

Article

Comparison of different label-free techniques for the semi-absolute quantification of protein abundance

Aarón Millán-Oropeza^{1*}, Mélanie Blein-Nicolas², Véronique Monnet¹, Michel Zivy² and Céline Henry¹

¹ PAPPSO, Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78350, Jouy-en-Josas, France;

² PAPPSO, Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190, Gif-sur-Yvette, France.

* Correspondence: aaron.millan-oropeza@universite-paris-saclay.fr

Abstract: In proteomics, it is essential to quantify proteins in absolute terms if we wish compare results among studies and integrate high-throughput biological data into genome-scale metabolic models. While labeling target peptides with stable isotopes allows protein abundance to be accurately quantified, the utility of this technique is constrained by the low number of quantifiable proteins that it yields. Recently, label-free shotgun proteomics has become the “gold standard” for carrying out global assessments of biological samples containing thousands of proteins. However, this tool must be further improved if we wish to accurately quantify absolute levels of proteins. Here, we used different label-free quantification techniques to estimate absolute protein abundance in the model yeast *Saccharomyces cerevisiae*. More specifically, we evaluated the performance of seven different quantification methods, based either on spectral counting (SC) or extracted-ion chromatogram (XIC), which were applied to samples from five different proteome backgrounds. We also compared the accuracy and reproducibility of two strategies for transforming relative abundance into absolute abundance: a UPS2-based strategy and the total protein approach (TPA). This study mentions technical challenges related to UPS2 use and proposes ways of addressing them, including utilizing a smaller, more highly optimized amount of UPS2. Overall, three SC-based methods (PAL, SAF, and NSAF) yielded the best results because they struck a good balance between experimental performance and protein quantification.

Keywords: label free; metabolic models; *Saccharomyces*; semi-absolute quantification; quantitative proteomics; TPA; UPS2

1. Introduction

Mass spectrometry-based proteomics has become an essential tool in the study of biological processes because it can provide an overall assessment of the proteomes of organisms, cells, organs, and tissues. Quantitative proteomics has made it possible to estimate the abundance of proteins coming from a given biological source and to compare the results obtained under multiple sets of conditions, with a view to assessing differences in protein features and protein involvement in particular processes or metabolic functions.

Depending on the final objective, proteins may be quantified in relative or absolute terms (as reviewed elsewhere [1]). Relative protein abundance is determined using a label-free shotgun approach, which can provide an overall view of proteomes across multiple situations and can detect thousands of proteins. The latter is possible because the approach is simpler and more versatile than label-based methods (e.g., ITRAQ, iCAT, TMT). However, the label-free shotgun approach cannot be used to compare the relative abundances of different proteins within the same sample because each peptide has unique ionization properties. Consequently, other quantification methods have been developed that can estimate the relative abundance and different biochemical properties of proteins (e.g., size, observable peptides, most intense peptides) [2]. Two methods based on spectral

counting (SC)—the protein abundance index (PAI) [3] and the exponentially modified PAI (emPAI) [4]—can be used to compare the levels of different proteins within samples since the values of these indices and molar protein concentrations are positively correlated [4]. However, compared to other quantification methods, SC-based methods, especially emPAI, consistently underperform because they overestimate levels of outlier proteins, notably those that are the most abundant or those with a single peptide spectrum match [5-6]. Another widely used method for comparing proteins within samples is intensity-based absolute quantification (iBAQ) [7], which is equivalent to PAI but uses peptide intensities instead of SC. To date, the accuracy of these methods remains to be evaluated.

In contrast, when proteins are quantified in absolute terms, protein abundances are expressed in clear units (e.g., grams, moles, molecules/cell) and can be easily compared within samples. The absolute quantification method (AQUA)[8-9] is among the most commonly used. It employs isotope-labeled standard synthetic peptides, which are added at known concentrations to cell lysates during digestion to determine the ratio between labeled and unlabeled peptides of interest. Previously used to carry out absolute quantification in mass spectrometry, this technique yields highly accurate measurements of protein abundance. Unfortunately, it has some key drawbacks. It is time consuming and costly, and, more importantly, it can only quantify a limited number of proteins at a time (in general, less than 100) [10-11]. The latter issue is quite limiting in the context of systems biology, given that absolute quantification is needed for large proteomics datasets (i.e., covering 1000s of proteins) if they are to be properly integrated into genome-scale metabolic models, where protein abundances may serve as upper limits [12-13]. In addition, by making available more proteomics datasets that use absolute values, it would be easier to carry out comparative studies across laboratories even when data have been acquired with different devices.

Label-free absolute quantification is therefore useful because it reduces the cost and complexity of sample preparation and allows larger numbers of proteins to be quantified. Indeed, all the peptides detected can be quantified, unlike in methods employing isobaric tags (iTRAQ, TMT) [14-16] that only quantify labeled peptides. Moreover, in label-free methods, there is no limit on the number of samples per experiment, which is not true for commercial label-based methods (e.g., TMTpro™ 16plex Label Reagent Set, Thermo Scientific).

There are two main label-free strategies for transforming the unitless measurements of intensity provided by the mass spectrometer into moles per gram of material. In the current terminology, use of these strategies results in semi-absolute quantification. First, there is the total protein approach (TPA)[17], which is rooted in the assumption that the total mass spectrometry signal (based on SC or extracted-ion chromatogram [XIC]) for all the proteins in a given sample reflects the total amount of protein present. Hence, the signal for a given protein should be proportional to its true abundance in the sample. This strategy has made it possible to carry out semi-absolute quantification without the need for external standards [18]. However, the protein abundances calculated using TPA might be wide of the mark since, in general, more than 60% of peptide fragments are not assigned at the protein identification stage. While numerous studies have applied various forms of TPA [19-23], it remains to be determined whether this strategy accurately estimates protein abundance. Second, there is a commonly employed strategy that relies on an external standard, the Universal Proteomics Standard 2 (UPS2). In this strategy, UPS2 proteins are added in known amounts to establish a standard of reference with which unitless intensities can be converted into concrete abundances. UPS2 contains a mixture of 48 human proteins at six different molar concentrations, where there are eight proteins of different molecular masses present at each concentration level. Multiple studies have found strong positive correlations between the expected and observed relative abundances of UPS2 proteins [7, 24-31]. However, in all cases, there was a need for massive amounts of UPS2 to carry out quantification (e.g., up to 3-10 µg per mass spectrometry run), which is a constraint if the goal is to analyze large cohorts. Indeed,

UPS2 is costly and is not available year round. Furthermore, there is no consensus on either the ratio (protein standard/ proteome background) or the amount of UPS2 required to develop a suitable standard of reference. However, there is a clear need to optimize this strategy by reducing the amount of UPS2 needed while maximizing the number of proteins that can be detected.

This situation therefore calls for the development of a practical, robust label-free technique for carrying out semi-absolute protein quantification that will yield large proteomics datasets of sufficient accuracy to be integrated into metabolic models. To this end, we evaluated fourteen different label-free semi-absolute quantification techniques (seven quantification methods x two transformation strategies). Our study system was chemostat cultures of the budding yeast *Saccharomyces cerevisiae* (CEN.PK113-7D) cultivated under five different sets of conditions.

2. Materials and Methods

2.1. Yeast cultures

Cultures of *Saccharomyces cerevisiae* CEN.PK113-7D were grown in 500-ml chemostats at a dilution rate of 0.1 h^{-1} . We used a synthetic medium containing 5 g/l of $(\text{NH}_4)_2\text{SO}_4$, 3 g/l of KH_2PO_4 , 0.5 g/l of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 7.5 g/l of glucose, trace elements, vitamins, and 1 g/l of pluronic PE6100 to reduce foaming. Cultures experienced one of five different sets of conditions: standard (30 °C, pH=5.5), low pH (30 °C, pH=3.5), high temperature (36 °C, pH=5.5), osmotic stress (30°C, pH=5.5, 1M Kcl), and anaerobic. Each set of conditions was independently replicated three times, resulting in a total of 15 samples (5 conditions x 3 replicates). Sampling was carried out as described elsewhere [32].

2.2. Total protein extraction and in-gel digestion

The protein preparation protocol is described in the Supplementary Methods. A 5- μg aliquot of total proteins was taken from each of the 15 yeast samples; these aliquots were then pooled to generate a representative sample (hereafter, “bulk” sample; Figure 1A). Then, we took 15 μg of total proteins from each of the 15 samples and 6 different bulk samples (the motivation of this choice is explained in the next section) and separated out their peptides on one-dimensional SDS-PAGE short-migration gels (1 × 1 cm lanes, Invitrogen, NP321BOX). Details on the digestion process are available in the Supplementary Methods. Extracted tryptic peptides were vacuum dried and resuspended in 75 μl of loading buffer containing 0.08% (v/v) of trifluoroacetic acid (TFA) and 2% (v/v) of acetonitrile (ACN) in water. The result was digested peptide mixtures at a concentration of 200 ng/ μl . Subsequently, 4 μl of each mixture was analyzed using a high-resolution mass spectrometer.

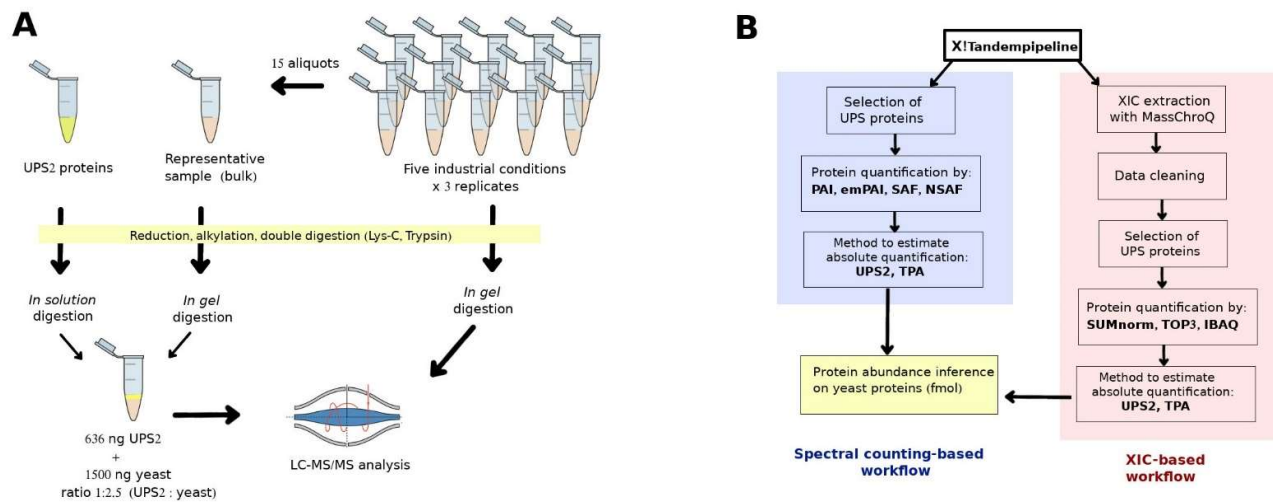


Figure 1. (A) Sample preparation and (B) the bioinformatic approaches used to determine absolute protein abundances.

2.3. Preparation of the UPS2 samples

UPS2 (Sigma) contains 48 human proteins with different molecular masses (6–83 kDa) at 6 different molar concentrations, ranging from 500 amoles to 50 pmoles. The contents of one vial of UPS2 (10.6 µg) underwent reduction, alkylation, and digestion as described in the Supplementary Methods. The extracted tryptic peptides were vacuum dried and resuspended in 25 µl of loading buffer. We took 1.5-µl samples from the mixture of UPS2 digested peptides (424 ng/µl) and spiked them into 7.5 µl of each of the six bulk samples (200 ng yeast peptides/µl) at a ratio of 1:2.35 (UPS2:yeast). Then, 4 µl of this mixture (949 ng) was analyzed using a high-resolution mass spectrometer (Figure 1A). These bulk samples were thus used as references when there was spiking with the digested UPS2 (Sigma). In this fashion, six replicates of UPS2-spiked bulk samples were randomly analyzed over the course of the experiment, and there was no need to spike each experimental sample with UPS2. This approach considerably reduced the quantity of UPS2 required to obtain reliable correlation results.

2.4. Sample preparation for method validation

An experiment was conducted to validate the quantification techniques. Additional bulk samples containing 15 µg of yeast proteins (four independent replicates) were supplemented with purified enzymes at different concentrations: insulin (Sigma, ref. I5500) at 51.9 fmol; alpha-lactalbumin (Sigma, ref. L5385) at 108.6 fmol; myoglobin (Sigma, ref. M0630) at 181.8 fmol; and ribonuclease A (Sigma, ref. R5500) at 342.6 fmol. The samples underwent in-gel digestion as described above. The extracted tryptic peptides were vacuum dried and resuspended in 75 µl of loading buffer. Then, 1.5 µl of digested UPS2 (424 ng/µl) was spiked into 7.5 µl of the digested peptides (from the yeast proteins and purified enzymes). The expected and observed protein abundances for the samples were compared for the different quantification techniques.

2.5. Mass spectrometry analysis

Mass spectrometry was performed using a Dionex U3000 RSLC system coupled to an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer (Thermo Fisher Scientific). Four µl of digested peptides were injected for each sample. A description of the liquid chromatography procedure and the MS/MS method can be found in the Supplementary Methods.

2.6. Protein identification

To identify the proteins, we used a custom-made database (see the Supplementary Methods), which contained genomic information for *S. cerevisiae* CEN.PK113-7D (UNIPROT, accessed on 14/09/2017) and *S. cerevisiae* s288c (accessed on UNIPROT, 03/11/2017); the UPS protein database (Sigma); and a database of common contaminants. Database searches were performed using the X!Tandem algorithm (Alanine 2017.02.01) implemented in the open-source search engine X!TandemPipeline (v. 3.4.3; <https://forgemia.inra.fr/pappso/xtcpp>) [33]. Enzymatic cleavage was defined as trypsin digestion with one possible miscleavage. The settings for the carboxyamidomethylation of cysteine residues and the oxidation of methionine residues were static modifications and potential modifications, respectively. Precursor and fragment mass tolerance was 10 ppm. The data filtering standards were as follows: peptide E-value < 0.01, protein log(E-value) < -3, and a minimum of one identified peptide per protein. Using such filtering criteria, the peptide and protein false discovery rates (FDR) were 0.04% and 0.68%, respectively. The mass spectrometry data were deposited online in the public database PROTEOMICS IDENTIFICATION DATABASE [34-36](repository: <http://pappso.moulon.inra.fr/protic/proticprod/angular/#/projects/199>) and in the Proteomics Identification Database [37] (PRIDE; dataset identifiers PXD014765 and PXD012836). PRIDE is a member of the ProteomeXchange Consortium.

2.7. Protein quantification

Protein abundance was quantified using SC- and XIC-based methods (Figure 1B). The R code that we used is available at <https://forgemia.inra.fr/aaron.millan-oropeza/protquant-saq/>. Four SC-based methods were employed. The first was the protein abundance index (PAI), in which the observed number of peptides is divided by the expected number of peptides [3]. The second was the exponentially modified PAI (emPAI), which is equal to $10^{\text{PAI} - 1}$ [4]. The third was the spectral abundance factor (SAF), in which the SC for a given protein is divided by the protein's length (L). The fourth was the normalized spectral abundance factor (NSAF), which is defined for a given protein i as follows:

$$NSAF_i = \frac{(SC/L)_i}{\sum_{j=1}^N (SC/L)} \quad (1)$$

where the number of SCs identified for protein i is divided by the protein's length (L) in amino acids, and the result is then divided by the sum of SC/L for all the proteins (N) in the experiment [38].

For the XIC-based methods, XIC extraction was carried out using MassChroQ [39] (v. 2.2.16); the peak detection threshold was between 30,000 and 50,000, and the range was 10 ppm. The resulting peptide intensities (i.e., the areas under the curve) were log10-transformed before carrying out further data analyses. We eliminated peptides when their standard deviation of retention time was higher than 20 s. A local normalization method, described elsewhere [40], was applied. Peptides that belonged to multiple proteins were removed, except in the method in which absolute abundance was intensity based (iBAQ, described below). Proteins with at least two peptides were kept for further quantification. Three XIC-based methods were used. The first was SUMnorm, in which the sum of the XIC results are normalized by amino acid number [41]. Second was TOP3, in which the mean intensity for the three most intense peptides is associated with a protein [42]. Third was iBAQ, in which the sum of all the MS1 peptide intensities is associated with a protein and divided by the expected number of observable peptides [7].

Two different strategies for transforming relative abundance into absolute abundance were applied to the results of the seven quantification methods: i) an adapted TPA [17] that takes into account total observed protein abundance (Equation. 2) and ii) a UPS2-based strategy in which a standard of reference is established using the linear

regression between the expected and observed abundances of UPS2 proteins (fmol). Here, the observed abundances were estimated via the different quantification methods (Figure 1B). A total of 70 datasets were generated (7 quantification methods \times 2 transformation strategies \times 5 sets of culture conditions).

$$Proteinconcentration_{ikm} = \frac{Relativeabundance_{ikm} Totalmass_k}{Totalrelativeabundance_{km} MW_i} \quad (2)$$

Where *relative abundance* is the abundance of protein *i* as estimated by method *m* using sample *k*; *MW* is the molecular weight (g/mol) of protein *i*; *total mass* is the total amount of digested proteins (g) in sample *k*; and *total relative abundance* represents the sum of all the abundances estimated by method *m* using sample *k*. *Protein concentration* is expressed in mol.

2.8. Data analysis

We used R [43] (v. 3.5.1) to quantify protein abundance and carry out the data analyses. The performance of the different techniques (the seven quantification methods \times the two transformation strategies) was assessed with various metrics. These metrics were i) accuracy intra-samples, as estimated by bootstrapping the results for the UPS2 proteins (see the Supplementary Methods for details); ii) the coefficient of determination (r^2) obtained from the linear regression between the expected and observed abundances of UPS2 proteins (fmol) for a given quantification method; iii) repeatability, as estimated among all the bulk samples using all the results for the UPS2 proteins (i.e., the CV among replicates); iv) the variability associated with the results for the UPS2 proteins, which was evaluated for each order of magnitude of concentration and reported as a median value (i.e., the CV among proteins); v) the number of proteins whose abundance was quantified; and vi) the ratio between the estimated total mass of all the quantified proteins and the mass of the digested proteins that were actually injected:

$$Massratio_{kt} = \frac{\sum_{j=1}^N Proteinconcentration_{it} MW_i}{Totalmass_k} \quad (3)$$

where *protein concentration* (mol) is the estimated abundance of a given protein *i* obtained using a given technique *t*; *MW* is the molecular weight (g/mol) of protein *i*; *total mass* is the total mass of digested proteins (g) in a sample *k*; and *N* is the total number of quantified proteins in a sample *k*.

3. Results and discussion

3.1. Implementation of the UPS2-based strategy in yeast

To determine the maximum number of UPS2 proteins that we could identify under our experimental conditions, a 400-ng dose of pure UPS2 peptides (i.e., not spiked in a yeast background) was analyzed in a single LC-MS/MS run. A total of 34 out of 48 UPS2 proteins were identified across 5 molar concentrations (ranging from 0.12 fmol to 1200 fmol; Figure S1), which indicates that not all the UPS2 proteins were detectable in our conditions, especially those at the lowest concentration level (0.5 fmol). These results fit with what was observed by Tsou et al, [44], who compared the performance of data-independent acquisition (DIA) methods and data-dependent acquisition (DDA) methods across multiple samples, including a UPS2 sample. Although the injection amount was not specified, it was indicated that the researchers were able to identify 34 proteins after combining three different search engines in the DIA pipeline; however, none of the UPS2 proteins at the lowest concentration level (0.5 fmol) were detected.

Using 1D SDS-PAGE stacking gels, we separated out the peptides in the samples of the *S. cerevisiae* CEN.PK113-7D cultures grown under the five sets of conditions (i.e., standard, low pH, high temperature, osmotic stress, and anaerobic) and in the bulk samples. After in-gel digestion, the UPS2 peptides were spiked into the bulk samples at a

ratio of 1:2.4 (UPS2:yeast). This ratio was chosen based on the results of previous assays, in which the use of lower relative quantities of UPS2 (ratios ranging from 1:5 to 1:647, Figure S2) resulted in only small numbers of UPS2 proteins being detected (<15), precluding the ability to perform regression analysis. Here, our UPS2:yeast ratio (1:2.4) resulted in 282 ng of total UPS2 peptides being analyzed and up to 28 UPS2 proteins across five molar concentrations being detected in a single LC-MS/MS run (File S1).

Other studies have spiked samples from different organisms with UPS2 before the digestion step. They detected the following number of UPS2 proteins: 32 across four molar concentrations for a full vial of UPS2 (10.6 µg) [24]; 25 across five molar concentrations for a 3.3-µg sample of UPS2 [30]; and 24 across four molar concentrations for a 4.24-µg sample of UPS2 [7]. In these studies, large amounts of UPS2 were used to detect a reasonable number of proteins spanning four orders of magnitude in concentration. Here, we optimized the amount of UPS2 that we used to spike the samples (636 ng, which represents up to 16 times less than in the studies mentioned above) while still aiming to detect an acceptable number of proteins across several molar concentrations. Since only the bulk samples, which served as our references, were spiked with UPS2, the operational costs were considerably reduced. As a result, it is possible to increase the number of bulk samples within an experiment, thus allowing additional experimental samples/conditions to be added.

Based on the recommended amount of UPS2 to be spiked (Figure 1A), the number of bulk samples that can be used climbs from 16 (as obtained from a single UPS2 vial in previous studies) to only 2 [7], 3 [30], or 10 [26] samples.

3.2. Performance of the quantification methods with the UPS2 proteins

The abundances of the 28 UPS2 proteins detected in the bulk samples were quantified using seven different methods (based on SC: PAI, emPAI, SAF, and NSAF; based on XIC: SUMnorm, Top3, and iBAQ; Figure 1B). Prior to carrying out the final regression models, a Cook's distance analysis [45] was performed to identify any outliers that could negatively affect the linear models. An outlier was defined as a point with a Cook's distance of more than three times the value of the mean (μ). Based on the results of this analysis, the proteins UBIQ (P62988), COS5 (P01031), and SYUG (O76070) were removed from the dataset. The short length of these proteins coupled with the proximity of their arginine and lysine residues resulted in small peptides that were poorly detected after mass fragmentation. In similar studies performed at our proteomics facility using different yeast strains and bacteria (data not shown), COS5 and SYUG were also systematically removed from datasets, suggesting that these proteins cannot be properly quantified via label-free shotgun techniques.

Linear regression analysis was thus carried out using the data for the remaining 25 UPS2 proteins. The relationships between the expected and observed protein abundances for the different quantification methods yielded various standards of reference (File S2). Overall, the regression parameters and statistical measures (i.e., intercept, slope, and coefficient of determination) were reproducible across the bulk samples for each quantification method. The six independent bulk sample replicates were employed to establish a final standard of reference that was used to determine the linear regression parameters for estimating the absolute abundance (in fmol) of the yeast proteins. Different metrics were used to evaluate the performance of each quantification method in tandem with the use of this standard of reference (Table 1).

Table 1. Metrics used to evaluate the performance of the quantification methods, based either on spectral counting (SC) or extracted ion chromatogram (XIC). The coefficients of determination (r^2) and variation (CV) among replicates were calculated from the bulk samples (median values; n=6 samples). The among-protein estimates of CV took into account the variability among UPS2 proteins at all the orders of magnitude of concentration (median values; n=25 proteins). The number of

proteins quantified per sample was determined using samples of yeast grown under five different sets of conditions (median; n=15).

| Quantification methods | | Linearity (r^2) | CV among proteins (%) | CV among replicates (%) | Proteins per sample | Proteins in the whole experiment (n=15) |
|------------------------|---------|---------------------|-----------------------|-------------------------|---------------------|---|
| SC-based | PAI | 0.89 | 48.8 | 10.2 | 1523 | 2204 |
| | emPAI | 0.61 | 161.4 | 59.1 | | |
| | SAF | 0.90 | 48.0 | 10.2 | | |
| | NSAF | 0.90 | 48.0 | 10.9 | | |
| XIC-based | SUMnorm | 0.96 | 52.9 | 10.0 | 1447 | 1556 |
| | TOP3 | 0.91 | 62.6 | 10.5 | 1423 | |
| | iBAQ | 0.96 | 51.3 | 10.0 | 1468 | |

For most of the methods, CV was low among replicates (<11%), with the exception of emPAI. The CV among proteins varied as a function of protein concentration (File S2). As expected, the XIC-based methods had higher r^2 values for their CVs among proteins and among replicates than did the SC-based methods, the same trend as seen elsewhere [2]. In another study, the CV among replicates ranged from 1 to 16% when AQUA [46] was used. The methods used in this study yielded similar CVs among replicates (Table 1). Furthermore, the total number of proteins for which absolute abundance could be estimated (>2000) was higher than that obtained with AQUA (<50).

3.3. Performance of the semi-absolute quantification techniques with the UPS2 proteins

While CV is a useful metric for quantifying repeatability among replicates and proteins, it cannot be used to assess the ability of quantification methods to accurately determine the abundance of a given protein. For this reason, the accuracy of semi-absolute quantification among samples was determined for the different techniques (seven quantification methods x two transformation strategies; Figure 2) by applying an iterative bootstrap to the results for each UPS2 protein (see the Supplementary Methods). Overall, the XIC-based techniques provided more accurate results than did the SC-based techniques. However, when a t-test was used to compare the results obtained using SUMnorm and SAF (i.e., the best XIC- and SC-based methods, respectively), no significant difference was found (p = 0.209).

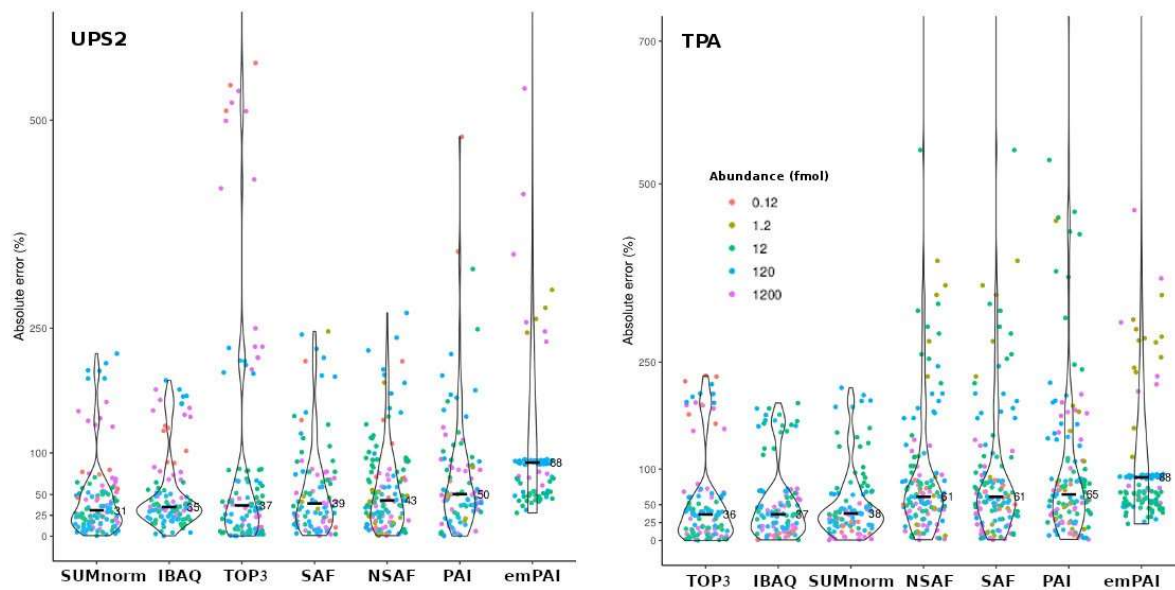


Figure 2. Absolute error among samples for the different quantification methods as obtained by bootstrapping the data for the spiked UPS2 proteins. Results for the two transformation strategies: the UPS2-based approach (left) and TPA (right). The points are displayed as a function of the different UPS2 protein concentrations (fmol). The inner marks indicate median absolute error ($n = 144$).

3.4. Performance of the semi-absolute quantification techniques with external proteins

We used an additional experiment to determine which transformation strategy most accurately converted relative abundance into absolute abundance. To this end, we compared the results obtained using the seven quantification methods in tandem with either TPA or the UPS2-based strategy (Figure 3A). In this experiment, bulk samples of yeast were supplemented with purified enzymes: insulin (INS) at 51.9 fmol; alpha-lactalbumin (LALBA) at 108.6 fmol; myoglobin (MYG) at 181.8 fmol; and ribonuclease A (RNAS1) at 342.6 fmol. The mixtures were separated and digested in gel; then, the digested UPS2 peptides were spiked into the samples before determining the absolute abundance of the enzymes (Figure S3).

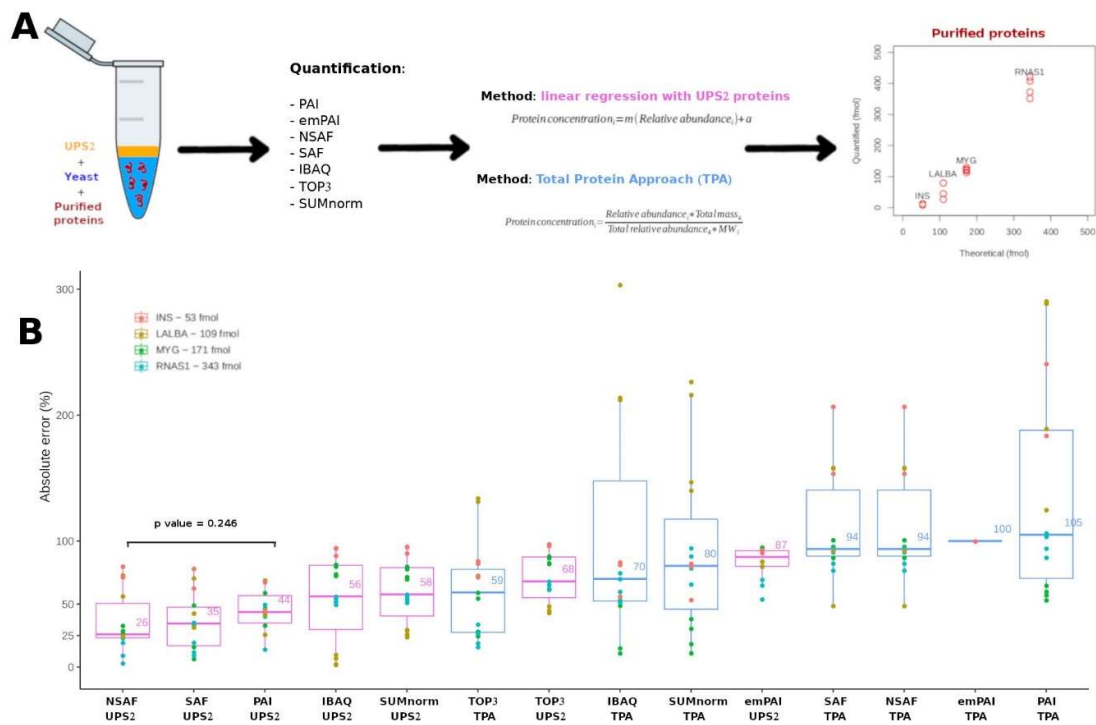


Figure 3. (A) Workflows used for validation. For TPA and the UPS2-based strategy, i refers to a given protein from a given LC-MS/MS sample k . For the standard of reference obtained using the UPS2-based strategy, m represents the slope and a represents the y-intercept. (B) Comparison of the different semi-absolute quantification techniques using the absolute error of the estimated abundance of purified proteins at known concentrations. Results of the t-test comparing the NSAF (UPS2) and PAI (UPS2) methods. The UPS2-based strategy is indicated in the pink boxes. TPA is indicated in the blue boxes.

A rough estimate of overall accuracy was obtained using the total mass ratio (Eq. 3). Using this metric, the XIC-based methods (SUMnorm, TOP3, iBAQ) displayed better performance than the SC-based methods (Table 2). This finding was consistent with the results of the previous experiment (Figure 2). When the transformation strategies were compared, TPA arrived at total mass ratios that were closer to the expected values (mass ratio = 1) than did the UPS2-based strategy. However, this metric should be interpreted carefully since, by definition, TPA calculations are already normalized using total mass.

Table 2. Total mass ratio calculated for purified proteins and the spiked UPS2 proteins (median values) using the two transformation strategies (UPS2 and TPA).

| Quantification methods | Purified proteins | | Spiked UPS2 proteins | |
|------------------------|-------------------|------|----------------------|------|
| | UPS2 | TPA | UPS2 | TPA |
| iBAQ | 0.15 | 0.65 | 0.69 | 0.96 |
| SUMnorm | 0.16 | 0.89 | 0.74 | 0.95 |
| TOP3 | 0.21 | 0.53 | 1.09 | 0.96 |
| NSAF | 0.21 | 0.16 | 1.21 | 0.92 |
| SAF | 0.22 | 0.16 | 1.19 | 0.92 |
| PAI | 0.17 | 0.15 | 1.17 | 0.92 |
| emPAI | 0.12 | 3.67 | 84.83 | 0.99 |

Absolute error was calculated by comparing the observed and expected abundances of the enzymes. The techniques with the lowest median absolute errors were NSAF (26%),

SAF (35%), and PAI (44%) used in tandem with the UPS2-based transformation strategy (Figure 3B). These absolute error values are comparable to those of the commercial kit READYBEADS™ (ANALYTICAL, France), which is used to quantify absolute abundance with a BSA standard at different concentrations [47]. Another study indicated that TPA could be preferentially employed because it yielded similar results to those of the UPS2-based strategy when used with iBAQ applied to a single yeast sample [19]. This finding fits with the accuracy among samples observed in our study (Figure 2); it contrasts with what we observed for the absolute error of the external proteins (Figure 2.B). Indeed, TPA offers the advantage of not relying on external standards.

Taking the results together, the NSAF-UPS2, PAI-UPS2, and SAF-UPS2 techniques displayed a good balance between performance (linearity, reproducibility, accuracy) and the number of proteins quantified. We performed a PCA on the protein abundances (fmol) estimated using the three semi-absolute quantification techniques with the lowest median absolute error with UPS2 approach (Figure S4). In the PCA, the five sets of culture conditions were clearly differentiated (>33% of variance considering the PC1 and PC2), indicating that these quantification techniques could distinguish among the biological features of datasets obtained from different proteome backgrounds. The same pattern was observed for a PCA of the results for the techniques SUMnorm-TPA, iBAQ-TPA, and TOP3-TPA (Figure S4).

Using any of the top performing SC-based methods (PAI, SAF, NSAF) with the UPS2-based transformation strategy represents a simple way to carry out semi-absolute quantification within a routine shotgun workflow. Nevertheless, the use of XIC-based methods like iBAQ or SUMnorm with either transformation strategy could prove interesting to implement within new workflows that rely exclusively on peptide intensity data, such as DIA methods.

Based on the previous findings, it could be helpful to replace relative quantification with semi-absolute quantification via the SUMnorm-TPA or iBAQ-TPA techniques in the following situations: i) the research aims involve the analysis of human proteins, since UPS2 contains human proteins; ii) UPS2 is not available; or iii) a large-scale experiment is being carried out that needs cost-effective and straightforward wet-lab procedures.

4. Conclusions

In this study, using five different proteome backgrounds, we demonstrated the feasibility of utilizing two transformation strategies—TPA and a UPS2-based approach—as the foundation for label-free semi-absolute quantification. Based on the performance of the fourteen different techniques that we tested, we recommend employing a SC-based quantification method—PAI, SAF, or NSAF—in conjunction with the UPS2-based transformation strategy when calculating protein abundances (fmol) in complex mixtures. The techniques tested in this study remain imperfect, given that the CVs we obtained were inferior to those associated with methods based on isotope labeling (e.g., AQUA, SILAC). However, they do represent a potentially helpful tool for performing semi-absolute quantification in studies where thousands of proteins are being examined and the goal is to generate datasets that can be integrated into metabolic models. Conversely, the XIC-based methods SUMnorm or iBAQ coupled with the TPA transformation strategy could provide a cheaper and simpler alternative for quantifying semi-absolute abundance in large-scale experiments. The proposed techniques could also be employed in fundamental research looking at the metabolic dynamics of microbes, plants, and other organisms. They could be particularly useful in cases where we know little about the stoichiometry of reference proteins, given that such information is needed for absolute quantification.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, **Figure S1.** UPS2 proteins detected in a 400-ng sample analyzed in a single mass spectrometry run (34/48) performed with an Orbitrap™ Fusion™ Lumos Tribrid™. **Figure S2.** Spectral counts of UPS2 proteins detected when spiked into the bulk samples at different ratios of UPS2:yeast. The ratios

(ups2:YEAST) were 1:20, 1:45, 1:90, 1:150, 1:302 and 1:647. **Figure S3.** Comparison of different methods to assess semi-absolute quantification in the purified proteins at known concentrations. The purified proteins are: Insulin from bovine pancreas (INS) at 51.9 fmol, Alpha-lactalbumin from bovine milk (LALBA) at 108.6 fmol, Myoglobin from equine skeletal muscle (MYG) at 181.8 fmol and Ribonuclease A from bovine pancreas (RNAS1) at 342.6 fmol (n=4). **Figure S4.** Principal component analysis of the absolute abundances (fmol) of yeast proteins from the selected SC-based (UP2 approach) quantification methods NSAF, SAF and PAI. As well as the XIC-based methods iBAQ, TOP3 and SUMnorm (TPA). **File S1.** Proteins identified using X!TandemPipeline. **File S2.** Results of the linear regressions for the seven different quantification methods performed using the data from the bulk samples spiked with UPS2 proteins. **File S3.** Protein abundances (fmol) for *S. cerevisiae* CEN.PK113-7D grown under the five sets of culture conditions. Estimates were obtained using the 14 different techniques (7 quantification methods x 2 transformation strategies).

Author Contributions: A.M.O. and C.H. performed the LC-MS/MS analysis. A.M.O. carried out the statistical analyses and wrote the manuscript. All the authors helped design the experiments and reviewed the manuscript.

Funding: This research was funded by the European Union's Horizon 2020 research programme and the CHASSY innovation program, grant number 720824.

Data Availability Statement: Mass spectrometry data were deposited online in the public database PROTEOMICS <http://pappso.moulon.inra.fr/protic/proticprod/angular/#/projects/199> and in the Proteomics Identification Database (PRIDE) dataset identifiers PXD014765 and PXD012836.

Acknowledgments: The proteomics analyses were performed at the PAPPISO platform (<http://pappso.inra.fr>), which is funded by INRAE (<http://www.inrae.fr>), the Ile-de-France regional council (<https://www.iledefrance.fr/education-recherche>), IBISA (<https://www.ibisa.net>), and CNRS (<http://www.cnrs.fr>). We thank Olivier Langella for making bioinformatic improvements to the search engine tools, and we wish to acknowledge the contributions of our colleagues in the CHASSY consortium.

Conflicts of Interest: "The authors declare no conflict of interest."

References

1. Ankney, J. A.; Muneer, A.; Chen, X. Relative and Absolute Quantitation in Mass Spectrometry-Based Proteomics. *Annu Rev Anal Chem.* **2018**, *11* (1), 49–77. DOI: 10.1146/annurev-anchem-061516-045357.
2. Blein-Nicolas, M.; Zivy, M. Thousand and One Ways to Quantify and Compare Protein Abundances in Label-Free Bottom-up Proteomics. *Biochim. Biophys. Acta.* **2016**, *1864* (8), 883–895. DOI: 10.1016/j.bbapap.2016.02.019.
3. Rappsilber, J.; Ryder, U.; Lamond, A. I.; Mann, M. Large-Scale Proteomic Analysis of the Human Spliceosome. *Genome Res.* **2002**, *12* (8), 1231–1245. DOI: 10.1101/gr.473902.
4. Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. Exponentially Modified Protein Abundance Index (EmPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Mol. Cell Proteomics.* **2005**, *4* (9), 1265–1272. DOI: 10.1074/mcp.M500061-MCP200.
5. Bubis, J. A.; Levitsky, L. I.; Ivanov, M. V.; Tarasova, I. A.; Gorshkov, M. V. Comparative Evaluation of Label-Free Quantification Methods for Shotgun Proteomics. *Rapid Commun. Mass Spectrom.* **2017**, *31* (7), 606–612. DOI: 10.1002/rcm.7829.
6. Dowle, A. A.; Wilson, J.; Thomas, J. R. Comparing the Diagnostic Classification Accuracy of ITRAQ, Peak-Area, Spectral-Counting, and EmPAI Methods for Relative Quantification in Expression Proteomics. *J. Proteome Res.* **2016**, *15* (10), 3550–3562. DOI: 10.1021/acs.jproteome.6b00308.
7. Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global Quantification of Mammalian Gene Expression Control. *Nature.* **2011**, *473* (7347), 337–342. DOI: 10.1038/nature10098.
8. Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P. Absolute Quantification of Proteins and Phosphoproteins from Cell Lysates by Tandem MS. *PNAS.* **2003**, *100* (12), 6940–6945. DOI: 10.1073/pnas.0832254100.
9. Kirkpatrick, D. S.; Gerber, S. A.; Gygi, S. P. The Absolute Quantification Strategy: A General Procedure for the Quantification of Proteins and Post-Translational Modifications. *Methods.* **2005**, *35* (3), 265–273. DOI: 10.1016/j.ymeth.2004.08.018.
10. Brun, V.; Dupuis, A.; Adrait, A.; Marcellin, M.; Thomas, D.; Court, M.; Vandenesch, F.; Garin, J. Isotope-Labeled Protein Standards: Toward Absolute Quantitative Proteomics. *Mol. Cell Proteomics.* **2007**, *6* (12), 2139–2149. DOI: 10.1074/mcp.M700163-MCP200.
11. Bär, C.; Mathis, D.; Neuhaus, P.; Dürr, D.; Bisig, W.; Egger, L.; Portmann, R. Protein Profile of Dairy Products: Simultaneous Quantification of Twenty Bovine Milk Proteins. *Int. Dairy J.* **2019**, *97*, 167–175. DOI: 10.1016/j.idairyj.2019.01.001.

12. Sánchez, B. J.; Nielsen, J. Genome Scale Models of Yeast: Towards Standardized Evaluation and Consistent Omic Integration. *Integr Biol.* **2015**, *7* (8), 846–858. DOI: 10.1039/c5ib00083a.
13. Monk, J.; Nogales, J.; Palsson, B. O. Optimizing Genome-Scale Network Reconstructions. *Nat. Biotechnol.* **2014**, *32* (5), 447–452. DOI: 10.1038/nbt.2870.
14. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-Reactive Isobaric Tagging Reagents. *Mol. Cell Proteomics.* **2004**, *3* (12), 1154–1169. DOI: 10.1074/mcp.M400129-MCP200.
15. Choe, L.; D'Ascenzo, M.; Relkin, N. R.; Pappin, D.; Ross, P.; Williamson, B.; Guertin, S.; Pribil, P.; Lee, K. H. 8-Plex Quantitation of Changes in Cerebrospinal Fluid Protein Expression in Subjects Undergoing Intravenous Immunoglobulin Treatment for Alzheimer's Disease. *Proteomics.* **2007**, *7* (20), 3651–3660. DOI: 10.1002/pmic.200700316.
16. Werner, T.; Becher, I.; Sweetman, G.; Doce, C.; Savitski, M. M.; Bantscheff, M. High-Resolution Enabled TMT 8-Plexing. *Anal. Chem.* **2012**, *84* (16), 7188–7194. DOI: 10.1021/ac301553x.
17. Wiśniewski, J. R.; Ostasiewicz, P.; Duś, K.; Zielińska, D. F.; Gnad, F.; Mann, M. Extensive Quantitative Remodeling of the Proteome between Normal Colon Tissue and Adenocarcinoma. *Mol Syst Biol.* **2012**, *8*, 611. DOI: 10.1038/msb.2012.44.
18. Wiśniewski, J. R.; Hein, M. Y.; Cox, J.; Mann, M. A "Proteomic Ruler" for Protein Copy Number and Concentration Estimation without Spike-in Standards. *Mol Cell Proteomics.* **2014**, *13* (12), 3497–3506. <https://doi.org/10.1074/mcp.M113.037309>.
19. Sánchez, B. J.; Lahtvee, P.-J.; Campbell, K.; Kasvandik, S.; Yu, R.; Domenzain, I.; Zelezniak, A.; Nielsen, J. Benchmarking Accuracy and Precision of Intensity-Based Absolute Quantification of Protein Abundances in *Saccharomyces cerevisiae*. *BioRxiv.* **2020**, 2020.03.23.998237. DOI: 10.1101/2020.03.23.998237.
20. Vildhede, A.; Wiśniewski, J. R.; Norén, A.; Karlgren, M.; Artursson, P. Comparative Proteomic Analysis of Human Liver Tissue and Isolated Hepatocytes with a Focus on Proteins Determining Drug Exposure. *J. Proteome Res.* **2015**, *14* (8), 3305–3314. DOI: 10.1021/acs.jproteome.5b00334.
21. Belouah, I.; Nazaret, C.; Pétriacq, P.; Prigent, S.; Bénard, C.; Mengin, V.; Blein-Nicolas, M.; Denton, A. K.; Balliau, T.; Augé, S.; Bouchez, O.; Mazat, J.-P.; Stitt, M.; Usadel, B.; Zivy, M.; Beauvoit, B.; Gibon, Y.; Colombié, S. Modeling Protein Destiny in Developing Fruit1. *Plant Physiol.* **2019**, *180* (3), 1709–1724. DOI: 10.1104/pp.19.00086.
22. Belouah, I.; Bénard, C.; Denton, A.; Blein-Nicolas, M.; Balliau, T.; Teyssier, E.; Gallusci, P.; Bouchez, O.; Usadel, B.; Zivy, M.; Gibon, Y.; Colombié, S. Transcriptomic and Proteomic Data in Developing Tomato Fruit. *Data Brief.* **2020**, *28*, 105015. DOI: 10.1016/j.dib.2019.105015.
23. Sabatier, P.; Saei, A. A.; Wang, S.; Zubarev, R. A. Dynamic Proteomics Reveals High Plasticity of Cellular Proteome: Growth-Related and Drug-Induced Changes in Cancer Cells Are Comparable. *Proteomics.* **2018**, *18* (24), e1800118. DOI: 10.1002/pmic.201800118.
24. Soufi, B.; Krug, K.; Harst, A.; Macek, B. Characterization of the E. Coli Proteome and Its Modifications during Growth and Ethanol Stress. *Front Microbiol.* **2015**, *6*, 103. DOI: 10.3389/fmicb.2015.00103.
25. Carpy, A.; Krug, K.; Graf, S.; Koch, A.; Popic, S.; Hauf, S.; Macek, B. Absolute Proteome and Phosphoproteome Dynamics during the Cell Cycle of *Schizosaccharomyces Pombe* (Fission Yeast). *Mol. Cell Proteomics.* **2014**, *13* (8), 1925–1936. DOI: 10.1074/mcp.M113.035824.
26. Ahrné, E.; Molzahn, L.; Glatter, T.; Schmidt, A. Critical Assessment of Proteome-Wide Label-Free Absolute Abundance Estimation Strategies. *Proteomics.* **2013**, *13* (17), 2567–2578. DOI: 10.1002/pmic.201300135.
27. Trudgian, D. C.; Ridlova, G.; Fischer, R.; Mackeen, M. M.; Ternette, N.; Acuto, O.; Kessler, B. M.; Thomas, B. Comparative Evaluation of Label-Free SING Normalized Spectral Index Quantitation in the Central Proteomics Facilities Pipeline. *Proteomics.* **2011**, *11* (14), 2790–2797. DOI: 10.1002/pmic.201000800.
28. Krey, J. F.; Wilmarth, P. A.; Shin, J.-B.; Klimek, J.; Sherman, N. E.; Jeffery, E. D.; Choi, D.; David, L. L.; Barr-Gillespie, P. G. Accurate Label-Free Protein Quantitation with High- and Low-Resolution Mass Spectrometers. *J. Proteome Res.* **2014**, *13* (2), 1034–1044. DOI: 10.1021/pr401017h.
29. Wu, Q.; Shan, Y.; Qu, Y.; Jiang, H.; Yuan, H.; Liu, J.; Zhang, S.; Liang, Z.; Zhang, L.; Zhang, Y. Improved Accuracy for Label-Free Absolute Quantification of Proteome by Combining the Absolute Protein EXpression Profiling Algorithm and Summed Tandem Mass Spectrometric Total Ion Current. *Analyst.* **2014**, *139* (1), 138–146. DOI: 10.1039/c3an01738a.
30. Smits, A. H.; Lindeboom, R. G. H.; Perino, M.; van Heeringen, S. J.; Veenstra, G. J. C.; Vermeulen, M. Global Absolute Quantification Reveals Tight Regulation of Protein Expression in Single *Xenopus* Eggs. *Nucleic Acids Res.* **2014**, *42* (15), 9880–9891. DOI: 10.1093/nar/gku661.

31. Kilani, J.; Davanture, M.; Simon, A.; Zivy, M.; Fillinger, S. Comparative Quantitative Proteomics of Osmotic Signal Transduction Mutants in *Botrytis Cinerea* Explain Mutant Phenotypes and Highlight Interaction with CAMP and Ca²⁺ Signalling Pathways. *J Proteomics*. **2020**, *212*, 103580. DOI: 10.1016/j.jprot.2019.103580.
32. Doughty, T. W.; Domenzain, I.; Millan-Oropeza, A.; Montini, N.; Groot, P. A. de; Pereira, R.; Nielsen, J.; Henry, C.; Daran, J.-M. G.; Siewers, V.; Morrissey, J. P. Stress-Induced Expression Is Enriched for Evolutionarily Young Genes in Diverse Budding Yeasts. *Nat Commun*. **2020**, *11* (1), 1–9. DOI: 10.1038/s41467-020-16073-3.
33. Langella, O.; Valot, B.; Balliau, T.; Blein-Nicolas, M.; Bonhomme, L.; Zivy, M. X!TandemPipeline: A Tool to Manage Sequence Redundancy for Protein Inference and Phosphosite Identification. *J. Proteome Res*. **2017**, *16* (2), 494–503. DOI: 10.1021/acs.jproteome.6b00632.
34. Ferry-Dumazet, H.; Houel, G.; Montalent, P.; Moreau, L.; Langella, O.; Negroni, L.; Vincent, D.; Lalanne, C.; Daruvar, A. de; Plomion, C.; Zivy, M.; Joets, J. PROTICdb: A Web-Based Application to Store, Track, Query, and Compare Plant Proteome Data. *Proteomics*. **2005**, *5* (8), 2069–2081. DOI: 10.1002/pmic.200401111.
35. Langella, O.; Valot, B.; Jacob, D.; Balliau, T.; Flores, R.; Hoogland, C.; Joets, J.; Zivy, M. Management and Dissemination of MS Proteomic Data with PROTICdb: Example of a Quantitative Comparison between Methods of Protein Extraction. *Proteomics*. **2013**, *13* (9), 1457–1466. DOI: 10.1002/pmic.201200564.
36. Langella, O.; Zivy, M.; Joets, J. The PROTICdb Database for 2-DE Proteomics. In *Plant Proteomics: Methods and Protocols*; Thiellement, H., Zivy, M., Damerval, C., Méchin, V., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2007; pp 279–303. DOI: 10.1385/1-59745-227-0:279.
37. Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dienes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; Binz, P.-A.; Xenarios, I.; Eisenacher, M.; Mayer, G.; Gatto, L.; Campos, A.; Chalkley, R. J.; Kraus, H.-J.; Albar, J. P.; Martinez-Bartolomé, S.; Apweiler, R.; Omenn, G. S.; Martens, L.; Jones, A. R.; Hermjakob, H. ProteomeXchange Provides Globally Co-Ordinated Proteomics Data Submission and Dissemination. *Nat Biotechnol*. **2014**, *32* (3), 223–226. DOI: 10.1038/nbt.2839.
38. Zybailov, B.; Mosley, A. L.; Sardi, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P. Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces cerevisiae*. *J. Proteome Res*. **2006**, *5* (9), 2339–2347. DOI: 10.1021/pr060161n.
39. Valot, B.; Langella, O.; Nano, E.; Zivy, M. MassChroQ: A Versatile Tool for Mass Spectrometry Quantification. *Proteomics*. **2011**, *11* (17), 3572–3577. DOI: 10.1002/pmic.201100120.
40. Millan-Oropeza, A.; Henry, C.; Blein-Nicolas, M.; Aubert-Frambourg, A.; Moussa, F.; Bleton, J.; Virolle, M.-J. Quantitative Proteomics Analysis Confirmed Oxidative Metabolism Predominates in *Streptomyces coelicolor* versus Glycolytic Metabolism in *Streptomyces lividans*. *J. Proteome Res*. **2017**, *16* (7), 2597–2613. DOI: 10.1021/acs.jproteome.7b00163.
41. Ning, K.; Fermin, D.; Nesvizhskii, A. I. Comparative Analysis of Different Label-Free Mass Spectrometry Based Protein Abundance Estimates and Their Correlation with RNA-Seq Gene Expression Data. *J. Proteome Res*. **2012**, *11* (4), 2261–2271. DOI: 10.1021/pr201052x.
42. 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* (1), 144–156. DOI: 10.1074/mcp.M500230-MCP200.
43. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
44. 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* (3), 258–264, DOI: 10.1038/nmeth.3255.
45. Cook, R. D. Detection of Influential Observation in Linear Regression. *Technometrics* **2000**, *42* (1), 65–68. DOI: 10.2307/1271434.
46. Muntel, J.; Fromion, V.; Goelzer, A.; Maaß, S.; Mäder, U.; Büttner, K.; Hecker, M.; Becher, D. Comprehensive Absolute Quantification of the Cytosolic Proteome of *Bacillus Subtilis* by Data Independent, Parallel Fragmentation in Liquid Chromatography/Mass Spectrometry (LC/MSE). *Mol Cell Proteomics*. **2014**, *13* (4), 1008–1019. DOI: 10.1074/mcp.M113.032631.
47. Trauchessec, M.; Enjalbert, Q.; Bardet, C.; Homo-Prault, X.; Jacquet, C.; Herment, L.; Fortin, T. A Universal Approach For individual Identification and Quantities Assessment of Host Cell Proteins with LC-MS. **2019**. https://www.anaquant.com/wp-content/uploads/2019/05/POSTER-ANAQUANT_HCP-analysis_cellandgenethrapy.pdf