, S.; Kuster, B.; Wilhelm, M. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* **2019**, *16*, 509–518.

(15) Gabriels, R.; Martens, L.; Degroeve, S. Updated MS(2)PIP web server delivers fast and accurate MS(2) peak intensity prediction for multiple fragmentation methods, instruments and labeling techniques. *Nucleic Acids Res.* **2019**, *47*, W295–w299.

(16) Zolg, D. P.; Wilhelm, M.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Delanghe, B.; Bailey, D. J.; Gessulat, S.; Ehrlich, H.-C.; Weininger, M.; Yu, P.; Schlegl, J.; Kramer, K.; Schmidt, T.; Kusebauch, U.; Deutsch, E. W.; Aebersold, R.; Moritz, R. L.; Wenschuh, H.; Moehring, T.; Aiche, S.; Huhmer, A.; Reimer, U.; Kuster, B. Building ProteomeTools based on a complete synthetic human proteome. *Nat. Methods* **2017**, *14*, 259–262.

(17) Van Puyvelde, B.; Willems, S.; Gabriels, R.; Daled, S.; De Clerck, L.; Vande Castele, S.; Staes, A.; Impens, F.; Deforce, D.; Martens, L.; Degroeve, S.; Dhaenens, M. Removing the Hidden Data Dependency of DIA with Predicted Spectral Libraries. *Proteomics* **2020**, *20*, 1900306.

(18) Searle, B. C.; Swearingen, K. E.; Barnes, C. A.; Schmidt, T.; Gessulat, S.; Küster, B.; Wilhelm, M. Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat. Commun.* **2020**, *11*, 1548.

(19) Searle, B. C.; Pino, L. K.; Egertson, J. D.; Ting, Y. S.; Lawrence, R. T.; MacLean, B. X.; Villén, J.; MacCoss, M. J. Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nat. Commun.* **2018**, *9*, 5128.

(20) Ting, Y. S.; Egertson, J. D.; Bollinger, J. G.; Searle, B. C.; Payne, S. H.; Noble, W. S.; MacCoss, M. J. PECAN: library-free peptide detection for data-independent acquisition tandem mass spectrometry data. *Nat. Methods* **2017**, *14*, 903–908.

(21) Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **2007**, *4*, 923–925.

(22) Tsou, C.-C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A.-C.; Nesvizhskii, A. I. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* **2015**, *12*, 258–264.

- (23) Fels, U.; Gevaert, K.; Van Damme, P. Proteogenomics in Aid of Host-Pathogen Interaction Studies: A Bacterial Perspective. *Proteomes* **2017**, *5*, 26.
- (24) Lau, N.; Haerberle, A. L.; O'Keefe, B. J.; Latomanski, E. A.; Celli, J.; Newton, H. J.; Knodler, L. A. SopF, a phosphoinositide binding effector, promotes the stability of the nascent Salmonella-containing vacuole. *PLoS Pathog.* **2019**, *15*, e1007959.
- (25) Li, Z.; Liu, Y.; Fu, J.; Zhang, B.; Cheng, S.; Wu, M.; Wang, Z.; Jiang, J.; Chang, C.; Liu, X. Salmonella Proteomic Profiling during Infection Distinguishes the Intracellular Environment of Host Cells. *mSystems* **2019**, *4*, e00314–18.
- (26) Liu, Y.; Zhang, Q.; Hu, M.; Yu, K.; Fu, J.; Zhou, F.; Liu, X. Proteomic Analyses of Intracellular Salmonella enterica Serovar Typhimurium Reveal Extensive Bacterial Adaptations to Infected Host Epithelial Cells. *Infect. Immun.* **2015**, *83*, 2897–2906.
- (27) Stapels, D. A. C.; Hill, P. W. S.; Westermann, A. J.; Fisher, R. A.; Thurston, T. L.; Saliba, A.-E.; Blommestein, I.; Vogel, J.; Helaine, S. Salmonella persists undermine host immune defenses during antibiotic treatment. *Science* **2018**, *362*, 1156–1160.
- (28) Westermann, A. J.; Förstner, K. U.; Amman, F.; Barquist, L.; Chao, Y.; Schulte, L. N.; Müller, L.; Reinhardt, R.; Stadler, P. F.; Vogel, J. Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions. *Nature* **2016**, *529*, 496–501.
- (29) Damron, F. H.; Oglesby-Sherrouse, A. G.; Wilks, A.; Barbier, M. Dual-seq transcriptomics reveals the battle for iron during *Pseudomonas aeruginosa* acute murine pneumonia. *Sci. Rep.* **2016**, *6*, 39172.
- (30) Govaert, E.; Van Steendam, K.; Willems, S.; Vossaert, L.; Dhaenens, M.; Deforce, D. Comparison of fractionation proteomics for local SWATH library building. *Proteomics* **2017**, *17*, 1700052.
- (31) Muntel, J.; Gandhi, T.; Verbeke, L.; Bernhardt, O. M.; Treiber, T.; Bruderer, R.; Reiter, L. Surpassing 10000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Mol. Omics* **2019**, *15*, 348–360.
- (32) Hoiseth, S. K.; Stocker, B. A. D. Aromatic-dependent Salmonella typhimurium are non-virulent and effective as live vaccines. *Nature* **1981**, *291*, 238–239.
- (33) Löber, S.; Jäckel, D.; Kaiser, N.; Hensel, M. Regulation of Salmonella pathogenicity island 2 genes by independent environmental signals. *Int. J. Med. Microbiol.* **2006**, *296*, 435–447.
- (34) Kröger, C.; Colgan, A.; Srikanth, S.; Händler, K.; Sivasankaran, S. K.; Hammarlöf, D. L.; Canals, R.; Grissom, J. E.; Conway, T.; Hokamp, K.; Hinton, J. C. D. An infection-relevant transcriptomic compendium for Salmonella enterica Serovar Typhimurium. *Cell Host Microbe* **2013**, *14*, 683–695.
- (35) Bourret, T. J.; Porwollik, S.; McClelland, M.; Zhao, R.; Greco, T.; Ischiroopoulos, H.; Vázquez-Torres, A. Nitric oxide antagonizes the acid tolerance response that protects Salmonella against innate gastric defenses. *PLoS One* **2008**, *3*, e1833.
- (36) Escher, C.; Reiter, L.; MacLean, B.; Ossola, R.; Herzog, F.; Chilton, J.; MacCoss, M. J.; Rinner, O. Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **2012**, *12*, 1111–1121.
- (37) 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.
- (38) Frewen, B.; MacCoss, M. J. Using BiblioSpec for creating and searching tandem MS peptide libraries. *Current Protocols in Bioinformatics*; Wiley, 2007, Chapter 13, Unit 13.7.
- (39) Reiter, L.; Claassen, M.; Schrimpf, S. P.; Jovanovic, M.; Schmidt, A.; Buhmann, J. M.; Hengartner, M. O.; Aebersold, R. Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol. Cell. Proteomics* **2009**, *8*, 2405–2417.
- (40) Lam, H.; Deutsch, E. W.; Eddes, J. S.; Eng, J. K.; King, N.; Stein, S. E.; Aebersold, R. Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* **2007**, *7*, 655–667.
- (41) Cheng, C.-Y.; Tsai, C.-F.; Chen, Y.-J.; Sung, T.-Y.; Hsu, W.-L. Spectrum-based method to generate good decoy libraries for spectral library searching in peptide identifications. *J. Proteome Res.* **2013**, *12*, 2305–2310.
- (42) Willems, P.; Fijalkowski, I.; Van Damme, P. Lost and found: re-searching and re-scoring proteomics data aids the discovery of bacterial proteins and improves proteome coverage. **2019**, bioRxiv:12.18.881375.
- (43) Maia, T. M.; Staes, A.; Plasman, K.; Pauwels, J.; Boucher, K.; Argentini, A.; Martens, L.; Montoye, T.; Gevaert, K.; Impens, F. Simple Peptide Quantification Approach for MS-Based Proteomics Quality Control. *ACS Omega* **2020**, *5*, 6754–6762.
- (44) Timmins-Schiffman, E.; Mikan, M. P.; Ting, Y. S.; Harvey, H. R.; Nunn, B. L. MS analysis of a dilution series of bacteria: phytoplankton to improve detection of low abundance bacterial peptides. *Sci. Rep.* **2018**, *8*, 9276.
- (45) Bouwmeester, R.; Gabriels, R.; Hulstaert, N.; Martens, L.; Degroeve, S. DeepLC can predict retention times for peptides that carry as-yet unseen modifications. **2020**, bioRxiv:03.28.013003.
- (46) Lou, R.; Tang, P.; Ding, K.; Li, S.; Tian, C.; Li, Y.; Zhao, S.; Zhang, Y.; Shui, W. Hybrid Spectral Library Combining DIA-MS Data and a Targeted Virtual Library Substantially Deepens the Proteome Coverage. *iScience* **2020**, *23*, 100903.
- (47) Muntel, J.; Gandhi, T.; Verbeke, L.; Bernhardt, O. M.; Treiber, T.; Bruderer, R.; Reiter, L. Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Mol. Omics* **2019**, *15*, 348–360.
- (48) Rosenberger, G.; Bludau, I.; Schmitt, U.; Heusel, M.; Hunter, C. L.; Liu, Y.; MacCoss, M. J.; MacLean, B. X.; Nesvizhskii, A. I.; Pedrioli, P. G. A.; Reiter, L.; Röst, H. L.; Tate, S.; Ting, Y. S.; Collins, B. C.; Aebersold, R. Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. *Nat. Methods* **2017**, *14*, 921–927.
- (49) Storey, J. D.; Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9440–9445.
- (50) Silva, J. C.; Gorenstein, M. V.; Li, G.-Z.; Vissers, J. P. C.; Geromanos, S. J. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **2006**, *5*, 144–156.
- (51) Navarro, P.; Kuharev, J.; Gillet, L. C.; Bernhardt, O. M.; MacLean, B.; Röst, H. L.; Tate, S. A.; Tsou, C.-C.; Reiter, L.; Distler, U.; Rosenberger, G.; Perez-Riverol, Y.; Nesvizhskii, A. I.; Aebersold, R.; Tenzer, S. A multicenter study benchmarks software tools for label-free proteome quantification. *Nat. Biotechnol.* **2016**, *34*, 1130–1136.
- (52) Wichmann, C.; Meier, F.; Virreira Winter, S.; Brunner, A.-D.; Cox, J.; Mann, M. MaxQuant.Live Enables Global Targeting of More Than 25,000 Peptides. *Mol. Cell. Proteomics* **2019**, *18*, 982–994.
- (53) C Silva, A. S.; Bouwmeester, R.; Martens, L.; Degroeve, S. Accurate peptide fragmentation predictions allow data driven approaches to replace and improve upon proteomics search engine scoring functions. *Bioinformatics* **2019**, *35*, 5243–5248.
- (54) Tiwary, S.; Levy, R.; Gutenbrunner, P.; Salinas Soto, F.; Palaniappan, K. K.; Deming, L.; Berndt, M.; Brant, A.; Cimermanic, P.; Cox, J. High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nat. Methods* **2019**, *16*, 519–525.
- (55) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **2020**, *17*, 41–44.
- (56) Kuharev, J.; Navarro, P.; Distler, U.; Jahn, O.; Tenzer, S. In-depth evaluation of software tools for data-independent acquisition based label-free quantification. *Proteomics* **2015**, *15*, 3140–3151.
- (57) Searle, B. C.; Swearingen, K. E.; Barnes, C. A.; Schmidt, T.; Gessulat, S.; Kuster, B.; Wilhelm, M. Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat. Commun.* **2020**, *11*, 1548.
- (58) 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, Ş.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*, D442–d450.

(59) Foster, E. D.; Deardorff, A. Open Science Framework (OSF). *J. Med. Libr. Assoc.* **2017**, *105*, 203–206.