Critical assessment of proteome-wide label-free absolute abundance estimation strategies

Erik Ahrné, Lars Molzahn, Timo Glatter and Alexander Schmidt

Proteomics Core Facility, Biozentrum, University of Basel, Basel, Switzerland

There is a great interest in reliable ways to obtain absolute protein abundances at a proteome-wide scale. To this end, label-free LC-MS/MS quantification methods have been proposed where all identified proteins are assigned an estimated abundance. Several variants of this quantification approach have been presented, based on either the number of spectral counts per protein or MS1 peak intensities. Equipped with several datasets representing real biological environments, containing a high number of accurately quantified reference proteins, we evaluate five popular low-cost and easily implemented quantification methods (Absolute Protein Expression, Exponentially Modified Protein Abundance Index, Intensity-Based Absolute Quantification Index, Top3, and MeanInt). Our results demonstrate considerably improved abundance estimates upon implementing accurately quantified reference proteins; that is, using spiked in stable isotope labeled standard peptides or a standard protein mix, to generate a properly calibrated quantification model. We show that only the Top3 method is directly proportional to protein abundance over the full quantification range and is the preferred method in the absence of reference protein measurements. Additionally, we demonstrate that spectral count based quantification methods are associated with higher errors than MS1 peak intensity based methods. Furthermore, we investigate the impact of miscleaved, modified, and shared peptides as well as protein size and the number of employed reference proteins on quantification accuracy.

Keywords:
Absolute protein quantification / APEX / Bioinformatics / emPAI / iBAQ / MS / Top3

1 Introduction

Over the last decade, MS-based proteomics has become an essential tool in biological research. It has been particularly useful in comparative quantitative studies to identify proteins or PTMs with altered abundance between samples of interest. Since the same protein and peptide molecules are compared, detecting quantitative differences is straightforward. By contrast, absolute quantification is more challenging, since different protein and peptide species with various MS ionization efficiencies have to be quantitatively compared within the same sample. To address this issue, most studies published so far employed spiked in heavy labeled reference proteins or peptides of known amounts to determine the absolute abundance for a given protein set. While enabling accurate quantification, the high costs of the required heavy labeled reference proteins and peptides impede the application of this method on a proteome-wide scale.

To circumvent this issue, several research groups have proposed methods to derive absolute protein abundances from
label-free shotgun proteomics datasets, where all identified proteins are assigned an estimated measure of concentration (e.g. copies/cell or mol/L) [1–6]. Although these methods only approximate protein abundances, they are becoming increasingly popular as their implementation is easy and low cost, and the resulting protein quantities are accurate enough to address system-wide biological questions [2, 4, 7–10].

Many variants of such pseudo-quantitative workflows have been proposed, which all employ different metrics extracted from the MS data to represent protein abundances, including the Absolute Protein Expression (APEX) method [2, 4], Exponentially Modified Protein Abundance Index (emPAI) [3], Intensity-Based Absolute Quantification (iBAQ) Index [4, 5, 11, 12], Top3 [5, 13], and MeanInt [11, 14, 15]. As the performance of each method has been demonstrated on separate datasets, experimentalists find little guidance in the literature, when selecting a method for the analysis of their data. To date, the limited number and scale of available reference MS datasets have prevented a meaningful comparison of the different approaches. The few benchmark studies published so far either involve artificial protein mixture containing a limited number of standard proteins [11, 12, 14] or use gene expression data as reference abundance measurements [4, 16].

In this study, we apply five popular label-free LC-MS/MS-based proteome-wide absolute quantification methods to a low complex standard protein mix and several complex biological samples in which absolute protein quantities of endogenous proteins have been accurately determined by Selected Reaction Monitoring (SRM) and stable isotope dilution [7, 15, 17]. Equipped with a high number of accurately quantified endogenous proteins across the various samples, a thorough evaluation of the different quantification methods is presented, for the first time, in a real biological environment.

2 Materials and methods

2.1 LC-MS datasets used for assessment of proteome-wide absolute abundance estimation methods

In total, the following seven LC-MS datasets, obtained from protein samples of different complexity, were employed in this study, including two published large LC-MS datasets from Schizosaccharomyces pombe [18, 19] and Escherichia coli [20].

2.1.1 Standard protein mix

A first LC-MS dataset was acquired from a mixture of 48 precisely quantified proteins spanning five orders of magnitude ranging from 50 pmol to 500 amol (Universal Proteomics Standard, UPS2, Sigma-Aldrich). The protein sample was dissolved in 20 μL digestion buffer (2 M urea, 0.1 M ammonium bicarbonate), reduced with 5 mM tris 2-carboxyethyl phosphine for 60 min at 37°C, and alkylated with 10 mM iodoacetamide for 30 min in the dark at 25°C. After quenching the reaction with 12 mM N-acetyl-cysteine, the protein sample was digested by incubation with sequencing grade Lys-C (1/200, w/w; Wako) for 4 h at 37°C followed by modified trypsin (1/50, w/w; Promega, Madison, Wisconsin) overnight at 37°C. Subsequently, the peptides were desalted on C18 RP spin columns according to the manufacturer’s instructions (Microspin, Harvard Apparatus), dried under vacuum, and stored at −80°C until further use.

Three more proteomic datasets were acquired from the following organisms: (i) Mycoplasma pneumoniae, cells prepared according to Maier et al. [8, 20], (ii) Leptospira interrogans, cells prepared according to Schmidt et al. [7, 19], and (iii) Drosophila melanogaster, cells prepared according to Glatter et al. [18, 21].

Supporting Information Table 1). Cells were harvested by centrifugation at 2000 × g resuspended in lysis buffer (8 M urea, 0.1% RapiGest™, 100 mM ammonium bicarbonate), and lysed by vortexing (3 × 1 min) and sonication (2 × 20 s). Then, protein concentrations in the extracts were determined by a colorimetric bicinchoninic acid assay (Thermo Fisher Scientific) and an aliquot of 1.06 μg of UPS2 standard protein mix was spiked into 5 μg of each protein extract. Protein alkylation, digestion, and peptide cleanup were carried as described above with one modification. Before trypsin digestion, the lysate was diluted 1:4 v/v with 0.1 M ammonium bicarbonate to reduce urea concentration to 1.6 M.

2.1.2 E. coli whole cell digest

LC-MS/MS files were obtained from a recently published study determining absolute protein levels for E. coli strain BW25113 grown under 19 different conditions (Schmidt et al., The quantitative condition-dependent Escherichia coli proteome, submitted for publication, MSB). Protein extraction and LC-MS/MS analysis were performed as described above. Forty-two proteins were absolutely quantified using stable isotope dilution and SRM analysis. Supporting Information Table 2 lists all relevant peptides used in this study.

2.1.3 Ashbya gossypii whole cell digest

A. gossypii (strain Agleu2Δthr4Δ) was grown on full medium containing 10 g/l Bacto Peptone (Pancreatic Digest of Casein, Difco), 10 g/l Yeast Extract (Micro Granulated, Formedium, Norwich, England), 1 g/l Myo-inositol (Merck), and 2% D(+)-Glucose-Monohydrate (Merck). For spore production, A. gossypii was grown on 35 × 10 mm Petri dish (Falcon 1000, Becton Dickinson Labware, New Jersey, USA) by adding 15 g/l agar (Formedium, Norwich, England) to solidify the medium. Cultures were incubated at 30°C in a temperature-controlled environment. Liquid cultures were specially
incubated in baffled flasks (Duran, Schott, Germany) on a rotary shaker (INFORS® AG) at 150 rpm, to guarantee oxygen supply. Spores were isolated according to Brachat et al. [2, 20], inoculated in liquid full medium, and assayed after five different time points. Cells were harvested and prepared for LC-MS analysis as described above.

2.1.4 S. pombe whole cell digest

Only MS data of the unfraccionated samples from a previous system-wide proteomics study of S. pombe [19, 22] were considered for our method evaluation. A list of employed peptides can be found in Supporting Information Table 3.

2.1.5 BSA dilution curve

Peptides from BSA (Sigma-Aldrich) were generated in the same manner as for the UPS2 standard protein mix described above. These were mixed with 1 µg of a whole yeast digest in decreasing amounts starting from 1 pmol to 1 amol using fivefold dilution steps. The yeast peptide sample was prepared as recently specified [3, 21].

2.2 Shotgun LC-MS/MS analyses

One micrograms of peptides (0.212 µg for UPS2 mix alone) were separated on a RP-LC column (75 µm × 38 cm) packed in-house with C18 resin (Magic C18 AQ 3 µm; Michrom BioResources, Auburn, CA, USA) using a linear gradient from 95% solvent A (98% water, 2% ACN, 0.15% formic acid) and 5% solvent B (98% ACN, 2% water, 0.13% formic acid) to 30% solvent B over 120 min at a flow rate of 0.2 µL/min. Each survey scan acquired in the Orbitrap at 60 000 full width half maximum was followed by 20 MS/MS scans of the most intense precursor ions in the linear ion trap with enabled dynamic exclusion for 30 s. Charge state screening was employed to select for ions with at least two charges and rejecting ions with undetermined charge state. The normalized collision energy was set to 32% and one microscan was acquired for each spectrum. CID was triggered when the precursor exceeded 500 ion counts. The ion accumulation time was set to 300 ms (MS) and 50 ms (MS/MS). All samples were measured in triplicates.

2.3 Protein identification and calculation of Protein Quantification Indices (PQIs)

All LC-MS raw data files were imported into Progenesis software (Nonlinear Dynamics, version 4.0) for label-free quantification using default parameters. Data in MASCOT generic format were exported directly from Progenesis and protein identification from the MS/MS data was performed using MASCOT (Matrix Science, version 2.4), where the final protein list was compiled in accordance with the two-peptide rule, upon exclusion of nonprotein-specific peptides, as well as a False Discovery Rate cutoff of 0.01, estimated using a target-decoy search strategy. MASCOT search parameters were set as follows: full tryptic specificity was required (cleavage after lysine or arginine residues unless followed by proline), up to two missed cleavages were allowed, carbamidomethyl (C) was set as fixed modification, oxidation (M) as variable modification, 5 ppm precursor mass tolerance, and 0.6 Da fragment mass tolerance. Results from the database search were imported into Progenesis, mapped to MS1 features, and the resulting feature table exported.

APEX values were obtained using the APEX Quantitative Proteomics Tool [2, 23] upon processing the MASCOT database search results with the Trans Proteomic Pipeline modules PeptideProphet and ProteinProphet [22, 24]. MeanInt, iBAQ, Top3, and emPAI values were calculated, for all confidently identified reference proteins, using an in-house written R-script.

2.4 Absolute quantification of selected endogenous A. gossypii proteins

For 18 proteins covering the entire abundance range, absolute quantification was done using SRM based on heavy labeled reference peptides that served as an internal standard (Supporting Information Table 4). The amount of the spiked reference peptides was adjusted to meet the abundance of the corresponding endogenous protein. Therefore, 5 or 0.5 pmol (depending on protein abundance, see Supporting Information Table 4 for details) of each heavy labeled reference peptide (AQUA grade, Thermo Fisher Scientific) was added to 50 µg of proteins of each sample, respectively, followed by C18 purification using spin columns (Microspin, Harvard Apparatus) according to manufacturer’s instructions. For each peptide the precursor-to-fragment ion transition was calculated for doubly and triply charged ion states of the y-ion series. Data derived from a spectral library generated based on acquired higher-energy collisional dissociation spectra of the standard peptide mix were imported into Skyline version 1.1 (https://skyline.gs.washington.edu/labkey/wiki/home/software/Skyline/page.view?name=default) to match the predicted peptide transitions. Up to five transitions per peptide were traced on a triple quadrupole mass spectrometer (TSQ Vantage, Thermo Electron) connected to an electrospray ione source (Proxeon Biosystems). Peptide separation was carried out using an easy nano-LC system (Proxeon Biosystems) equipped with a RP-HPLC column (75 µm × 38 cm) packed in-house with C18 resin (Magic C18 AQ 3 µm; Michrom BioResources) using a linear gradient from 95% solvent A (0.15% formic acid, 2% ACN) and 5% solvent B (98% ACN, 0.15% formic acid) to 35% solvent B over 90 min at a flow rate of 0.2 µL/min. All MS raw files were imported into the Skyline software for
transition detection, peptide ratio calculation, and absolute protein abundance determination.

3 Results and discussion

3.1 Selected PQIs

In a typical shotgun LC-MS experiment, the results output includes a list of proteins where the presence of each protein is established on the basis of a list of confidently identified peptides. Every peptide map to one or more MS/MS spectra acquired from MS1 features of varying amplitude. Proteome-wide label-free absolute quantification relies on compiling these data into an appropriate quantification index for each protein correlating with its abundance. Subsequently, a simple mathematical model can be derived explaining protein absolute abundance as a function of PQI.

Several methods have been proposed to calculate PQIs, and can be divided into two main categories. The first category correlates the number of Peptide Spectrum Matches, PSMs, or unique parent ions, per protein, with protein abundance. Such PQIs are from now on referred to as spectral count (MS2) based PQIs. The second category of PQIs is based on the integrated peptide ion intensities extracted from the MS1 spectra, here referred to as intensity-based PQIs. We evaluate the most commonly used spectral count and intensity-based methods for label-free protein quantification, utilized for proteome-wide absolute quantification. An overview of the different workflows is shown in Fig. 1.

The emPAI is calculated as $10^{\text{PAI}} - 1$, where PAI is defined as the number of unique parent ions per protein divided by the number of possible peptide precursors per protein [1, 3]. The APEX method, scales the number of confidently identified PSMs per protein with a protein-specific normalization factor, obtained using a classification algorithm predicting the detectability of a protein’s constituting peptides [2, 23]. The emPAI values can be easily obtained using the emPAI calculator [4, 24], alternatively the MS/MS database search algorithm MASCOT (Matrix Science) provides emPAI values per protein as a part of the search results output. APEX scores can be retrieved from database search results processed using the TPP [1, 14] in conjunction with the APEX Quantitative Proteomics Tool [2, 5].

Additionally, different roll-up methods have been proposed to compile peptide feature intensities into PQIs. The iBAQ is the sum of the extracted ion intensities of all identified peptides per protein, normalized by the number of theoretically observable peptides [4, 13]. The PQI referred to as MeanInt, in this manuscript, is simply the average peptide feature intensity per protein [11, 14], while Silva et al. [5, 14] proposed to represent the abundance of each protein by the average intensity of its three best ionizing peptides, Top3. The MaxQuant software suite [4, 13] includes a module for calculating iBAQ values, while the Top3 and MeanInt methods have been implemented in the T3PQ [7, 11, 19, 25, 26] and the Progenesis Post-Processor (NonLinear Dynamics) [14, 26] software tools.

3.2 Translating PQI to protein absolute abundance

Two methods have been proposed to translate PQIs into absolute protein abundances. The first method relies on deriving a calibration curve explaining protein absolute abundance as a function of PQI for a subset of reference proteins of known abundances and extrapolating this model to all identified proteins. This calibration is either based on a set of spiked in unlabeled proteins with known concentrations [4] or endogenous proteins quantified using spiked in heavy reference proteins. This calibration can be obtained upon linear regression of log-transformed protein abundances and PQIs. Alternatively, in the absence of reference proteins, protein absolute abundance is assumed to be directly proportional to PQI values. Here, a measurement of the total protein amount in the sample is required in order to derive the proportionality constant ($\alpha$). (4) Quantification of all identified proteins. The quantification model is applied to all confidently identified proteins.
abundances of the spiked in reference proteins (Eq. 1). The log transformation was applied to stabilize the variance of high MS signals.

\[
\log_{10} C_i = \alpha + \beta \times \log_{10} P_i
\]  

(1)

Note that this can be rewritten as

\[
C_i = 10^\alpha \times P_i^\beta
\]  

(2)

where an exponent parameter \( \beta \) of 1 corresponds to direct proportionality between protein absolute abundance and PQI.

The second nonreference protein-dependent and, therefore, cheaper translation method simply assumes direct proportionality between PQI and protein amount \([3,5,6,9,23,26,27]\). Thus, given a measurement of total protein abundance, which can be obtained in a bicinchoninic acid assay, the abundance of an individual protein is calculated as given in Eq. (2).

\[
C_i = \frac{P_i \times C_{\text{TOT}}}{\sum_{k=1}^{N} P_k}
\]  

(3)

where \( C_i \) is the absolute abundance per protein (e.g. in copies/cell or mol/L), \( P_i \) the PQI per protein, and \( C_{\text{TOT}} \) the total protein abundance in the sample.

For evaluation purposes, we applied the direct proportionality based quantification model to the set of reference proteins of each LC-MS dataset. Therefore \( C_{\text{TOT}} \) (Eq. 2) was not measured, but calculated as the summed known absolute abundance of all identified reference proteins. The two quantification models predicting absolute abundance as a function of PQI are from now on referred to as the calibration curve model and the direct proportionality model, respectively (Fig. 1).

### 3.3 Assessment of the different PQIs

We applied the PQIs described above to a total of seven datasets from samples of varying proteome complexity. Four of the investigated datasets included spiked in reference proteins of known amounts and another three datasets included spiked in AQUA peptides used to obtain absolute abundances of endogenous reference proteins. The expected quantification error associated with each PQI and dataset was assessed for the calibration curve and direct proportionality based quantification models. The combined results reveal whether a given PQI is compatible with the assumption of direct proportionality, that is, the PQI can be applied in the absence of a set of reference proteins used to derive a calibration curve. The accuracy of each PQI was further assessed, with respect to the number of MS-compatible peptides available for each protein, and when discarding miscleaved, modified peptides and peptides shared between multiple proteins.

#### 3.3.1 UPS2 datasets

Analysis of the UPS2 standard protein mix samples consistently showed a higher correlation (Pearson’s R-squared, \( R^2 \)) between MS1 intensity based PQIs and absolute protein abundance than spectral count based PQIs and absolute protein abundance (Fig. 2). This is in accordance with previous studies that found higher quantification accuracies for intensity over spectral count based label-free approaches for relative protein quantification \([28–31,33]\).

Figure 2 also shows that APEX and emPAI values saturated at higher protein concentrations across all datasets. This effect might be caused by the dynamic exclusion during LC-MS analysis that limits the number of PSMs as the degree of sampling reaches maximum sequence coverage. This is a general issue affecting all data-dependent-acquisition LC-MS datasets \([28,32,34]\).

Furthermore, we observed that the Top3 calibration curves were in good agreement with the direct proportionality model, regardless of sample background (Fig. 2B). This means that the Top3 method can be applied in the absence of a set of reference proteins, without introducing a biased quantification error (i.e. the abundances of low- or high-abundant proteins would be consistently over- or underestimated). More important differences between the two quantification models are observed for the other four PQIs, indicating that they are less compatible with the assumption of direct proportionality. It is noticeable that the slope of the calibration curves of these four PQIs seems to depend on sample background complexity. Thus, the application of these PQIs will benefit from a properly calibrated, sample-specific quantification model.

The consequence of incautiously assuming direct proportionality between a PQI and protein abundance is demonstrated in Fig. 2G–I. Here we estimated the proteome-level average quantification fold error employing a bootstrapping algorithm similar to Malmström et al. \([26,28]\). Firstly, in agreement with the Pearson’s R-squared coefficients presented previously, intensity based PQIs were generally associated with lower fold errors than spectral count based PQIs. More particularly, we noticed dramatically higher mean fold error estimates for the iBAQ method when assuming direct proportionality rather than making use of a calibration curve. This effect was prominent for the \( M. \) pneumoniae, \( L. \) interrogans, and \( D. \) melanogaster background samples, but absent in the UPS2 sample alone, once again emphasizing the importance of a properly calibrated, sample-specific quantification model. Supporting Information Fig. 1 displays PQI versus absolute abundance for all four UPS2 datasets, superimposing the direct proportionality model. Here, we see, for example, that the increased quantification error associated with the iBAQ method in combination with the direct proportionality model is explained by a general underestimation of low-abundant proteins. It is important to note that, under normal experimental circumstances, the direct proportionality model does not allow for an assessment...
of an estimated proteome-wide quantification error, as this relies on reference protein measurements.

3.3.2 AQUA peptide spike-in datasets

To further consolidate the findings described above, we applied the five PQIs to three datasets generated from whole cell lysates of *E. coli*, *A. gossypii*, and *S. pombe*, respectively, grown under various conditions. These large datasets comprise substantially more reference protein abundances and therefore are well suited for additional method assessments.

The *E. coli*, *A. gossypii*, and *S. pombe* datasets were analyzed in the same manner as the UPS2 spike-in datasets described above. Globally, the results are in good agreement with those obtained for the UPS2 datasets. Figure 3 provides an overview of the *E. coli* dataset analysis results, while similar overviews for the *A. gossypii* and *S. pombe* datasets are available in the Supporting Information (Supporting Information Figs. 2 and 3).
Figure 3. PQI values versus absolute abundances for all E. coli reference proteins are plotted on a log–log scale, superimposing the calibration curve (blue line), and direct proportionality model (red line). The Pearson’s R-squared of log-transformed PQI and log-transformed absolute abundance is given in upper left-hand corner.

As for the UPS2 samples, we observed that the intensity based PQIs (Fig. 3A–C) were generally associated with higher Pearson’s correlation coefficients than the two spectral count based PQIs (Fig. 3D and E). Accordingly, the fold error distribution of intensity-based PQIs was associated with lower relative errors than spectral count based PQIs, when employing a calibration curve. In the case of the E. coli dataset, the best performing PQIs, that is, iBAQ and MeanInt, had a median fold error of 1.4 while 90% of all reference proteins had a relative error lower than 2.4 (Fig. 4A). For the emPAI PQIs, those metrics were 1.8 and 4.5, respectively. It is once again noticeable that the iBAQ method performed dramatically worse when assuming direct proportionality to absolute abundance.

Figure 4. The per protein quantification error was assessed by leave-one-out cross-validation. (A) The fraction of all E. coli reference proteins quantified with an absolute fold error smaller than the values given on the x-axis, per PQI and quantification model. (B) The absolute fold error as a function of reference protein abundance. (C) The directional fold error as a function of reference protein abundance. The error bars give the SEM.
Given the large number of reference proteins at hand, it is meaningful to study the distribution of relative quantification errors as a function of absolute protein abundance. In this respect, we found that calibration curve-based estimation of protein abundances was associated with an increasing fold error as the reference protein amount decreases (Fig. 4B). For example, the average fold error of a protein present in the sample at 1–5 fmol on column was approximately 1.9 while the average fold error of a 500–1000 fmol on column protein was only around 1.4, for the three intensity-based PQIs when employing a calibration curve. Higher quantification error of low-abundant proteins is to be expected, since their quantification relies on relatively few detected peptides [9,33], which in turn are measured with a relatively high MS signal coefficient of variance [34–36].

In the case of direct proportionality based protein abundance estimation, this trend was much stronger for the iBAQ, emPAI, and APEX PQIs, that is, quantitative estimates of low-abundant proteins were far less accurate than for high-abundant proteins. Looking at the directional fold error (Fig. 4C) of direct proportionality based protein abundance estimation, we observed that the iBAQ method strongly underestimated the protein amounts in the low-abundance range. On the contrary, using the MeanInt, emPAI, and APEX methods led to an overestimation of low-abundant proteins. For instance, in the 1–5 fmol range, the iBAQ method underestimated protein abundances by approximately five folds while the emPAI method overestimated protein abundances with a similar fold error. This observation will be further discussed below.

3.3.3 Evaluation of the direct proportionality assumption

The quantification biases described above were clearly observed for the iBAQ, MeanInt, emPAI, and APEX PQIs in all seven analyzed datasets, but much less prominent for the Top3 PQI where we consistently observed a close agreement between the proportionality model and the calibration curve, that is, a calibration curve β parameter close to one (Eq. 1). Figure 5 presents the calculated β parameters of all analyzed datasets, which is an indicator of how well each PQI meets the assumption of direct proportionality. Based on these findings, we would recommend experimentalists to use the Top3 method unless they are equipped with a set of reference proteins, with known absolute abundances, which allow for appropriate calibration of the mathematical model explaining protein absolute abundance as a function of PQI.

While the iBAQ and MeanInt PQIs outperformed the Top3 method for several of the presented datasets, in terms of mean fold error, it should be noted that they increase in a nonlinear fashion with respect to protein abundance. The extracted MS1 peak intensity of a peptide can be assumed to increase linearly with protein abundance [28,37], but the number of identified peptides per protein will saturate at high protein concentrations. Consequently, the summed peptide intensities per protein (as in iBAQ) should not be expected to increase in direct proportionality with protein abundance over the full quantification range. Neither should one assume direct proportionality between protein abundance and the average peptide intensity per protein. While the average intensity of high-abundant proteins is calculated from a large pool of peptides with different ionization efficiencies, the average intensity of low-abundant proteins will be based on the average intensity of those peptides with the highest ionization efficiency. Consequently, the average peptide intensity of low-abundant proteins will be relatively higher than the peptide intensity of high-abundant proteins.

This is further illustrated in Supporting Information Fig. 4 displaying the calculated MeanInt values, Top3 values, and peptide intensity sums, for a single protein (ALBU_BOVIN) diluted over five orders of magnitude in a constant complex peptide background. We see how the Top3 regression model shares the same slope (calibration curve parameter β of close to 1), as most individual peptides, while the slopes of the iBAQ and MeanInt regression lines are steeper and flatter, respectively. These deviations occur as the number of detected peptide features decreases dramatically towards low protein concentrations, whereas in the upper concentration range, the MeanInt and iBAQ PQIs increase almost parallel to the individual peptides and the Top3 method.

Hence, we advise against using the MeanInt or iBAQ methods without properly calibrating the quantification model. For completeness and experimental cost estimations, we evaluated the quantification error as a function of the number of reference proteins used to derive a calibration curve (Supporting Information Fig. 5). We find that at least 20 protein measurements are required for a stable quantification model. Reference sets of more than 30 proteins do not significantly improve quantification accuracy.

3.3.4 Quantification of small proteins

Different proteins of the same organism display important variability in terms of size. For example, the 10% largest proteins of the human proteome include, on average, almost seven times as many tryptic peptides as the smallest decile of proteins. Consequently the quantitative MS signal associated with a given protein is expected to depend on protein size, which is explicitly accounted for by three of the evaluated quantification methods, that is, iBAQ, emPAI, and APEX.

To further assess our selected PQIs, we studied the quantification error as a function of protein size, here represented by the number of possible tryptic peptides per protein. In order to augment the pool of small-size proteins, our reference proteins were artificially tryptic truncated. Starting from the protein n-terminus, each protein was cut in silico to include up to 5, 10, 15 etc. tryptic peptides. Subsequently, PQI values were calculated for these artificially shortened proteins.
Next, we applied the calibration curves, per PQI, previously derived for full-length reference protein datasets, and calculated the mean fold error per protein size group (Supporting Information Fig. 6). For practical reasons, this evaluation only includes the iBAQ, Top3, MeanInt, and emPAI PQIs. We see a general tendency of higher fold errors for short proteins. This effect is minor for the iBAQ method, while prominent for the Top3 and emPAI methods. For these two PQIs, especially high quantification errors were observed for proteins with fewer than 50 possible tryptic peptides (allowing for two miscleavages) and a molecular weight smaller than 25 kDa on average, which corresponds to approximately 25% of the E. coli proteome. This group of short proteins was generally underestimated when applying the Top3 method and overestimated when applying the emPAI method. Similar results were observed for the A. gossypii and S. pombe datasets (Supporting Information Fig. 6).

It is not surprising that short proteins are quantified with less accuracy when using the Top3 method. This quantification approach is based on the assumption that the most intense peptides per protein have similar MS ionization efficiencies, but clearly one would expect higher Top3 values for the larger protein when comparing two proteins of the same abundance, but different sizes. A simple simulation where peptide intensities were sampled from a realistic intensity distribution, showed a nonlinear relationship between Top3 values and the number of simulated peptides per protein (data not shown). This would have to be adjusted for in order to improve the quantification accuracy of the Top3 method.

Finally, an overestimation of small proteins using the emPAI approach was also observed previously [9, 38, 39]. In this dataset, ribosomal proteins had very high concentration estimates of up to $10^8$ copies/cell, particularly for small proteins of less than 7 kDa. In this regard, the mass of the most abundant 50S ribosomal protein L33 (rpmG) alone (1928 fg) would exceed the expected total protein mass determined for an E. coli cell (~200–300 fg) by several folds [7, 8, 35, 36], indicating a clear overestimation of these small ribosomal proteins. Based on our results, these obvious quantification biases can be explained, at least in part, by the general overestimation of short proteins using the emPAI PQI.

### 3.3.5 Dealing with miscleaved, modified, and shared peptides

All data presented above included peptides containing up to two missed tryptic cleavages and oxidation of methionine. In
the E. coli dataset, miscleaved peptides made up 20% of the total MS1 precursor intensity and 21% of the spectral count sum. For the same dataset, oxidized methionine peptides made up 8% of the total MS1 precursor intensity and 15% of the spectral count sum. Similar numbers were noted for the other analyzed datasets.

Sample preparation induced protein modifications and the efficiency in protein digestion of different cleavage sites can be an important source of MS-signal variation [37, 40]. Thus, given the considerable detection of miscleaved and modified peptides, we evaluated the impact on the mean fold error per PQI when excluding such peptides. This analysis was done for the three AQUA peptide datasets (Supporting Information Fig. 7). Looking at the results of all three datasets no clear trend stands out, however, we see a tendency of spectral count based PQIs being more affected by the exclusion of miscleaved and modified features than the intensity-based methods. Generally, the bootstrapped mean fold error increases for the emPAI and APEX methods as miscleaved and modified peptides are excluded. In conclusion, we find no obvious reason to discard miscleaved and modified peptides from the quantitative analysis.

Along the same lines, we assessed the impact on quantification accuracy when removing peptides shared by multiple proteins. Due to the low number of such peptides present in our datasets, no significant changes were observed (data not shown). However, in complex samples from higher organisms, peptides will be shared at a higher frequency, between entirely different proteins and especially between different protein isoforms. Here, the choice of protein inference method has been shown to affect quantification [38, 39]. While two or more proteins contribute to the intensity of shared peptides, they are detected with relatively intense MS signals. Consequently, we would expect the Top3 method to be more influenced by the removal of shared (and intense) peptides, than the other PQIs, which generally employ more peptides for quantification. This was also observed above upon shortening proteins and thereby reducing the number of peptides available for quantification (Supporting Information Fig. 6).

4 Concluding remarks

We have evaluated five PQIs, iBAQ, MeanInt, Top3, emPAI, and APEX, in the context of proteome-wide absolute quantification estimation. For this purpose we made use of seven datasets of varying complexity, each including a set of accurately quantified reference proteins. Two methods for translating PQI values into absolute abundances were assessed. We studied the estimated quantification error as a function of protein size and investigated the influence of miscleaved, modified, and shared peptides on protein quantification. In summary, the following conclusions can be drawn:

Firstly, we consistently observed a better correlation between intensity-based PQIs and absolute abundance than spectral count based absolute abundance estimation approaches such as the APEX or emPAI method. Although spectral count based absolute abundance estimation approaches such as the APEX or emPAI method are easy to implement, they were associated with comparably large as well as biased quantification errors. It should be stressed that these findings apply to a 1D-LC-MS/MS workflow. When employing a multidimensional peptide separation setup, the computational task of MS1 peak integration is more challenging and thus disfavoring peak intensity based PQIs. Here, spectral count based quantification is likely a more attractive option [41]. In the choice between the two spectral count methods we have evaluated in this study, our results showed that the APEX method generally provided better quantification accuracy than the emPAI method.

Secondly, whenever experimentalists are equipped with a set of reference proteins of known absolute abundances, the quantification model explaining protein abundance as a function of PQI can be appropriately calibrated in order to limit quantification biases. In this set up, the three MS1 peak intensity methods (iBAQ, MeanInt or Top3) had comparable global quantification accuracy. However, the MeanInt and to a greater extent the Top3 methods were associated with large quantification errors of small proteins (<25 kDa).

Thirdly, an alternative and more cost-efficient approach to translate PQIs into absolute abundances is to assume direct proportionality between the two variables. Here, we advocate against using the iBAQ method, since it led to large and biased quantification errors, that is, low-abundant proteins were dramatically underestimated, across all analyzed datasets. The Top3 and MeanInt PQIs came with substantially lower proteome-wide quantification errors, when applying this PQI translation method. In a dilution series experiment carried out to study the relationship between PQI and protein abundance across the full quantification range, we demonstrated that the Top3 method is more compatible with the assumption of direct proportionality. Since this method only requires the analysis of the three best ionizing peptides per protein, it is the only approach applicable for directed [7, 8] or targeted [40] LC-MS approaches. In the latter case, Ludwig et al. recommended a variant of the Top3 method where the two most intense transitions of the three best-flying peptides per protein were used for quantification. When implementing this method, an appropriate prior selection of the best ionizing peptides is crucial to achieve accurate quantification. The observed quantification errors and correlation coefficients in this study were comparable to those obtained for the Top3 approach in our data.

Finally, we found that filtering out modified or miscleaved peptides from our datasets did not lead to a significant improvement of quantification accuracy, for any of the PQIs. We also speculate that the Top3 method would be the least robust to the removal of shared peptides.

In addition to elucidating the different commonly used label-free quantification indices, our study provides a general assessment of the accuracy and precision of LC-MS/MS-based proteome-wide absolute quantification. Across the seven datasets included in this work, the majority of the
quantified proteins, had a quantification error lower than 1.5-fold and 90% of the proteins had a maximum twofold quantification error (when applying a MS1 intensity based PQI and a quantification model calibrated on a sufficiently large set of reference proteins). This level of accuracy and precision is compatible with system-level interpretation relying on protein abundances of distinct functional protein classes and cellular compartments [7, 8, 10]. The accuracy of the obtained protein estimates is also comparable to that achieved by system-wide transcriptome quantification using RNA-sequencing [19]. Thus, label-free protein absolute quantification allows for a meaningful comparison of absolute transcriptome and proteome absolute abundance values in different biological processes and organisms, as demonstrated recently [4, 8, 19].

However, the accuracy is on average not sufficient to determine stoichiometries of protein complexes. Therefore, if accurate protein abundances are needed, SID-based quantification strategies are preferable. With little additional experimental efforts as thoroughly described above, the protein abundance measures obtained in such a targeted MS experiment can be employed to quantify the full proteome, and thus placing the protein set of interest in a system-wide context.

The authors have declared no conflict of interest.

5 References


