JOURNAL OF PROTEOMICS 109 (2014) 322–331

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Multi-enzyme digestion FASP and the 'Total Protein Approach'-based absolute quantification of the *Escherichia coli* proteome

Jacek R. Wiśniewski[a,*], *Dariusz Rakus*[b]

[a]*Biochemical Proteomics Group, Department of Proteomics and Signal Transduction, Max-Planck-Institute of Biochemistry, Martinsried, Germany*
[b]*Department of Animal Molecular Physiology, Wroclaw University, Wroclaw, Poland*

ARTICLE INFO

Article history:
Received 9 April 2014
Accepted 14 July 2014
Available online 22 July 2014

Keywords:
Escherichia coli proteome
'Total Protein Approach'
Filter-aided sample preparation
FASP
Absolute protein quantification
Protein copy number

ABSTRACT

We describe a proteomic approach combining the multi-enzyme digestion FASP-sample processing strategy and the 'Total Protein Approach' applied to absolute quantification of proteins in *Escherichia coli*. Consecutive digestion of whole cell lysates with LysC and trypsin allowed the generation of two populations of peptides at a yield of 76%. Subsequent two 4-hour LC–MS/MS analyses allowed the identification of 19,000 unique peptides per sample. Notably, only 1.2 and 2.4% of the identified peptides were found to be incompletely cleaved by the LysC and trypsin, respectively. The analysis resulted in the identification of 2200 proteins per sample. We show high reproducibility of the approach, allowing the accurate estimation of cellular protein concentrations. Quantitative analysis of the DNA content per sample enabled the calculation of the protein content per bacterial cell and, as a result, estimation of protein copy numbers. The accuracy of these estimations was confirmed by analyzing protein complexes with known subunit stoichiometry and cellular abundances. In stationary culture, a single bacterium contains about 6500 copies of ribosomes, 300 molecules of RNA polymerase and 10 DNA polymerase assembles. The here presented experimental and computational workflow offers an easy way to analyze proteomes quantitatively.

Biological Significance
We demonstrate a proteomic workflow for in-depth analysis of small proteomes with minimal fractionation extent and mass spectrometry measuring time. For the first time we provide the quantitative picture of the *Escherichia coli* proteome at protein copy number.

1. Introduction

Escherichia coli, a gram negative bacterium abundantly occurring in the lower intestine of mammals, belongs to the most extensively studied organisms. It serves as a prokaryotic model species and plays an important role in biotechnology. During the recent decade the *E. coli* proteome has been the subject of numerous investigations involving protein mapping

DOI of original article: http://dx.doi.org/10.1016/j.dib.2014.08.004.
* *Corresponding author at:* Department of Proteomics and Signal Transduction, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany. Tel.: +49 89 8578 2502.
E-mail address: jwisniew@biochem.mpg.de (J.R. Wiśniewski).

http://dx.doi.org/10.1016/j.jprot.2014.07.012

studies [1–3] and analyses of relative protein changes under different growth conditions [4–6]. Using extensive fractionation strategies, up to 2600 proteins were identified in the bacterium [3,5]. Regardless of these achievements, the quantitative organization of the *E. coli* proteome remains only partly characterized, and our knowledge is mainly based on biochemical data published over the past decades. Lu et al. made the first attempt to estimate protein abundances in a system-wide manner, developing a method for large-scale absolute protein expression measurements (APEX) [7,8]. This study provided estimates of protein abundances in copy number per cell for 450 proteins identified in a proteomic experiment. An indirect proteomic approach using a yellow fluorescent protein fusion library for *E. coli* provided estimates of abundances for 1018 proteins [9].

Here we describe an analysis of the *E. coli* proteome which combines the multi-enzyme digestion FASP (MED-FASP) [10] sample preparation and the Total Protein Approach (TPA) [11] for calculation of absolute protein abundances. The procedure integrates determination of the DNA and RNA contents in the sample, which allows scaling of protein abundances and calculation of protein copy numbers per cell.

2. Materials and methods

2.1. Bacterial lysate

E. coli strain ATCC 25922 was cultured at 37 °C in Luria-Bertani broth medium with shaking at 250 rpm for approximately 15 h. The bacteria were harvested by centrifugation at 5000 × *g* and then lysed within 2% SDS in 0.1 M Tris–HCl, pH 7.8 containing 0.1 M DTT at 100° for 5 min. Three samples were prepared from a single bacterial suspension. The non-soluble material was removed by centrifugation at 16,000 × *g*, and the supernatants were used for analysis.

2.2. Filter-aided protein and nucleic acid digestion

Three whole cell lysates were processed according to the MED-FASP [10] protocol that was extended with nucleic acid digestion steps. Briefly, aliquots containing 50 μg total protein were mixed with 200 μL of 8 M urea in 0.1 M Tris/HCl, pH 8.5 (UA) in centrifugal ultrafiltration units with a nominal molecular weight cutoff of 30,000 (Cat No. MRCF0R030, Millipore), and then centrifuged at 14,000 × *g*, 20 °C, for 15 min. The eluates were discarded, 100 μL of UA was pipetted into the filtration unit, and the units were centrifuged again. Then 50 μL of 0.05 M iodoacetamide in UA was added to the filters, and samples were incubated in darkness for 20 min. Filters were washed twice with 100 μL of UA followed by two washes with 100 μL of 0.05 M Tris/HCl pH 8.5. Proteins were digested in 40 μL 0.05 M Tris/HCl pH 8.5 at 37 °C for 18 h, using endoproteinase LysC, at an enzyme to protein ratio of 1:50. The released peptides were collected by centrifugation at 14,000 × *g* for 10 min followed by two washes with 0.05 M Tris/HCl pH 8.5. After isolation of the peptides, material remaining on the filter was digested with trypsin using the above conditions, except that the cleavage reaction was performed for only 2 h. After collection of the peptides released by trypsin, the material remaining on the filter was washed once with TE buffer (10 mM Tris–HCl, pH 8.0) and then the RNA was cleaved with 0.5 μL (0.5U) of RiboShredder (Epicentre, Madison, WI) in 60 μL of TE at 37 °C for 1 h. The ribonucleotides were collected by centrifugation and two consecutive, 80 μL each, washes with the TE buffer. Next the filtration units were assembled in new tubes and the DNA was digested with 6 μg DNase (DN25, Sigma, St. Louis) in 60 μL of 10 mM Tris–HCl, pH 7.8 buffer containing 2.5 mM $MgCl_2$ and 0.5 mM $CaCl_2$, at 37 °C for 1 h. The DNA-digests were collected by centrifugation and the elution was completed by subsequent eluting with two 80 μL aliquots of the buffer. The DNA and RNA yields were better than 95% as observed in preliminary experiments in which defined amounts of DNA and RNA were added to protein lysates before FASP and the nucleic acid cleavages.

2.3. Determination of the total protein and nucleic acid content

Total protein and total peptide content was determined using a tryptophan-fluorescence assay as described previously [12]. The DNA and RNA content was determined by UV spectrometry using the extinction coefficient of 0.025 $(\mu g/mL)^{-1}\,cm^{-1}$ at 260 nm for ribonucleotides and 0.030 $(\mu g/mL)^{-1}\,cm^{-1}$ at 260 nm for deoxyribonucleotides.

2.4. LC–MS/MS and data analysis

Aliquots containing 6 μg of LysC peptide or 4 μg of tryptic peptides were separated on a reverse phase column and analyzed on QExactive mass spectrometer as described previously [13]. Briefly, the peptides were separated on a reverse phase column (20 cm × 75 μm inner diameter) packed with 1.8 μm C18 particles (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) using a 4 h acetonitrile gradient in 0.1% formic acid at a flow rate of 250 nl/min. The liquid chromatography was coupled to a Q Exactive mass spectrometer (Thermo Fisher Scientific, Germany) via a nanoelectrospray source (Proxeon Biosystems, now Thermo Fisher Scientific). The QExactive was operated in data dependent mode with survey scans acquired at a resolution of 50,000 at m/z 400 (transient time 256 ms). Up to the top 10 most abundant isotope patterns with charge m/z 2 from the survey scan were selected with an isolation window of 1.6 Th and fragmented by HCD with normalized collision energies of 25. The maximum ion injection times for the survey scan and the MS/MS scans were 20 ms and 60 ms, respectively. The ion target value for both scan modes were set to 10. The MS data were analyzed within the software environment MaxQuant [version 1.2.6.20], using the Andromeda search engine [14,15]. The spectra of the LysC and tryptic fractions were combined in MaxQuant unless indicated otherwise. Proteins were identified by searching MS and MS/MS data of peptides against UniProtKB *E. coli* (K12). The FDR threshold was derived by analyzing the decoy database. Carboamidomethylation of cysteines was set as fixed modification. The maximum false peptide and protein discovery rates were specified as 0.01. Spectra were searched with K-specificity for LysC and K/R but not K/RP for trypsin.

2.5. Absolute protein quantification

Protein abundance was calculated on the basis of spectral protein intensity (raw intensities, not LFQ intensities) using

the TPA concept [11]. The total protein content was defined as a sum of peptide intensities integrated over the elution profile of each peptide. The amount of individual proteins was calculated as the ratio of their intensity (MS_signal) to the sum of all intensities (total MS_signal) in the measured sample:

$$Total\ protein(i) = \frac{MS\_signal\ (i)}{Total\ MS\_signal} \quad (1)$$

Molar concentrations of proteins per g total protein were calculated from:

$$Protein\ concentration(i) = \frac{MS\_signal\ (i)}{Total\ MS\_signal\ \times MW(i)} \left[\frac{mol}{g\ total\ protein}\right] \quad (2)$$

where *MS_signal* and *Total MS_signal* refer to total MS1 signal intensity of the protein i and the total protein MS1 signal.

Total protein content per cell was calculated from:

$$Total\ protein\ mass\ per\ cell = \frac{DNA\ per\ cell}{Total\ DNA\ per\ sample} \times protein\ mass\ per\ sample \quad (3)$$

Where *DNA per cell* equals 5 fg (mass of the *E. coli* genome of 4.6 Mb DNA).

The protein copy number per cell was obtained from:

$$Copy\ number\ (i) = Protein\ concentration\ (i) \times Total\ protein\ mass\ per\ cell \times N_A \quad (4)$$

where N_A is the Avogadro constant.

The calculations were performed in Microsoft Excel.

3. Results

3.1. Peptide and protein identifications

Whole *E. coli* cell lysates were processed using the MED-FASP method in which the proteins are first digested with LysC and, after elution of the released peptides, the material remaining in the filter is cleaved using trypsin (Fig. 1A). Both peptide populations are then analyzed in two separate LC–MS/MS runs [10]. In each of the three experiments, aliquots of the lysate containing 50 μg of total protein were processed. The protein to peptide conversion yield was on average 76 ± 1% (Table 1). The LC–MS/MS analyses allowed the identification of 8206 ± 270 unique peptides in the LysC fraction, and 10,728 ± 319 tryptic peptides per sample (Fig. 1B; Supplementary Table 1). These peptides corresponded to 2200 ± 11 proteins per sample (Supplementary Table 2). 88 ± 1% of the proteins were identified by at least two peptides.

A low number of peptides with missed cleavages is an indicator of the completeness of the digestion process. Inspection of the peptides derived from the LysC cleavage revealed that only 1.2% of the peptides had a lysyl residue at a non-C-terminal position, whereas in the fraction of tryptic peptides we found 2.4% of partially digested peptides (Fig. 1B; Supplementary Table 1). More than half of the not fully cleaved tryptic peptides were also identified in the LysC fraction. Since more than 75% of these peptides were several-fold less abundant in the fraction generated by trypsin than in the LysC derived fraction, it is likely that their occurrence in the tryptic fraction resulted from a carryover from LysC to trypsin fraction (Fig. 1C), rather than originating from non-complete digestion by trypsin. Therefore we estimate that the portion of partially digested peptides in the samples was between 1 and 1.5%.

3.2. Determination of nucleic acids

In contrast to the in-solution and in-gel protein digestion methods, in FASP, peptides are separated from undigested material by ultrafiltration. The material retained in the filter contains peptides that are too large to pass through the ultrafiltration membrane, as well as non-proteinaceous substances, such as nucleic acids. Since determination of the amounts of DNA and RNA can provide quantitative information on the analyzed sample, such as equivalents of number of cells or ribosomes, we consecutively cleaved the remaining material with RNase and DNase (Fig. 1A). Following the separate elution-steps, ribonucleotides and deoxyribonucleotides were quantified by UV-absorption measurements to determine RNA and DNA content in the sample (Table 1).

For the calculation of the total protein content per cell (using Eq. (3)), we assumed that under stationary culture conditions an *E. coli* cell contains one copy of the 4.6 Mb DNA. This corresponds to 5.0 fg of the nucleic acid per cell. Therefore a single cell used in the experiments contained 75 ± 7.4 fg of total protein and 20 ± 1.6 fg of RNA (Table 2). These values are close to the estimates provided by Bremer and Dennis [16] for a slowly dividing *E. coli* (0.6 doublings per h) (Table 2).

3.3. Calculation of the absolute proteins abundances

Next, we estimated total protein contents and concentrations of proteins by means of the TPA method [11]. This approach, similar to various other label free absolute protein quantitation methods [7,17,18], assumes proportionality between protein abundance and its measured spectral intensity. But in contrast to the other approaches, in the TPA method the absolute protein values, such as a percent of the total protein or molar concentration per g total protein, are directly derived from the MS outputs without using any standard [11]. However, the calculation of protein copy numbers requires information on the total protein content per cell (Fig. 1A). Using the total protein content of 75 fg per cell (see above) we computed protein copy numbers per cell. We estimated the total number of protein molecules per cell to be $1.32 \pm 0.15 \times 10^6$. This value is about 2 times lower than the value reported for the exponentially growing cells at a doubling rate of 40 min [19]. It is likely that the difference reflects the fact that rapidly dividing *E. coli* cells contain several-fold more total protein than slowly growing cells [16].

Recently Peng et al. have demonstrated that absolute protein quantitation data are heavily biased by the protease's specificity and irreproducibility in peptide formation [20]. In contrast to that, we observed high correlation in copy numbers between the 3 experiments ($r^2 > 0.97$) (not shown). We also compared copy numbers calculated from the two step digestion with those from the analysis of the LysC peptides (Fig. 2). Considering 1580 proteins identified in both analyses we found a relatively

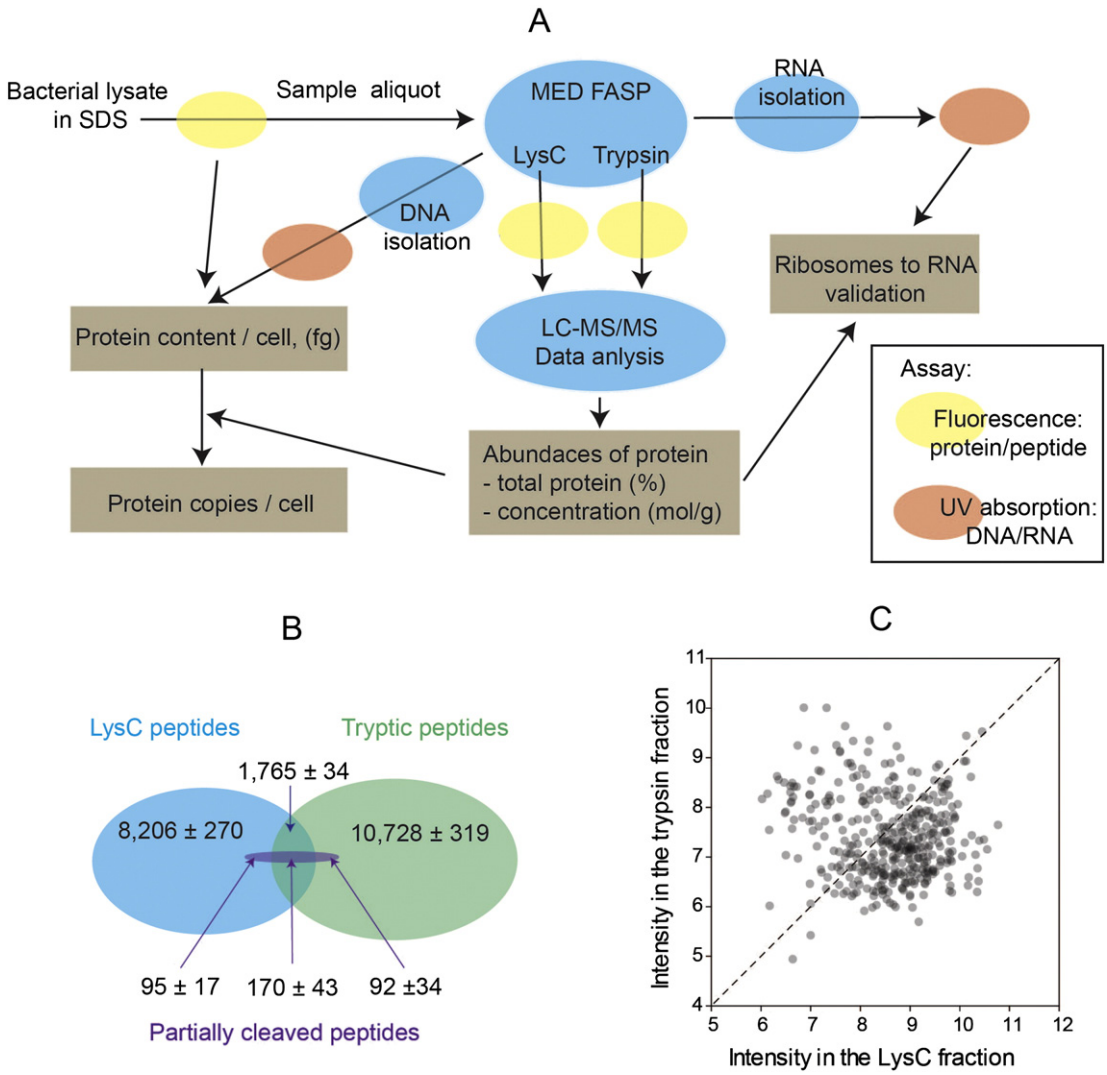

Fig. 1 – (A) Overview on the proteomic workflow applied to the *E. coli* lysates. Bacterial lysates were processed according to MED-FASP protocol. Peptides were analyzed by LC–MS/MS and the resulting spectra were handled by MaxQuant software. Absolute protein contents and protein concentrations were calculated using the Total Protein Approach. RNA and DNA were digested with nucleases and the released nucleotides were quantified. The total protein content of the single bacterial cell was calculated from the total DNA and total protein of the sample. The total protein content of the single cell was used for computation of protein copy numbers per cell. RNA content was used for validation of the number of ribosomes per cell. (B) Identification of peptides in the LyC and tryptic fractions. (B) Spectral abundances of the miscleaved peptides.

high correlation (r^2 = 0.88) of the copy numbers between the datasets (Fig. 2A). When the list of proteins common to both sets was limited to those identified with at least 3 peptides (1180 proteins) the correlation coefficient r^2 was as high as 0.92 (Fig. 2B). We attribute these remarkable good correlations to the digestion efficiency and reproducibly achieved using the MED-FASP approach.

Table 1 – Processing of whole lysates and analysis of their contents.

Sample (total protein)	LysC + Tryptic peptides (μg) (yield)	RNA (μg)	DNA (μg)	DNA copies (cells) (10^8)	Total protein per cell (fg)
A (50 μg)	29.8 + 7.8 (75%)	14.7	3.5	7.0	71
B (50 μg)	28.6 + 8.7 (75%)	16.0	4.4	8.8	57
C (50 μg)	29.1 + 8.7 (76%)	16.0	3.8	7.6	66

We found that protein copy number values span 5 orders of magnitude form 0.1 to 10^4 molecules per cell, with a median value of 67 copies (Fig. 3A). A similar range of protein abundances was previously estimated using a yellow fluorescent protein fusion library [9]. The median abundance of the *E. coli* proteins which are essential for cell growth [21] was 380 copies per cell. The most abundant proteins were the major outer membrane lipoprotein Lpp, the elongation factor Tu, and the DNA-binding protein HUα, with 38,000 (or 0.52 fg total protein per cell), 22,300 (0.29 fg), and 18,800 (0.30 fg) copies per cell, respectively. Lu et al. [8] previously estimated the presence of 87,000 copies per cell of the elongation factor, and 15,000 copies of HUα, assuming a total

Table 2 – Basic macromolecular components of *E. coli* determined in this study and those given by Bremer and Dennis [16] and Neidhardt and Umbarger [19].

Parameter	This study	*E. coli* (literature)
Total protein per cell (fg)	75 ± 7.4	100
RNA per cell (fg)	20 ± 1.6	20
DNA per cell (fg)	5*	7.6
RNA Polymerase total protein (%)	0.27 ± 0.03	0.9
RNA Polymerase copies per cell (α subunit)	589 ± 91	1500
Ribosomal total protein (%)	13.8 ± 1.8	9
Ribosomes per cell	6500 ± 3100	6800
Total number of protein molecules per cell	1.32 ± 0.15 × 10^6	2–3 × 10^6

* This value was set for calculations assuming one copy of DNA per cell in the stationary phase.

of 2.5 million molecules per cell (which is twice as high as the number determined by us). These values are in good agreement with our results, further confirming the accuracy of the approach. However, the accuracy of the protein abundances calculation is lower for proteins identified with a low number of peptides (Fig. 2 C), which may reflect biases of dynamic range and occurrence of peptides with poor ionization efficiency. Therefore, protein copy numbers calculated for these proteins have to be considered as a rough estimation. But importantly, we did not observe any systematic error in the TPA approach leading to under- or overestimation of the copy numbers for small proteins observed for other quantitation methods, such as for Top3 and emPAI [22]. Fig. 2D shows that there is no correlation between the copy number and the protein mass, with the exception for the most abundant proteins which copy numbers decrease for larger proteins. The latter effect has been previously observed for yeast proteome by Warringer and

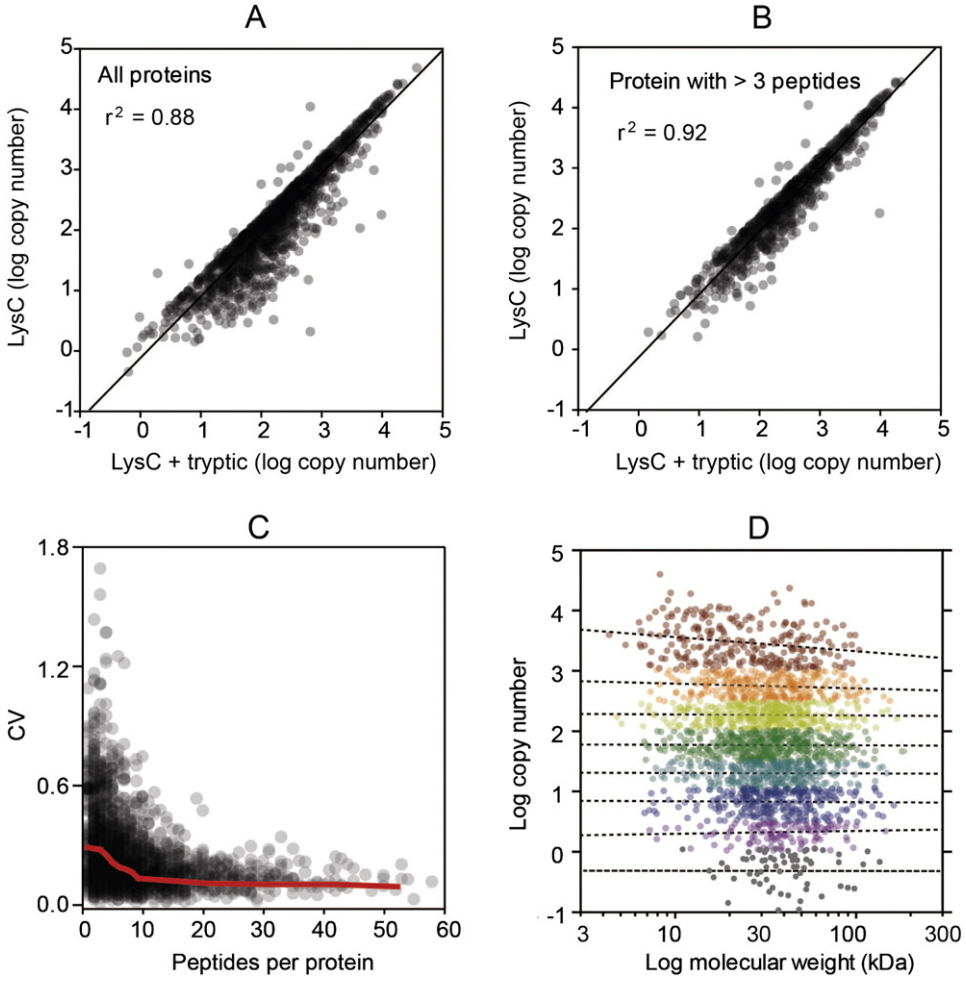

Fig. 2 – (A) and (B) correlation of protein copy numbers calculated using intensities of LysC fraction and the combined intensities of LysC and trypsin fractions. In A all matching proteins were compared whereas in B only those with at least 3 peptides. (C) Reproducibility between individual analyses plotted against number of identified peptides per protein. CV, coefficient of variation equals the ratio of standard deviation divided by the mean value. The red line shows the change of the median of the CV value. (D) Protein abundance versus molecular weight. Proteins were grouped according to their abundance into bins: >1000, 300–1000, 100–300, 30–100, 10–30, 3–10, 1–3 and 0.1–1 copies per cell. The *dashed lines* show linear regression for each group.

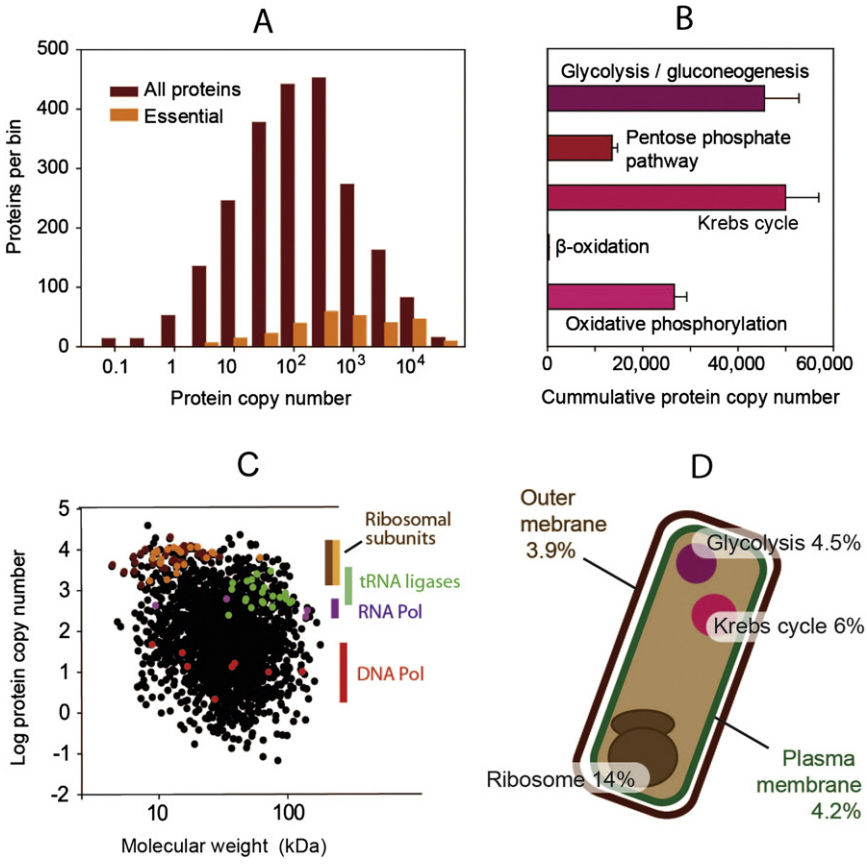

Fig. 3 – Quantitative insights in the *E. coli* proteome. (A) Distribution of the abundances of all proteins essential for bacterial growth. Proteins were grouped according to their abundance into bins. (B) Cumulative abundance of proteins of selected major metabolic processes (Supplemental Table 3). Bars show the summarized number of copies of proteins directly involved in the metabolism. (C) Functionally related proteins occur at similar copy numbers irrespectively their molecular weights. Proteins of the ribosomal subunits, tRNA-amino acid ligases, RNA, and polymerase complexes were highlighted (D) Abundance of basic elements of the architecture and composition of the *E. coli* cell. The values refer to the percent of total cellular protein.

Blomberg [23], who found an inverse relationship between protein size and protein expression such as highly expressed proteins tend to be of smaller size.

Quantitative proteomic data give insights into the abundance of proteins involved in biochemically defined processes. For example, one of the major functional group of processes in our dataset is energy metabolism, catabolizing nutrients to generate ATP. Roughly, it comprises glycolysis/the Entner–Doudoroff pathway and gluconeogenesis, the pentose phosphate pathway, the Krebs' cycle, β-oxidation and oxidative phosphorylation (Fig. 3B). The entire machinery comprises about 120–150 thousands of molecules per cell, e.g. about 10% of all molecules per cell, a value about 2.5 times lower than the number of ribosomal proteins.

In contrast to other catabolic processes, the titer of β-oxidation enzymes was several orders of magnitude times lower. This is in agreement with the hypothesis that *E. coli* regards fatty acids as energetic substrates of minor status. Detailed analysis of the data presented here demonstrated that the low concentration of fatty acid (FA) catabolism enzymes correlated with the high abundance of FadR (Supplementary Table 2), which represses transcription of genes essential for FA transport, and activation and β-oxidation (such as: fadA, fadB, fadD, fadE, fadI, fadJ and fadL) and activates the genes encoding enzymes of FA synthesis (e.g. fabA and fabB) [24].

Plotting copy number of proteins against their molecular weights demonstrates that proteins involved in specific processes have similar abundances irrespectively to their molecular masses (Fig. 3C). This applies to proteins organized in complexes such as ribosomal proteins and RNA and DNA polymerases as well as to proteins with similar functions but different activities such as amino acid RNA ligases.

Absolute protein abundances expressed as total protein allow quantitative insights in the composition and architecture of cells (Fig. 3D). The outer and inner membrane proteins compose together 8% of the total protein. The most abundant classes of proteins, ribosome subunits, glycolytic and citric acid cycle enzymes compose 14, 4.5 and 6% of the proteome, respectively.

3.4. Validation of the protein abundances with macromolecular assemblies

In cells, proteins are often organized in functional complexes. Therefore it is expected that the concentrations of proteins engaged in such multiprotein-assemblies occur at similar cellular concentrations [25,26].

E. coli ribosomal proteins are arranged into two subcomplexes, the large (50S) and the small (30S) subunits consisting of about 30 and 20 proteins respectively. We identified the complete set of these proteins. The average copy numbers of the proteins of the large and small subunits were $7.0 \pm 3.2 \times 10^3$ and $6.3 \pm 2.7 \times 10^3$, respectively (Fig. 4A). This agrees well with the 1:1 stoichiometry of the ribosomal subunits. We also found that the mass of the ribosomal proteins was 13.8% of the total protein (or 10.3 fg of protein per cell; 350,000 total protein copies) (Table 2), a value close to the 9% reported for slowly growing bacteria [16]. Next, we compared the cellular contents of the total ribosomal protein and RNA. In bacterial ribosomes, the mass ratio of rRNA to total ribosomal protein is about 2. This translates to about 20 fg of RNA per cell. Since in *E. coli* 85% of RNA is rRNA [16], the experimentally determined RNA content of 20 fg correlates well with the expected amount (Table 2).

The F1 subunit of the ATP synthase, the major ATP producing enzyme, is composed of the α, β, γ, δ, and ε subunits in the stoichiometric ratio 3:3:1:1:1. Our copy number estimates of these subunits reflect this composition. Whereas the α and β subunits occurred at 2700–3700 copies, the γ, δ, and ε ones were present at 600–1700 subunits (Fig. 4B).

The pyruvate dehydrogenase complex links the glycolytic pathway with the Krebs cycle by converting pyruvate into acetyl-CoA. Bacterial pyruvate dehydrogenase is assembled from 3 types of subunits: the pyruvate dehydrogenase E1 (aceE), dihydrolipoyl transacetylase (aceF), and the dihydrolipoyl dehydrogenase (lpdA). In the complex, the central core is assembled from up to 24 copies of aceF, whereas up to 24 molecules of aceE and 12 molecules of Ipds are arranged outside of the core. Our estimates show nearly equal number of copies of the aceE and aceF, and about 1.5-fold more molecules of IpdA (Fig. 4C).

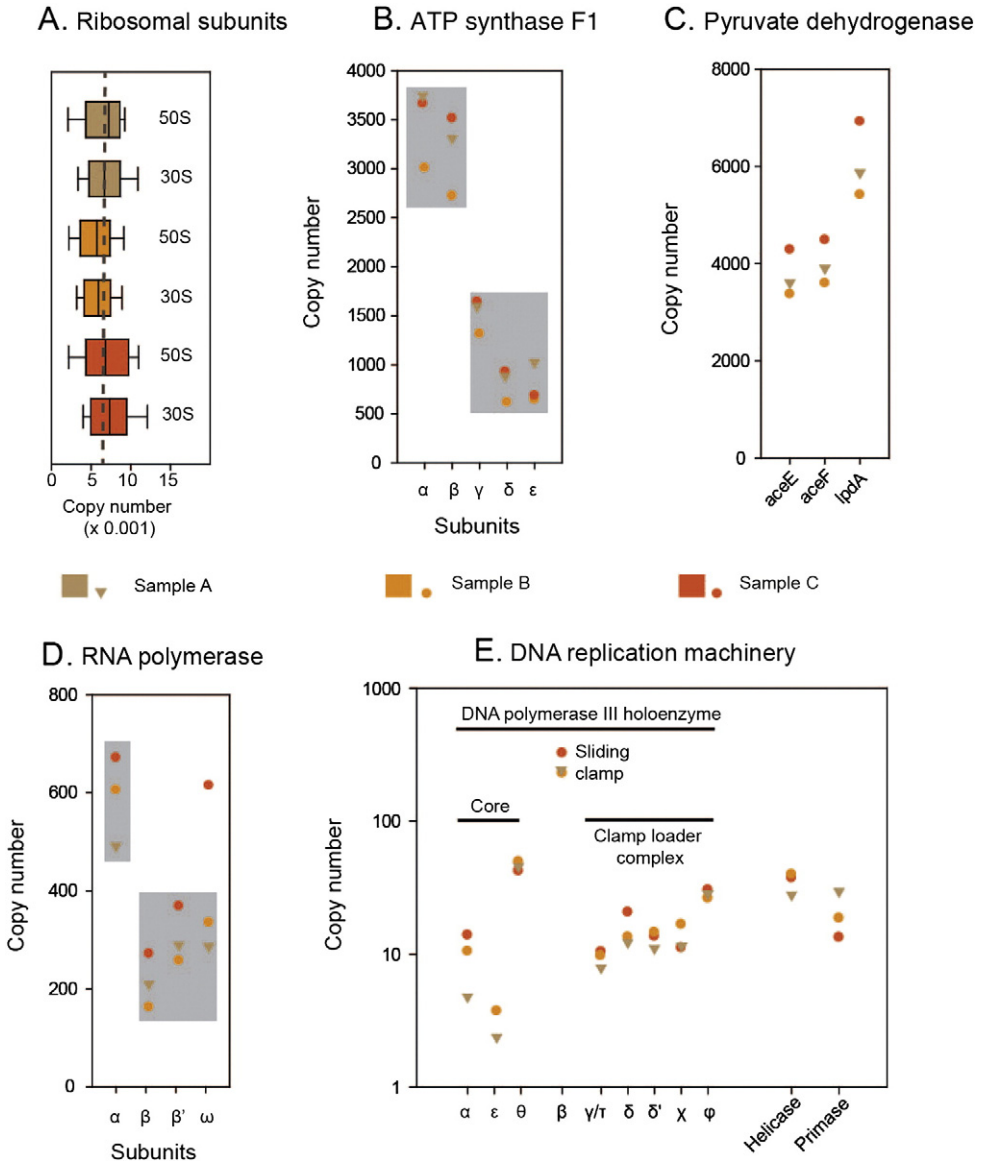

Fig. 4 – Absolute protein abundances in known protein assemblages. (A) ribosomal subunits; (B) F1 complex of the ATP synthase; (C) Pyruvate dehydrogenase complex; (D) RNA polymerase; (E) DNA replication machinery including the holoenzyme, DNA-primase, and helicase.

Bacterial RNA polymerase is composed of two α-subunits, and one copy of the β, β' and ω subunits [27]. Our copy number estimates fully reflect the stoichiometry of the complex (Fig. 4D). The α subunit was present at about 600 copies per cells, whereas the other components had on average 300 copies. This about 2.5-fold lower abundance of the RNA polymerase compared to the to the estimates of Bremer and Dennis [16] may reflect the different status of bacterial cells: stationary culture vs. slowly growing.

Finally, we looked at proteins involved in the DNA replication. In *E. coli*, DNA polymerase III is the major enzyme involved in DNA replication. The holoenzyme consists of 10 subunits organized in 3 subassemblies: the core enzyme with subunits α, ε, and θ, the sliding clamp factor β, and the clamp-loading complex composed of the τ/γ, δ, δ', χ, and ψ subunits. Beside the DNA polymerase III, DNA primase and DNA helicase are two additional key components of the replication machinery. Our analysis revealed that the concentrations of these components of the replication machinery were within a narrow range of 10–30 molecules per cell. Only the factor β was much more abundant with 320 copies per cell (Fig. 4E). These values are in an excellent agreement with previously published biochemical data that estimated 10–20 copies per cell of the holoenzyme except for the sliding clamp factor β, which was estimated at 300 copies per cell [28].

4. Discussion

Complete protein extraction and effective protein digestion are prerequisites for accurate quantitative proteomic analysis. A comparison of a variety of reagents commonly used in proteomics showed that SDS is the most effective in protein solubilization [29]. Since peptides must be depleted of the detergent prior to mass spectrometric analysis, different methods for its removal were developed, based on precipitation, detergent adsorption, and ultrafiltration. The latter approach initially proposed by Manza et al. [30] was further developed in our laboratory into the FASP method allowing efficient detergent depletion and protein digestion [31]. The chemical reactor principle of the FASP method makes it versatile, enabling stepwise protein cleavage with the consecutive isolation of peptides (MED-FASP) [10] as well as digestion of nucleic acids for the purpose of determination of their amount per sample. Here, we show that MED-FASP allows the generation of peptides with a very low content of 1–1.5% of missed cleavages, indicating that the digestion process was close to completeness. In our opinion, this is the result of the applied digestion conditions; a high pH of 8.5 for both LysC and trypsin, and the absence of chaotropic substances, such as urea or guanine HCl in the digestion buffer. Another important factor is avoiding an overload of the filtration unit. Our experience is that loads of more than 100 μg of total protein per 0.5 mL ultrafiltration unit often result in lower peptide yields in comparison to the yields when smaller amounts of sample are processed.

In contrast to many proteomic sample preparation techniques, including 'in-gel' and 'in-solution' approaches, the FASP procedure involves determination of the peptide yield. Depending on the sample the yields vary between 60 and 95% of the total protein. The 'incomplete yields' can be attributed to a leaking of small peptides present in the undigested sample and to a retention of tryptic peptides that are too large to pass the ultrafiltration membrane. On the other hand, the observed losses cannot be explained by partial protein digestion, because the content of peptides with missing cleavages is always low. In addition, the sample losses are not explainable by sample adsorption to the filter material, because it has been shown by several authors that at low sample loads FASP outperforms other sample preparation methods [32–34].

Recently Ahrné et al. [22] compared the accuracy of different methods for label free absolute protein quantification, concluding that approaches using standards provide more accurate values. We cannot judge to which extent these observations apply to TPA. Previously, we have demonstrated good correlation between the experimentally set and the calculated absolute values using the UPS2 standard [11]. In addition we showed a reasonably accurate correlation between protein copy derived by TPA and those obtained in study using stable isotope labeled standard proteins [35]. In contrast to the methods compared by Ahrne et al., TPA does not require neither spike-in standards nor biochemical determination of the sample size (total protein) and the protein concentrations (mol/g total protein or copies/g total protein) are directly derived from the MS intensity outputs. However, the prerequisite for TPA calculation is a substantial depth of proteomic analysis. The latter can be significantly enhanced by applying sequential protein digestion in the MED-FASP format [10].

In the described experiments, the entire LC–MS/MS analysis required 8 h. During this measuring time we identified 2200 proteins per sample. This is 85% of the proteins identified using triplicate 41 h peptide fractionation over a 3 m long reverse columns [2] or by cumulating data from different experiments involving strong anion exchange chromatography, off-gel isoelectric focusing, and gel-based LC–MS [3]. In contrast to those methods, protein digestion into two fractions with relatively low peptide overlap offers a simple peptide separation method that significantly increases the number of identifications, without anticipated material losses by column or cartridge fractionation. In this study, we identified about 19,000 peptides per sample. Using the same experimental design we found 28,000 and 55,000 peptides per sample of the budding yeast and human lung adenocarcinoma A549 cells (unpublished). This suggests that identification of the *E. coli* proteins beyond the 2200 found in this study was not possible due their absence or very low abundance (presumably below 1 or close to few copies per cell). The key limiting factor seems to be the dynamic range of the mass spectrometer, rather than the rate of MS/MS sequencing of the instruments.

The TPA is a simple absolute protein quantification method that is applicable to large scale proteomic analyses. We have already shown that the TPA copy number estimates are close to the values achieved using spike-in standards in HeLa cells [11]. Here we applied it to quantitate the bacterial proteome and show that the estimates of protein abundances reflect the stoichiometry of several well-characterized protein assemblages. The absolute concentrations determined by us match the literature values, where available. Notably, we demonstrate

that the accurate abundance estimates span from the highly abundant ribosomal complexes present at 6500 copies down to the DNA-replication machinery with proteins occurring at 10–20 copies per cell. The proteomic workflow described here offers an easy way to study proteomes quantitatively, providing protein copy number estimates without labeled spike-in standards and sophisticated computational analysis, therefore facilitating proteomic analysis in terms of time and resources.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2014.07.012.

Conflict of interest statement

The authors have declared no conflict of interest.

Acknowledgments

We thank Prof. Matthias Mann for the continuous support and Katharina Zettl for the technical assistance, and Dr. Georg Borner (MPI Biochemistry) and Dr. Agnieszka Gizak (Wroclaw University) for the critical reading of the manuscript. This work was supported by the Max-Planck Society for the Advancement of Science.

REFERENCES

[1] Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl FU, Kerner MJ, Frishman D. Protein abundance profiling of the *Escherichia coli* cytosol. BMC Genomics 2008;9:102.
[2] Iwasaki M, Miwa S, Ikegami T, Tomita M, Tanaka N, Ishihama Y. One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the *Escherichia coli* proteome on a microarray scale. Anal Chem 2010;82(7):2616–20.
[3] Krug K, Carpy A, Behrends G, Matic K, Soares NC, Macek B. Deep coverage of the *Escherichia coli* proteome enables the assessment of false discovery rates in simple proteogenomic experiments. Mol Cell Proteomics 2013;12(11):3420–30.
[4] Ping L, Zhang H, Zhai L, Dammer EB, Duong DM, Li N, Yan Z, Wu J, Xu P. Quantitative proteomics reveals significant changes in cell shape and an energy shift after IPTG induction via an optimized SILAC approach for *Escherichia coli*. J Proteome Res 2013;12(12):5978–88.
[5] Soares NC, Spat P, Krug K, Macek B. Global dynamics of the *Escherichia coli* proteome and phosphoproteome during growth in minimal medium. J Proteome Res 2013; 12(6):2611–21.
[6] Lopez-Campistrous A, Semchuk P, Burke L, Palmer-Stone T, Brokx SJ, Broderick G, Bottorff D, Bolch S, Weiner JH, Ellison MJ. Localization, annotation, and comparison of the *Escherichia coli* K-12 proteome under two states of growth. Mol Cell Proteomics 2005;4(8):1205–9.
[7] Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics 2006;5(1):144–56.
[8] Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol 2007;25(1):117–24.
[9] Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science 2010; 329(5991):533–8.
[10] Wisniewski JR, Mann M. Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. Anal Chem 2012;84(6):2631–7.
[11] Wisniewski JR, Ostasiewicz P, Dus K, Zielinska DF, Gnad F, Mann M. Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. Mol Syst Biol 2012;8:611.
[12] Wisniewski JR. Proteomic sample preparation from formalin fixed and paraffin embedded tissue. J Vis Exp 2013;79.
[13] Wisniewski JR, Dus K, Mann M. Proteomic workflow for analysis of archival formalin-fixed and paraffin-embedded clinical samples to a depth of 10 000 proteins. Proteomics Clin Appl 2013;7(3–4):225–33.
[14] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 2008; 26(12):1367–72.
[15] Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res 2011; 10(4):1794–805.
[16] Bremer H, Dennis PP. Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhardt ea, editor. *Escherichia coli* and *Salmonella* typhimurium: cellular and molecular biology. Washington, DC: ASM Press; 1996. p. 1559–85.
[17] Schwanhausser 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):337–42.
[18] Grossmann J, Roschitzki B, Panse C, Fortes C, Barkow-Oesterreicher S, Rutishauser D, Schlapbach R. Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. J Proteomics 2010;73(9):1740–6.
[19] Neidhardt FC, Umbarger HE. *Escherichia coli* and *Salmonella* Typhimurium: cellular and molecular biology. vol. 1. Washington, DC: ASM Press; 1996.
[20] Peng M, Taouatas N, Cappadona S, van Breukelen B, Mohammed S, Scholten A, Heck AJ. Protease bias in absolute protein quantitation. Nat Methods 2012;9(6):524–5.
[21] Kato J, Hashimoto M. Construction of consecutive deletions of the *Escherichia coli* chromosome. Mol Syst Biol 2007;3:132.
[22] Ahrne E, Molzahn L, Glatter T, Schmidt A. Critical assessment of proteome-wide label-free absolute abundance estimation strategies. Proteomics 2013;13(17):2567–78.
[23] Warringer J, Blomberg A. Evolutionary constraints on yeast protein size. BMC Evol Biol 2006;6:61.
[24] Iram SH, Cronan JE. The beta-oxidation systems of *Escherichia coli* and *Salmonella enterica* are not functionally equivalent. J Bacteriol 2006;188(2):599–608.
[25] Ovadi J, Saks V. On the origin of intracellular compartmentation and organized metabolic systems. Mol Cell Biochem 2004;256–257(1–2):5–12.
[26] Ovadi J, Srere PA. Macromolecular compartmentation and channeling. Int Rev Cytol 2000;192:255–80.
[27] Ebright RH. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. J Mol Biol 2000;304(5):687–98.
[28] Kelman Z, O'Donnell M. DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. Annu Rev Biochem 1995;64:171–200.
[29] Iwasaki M, Masuda T, Tomita M, Ishihama Y. Chemical cleavage-assisted tryptic digestion for membrane proteome analysis. J Proteome Res 2009;8(6):3169–75.

[30] Manza LL, Stamer SL, Ham AJ, Codreanu SG, Liebler DC. Sample preparation and digestion for proteomic analyses using spin filters. Proteomics 2005;5(7):1742–5.
[31] Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods 2009;6(5):359–62.
[32] Sharma R, Dill BD, Chourey K, Shah M, VerBerkmoes NC, Hettich RL. Coupling a detergent lysis/cleanup methodology with intact protein fractionation for enhanced proteome characterization. J Proteome Res 2012;11(12):6008–18.
[33] Wisniewski JR, Ostasiewicz P, Mann M. High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. J Proteome Res 2011;10(7):3040–9.
[34] Maurer M, Muller AC, Wagner C, Huber ML, Rudashevskaya EL, Wagner SN, Bennett KL. Combining filter-aided sample preparation and pseudoshotgun technology to profile the proteome of a low number of early passage human melanoma cells. J Proteome Res 2013;12(2):1040–8.
[35] Zeiler M, Straube WL, Lundberg E, Uhlen M, Mann M. A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. Mol Cell Proteomics 2012;11(3):O111[009613].