Comprehensive Absolute Quantification of the Cytosolic Proteome of *Bacillus subtilis* by Data Independent, Parallel Fragmentation in Liquid Chromatography/Mass Spectrometry (LC/MS<sup>E</sup>)*

Jan Muntel†§¶, Vincent Fromion∥, Anne Goelzer∥, Sandra Maaβ‡, Ulrike Mäder‡**, Knut Büttner‡, Michael Hecker‡, and Dörte Becher‡‡‡

In the growing field of systems biology, the knowledge of protein concentrations is highly required to truly understand metabolic and adaptational networks within the cells. Therefore we established a workflow relying on long chromatographic separation and mass spectrometric analysis by data independent, parallel fragmentation of all precursor ions at the same time (LC/MS<sup>E</sup>). By prevention of discrimination of co-eluting low and high abundant peptides a high average sequence coverage of 40% could be achieved, resulting in identification of almost half of the predicted cytosolic proteome of the Gram-positive model organism *Bacillus subtilis* (>1,050 proteins). Absolute quantification was achieved by correlation of average MS signal intensities of the three most intense peptides of a protein to the signal intensity of a spiked standard protein digest. Comparative analysis with heavily labeled peptides (AQUA approach) showed the use of only one standard digest is sufficient for global quantification.

The quantification results covered almost four orders of magnitude, ranging roughly from 10 to 150,000 copies per cell. To prove this method for its biological relevance selected physiological aspects of *B. subtilis* cells grown under conditions requiring either amino acid synthesis or alternatively amino acid degradation were analyzed. This allowed both in particular the validation of the adjustment of protein levels by known regulatory events and in general a perspective of new insights into bacterial physiology. Within new findings the analysis of “protein costs” of cellular processes is extremely important. Such a comprehensive and detailed characterization of cellular protein concentrations based on data independent, parallel fragmentation in liquid chromatography/mass spectrometry (LC/MS<sup>E</sup>) data has been performed for the first time and should pave the way for future comprehensive quantitative characterization of microorganisms as physiological entities. *Molecular & Cellular Proteomics* 13: 10.1074/mcp.M113.032631, 1008–1019, 2014.

In contrast to the rather static genome, composition of the proteome greatly varies with respect to environmental conditions (availability of nutrients, medium composition, stress, etc.) reflecting its key role in the adaptation of cells (1). Hence, proteome data for varying growth conditions should help to reach a comprehensive understanding of the physiology of adaptation to different nutritional conditions, which is the typical situation of bacterial cells in nature (2). In this context the availability of high quality absolute protein quantification data is of outstanding importance for the emerging field of systems biology because (a) proteins are major players for most biological processes and (b) their abundances decisively determine the adaptation rate of cellular processes. Additionally, an emerging set of theoretical and experimental works (reviewed in (3)) recently emphasized the importance of resource allocation in the growth rate management. Bacterial cells have to invest an available set of limited resources into biological processes to ensure growth and survival. Protein costs (or protein burden) of a biological process, defined as the total mass of proteins invested in the biological process, is then critical and must be finely tuned to sustain growth of bacteria. The determination of protein costs of different biological processes using genome-scale absolute protein quantification should thus represent a major breakthrough in understanding bacterial physiology and cellular design.
For many years the gold standard for absolute protein quantification has been quantitative Western blotting and has been successfully applied, for example, to the yeast proteome (4). In recent years mass spectrometry based absolute proteome quantification techniques have become available allowing determination of cellular protein concentrations. The absolute protein amount can be precisely determined by spiking defined amounts of isotopically labeled synthetic peptides into a protein digest (5). Absolute protein amounts become available by detection and comparison of signal intensities of heavy and light peptides, but only for proteins related to the added synthetic peptides. This method was extended to a more global absolute quantification (AQUA)1 by calibrating 2D gels with anchor proteins (6). Although the use of internal labeled standards for absolute protein quantification is very precise, availability and costs for such reference peptides are surely limiting. Therefore label free quantification techniques emerged. One of these methods is based on spectral counting. There, the number of sequenced peptides per protein is used to calculate the absolute quantity of one single protein in a complex sample (emPAI) (7). This can be refined by consideration of physicochemical properties of its peptides (APEX) (8, 9). Absolute protein quantification can also be achieved by comparing average signal intensities of the three most intense peptides per protein to an internally digested standard protein (Hi3 approach). Previous results showed that these average signal intensities per mole protein are constant within a tolerance of 10% (10). Recently the smart combination of AQUA technique and APEX approach was successfully applied to *Leptospira interrogans* covering about half of the proteome with an error less than threefold (11).

In order to quantify the highest possible number of expressed proteins in an absolute manner mass spectrometry based methods seem to be the method of choice. Within the field of proteomics MS is often coupled with liquid chromatography to reduce sample complexity prior to MS analysis (LC/MS). Commonly applied data dependent acquisition (DDA) methods for peptide identification suffer from some limitations. Often low abundant peptides with a low MS signal intensity are discriminated and their isobaric precursor ions cannot be isolated leading to low scores in database search and wrong assignments (12, 13). These obstacles lead to lower protein sequence coverage in general and higher numbers of protein identifications based on a single peptide only. In contrast, with data independent acquisition (DIA) methods like LC/MSE (14) all available precursor ions are fragmented in parallel without any selection by switching between low and high collision energy scans in high frequency. Therefore DIA can circumvent the disadvantages of DDA mentioned above. LC/MSE utilizes chromatographic elution profiles of precursor masses to track the fragment ion spectra. Because all charge states and isotopic peaks of precursor ions are included for fragmentation (15) the LC/MSE technique enables higher sequence coverage and has large advantages in the analysis of highly complex samples consisting of numerous co-eluting peptides (16). Combination of DIA methods with the approach based on Hi3 signal intensities was shown to be of potentially high performance for absolute protein quantification at global scale (17, 18) and is therefore used in this study.

In this article, we applied a global absolute quantification approach based on the Hi3 method and data independent acquisition to the Gram-positive model bacterium *Bacillus subtilis* grown under two conditions for which large differences both in absolute protein amount per cell and in the predicted configurations of metabolic pathways were expected (19): a glucose and ammonia salts minimal medium (condition S) and a solely amino acid based medium (condition CH). This experimental set up enabled the physiologically meaningful comparison of profound consequences of growth under conditions requiring amino acid synthesis (S) or amino acid degradation (CH), as well as concerning the change between glycolytic (S) and gluconeogenetic growth (CH). Moreover, as a model bacterium closely related to very important pathogens, *B. subtilis* is one of the best studied microorganisms. Particularly relevant for our study, the genomic organization of the chromosome, the regulatory network and metabolic pathways are well characterized. Based on this existing knowledge, global absolute protein quantification exemplary enabled (a) the large-scale investigation of protein distribution between cellular processes, (b) the systematic analysis of differential protein abundances for genes belonging to an operon (referred to as operon heterogeneity); and (c) the computation of the protein costs of cellular processes and of metabolic pathways in particular.

**EXPERIMENTAL PROCEDURES**

*Bacterial Growth and Sample Preparation—* *B. subtilis* BSB 168 trp+/ strain (20) was grown in minimal medium with glucose as carbon source (condition S, 2 g/l (NH₄)₂SO₄, 1.4 g/l KH₂PO₄, 0.6 g/l KH₃PO₄, 0.1 g/l Na₂citrate, 0.2 g/l MgSO₄, 0.001 g/l MnSO₄, 5 g/l glucose, 0.001 g/l FeCl₃); and in defined amino acid based medium (condition CH, 10 g/l Casemihydrolysate, 3.9 g/l L-glutamic acid, 1.3 g/l L-alanine, 1.5 g/l L-asparagine, 1.4 g/l KH₂PO₄, 1.4 g/l NH₄Cl, 0.1 g/l Na₂SO₄, 0.1 g/l NH₄NO₃, 0.1 g/l MgSO₄, 0.02 g/l CaCl₂, 0.075 g/l MnSO₄, 0.001 g/l FeCl₃) as described previously (21). Both media contain ammonia salts therefore nitrogen supply is not affected. Cells were harvested during exponential growth (OD₆₀₀ nm 0.5, maximal OD₆₀₀ nm 4 [condition S] and 4.5 [condition CH]) by centrifugation (8000 × g, 10 min, 4 °C) and samples were prepared as described previously (6).

For AQUA analysis heavy peptides (Thermo Fisher Scientific) were added to a final concentration of 5–25 fmol/µl. For LC/MSE analysis samples were cleaned up using C18 - Stage Tips (Proxeon) and complete tryptic digest of alcohol dehydrogenase of yeast

---

1 The abbreviations used are: AQUA, absolute quantification; LC/MS, data independent, parallel fragmentation in liquid chromatography/mass spectrometry; ADH, alcohol dehydrogenase; APEX, absolute protein expression; DIA, data independent acquisition; emPAI, exponentially modified protein abundance index; FWHH, full width at half height; m/c, molecules per cell; SRM, selected reaction monitoring; SCX, strong cation exchange chromatography; TEAB, triethylammoniumbicarbonate.
Absolute Proteome Quantification of *B. subtilis*

(ADH1, YEAST, Waters) was added to a final concentration of 50 fmol/μl.

All experiments were carried out in triplicate.

**Absolute Quantification** –

**Absolute Quantification by AQUA Technique**—Absolute quantification by the AQUA method was carried out for six proteins (CitZ, Icd, Mbi, PyrR, SdhA, and Upp) as described previously (6). The selected proteins cover three orders of magnitude from a couple of hundred protein copies per cell to more than ten thousand. Additionally four proteins were targeted that showed significant differences between the growth conditions (Ald, GapA, RocD, PckA). Briefly, three with heavy isotope labeled peptides per protein were spiked into the protein digests and analyzed by LC-SRM. Peptides were separated via reverse phase nanoHPLC (Ettan MDLC, GE Healthcare) by a 90 min gradient. SRM measurements were performed by 4000 QTRAP (ABSciex). Data analysis was performed in MultiQuant 1.2 (ABSciex).

Detailed MS-parameters can be found in supplemental material and Material of reference (6).

LC-SRM for the four additional proteins was carried out with a reverse phase nanoHPLC (easy-nLCII, Thermo Fisher) using a 80 min gradient. SRM measurements were performed by a TSQ Vantage (Thermo Fisher). Data analysis was performed with Pinpoint 1.0. Detailed MS-parameters are provided as supplemental Table S9.

All SRM samples were analyzed in triplicate.

**Global Absolute Quantification**—Digested protein mixture (5 μg) was separated using the nanoACQUITY™ UPLC™ system (Waters) by direct injection without a trapping column. Sample was loaded in 35 min on column (nanoACQUITY™ UPLC™ column, BEH130 C18, 1.7 μm, 75 μm x 200 mm, Waters) with 99% (v/v) buffer A, 1% (v/v) buffer B at a flow rate of 300 nl/min. Peptides were separated in 265 min applying the following gradient: in 165 min to 18% (v/v) buffer B, in 60 min to 26% (v/v) buffer B, in 40 min to 60% (v/v) buffer B, in 1 min on 99% (v/v) buffer B for 10 min and equilibration for 15 min with 99% (v/v) buffer A. The nanoHPLC was coupled online to Synapt G2 mass spectrometer (Waters) equipped with NanoLockSpray source and operated with the MassLynx V4.1 software (Waters). For all measurements a mass range of 50–2000 Th was used and analyzer was set in resolution mode. For lock mass correction [Glu1]fibrinopeptide B solution (GluFib, m/z: 785.8426 Th, Sigma, 500 fmol/μl) at 50% (v/v) acetonitrile, 0.1% (v/v) formic acid) was infused through the reference fluidics system of the Synapt G2 at a constant flow rate of 500 nl/min and sampled every 30 s. The mass spectrometer was run in LC/MSF mode in which collision energy was adjusted between 5 eV in the precursor ion trace and a ramp 10–40 eV for fragment ion trace. All scan times were set to 2 s.

Measurements were carried out in triplicate.

**Data Mining of LC/MSF Data**—For data processing and protein identification, raw data were imported into ProteinLynx Global Server 2.4 (PLGS; Waters, Milford, MA) and processed via Apex3D algorithm with following parameters: chromatographic peak width: automatic, MS ToF resolution: automatic, lock mass for charge 2: 785.8426 Da/e, lock mass window: 0.25 Da, low energy threshold: 250 cts, elevated energy threshold: 30 cts, retention time window: automatic, intensity threshold of precursor/fragment ion cluster: 1000 cts.

Database search was carried out by the ion accounting algorithm (22) against randomized *B. subtilis* 168 database (23) with added common laboratory contaminants and yeast ADH1 sequence (8,438 entries) using following parameters: peptide tolerance: automatic, protein tolerance: automatic, min fragment ions matches per peptide: 1, min fragment ions matches per protein: 5, min peptide matches per protein: 2, primary digest reagent: trypsin, missed cleavages: 2, fixed modification: carbamidomethylation C (+57.0215), variable modifications: deamidation N, Q (+0.9840), oxidation M (+15.9949), pyrrolidonecarboxyllic N-TERM (+27.9949), false discovery rate (FDR): 5%, calibration protein: yeast ADH1. All identification and quantitation data are summarized in supplemental Table S8 and are available at the PRIDE database.

For additional filtering of database search results, all proteins identified in only one of the nine replicates of one sample (three technical replicates of each of the three biological replicates) were discarded leading to a FDR between 3 and 4% on protein level. These FDR thresholds were applied for the global analysis of the data. For comparison of the dataset to other published data set, all proteins had to be identified in three out of the nine replicates, which led to a FDR below 1%. Replicate measurements were utilized for calculation for confidence intervals and statistical significance analysis.

**Calculation of Proteins per Cell**—After estimation of total amount of proteins (in fmol) in the digests (determined by AQUA or global Hi3 approach) the number of molecules per cell was calculated. Therefore number of cells per OD was counted in Thoma counting chamber using a light microscope (Axioskop re, Carl Zeiss). In S medium 6.3 × 10¹⁰ cells per ml per OD were counted and in CH medium 4.1 × 10⁸ cells per ml per OD. These values allowed determining the average protein amount per cell and the number of cells, which were used for one measurement.

For comparison of the absolute quantification results between the AQUA and Hi3 method, the results were standardized (molecule per cell level) and a Wilcoxon-Rank-Sum test was applied.

**RESULTS AND DISCUSSION**

**Protein Identification**—For global absolute protein quantification, *B. subtilis* cells were grown either in glucose and ammonia salts based minimal medium (condition S) or in amino acid based defined medium (condition CH). Tryptically digested extracts of these cells were analyzed by a data-independent-acquisition workflow (LC/MSF) in three technical replicates for each of the three biological replicates. After database search using the ion-accounting algorithm (24) an additional replicate filter was applied to exclude false positive identifications randomly occurring across replicate measurements. Each protein had to be identified for acceptance at least in three out of nine replicate measurements. The replicate filter criterion rendered a FDR of 0.8% for condition S and 0.4% for condition CH. A very small number of proteins (less than 1%) was identified based on one peptide only. Therefore these proteins were not considered in the downstream analysis, resulting in 1079 identified proteins with an average of 40% sequence coverage. Most of them were identified in both conditions analyzed (67%) whereas 23% were exclusively found in condition S and 10% in condition CH. The localization of the proteins within the *B. subtilis* cell was computed by LocateP (25). This analysis revealed that the majority (88%) of the proteins is located in the cytoplasm covering more than 40% of the predicted cytosolic proteome. Only a small number of identified proteins were predicted as membrane (8%) or extracellular proteins (4%). This result is in agreement with the applied workflow because no enrichment for other cellular compartments was performed (all identification and quantification results are summarized in supplemental Table S8).

Thus the LC/MSF technique enabled an identification result quantitatively comparable to other global proteomic studies in...
B. subtilis (26) but qualitatively superior because it was achieved without any pre-fractionation steps that are always critical for quantitative analyses and with a considerably higher average sequence coverage (16). These findings demonstrate the high potential of DIA methods for global proteomic studies.

Global Absolute Protein Quantification and Verification by an Orthogonal MS Approach—By comparison of the average signal intensities of the three most intense tryptic peptides to the internal standard protein (ADH) the absolute amount of identified proteins could be determined (Hi3 approach (10)).

![Graph showing dynamic range of protein copy numbers per cell](image)

**Fig. 1.** Overview of the dynamic range of protein copy numbers per cell in B. subtilis cells under condition S.

The results cover almost four orders of magnitude starting from 18 molecules per cell [m/c] (nitric oxide synthase, Nos) to more than 150,000 (flagellin protein Hag) as illustrated in Fig. 1 for condition S. Technical replicates were utilized to calculate the coefficient of variance (CV). The average CV of technical replicates was below 30% and even below 20% for more than half of the quantified proteins (summarized results of absolute and relative quantification in supplemental Table S1).

We first investigated the precision of the Hi3 method on LC/MS²-data by comparing our results with previously published copy numbers per cell determined by Western blot quantification. Although the growth conditions in the chosen publications were not the same like in the presented work, the quantifications are still within a twofold range (Table I).

For further verification of absolute quantification results, the same samples were independently analyzed via an additional MS-based highly precise absolute quantification method (AQUA approach (5)). Ten proteins, namely for CitZ, Icd, Mbl, SdhA, PyrR, Upp, Ald, GapA, PckA, and RocD were analyzed by this orthogonal approach (Table II). Therefore three heavily labeled peptides per protein were spiked into the samples in a defined amount enabling comparison of heavy (spiked) and light (endogenous) peptide intensities leading to the quantification of the peptide and subsequent protein concentration within the sample and within the cell. As expected, the AQUA approach allows a more reliable absolute quantification with

| Protein | Condition S | Condition CH |
|----------------|-------------|--------------|
| GapA | 19,200 | 8780 |
| CcpA | 460 | 1370 |
| CheA | 710 | 1580 |
| CheW | 840 | 1420 |
| YfmS | 1590 | 4050 |

**Table I**

Comparison of MS-based absolute quantification results with previously published Western blot-based quantification for selected proteins. Publications are mentioned in the table

| Protein | Condition S | Condition CH |
|----------------|-------------|--------------|
| GapA | 19,200 | 8780 |
| CcpA | 460 | 1370 |
| CheA | 710 | 1580 |
| CheW | 840 | 1420 |
| YfmS | 1590 | 4050 |

**Table II**

Overview of quantification results for ten proteins of B. subtilis, quantified by AQUA and Hi3 method. The cells were grown either under condition S or under condition CH. Data from the condition CH were not volume-corrected for method comparison

| Protein | Condition S | Condition CH |
|----------------|-------------|--------------|
| GapA | 19,200 | 8780 |
| CcpA | 460 | 1370 |
| CheA | 710 | 1580 |
| CheW | 840 | 1420 |
| YfmS | 1590 | 4050 |

**Table I**

Comparison of MS-based absolute quantification results with previously published Western blot-based quantification for selected proteins. Publications are mentioned in the table

| Protein | Condition S | Condition CH |
|----------------|-------------|--------------|
| GapA | 19,200 | 8780 |
| CcpA | 460 | 1370 |
| CheA | 710 | 1580 |
| CheW | 840 | 1420 |
| YfmS | 1590 | 4050 |

**Table II**

Overview of quantification results for ten proteins of B. subtilis, quantified by AQUA and Hi3 method. The cells were grown either under condition S or under condition CH. Data from the condition CH were not volume-corrected for method comparison

| Protein | Condition S | Condition CH |
|----------------|-------------|--------------|
| GapA | 19,200 | 8780 |
| CcpA | 460 | 1370 |
| CheA | 710 | 1580 |
| CheW | 840 | 1420 |
| YfmS | 1590 | 4050 |
an average CV of technical replicates below 10% in our experiments. Despite its high precision, the AQUA approach cannot be easily applied to large scale protein quantification because it requires the addition of expensive proteotypic peptides for each protein of interest. Therefore, this approach can be seen as gold standard for absolute protein quantification in small scale experiments and as a good orthogonal method to verify the results from the Hi3 method.

Quantification results of AQUA and Hi3 approach were in good agreement (Table II). The only big discrepancy in the quantification results was the quantification of RocD in condition CH (AQUA: 1184 m/c and Hi3 113,714 m/c), which might be explained by a wrong peptide assignment during the database search of the LC/MS<sup>E</sup> data. However, the two datasets didn’t show a significant difference in a Wilcoxon-Rank-Sum test (p = 0.44 for condition S and p = 0.46 for condition CH) and an average error of about 1.4-fold. Given this degree of consistence the Hi3 approach presented here can accomplish a similar accuracy as other current methods for estimating absolute protein abundance on a proteome-wide scale (6, 11). Genome-scale absolute abundance in general results in a higher variance in the protein abundance compared with the single peptide/protein quantification by AQUA. A way to improve the precision of the Hi3 approach is to increase the number of technical/biological replicates (in our case nine replicates per condition were analyzed).

**Statistical Analysis: Reproducibility and Quality of LC/MS<sup>E</sup>-Hi3 Method**—The dataset, composed of nine replicates, has been deeply analyzed in both conditions to evaluate (a) the reproducibility of LC/MS<sup>E</sup> method, (b) the bias in protein quantification that should arise from protein abundance, localization or the nature of protein and (c) the relative error in protein quantification.

For both conditions, the correlation between biological and technical replicates was computed. The correlation between biological replicates ranges from 0.77 to 0.97 (condition S) and from 0.81 to 0.96 (condition CH) (supplemental Table S2, supplemental Table S3). The correlation between technical replicates in both conditions is always higher than 0.88 and is therefore especially high.

For each condition, we investigated if the number of replicates, the size of proteins, or the localization of proteins, lead to systematic bias in protein quantification. No real coupling was found except for the number of replicates. The median value of protein abundance slightly increases with the number of replicates (actually low abundant proteins were less detected than others) (supplemental Figs. S1 and S2).

We finally investigated the dependence between variance and mean of each protein. As expected, the variance of protein increases with the mean value. Therefore, we applied a variance stabilizing transformation (Log2 transformation) on the original dataset and then computed the empirical distribution of the residual error. Both distributions for condition S and CH are close to normal distribution with longer left and right tails than standard normal distribution (supplemental Fig. S3 for condition S and CH). In fine, 94.8% of all residual errors in condition S (and 95.02% in condition CH) belong to a ± 1.96 σ band around protein average (supplemental Figs. S4 and S5) where σ is the standard deviation of the error residues. We computed the 95% confidence intervals of log data by bootstrapping the set of error residues. The converse transformation was finally applied on the mean and on each bound of the confidence intervals to deduce the final interval values. For each protein, the 95% confidence interval on protein abundance can be easily expressed as the product between geometric mean of the replicates and two relative coefficients, which are comparable for both conditions (supplemental Table S4). For nine replicates, 95% confidence interval for protein abundance is 30% around the mean value ([0.72 m, 1.39 m]).

Remarkably, the two multiplicative coefficients, corresponding to the relative errors on each protein, are independent of the mean values. Since the relative error is systematic, the LC/MS<sup>E</sup>-Hi3 method is thus highly suitable to detect and compare protein abundance even the ones having a large range of variation.

**Limitation of the Method**—For *B. subtilis* 259 genes are described as being essential (27, 28) of which 192 could be quantified absolutely (60% of the cytosolic essential proteins). The average concentration of these proteins (2902 m/c in condition S) is higher compared with the average amount of all proteins (1557 m/c). As shown in supplemental Fig. S6 protein abundances of essential proteins are consistently shifting toward higher copy numbers. This is in accordance to studies in *Escherichia coli* (29). Therefore the essential proteins of *B. subtilis* should be easily accessible for analysis and well-suited to elucidate the limitations of the workflow.

Missed essential proteins could be divided into two subgroups. The first group contains low-molecular-weight proteins (<8 kDa). The second group consists of membrane proteins as proteins involved in secretion or Na<sup>+</sup>/H<sup>+</sup> antiporters, which was indicated by the localization analysis of the identified proteins. Additionally, difficulties to quantify membrane proteins can be illustrated by a detailed view on the ATP synthase complex, which is partly located in the membrane and whose stoichiometry is well known. The intramembrane proteins of the complex could not be identified in this study and membrane associated proteins were quantified in lower concentrations than expected (supplemental Fig. S7).

Possibly, cytoplasmically located precursors of these proteins were quantified or the proteins were incompletely released from their original localization during sample preparation. Probably a subpopulation could only be analyzed. This assumption is strengthened by previously monitored spreading of proteins over several sub-proteomes (30). Therefore the absolute quantification workflow has to be modified for membrane proteins as well as for small proteins. This might be difficult because of lower probability to generate peptides by
tryptic digestion, which fit in the analytical window of the mass spectrometer (compared with larger and more hydrophilic proteins).

Therefore all quantified proteins with known localization in the membrane and or with a low molecular weight were marked in the results to indicate uncertainties.

Physiological Impacts in Both Conditions—The quantification workflow presented here allows a global view on cellular physiology of B. subtilis for two physiological conditions with different growth rate values (μ). Cells were grown either on glucose and ammonia salts based minimal medium (condition S with μ = 0.47 h⁻¹) or on amino acid based defined medium (condition CH with μ = 1.01 h⁻¹). From a physiological viewpoint, condition S corresponds to a glycolytic condition for which all amino acids have to be synthesized de novo whereas condition CH is a gluconeogenic condition under which amino acids have to be degraded and used as carbon sources. Moreover, cell volume in condition CH is 1.94 times higher than in condition S (21). Thus, to have comparable protein abundances over, cell volume in condition CH is 1.94 times higher than in condition S (21). Therefore all quantified proteins with known localization in the membrane and or with a low molecular weight were marked in the results to indicate uncertainties.

Physiological Impacts in Both Conditions—The quantification workflow presented here allows a global view on cellular physiology of B. subtilis for two physiological conditions with different growth rate values (μ). Cells were grown either on glucose and ammonia salts based minimal medium (condition S with μ = 0.47 h⁻¹) or on amino acid based defined medium (condition CH with μ = 1.01 h⁻¹). From a physiological viewpoint, condition S corresponds to a glycolytic condition for which all amino acids have to be synthesized de novo whereas condition CH is a gluconeogenic condition under which amino acids have to be degraded and used as carbon sources. Moreover, cell volume in condition CH is 1.94 times higher than in condition S (21). Thus, to have comparable protein abundances in both conditions, we corrected protein abundances (m/c) in condition CH by factor 1.94 (Table III, supplemental Table S1).

In condition S, abundance of 35 proteins is greater than 10,000 m/c including the most abundant flagellin protein Hag (~150,000 m/c), which is almost three times more abundant than the second abundant protein elongation factor EF-Tu (~57,000 m/c). Copy numbers of all other highly abundant proteins were determined from 10,000 to 50,000 m/c. These proteins are involved in central carbon metabolism (Icd, Mdh, GapA, FbaA, Fbg, PdhB, PdhC, PdhB, GndA, Eno), amino acid synthesis (IlvC, AroA, GlyA, MtnA, AspB, MetE), translation apparatus and chaperones (RplL, RpsG, RpsE, EF-Ts, GroEL, EF-G, RplK, RpmC, RpsP), and other processes (AhpC, Hbs, YvzB, DnhB, YepD, AcpA, ThiC, AtpD). Especially the result for YvzB is very interesting, as this protein is only very highly expressed on mRNA level (31). This indicates that further analysis regarding protein synthesis, stability, and degradation has to be carried out to get deeper insights into regulatory processes and to understand these processes in more detail.

By contrast, in condition CH, abundance of 51 proteins is greater than 10,000 m/c. Proteins of highest abundances include Hag (~145,000 m/c), EF-Tu (~123,000 m/c), RplL (~74,000 m/c), and RocD (~59,000 m/c) involved in arginine degradation. Most of the proteins already mentioned above are still highly abundant (~10,000 m/c). As expected, abundances of proteins involved in amino acid synthesis (except for GlyA) and glycolysis (except for FbaA) decrease over a wide range from at least twofold to more than one order of magnitude and are now below 10,000 m/c (Table III). Conversely, abundances of proteins involved in amino acid degradation (RocD, Ald, RocA, PutC) and gluconeogenesis (PckA) are high (~10,000 m/c).

Computation of protein abundance ratios (CH versus S) revealed that 127 proteins are differentially expressed with a large range of protein ratios from 0.07 (IlvC) to 400 (PckA) (supplemental Table S1, p value <0.01). However, more proteins should be differentially expressed because ratios cannot be computed in some cases. Some proteins, such as GapB, GltA or GltB, are not detected in one condition whereas they are still highly abundant (~10,000 m/c). As expected, abundances of proteins involved in amino acid synthesis (except for GlyA) and glycolysis (except for FbaA) decrease over a wide range from at least twofold to more than one order of magnitude and are now below 10,000 m/c (Table III). Conversely, abundances of proteins involved in amino acid degradation (RocD, Ald, RocA, PutC) and gluconeogenesis (PckA) are high (~10,000 m/c).

Computation of protein abundance ratios (CH versus S) revealed that 127 proteins are differentially expressed with a large range of protein ratios from 0.07 (IlvC) to 400 (PckA) (supplemental Table S1, p value <0.01). However, more proteins should be differentially expressed because ratios cannot be computed in some cases. Some proteins, such as GapB, GltA or GltB, are not detected in one condition whereas they are still highly abundant (~10,000 m/c). As expected, abundances of proteins involved in amino acid synthesis (except for GlyA) and glycolysis (except for FbaA) decrease over a wide range from at least twofold to more than one order of magnitude and are now below 10,000 m/c (Table III). Conversely, abundances of proteins involved in amino acid degradation (RocD, Ald, RocA, PutC) and gluconeogenesis (PckA) are high (~10,000 m/c).

Computation of protein abundance ratios (CH versus S) revealed that 127 proteins are differentially expressed with a large range of protein ratios from 0.07 (IlvC) to 400 (PckA) (supplemental Table S1, p value <0.01). However, more proteins should be differentially expressed because ratios cannot be computed in some cases. Some proteins, such as GapB, GltA or GltB, are not detected in one condition whereas they are still highly abundant (~10,000 m/c). As expected, abundances of proteins involved in amino acid synthesis (except for GlyA) and glycolysis (except for FbaA) decrease over a wide range from at least twofold to more than one order of magnitude and are now below 10,000 m/c (Table III). Conversely, abundances of proteins involved in amino acid degradation (RocD, Ald, RocA, PutC) and gluconeogenesis (PckA) are high (~10,000 m/c).

Computation of protein abundance ratios (CH versus S) revealed that 127 proteins are differentially expressed with a large range of protein ratios from 0.07 (IlvC) to 400 (PckA) (supplemental Table S1, p value <0.01). However, more proteins should be differentially expressed because ratios cannot be computed in some cases. Some proteins, such as GapB, GltA or GltB, are not detected in one condition whereas they are still highly abundant (~10,000 m/c). As expected, abundances of proteins involved in amino acid synthesis (except for GlyA) and glycolysis (except for FbaA) decrease over a wide range from at least twofold to more than one order of magnitude and are now below 10,000 m/c (Table III). Conversely, abundances of proteins involved in amino acid degradation (RocD, Ald, RocA, PutC) and gluconeogenesis (PckA) are high (~10,000 m/c).

Computation of protein abundance ratios (CH versus S) revealed that 127 proteins are differentially expressed with a large range of protein ratios from 0.07 (IlvC) to 400 (PckA) (supplemental Table S1, p value <0.01). However, more proteins should be differentially expressed because ratios cannot be computed in some cases. Some proteins, such as GapB, GltA or GltB, are not detected in one condition whereas they are still highly abundant (~10,000 m/c). As expected, abundances of proteins involved in amino acid synthesis (except for GlyA) and glycolysis (except for FbaA) decrease over a wide range from at least twofold to more than one order of magnitude and are now below 10,000 m/c (Table III). Conversely, abundances of proteins involved in amino acid degradation (RocD, Ald, RocA, PutC) and gluconeogenesis (PckA) are high (~10,000 m/c).

Computation of protein abundance ratios (CH versus S) revealed that 127 proteins are differentially expressed with a large range of protein ratios from 0.07 (IlvC) to 400 (PckA) (supplemental Table S1, p value <0.01). However, more proteins should be differentially expressed because ratios cannot be computed in some cases. Some proteins, such as GapB, GltA or GltB, are not detected in one condition whereas they are still highly abundant (~10,000 m/c). As expected, abundances of proteins involved in amino acid synthesis (except for GlyA) and glycolysis (except for FbaA) decrease over a wide range from at least twofold to more than one order of magnitude and are now below 10,000 m/c (Table III). Conversely, abundances of proteins involved in amino acid degradation (RocD, Ald, RocA, PutC) and gluconeogenesis (PckA) are high (~10,000 m/c).

Computation of protein abundance ratios (CH versus S) revealed that 127 proteins are differentially expressed with a large range of protein ratios from 0.07 (IlvC) to 400 (PckA) (supplemental Table S1, p value <0.01). However, more proteins should be differentially expressed because ratios cannot be computed in some cases. Some proteins, such as GapB, GltA or GltB, are not detected in one condition whereas they are still highly abundant (~10,000 m/c). As expected, abundances of proteins involved in amino acid synthesis (except for GlyA) and glycolysis (except for FbaA) decrease over a wide range from at least twofold to more than one order of magnitude and are now below 10,000 m/c (Table III). Conversely, abundances of proteins involved in amino acid degradation (RocD, Ald, RocA, PutC) and gluconeogenesis (PckA) are high (~10,000 m/c).
transhydrogenation cycle (32) or a post-translational regulation such as phosphorylation (33) turning off GapA activity.

Consistency of the Data Set With the Known Regulatory Network—Data analysis was mainly performed on the central metabolism because most of the regulations active in different growth conditions modulate expression of proteins involved in these metabolic pathways. Despite the limitations of the protein quantification method mentioned above, an almost complete coverage of metabolic pathways could be obtained (supplemental Fig. S8, definition of pathways in supplemental Table S5). We then performed a detailed comparison between dataset and predictions of altered protein abundances based on the model and the method provided by Goelzer and coworkers (19) (supplemental Table S1, S6). We used functional categories of the enzymes that are very similar to categories of SubtiWiki (34) but not identical in order to facilitate the description of the cellular processes for our analytical tools.

Most of the experimental results were in agreement with the known regulatory network. Gluconeogenic proteins PckA and GapB are strongly accumulated (up to 400 times increase for PckA) (35) whereas members of the gapA-operon (GapA, Tpi, Pgk, Pgm, Eno) appear in lower amounts in condition CH compared with condition S (36, 37) (Table III). Most of the amino acids synthesis pathways for which transcriptional regulations have been experimentally characterized (branched-chain amino acids, methionine, cysteine, glutamate and proline) are repressed in condition CH as expected (19). The opposite occurs for degradation pathways of amino acids (glutamate, arginine, branched-chain amino acids, asparagine and aspartate) (19) (supplemental Table S1). Here we have to point out the particular status of arginine in condition CH. Proteins involved in arginine degradation are strongly accumulated in condition CH (Table III, supplemental Table S1). This is in agreement with the known regulatory mechanisms. Indeed, in the presence of arginine, the transcription factor AhrC is known (a) to repress the expression of genes coding for arginine synthesis and (b) to induce the degradation of arginine. In addition, genes for arginine degradation are transcribed from a σ70-dependent promoter and activated by RocR (38). Considering only these regulations, arginine synthesis should be repressed. In fact, on transcriptional level arginine synthesis is strongly (~50-fold) repressed, whereas arginine degradation is induced (~20-fold), (personal communication U. Mäder). But in contrast, the absolute concentration of arginine synthesis enzymes seems to be unaffected. This might be caused by other regulatory mechanisms like translational control via small RNA SR1 (39), post-translational protein control, and differences in protein stability and therefore will need further investigation. Thus our method can be systematically used to detect unknown post-transcriptional regulations.

Discrepancies between the data and the known regulatory mechanisms occurred in the case of proline degradation (supplemental Fig. S9). In condition CH the proline degradation pathway was predicted to be repressed in presence of branched chain amino acids through repression by the global regulator CodY (40, 41), whereas the corresponding enzymes of the degradation pathway, PutB, PutC, were strongly accumulated to ~790 m/c and ~20,000 m/c (200 times increase), respectively. As the transcriptional activator PutR is known to induce the expression of the putBCP-operon in the presence of proline (thus in condition CH) (41), observed accumulation of PutB and PutC could result from the competition between CodY and PutR for the promoter (41).

Metabolic pathways that were reported to be not regulated or for which no regulatory mechanism are known yet would be predicted to show comparable amounts of the corresponding enzymes between conditions S and CH. However, some of them showed differential protein expression between conditions S and CH, and might thus be regulated by an unknown mechanism (supplemental Figs. S8 and S9). Based on the medium composition, sensing signals of these unknown mechanisms can already be postulated. In particular, (a) histidine synthesis could be repressed in presence of histidine (Table III); (b) chorismate synthesis could be repressed in presence of tyrosine (42).

Operon Heterogeneity—Absolute quantification provides insights into the heterogenous levels of protein amounts for proteins whose encoding genes belong to the same operon. Operon heterogeneity was analyzed using a set of well-defined operons associated to the core metabolism and to cofactor synthesis pathways (definition of operons: supplemental Table S7). Only operons with at least three quantified proteins were considered. In general two different types of operons are to be distinguished. For some operons such as the his-operon (op-50), the corresponding protein amounts are quite similar for all genes, whereas highly differentiated protein amounts within an operon are observed for others, such as the ilv-leu-operon (op-1, Fig. 2), which is in agreement with previous works (43). In case of the ilv-leu-operon, the differential protein expression is mainly because of post-transcriptional modifications of the transcript (43). Thus, absolute protein quantification can raise fundamental questions on mechanisms enabling in fine to obtain such heterogeneity in protein expression. More broadly, it also paves the way to the systematic characterization of post-transcriptional regulations at a genome scale opening deeper insights into cellular physiology.

Computation of Protein Cost of Biological Processes—Global View on Metabolism—Resource allocation (especially proteins) appears to be central in growth rate management of bacteria (44–46). Key aspects of resource allocation are first to save resources (especially proteins) by mechanisms of regulation such as the repression of certain metabolic pathways, and secondly to invest the saved resource in biological processes required to increase the growth rate, for example, appropriate metabolic pathways, translational apparatus etc. Protein cost of a biological process defined as
**Fig. 2.** Example of variations of protein amounts through operons (operon heterogeneity) of *B. subtilis* grown under (A) condition S and (B) condition CH. The number of molecules per cell of each operon protein was divided by the minimum value of the operon in order to have ratios always greater than 1. Only operons are depicted with at least three quantified proteins. The protein names with a white background are not detected by the Hi3 method (red protein name corresponds to protein with a mass less than 8 kDa), '*' correspond to the protein with a localization related to the membrane (but without predicted multi-pass domain), and '**' correspond to the proteins predicted to have a multi-pass domain.
the total mass of proteins invested into the biological process is then critical and must be finely tuned to sustain growth of bacteria. The cost of biological processes (and particularly the ones of metabolic pathways) can be computed at a genome scale as the sum of the cost of individual proteins involved in the biological process. The cost of an individual protein corresponds to the amount of the protein (determined by absolute protein quantification in copy numbers) multiplied by its individual mass in Daltons.

Fig. 3 provides a global view on the mass distribution with respect to the main biological functions. About 20% of the mass of all cytosolic proteins are dedicated to central carbon metabolism (glycolysis, TCA cycle, pentose phosphate pathway, and overflow metabolism), which also represents the most costly pathway of metabolism (Fig. 3). For higher growth rates (condition S: $\mu = 0.47$ h$^{-1}$, condition CH: $\mu = 1.01$ h$^{-1}$), the portion of translational apparatus is increasing from 22 to 27% of the total protein mass in agreement with theoretical expectations and experimental results (44, 46) (Table III). As expected, the investment of *B. subtilis* in amino acid synthesis is decreasing from 13% in condition S to 5% in condition CH because of presence of amino acids in the growth medium while the proportion of amino acid degradation is increasing from 1 to 10%. However, the portion of membranous proteins belonging to the respiratory chain and ATPase seems to be underestimated because of the limitation of the workflow. Thus, the weight of aerobic respiration and ATPase on the total mass distribution should be underestimated, which could change the relative investment of processes on Fig. 3. In general, one can also notice the high portion of flagella and motility proteins. In particular, flagellin protein Hag, which composes flagella, is the most abundant single protein in both conditions. Since flagella and motility proteins typically correspond to “unneeded protein”, for example, not involved in metabolic networks or the translation apparatus, a strain lacking these proteins should have higher growth rates in agreement with previous experimental works (44–46). This assumption has been confirmed using a $\Delta$sigD mutant strain, which is not able to express the flagella and motility proteins. Compared with the wild type strain, the mutant strain showed higher growth rates in glucose minimal medium (47) and personal communication A. Goelzer).

**Particular Distribution to Processes and Pathways**—To go further, we investigated the distribution of protein mass among particular biological processes and metabolic pathways (supplemental Fig. S9). Each protein has been assigned to a single biological function (in contrast to supplemental Fig. S8; see supplemental Table S6 for description). This view allows better assessing of the “slight” effect of some gene expression regulations, which is hard to understand on basis of transcriptional and translational control alone. TCA cycle appears to be one of the most expensive pathways of the cell as in *E. coli* (48). Therefore modest changes therein are causing large consequences and fine tuning of the amount of these
enzymes is essential for the cell. A difference between conditions S and CH of factor 1.5 was observed for the TCA cycle, which usually corresponds to “weak” repression (supplemental Fig. S9, Table III). However, this weak repression also leads to save 4% of total protein mass, which can be efficiently used then to increase the growth rate. Like for the TCA cycle, “slight” modulations of abundant enzyme synthesis should have great impacts on global protein costs and growth rate management. Conversely, some biosynthesis pathways of amino acids such as for alanine or aspartate that represent less than 1% of total protein mass, are most likely not regulated at the level of gene expression in agreement with previous results (see (19) and references therein). Their constitutive expression could be explained by their relatively low cost compared with other processes. Since designing transcriptional regulation is also a costly process, for example, a transcription factor is also a protein, the benefit obtained by the repression of these pathways may not counterbalance the protein cost associated to their regulation. Constitutive expression could also reflect the adaptation of B. subtilis to its ecological niche. Serious prognosis about ecology and or evolution becomes possible.

Finally, absolute protein quantification enables the computation of the protein cost for all metabolic pathways. The diversity in protein abundance between metabolic pathways (supplemental Fig. S9) and inside a specific metabolic pathway (supplemental Fig. S8) can now be investigated. Costly or conversely low costly metabolic pathways of B. subtilis are revealed. The relative cost of methionine metabolism and particularly the one of methionine/salvage pathway/polyamine (supplemental Fig. S9) is higher than the one of most other amino acids synthesis pathways. This pathway corresponds to the recycling of methionine and the synthesis of polyamines (see supplemental Fig. S8 for the pathway composition). Since the amounts of polyamines in biomass requirements are usually weak, the apparent cost for this pathway is surprisingly high in B. subtilis. In E. coli, the cost of methionine metabolism is also very high compared with other metabolic pathways (48). Altogether, the protein investment in metabolic pathways seems to be comparable in these two organisms.

Unexpectedly, the protein cost of metabolic pathways does not seem to depend on the number of enzymes that are involved. For example, the cost of arginine synthesis pathway (composed of 10 enzymes) is similar to that one for threonine (composed of 3 enzymes) pointing to similar requirements of both pathways for biomass formation (49) (supplemental Table S5). The cost of arginine synthesis is thus quite low compared with its length.

Inside metabolic pathways, absolute protein quantification also revealed high variability in protein abundance (supplemental Fig. S8). For example, GndA is four times more abundant than Zwf in the oxidative part of pentose phosphate pathway under both conditions S and CH. Therefore, GndA could have either an additional function or more probably a lower efficiency at catalyzing a complex chemical reaction than the homologous enzymes of E. coli (48) and Saccharomyces cerevisiae (8). More broadly, the systematic comparison of absolute protein quantifications between species should help us to target these apparently inefficient enzymes and the specificities of their associated complex chemical reactions.

Regulation of gene expression ensures that each single protein is allocated at the appropriate time and biological process in the appropriate concentration. Thus, global absolute protein quantification enables a better understanding of bacterial physiology by visualizing regulation of gene expression by analysis of protein repartition and resource allocation in different growth conditions and operon heterogeneity on protein level. Compared with transcriptome data (even high-resolution data), absolute protein quantification allows to observe the impact of any regulation (transcriptional, translational, and post-translational) occurring at the level of protein production. Thus, absolute protein quantification will pave the way to deeper investigations in the management of cellular functions in bacteria.

CONCLUSIONS

The presented workflow enabled the identification of more than 1,050 proteins of B. subtilis with an unusually high quality implied by average sequence coverage of 40% and no protein identification based on a single peptide. Although not all sub-proteomes are suitable for analysis almost half of the predicted cytosolic proteins could be quantified absolutely. Considering that not all proteins are expressed an even higher coverage was achieved. In comparison to other absolute quantification approaches for B. subtilis (6) the dynamic range could be expanded from three to four orders of magnitude and the number of quantified proteins was more than twofold higher (1050 to 467). Spectral counting methods for the Gram-negative model organism E. coli revealed 1,103 absolutely quantified proteins (7). These results are within the range of the presented work, but relied on a technically problematic pre-fractionation of the peptide mixture by SCX. Comparative analysis with the AQUA technique, which is superior in terms of reproducibility and was therefore applied as gold standard for absolute quantification clearly showed the accordance of absolute protein quantification using the MS²–Hi3 methods based only one internal standard digest. With an average CV between technical replicates below 35%, the MS²–Hi3 approach has thus a sufficient reproducibility for global-scale absolute protein quantification (8). Although the CV of the MS²–Hi3 approach was threefold higher compared with the AQUA method, the workflow still provides a sufficient reproducibility for absolute quantification on a global scale and is applicable as a general approach for systems biology purposes.

Comprehensive absolute protein quantification should contribute to systems biology at various levels such as (a) the
refinement of dynamic models integrating protein concentration, (b) the detection of new mechanisms of regulation (transcriptional, translational or post-translational), (c) the real cost of proteins in different biological processes, and in particular of the costly metabolic pathways, (d) the distribution of cellular resources. However these comprehensive and accurate quantification data represent only the basis for quantitative assessment of functional biological relations by including further enzyme parameters like biochemical reactivity, post-translational status or spatial amenability. In summary the presented workflow delivers a reliable basis for systems biology approaches and might be the method of choice for similar global investigations in future.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomcentral.proteomeexchange.org) via the PRIDE partner repository (50) with the data set identifier PXD000154. Reviewers are suggested to use the reviewer account with PRIDE Inspector (http://code.google.com/p/pride-toolsuite/wiki/PRIDEInspector#Getting_PRIDE_Inspector) in order to get access to all of the files belonging to our submission. To get access PRIDE Inspector needs to be opened. The download can be started using option Private Download/ProteomeXchange/PX reviewer account details. PX reviewer account details are as follows: ProteomeXchange accession: PXD000154.

Acknowledgments—We thank Daniela Wieseler for excellent technical assistance. We are grateful to Marc Kippinc and Holger Kock for critical reading the manuscript.

1. Hecker, M. (2003) A proteomic view of cell physiology of Bacillus subtilis–bringing the genome sequence to life. Adv. Biochem. Eng. Biotechnol. 83, 57–92
2. Souchelnytskiy, S. (2005) Bridging proteomics and systems biology: What are the roads to be traveled? Proteomics 5, 4123–4137
3. Goessler, A., and Fromion, V. (2011) Bacterial growth rate reflects a bottleneck in resource allocation. Biochim. Biophys. Acta 1810, 978–988
4. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dehoure, N., O’Shea, E. K., and Weissman, J. S. (2003) Global analysis of protein expression in yeast. Nature 425, 737–741
5. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc. Natl. Acad. Sci. U.S.A. 100, 6940–6945
6. Maas, S., Sievers, S., Zülke, D., Kuzinski, J., Sappa, P. K., Muntel, J., Hessling, B., Bernhardt, J., Sietmann, R., Völker, U., Hecker, M., and Becher, D. (2011) Efficient, global-scale quantification of absolute protein amounts by integration of targeted mass spectrometry and two-dimensional gel-based proteomics. Anal. Chem. 83, 2677–2684
7. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005) Exponentially modified protein abundance index (EMPA) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol. Cell. Proteomics 4, 1265–1272
8. Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat. Biotechnol. 25, 117–124
9. Braisted, J. C., Kuntumalla, S., Vogel, C., Marcotte, E. M., Rodrigues, A. R., Wang, R., Huang, S.-T., Ferlanti, E. S., Saeed, A. I., Fleischmann, R. D., Peterson, S. N., and Pieper, R. (2008) The APEX Quantitative Proteomics Tool: generating protein quantitation estimates from LC-MS/MS proteomics results. BMC Bioinformatics 9, 529
10. Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C., and Geromanos, S. J. (2006) Absolute quantification of proteomes by LCMS²: A virtue of parallel MS acquisition. Mol. Cell. Proteomics 5, 144–156
11. Maimström, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E. W., and Aebersold, R. (2009) Proteome-wide cellular protein concentrations of the human pathogenic Lactospora intergenog. Nature 460, 762–765
12. Hoopmann, M. R., Finney, G. L., and MacCoss, M. J. (2007) High-speed data reduction, feature detection, and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass spectrometry. Anal. Chem. 79, 5620–5632
13. Luethy, R., Kessner, D. E., Katz, J. E., Maclean, B., Grothe, R., Kani, K., Faça, V., Pitteri, S., Hanash, S., Aguas, D. B., and Mallick, P. (2008) Precursor-ion mass re-estimation improves peptide identification on hybrid instruments. J. Proteome Res. 7, 4031–4039
14. Bateman, R. H., Carruthers, R., Hoyes, J. B., Jones, C., Langridge, J. I., Millar, A., and Vissers, J. P. C. (2002) A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. J. Am. Soc. Mass Spectrom. 13, 792–803
15. Geromanos, S. J., Vissers, J. P., Silva, J. C., Dorschel, C. A., Li, G.-Z., Gorenstein, M. V., Bateman, R. H., and Langridge, J. J. (2009) The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependent LC-MS/MS. Proteomics 9, 1683–1695
16. Blackburn, K., Mbeunkfi, F., Mitra, S. K., Mentzel, T., and Goshe, M. B. (2010) Improving protein and proteome coverage through data-independent multiplexed peptide fragmentation. J. Proteome Res. 9, 3621–3637
17. Cheng, F., Blackburn, K., Lin, Y., Goshe, M. B., and Williamson, J. D. (2009) Absolute protein quantification by LC/MS/MS for global analysis of sialic-acid-induced plant protein secretion responses. J. Proteome Res. 8, 82–93
18. Röwer, C., Vissers, J. P., Koy, C., Kipping, M., Hecker, M., Reimer, T., Gerber, B., Thiessen, H.-J., and Glocker, M. O. (2009) Towards a proteomic signature for invasive ductal breast carcinoma derived from label-free nanoscale LC-MS protein expression profiling of tumour and glandular tissue. Anal. Bioanal. Chem. 395, 2443–2456
19. Goelzer, A., Bekkal Biriki, F., Martin-Verstraete, I., Nairoit, P., Bessières, P., Aymerich, S., and Fromion, V. (2008) Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of Bacillus subtilis. BMC Syst. Biol. 2, 20
20. Jules, M., Le Chat, L., Aymerich, S., and Le Coq, D. (2009) The Bacillus subtilis ywlJ (gplX) gene encodes a class II fructose-1,6-bisphosphatase, functionally equivalent to the class III Fbp enzyme. J. Bacteriol. 191, 3168–3171
21. Sharpe, M. E., Hauser, P. M., Sharpe, R. G., and Errington, J. (1998) Bacillus subtilis cell cycle as studied by fluorescence microscopy: constancy of cell length at initiation of DNA replication and evidence for active nuclear partitioning. J. Bacteriol. 180, 547–555
22. Li, G.-Z., Vissers, J. P., Silva, J. C., Golick, D., Gorenstein, M. V., and Geromanos, S. J. (2009) Database searching and analysis of multiplexed precursor and product ion spectra from data independent LC-MS with data dependent LC-MS/MS. Proteomics 9, 1696–1719
23. Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P., Meurice, G., Sekowska, A., Vallenet, D., Wang, T., Moszer, I., Medigue, C., and Danchin, A. (2009) From a consortium sequence to a unified sequence: the Bacillus subtilis 168 reference genome a decade later. Microbiology 155, 1758–1775
24. Vissers, J. P., Langridge, J. J., and Aerts, J. M. F. G. (2007) Analysis and quantification of diagnostic serum markers and protein signatures for Gaucher disease. Mol. Cell. Proteomics 6, 755–766
25. Zhou, M., Boekhorst, J., Francke, C., and Siezen, R. J. (2008) LocateP–
29. Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., Ruhr, M., Le Coq, D., Aymerich, S., and Sauer, U. (2012) 13C-flux analysis of Bacillus subtilis salt stress adaptation. J. Bacteriol. 192, 870–882.

30. Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Nicolas, P., Maeder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonnau