

What is the total number of protein molecules per cell volume? A call to rethink some published values

Ron Milo

Novel methods such as mass-spectrometry enable a view of the proteomes of cells in unprecedented detail. Recently, these efforts have culminated in quantitative measurements of the number of copies per cell for most expressed proteins in organisms ranging from bacteria to mammalian cells. Here, we estimate the expected total number of proteins per unit of cell volume using known parameters related to the composition of cells such as the fraction of cell mass that is protein, and the average protein length. Using simple arguments, we estimate a range of 2-4 million proteins per cubic micron (i.e. 1 fL) in bacteria, yeast, and mammalian cells. Interestingly, we find that measured values that are reported for fission yeast and mammalian cells are often about 3-10 times lower. We discuss this apparent discrepancy and how to use the estimate as benchmark to recalibrate proteome-wide quantitative censuses or to revisit assumptions about cell composition.

Keywords:

cell biology by the numbers; mass spectrometry; protein abundance; proteomic calibration; protein copy numbers; proteins per cell; quantitative proteomics

Introduction

Proteins are dominant players in the cell in terms of both functionality and biomass, accounting for about half of the total dry mass. They are thus a focus of attention in biological research. With the advent of quantitative proteomics

there has been a growing capability to report the copy numbers of proteins in a wide variety of cell types. However, the explicit question of how many protein molecules are in a cell in total often baffles even experienced researchers. Here, we show how it is possible to estimate this value from properties of

the cell for which we have quite accurate measurements, and compare it with reported values based on different measurement techniques. Various techniques were used throughout the years to measure protein abundances ranging from colorimetric [1] and spectroscopic [2] to amino acid analysis and radioimmunoassays. Those methods are focused on either total protein or specific protein quantification after proper purification, identification, and calibration. The pioneering efforts to quantify a significant fraction of the proteome using radio isotopic labeling and 2D gels [3] has since given way to a torrent of information from mass spectrometry, which enables comprehensive proteome-wide quantification [4-14].

Discussions of the cell content in terms of absolute values are becoming more common as techniques facilitating such measurements become ever more advanced and the functional implications such as the stochasticity that accompanies the many proteins that exist at very low copy numbers, become exposed [15]. Absolute values can also serve as sanity checks on measured reports where, for example, the number of histones in a cell can be constrained by the known DNA length. Similarly, the number of sugar transporters can be constrained, given the growth rate and turnover numbers, as their flux has to suffice to build the cell biomass [16]. Ribosomes have been quantified in various studies and can also serve as benchmarks for the absolute copy numbers per cell at a given growth condition. Careful analysis of the

DOI 10.1002/bies.201300066

Department of Plant Sciences, Weizmann Institute of Science, Israel

Corresponding author:

Ron Milo
E-mail: ron.milo@weizmann.ac.il

Abbreviations:

BNID, BioNumbers website IDentification number; **IBAQ**, intensity based absolute quantification; **aa**, amino acids; **MW**, molecular weight; **kDa**, kilodaltons.

estimating the number of proteins per cell volume

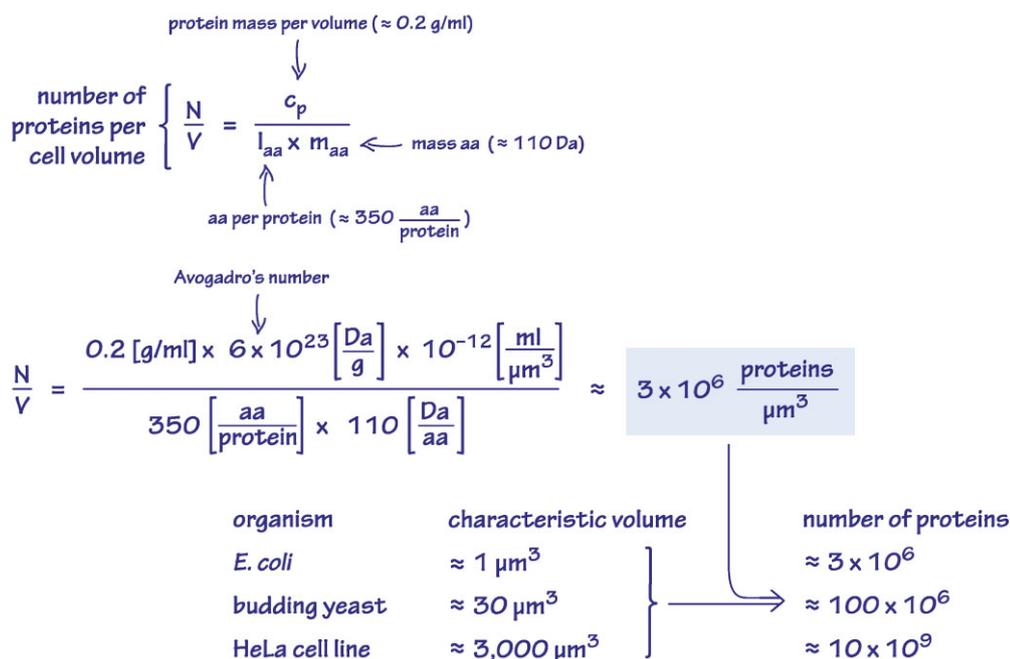


Figure 1. A back of the envelope calculation of the number of proteins per cell volume. Application for selected model organisms based on their characteristic cell volumes is also given. Estimate is based on generic parameter values. For more accurate organism specific values see main text.

absolute values can thus reveal biases in quantification [17]. Recent work showed how values from even the most reliable mass spectrometry methods such as iBAQ [18] should be normalized [19] to give reasonable overall cellular concentrations. While experimental measurements are the foundation of our knowledge of biological systems, estimates serve as predictions that are useful in testing where we might be holding misconceptions about cells or about methods that measure their properties.

We begin by deriving an expression for the number of proteins per cell volume. We then use known parameter values to arrive at concrete predictions for the number of proteins per cell volume. These predictions are contrasted with a survey of the reported values in the literature. It is shown how these values differ markedly, and possible reasons for these seeming discrepancies are discussed. Renormal-

ization factors are given that can be used to reinterpret previously reported protein copy numbers to be consistent with the analysis given here.

Estimating the total number of proteins per cell volume

Protein content scales roughly linearly with cell mass and volume. Given that cell volume can change several fold as a consequence of growth conditions or strain identity, we choose to discuss protein content per unit cell volume. Later, we can multiply by cell volume to arrive at the total protein number per cell under a given growth condition.

Our first method of estimate is shown as a back-of-the-envelope calculation in Fig. 1 using rounded “generic” parameter values. The estimate relies on knowledge of the protein mass per unit volume (denoted by c_p). The units of c_p are [g protein]/[mL cell volume] and this parameter has been reported for different cell types (see e.g. [20] and also BNID 105938, where here and later in this manuscript pointers for directly assessing values from the literature are given using the BioNumbers database

identifiers [21], www.bionumbers.hms.harvard.edu). We denote by l_{aa} the average length, in amino acids, of a protein, and the average mass of an amino acid by m_{aa} . The number of proteins per unit volume is equal to: $N/V = c_p / (l_{aa} \times m_{aa})$.

In *Escherichia coli* and other bacteria, we use an average protein length, l_{aa} , of 300 aa/protein and in budding yeast, fission yeast, and human cells, we adopt the value of 400 aa/protein (where values are rounded to within about 10-20% accuracy). These average lengths were calculated by weighting the lengths of proteins by their abundance in the cell. This takes into account that high abundance proteins tend to be smaller than low abundance proteins. This calculation used published datasets mentioned in Table 1 (see also [22, 23] and BNID 107049, 108983).

Moving on to the protein concentration in the cells, reports are surprisingly scarce: old measured values for c_p are 0.24 g/mL for *E. coli* and 0.28 g/mL for budding yeast (BNID 105938, 108879, 108263, see also [24]). Values are expected to be similar when the concentration values refer to either the total cell volume and protein complement including membrane associated proteins, or solely to cytoplasmic volume

Table 1. Range of values for the number of proteins per cell based on published studies.

Reported proteins per cell	Cell volume (μm^3)	Inferred proteins per volume ($10^6/\mu\text{m}^3$)	Mismatch from calculation	BNID, Ref, Method
<i>M. pneumoniae</i> 0.05×10^6	0.015	3	<2-fold	Kühner et al. [6]. Using cryoEM counting of ribosomes for normalization. Volume calculated based on pear shaped 700 nm \times 300 nm.
<i>L. interrogans</i> ^a $(1.0-1.2) \times 10^6$	0.22	5	<2-fold	Schmidt et al. [7]. Summing table SV. Volume given in Maier et al. [36] citing Beck et al. [37].
<i>E. coli</i> 2.36×10^6	0.86	2.7	<2-fold	100088, Neidhardt & Umbarger, EcoSal Ch. 3, 1996, 40 min doubling time, 0.95 pg cell total weight and assuming cell density of 1.1 (used for normalization by Lu [30]). Original calculation based on average protein MW = 40 kDa but weighing by abundance gives MW \approx 30 kDa that will make value higher by a third.
2.3×10^6	0.7	3.3	<2-fold	Arike et al. [18]. Volume not specified
^a 0.1×10^6		N/A	N/A	Taniguchi et al. [32], cover about 1/4 of genome. Growth at 30°C (based on fluorescent protein).
^a 280×10^6		N/A	N/A	Ishihama et al. [8], they report a massive overestimate in ribosomal protein counts
<i>B. subtilis</i> ^a 2.3×10^6	1.13	2.0	<2-fold	Maass et al. [33], exponential growth, early stationary, and late stationary, respectively. Only cytosolic proteins with isoelectric point at pH 4-7 quantified.
^a 1.3×10^6	0.62	2.1	<2-fold	
^a 1.8×10^6	0.85	2.1	<2-fold	
<i>S. aureus</i> ^a 0.35×10^6	0.33	1.1	\approx 3-fold	Maass et al. [33], exponential growth, early stationary, and late stationary, respectively. Only cytosolic proteins with isoelectric point at pH 4-7 quantified.
^a 0.27×10^6	0.23	1.2	\approx 3-fold	
^a 0.26×10^6	0.23	1.1	\approx 3-fold	
<i>S. cerevisiae</i> (haploid) 50×10^6	\approx 30-40	1-2	\approx 2-fold	106198, Futcher et al. [34], based on 1977 paper measuring 4 fg of protein per cell (used for normalization by Lu [30])
^a 47×10^6				Ghaemmaghmi et al. [31], summing up all measured proteins (based on TAP tag).
53×10^6 (30×10^6 - 80×10^6)				104313, von der Haar [35], the author merged various high throughput measurements.
<i>S. pombe</i> 60.3×10^6	\approx 100	0.6	\approx 5-fold	Marguerat et al. [9]. Mass Spectrometry, exponential growth.
<i>M. musculus</i> (NIH3T3 cells) ^a 3×10^9	\approx 2,000	1.5	<2-fold	Schwanhäusser et al. [10] updated in Nature 2013. SILAC medium. Volume based on BNID 108979.
<i>H. sapiens</i> (U2OS) ^a $(0.95-1.7) \times 10^9$	\approx 4,000	0.2-0.4	\approx 10-fold	Beck et al. [11]. Range stems from the 11 most highly abundant proteins for which the authors could not calibrate accurately and originally reported as >20 million copies per cell. Assuming 20 million gives the lower value and using the original measured higher values gives the upper bound.
<i>H. sapiens</i> (HeLa) ^a 2.0×10^9	\approx 2,000	1	\approx 3-fold	Nagaraj et al. [12].
^a 2.3×10^9	\approx 2,000	1	\approx 3-fold	Finka and Goloubinoff [19]. Analyzing data from Geiger et al. [13].

In some cases the number is inferred from supplementary information and was not reported as such. When cell volume was not reported in the study, literature values under similar conditions were used. Mismatch between values inferred from the literature per unit volume and estimates given in this paper (2-4 million proteins per micron cubed) is calculated.

^a Value for total proteins per cell was not explicitly reported and is based on summing the abundance values as reported in the supplementary material across the proteome.

and proteins). Assuming an average amino acid mass of 110 Da - and with some unit conversions - we arrive at:

$$\left(\frac{N}{V}\right)_{\text{coli}} = \frac{0.24 \text{ g/mL} \times 6 \times 10^{23} \text{ Da/g} \times 10^{-12} \text{ mL}/\mu\text{m}^3}{300 \text{ aa/protein} \times 110 \text{ Da/aa}} \approx 4.4 \times 10^6 \text{ proteins}/\mu\text{m}^3$$

and

$$\left(\frac{N}{V}\right)_{\text{yeast}} = \frac{0.28 \text{ g/mL} \times 6 \times 10^{23} \text{ Da/g} \times 10^{-12} \text{ mL}/\mu\text{m}^3}{400 \text{ aa/protein} \times 110 \text{ Da/aa}} \approx 3.8 \times 10^6 \text{ proteins}/\mu\text{m}^3$$

The reader might be wondering about the accuracy of the value of c_p used. We can derive it based on other, better known, properties: cell density, water content, and protein fraction of dry mass. The total cell density, d , is about 1.1 g/mL (BNID 103875, 102239, 106439). The water content, which we denote by w , is in *E. coli* $\approx 70\%$ and in budding yeast $\approx 60\%$ by mass (BNID 103689). The protein fraction of the dry mass, p , at characteristic exponential growth conditions was measured at $\approx 55\%$ in *E. coli* and $\approx 40\%$ in yeast. The relationship between these quantities is: $c_p = d \times (1 - w) \times p$. Plugging in the parameter values:

$$C_{p,\text{coli}} = 1.1 \text{ g/mL} \times (1 - 0.7) \times 0.55 = 0.19 \text{ g/mL}$$

and

$$C_{p,\text{yeast}} = 1.1 \text{ g/mL} \times (1 - 0.6) \times 0.4 = 0.18 \text{ g/mL}$$

The resulting values are smaller than those quoted above by 20-40%, and they lead to estimates for *E. coli* and budding yeast of, respectively:

$$\left(\frac{N}{V}\right)_{\text{coli}} = \frac{0.19 \text{ g/mL} \times 6 \times 10^{23} \text{ Da/g} \times 10^{-12} \text{ mL}/\mu\text{m}^3}{300 \text{ aa/protein} \times 110 \text{ Da/aa}} \approx 3.5 \times 10^6 \text{ proteins}/\mu\text{m}^3$$

and

$$\left(\frac{N}{V}\right)_{\text{yeast}} = \frac{0.18 \text{ g/mL} \times 6 \times 10^{23} \text{ Da/g} \times 10^{-12} \text{ mL}/\mu\text{m}^3}{400 \text{ aa/protein} \times 110 \text{ Da/aa}} \approx 2.5 \times 10^6 \text{ proteins}/\mu\text{m}^3$$

The estimates for the total number of proteins per cell unit volume are calculated based on the overall cell density, the average protein length, and the overall water content. These parameters refer not merely to the cytoplasmic

volume but are rather a weighted average that also includes the volumes of organelles (including vacuoles),

membranes, the nucleus, and the cell wall. The estimates are therefore representative of the overall average cellular protein content.

For mammalian cells, a value for protein density of $c_p \approx 0.2 \text{ g/mL}$ was reported (BNID 105938, see also [25, 26]) leading to similar estimates as for budding yeast:

$$\left(\frac{N}{V}\right)_{\text{human}} = \frac{0.2 \text{ g/mL} \times 6 \times 10^{23} \text{ Da/g} \times 10^{-12} \text{ mL}/\mu\text{m}^3}{400 \text{ aa/protein} \times 110 \text{ Da/aa}} \approx 2.7 \times 10^6 \text{ proteins}/\mu\text{m}^3$$

Cell type characteristics can lead to variations, but these are not expected to lead to differences of more than twofold except for very unique cell types. An outlier low protein density of 60 mg/mL has been mentioned [19] yet this is not considered a representative value in the literature. The calculation here does not account for secreted proteins, which are expected to be a small fraction in most cell lines (some immune cells might be notable exceptions).

We can now use characteristic volumes to reach the number of proteins per

cell. For an *E. coli* cell of $1 \mu\text{m}^3$ volume (average values often vary between 0.5 and $2 \mu\text{m}^3$ depending on growth rate and conditions), the estimates give a range of 3-4 million proteins per cell. For a haploid budding yeast cell of characteristic volume

$40 \mu\text{m}^3$, the two estimates give a range of 100-150 million proteins per cell. Applying the estimated protein densities to mammalian cell volumes yields a value of about 10^{10} proteins per cell for cell lines with characteristic volumes of 2,000-4,000 μm^3 . Yet because cell volume can change several fold under different growth conditions, it is usually much more accurate to use values per unit volume rather than a total protein count per cell.

Comparison to values reported in the literature and discussion of possible sources of mismatch

How do these values compare to previous reports in the literature? Table 1 shows a compilation of values based on published proteome-wide studies. Notably, in many

cases a total sum over all proteins was not reported and was inferred for our purposes by summing all measured abundances. While many of the total sums are within twofold of the estimates above, some values - most notably for eukaryotic cells including yeast and mammalian cells - are lower than predicted by as much as a factor of 5-10. This is beyond the uncertainty that can be explained by the ranges of the parameters that went into the estimates. What can possibly explain the marked differences between some reported values and the estimates above?

In mass spectrometry measurements, a fraction of the proteome is not reported because the measured values are below the reliable detection limit. Even though the number of such genes can be significant (measuring in thousands), current sensitivity is so good that the contribution to the overall sum is expected to be negligible, amounting to much less than 20% of the overall count. The small quantitative effect of the proteins that are below the detection limit is in line with the fact that the top 1,000 most highly expressed proteins in a cell make up over 80% of the proteome mass as calculated by the author from

the data published in papers referred to in Table 1. In most cells, the fraction of the proteome covered by the top 1,000 proteins is actually over 90% of the protein copies (as well as amino acids), see also [12, 14, 19]. The fraction of unmeasured protein thus cannot explain the differences observed between the reported values and the estimate prediction.

Vacuoles are a significant constituent of eukaryotic cells. In some plant cells, they can often occupy 80-90% of the cell volume (BNID 103442, 104368) and tend to have a lower protein concentration. Can this explain the seeming discrepancy observed here for yeast and mammalian cells? For budding yeast, the vacuole can be >50% of the cell volume under conditions of stress; however, it is generally much smaller - 10-20% of cell volume - under exponential growth with glucose. Moreover, the measured parameter of protein concentration per unit volume, c_p , which was used in the estimates above applies to intact cells, including any vacuole fraction. If there is, indeed, within the vacuole a lower abundance of proteins than in the other parts of the cell, the concentration should be even higher in the other parts of the cell in order to achieve the overall measured concentration. This should not have an effect on the overall estimate for the number of proteins per cell unit volume. Finally, in mammalian cells vacuoles are much more rare. It is therefore hard to see how vacuoles explain the observed mismatches between some reported values for the overall number of proteins per unit volume and the estimates provided here.

In label-based approaches, such as fluorescent tagging, different groups of proteins can be resistant to tagging, and thus not quantified. Ribosomal proteins, which can account for over a fifth of the proteome under fast growth rates [27], often fall into this category. It is also known that there can be major biases in measuring non-cytoplasmic proteins. Mass spectrometry can be limited in measuring membrane-associated proteins. For example, the highly abundant (BNID 100082, [28]) hydrophobic lipoprotein Lpp connected to the outer membrane and peptidoglycan of *E. coli* was not observed in some mass spectrometry studies. The higher num-

ber of protein modifications in eukaryotes can reduce the measured protein abundances. Though these technical biases can cause underestimates of the cell protein count, it is difficult to see how this can account for the major part of the 5-10-fold mismatch shown in Table 1 for fission yeast and a U2OS human cell line. A more "global" scaling factor is probably required based on technical issues in the quantification (or in the calculation above). Such a global factor can be loss of material during cell lysis, tryptic digest, and preparation for mass spectrometry, or any other step that is not adequately corrected via standards. Another explanation can be an erroneous factor applied somewhere in the very elaborate set of calculations used to derive absolute quantification values.

The need to measure cell volumes and key physiological parameters

The values of characteristic cell parameters used here for the calculations of protein content per cell dry mass or per cell volume are mostly based on physiological studies from decades ago. Probably the least robust value is that of the protein content per cell unit volume (usually reported in mg of protein per mL of cell volume). It is worthwhile revisiting these values with the most modern techniques. Yet, it is hard to imagine that a modernized version of the physiology experiments for fission yeast and human cell lines will explain the 5-10-fold difference seen between simple estimates and measured values. If - and that would be surprising - newer measurements will show the values to be so significantly different from those found in previous measurements, this would be an important step forward in our understanding of cell composition.

Another parameter that is suspected is the value used for cell sizes. Measurements of the average cell volume for a population of cells can have accuracy of better than 10%, for example by using the standard Coulter cell counter. Yet in many of the studies on quantitative proteomics there is no concurrent measurement of cell volume reported.

In some cases, there is no value for the cell volume at all and in other cases inference is made based on reported measurements of cell diameters. In these cases, it is possible that the cells whose proteomes were quantified were somehow much smaller than those previously reported in the literature, and whose volume served as the basis for our estimates of the mismatch from the calculations as shown in Table 1. It is known that cell volume can change several fold in response to subtle growth conditions; an example is HeLa cell volume as a function of time since splitting and re-plating [29]. However, we do not see a reason for which such biases in cell size would be consistently in the direction required to explain the mismatch between the reports and the estimate given here. Moreover, this explanation can also be tested by looking at copy numbers of proteins that should not depend on cell size, e.g. histones. A major take-home message of the current analysis is the crucial importance of measuring cell volumes (using the Coulter cell counter, microscopy, or some other method) in any future proteomics study that seeks to report absolute protein copy numbers.

Conclusions and outlook

Mass spectrometry is the most common method for quantification, and is the method used in the studies showing the most marked mismatches with the estimates. The source of the mismatch from the calculation presented here is unclear at this time. Some methods of analysis result in measured values close to the estimate, whereas others deviate significantly. Hence, it seems that a careful comparative analysis of the calibration steps, as well as a direct measurement of the protein mass per cell volume, will clarify the underlying reasons for the mismatch. It may take time - hopefully not too long - to reveal the sources of the seeming discrepancies and report updated values or to re-measure and revise literature reported cell composition properties used as assumptions in the estimate. In the meantime, the information in Table 1 can be used to rescale reported values of protein abundance to fit the estimates for the overall cell

content. The multiplicative scaling factor needed is the ratio between the estimated $2\text{--}4 \times 10^6$ protein/ μm^3 and the value calculated in Table 1, protein per unit volume column. For example, the fission yeast protein counts can be multiplied ≈ 5 -fold (between 2/0.6 and 4/0.6). The need for such “normalization” was recently pointed out [19] after noting that summing protein abundances results in nonrealistic overall protein concentrations. It is advised that researchers performing future proteome-wide quantitative studies compare their summed abundances to the estimates given here as a crude sanity check. Such a simple analysis will ensure consistency across studies of the values attained through the impressive technical achievement of proteome-wide abundance datasets.

Novel high-throughput methods in proteomics hold the promise for a detailed census of the proteome. This study indicates that there remain important challenges for careful calibration in order to achieve definitive answers for those interested in quantitatively mapping the cell’s contents.

Acknowledgements

I am grateful to Amir Aharoni, Liisa Arike, Uri Barenholtz, Arren Bar Even, Andrija Finka, Tamar Geiger, Emmanuel Levy, Gene-Wei Li, Wolfram Liebermeister, Sandra Maass, Tobias Maier, Uri Moran, Elad Noor, Alessandro Ori, Rob Phillips, Noa Rippel, Alex Schmidt, and Matthias Selbach for discussions of the issues detailed in this manuscript. RM is the incumbent of the Anna and Maurice Boukstein Career Development Chair and is supported by the European Research Council (Project SYMPAC - 260392); Israel Science Foundation (Grant 750/09) and the Helmsley Charitable Foundation.

References

1. Sapan CV, Lundblad RL, Price NC. 1999. Colorimetric protein assay techniques. *Bio-technol Appl Biochem* **29**: 99-108.
2. Aitken A, Learmonth M. 2002. Protein determination by UV absorption. *Notes* **205**: 3-6.

3. Pedersen S, Bloch PL, Reeh S, Neidhardt FC. 1978. Patterns of protein synthesis in *E. coli*: a catalog of the amount of 140 individual proteins at different growth rates. *Cell* **14**: 179-90.
4. Cox J, Mann M. 2011. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem* **80**: 273-99.
5. Aebersold R, Mann M. 2003. Mass spectrometry-based proteomics. *Nature* **422**: 198-207.
6. Kühner S, Van Noort V, Betts MJ, Leo-Macias A, et al. 2009. Proteome organization in a genome-reduced bacterium. *Science (New York, NY)* **326**: 1235-40.
7. Schmidt A, Beck M, Malmström J, Lam H, et al. 2011. Absolute quantification of microbial proteomes at different states by directed mass spectrometry. *Mol Syst Biol* **7**: 510.
8. Ishihama Y, Schmidt T, Rappsilber J, Mann M, et al. 2008. Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* **9**: 102.
9. Marguerat S, Schmidt A, Codlin S, Chen W, et al. 2012. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* **151**: 671-83.
10. Schwanhäusser B, Busse D, Li N, Dittmar G, et al. 2011. Global quantification of mammalian gene expression control. *Nature* **473**: 337-42.
11. Beck M, Schmidt A, Malmström J, Claassen M, et al. 2011. The quantitative proteome of a human cell line. *Mol Syst Biol* **7**: 549.
12. Nagaraj N, Wisniewski JR, Geiger T, Cox J, et al. 2011. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* **7**: 548.
13. Geiger T, Wehner A, Schaab C, Cox J, et al. 2012. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics* **11**: M111.014050.
14. Geiger T, Velic A, Macek B, Lundberg E, et al. 2013. Initial quantitative proteomic map of twenty-eight mouse tissues using the SILAC mouse. *Mol Cell Proteomics* **12**: 1709-22.
15. Eldar A, Elowitz MB. 2010. Functional roles for noise in genetic circuits. *Nature* **467**: 167-73.
16. Phillips R, Milo R. 2009. A feeling for the numbers in biology. *Proc Natl Acad Sci USA* **106**: 21465-71.
17. Schwanhäusser B, Busse D, Li N, Dittmar G, et al. 2013. Corrigendum: global quantification of mammalian gene expression control. *Nature* **495**: 126-7.
18. Arike L, Valgepea K, Peil L, Nahku R, et al. 2012. Comparison and applications of label-free absolute proteome quantification methods on *Escherichia coli*. *J Proteomics* **75**: 5437-48.
19. Finka A, Goloubinoff P. 2013. Proteomic data from human cell cultures refine mechanisms of chaperone-mediated protein homeostasis. *Cell Stress Chaperones* **18**: 591-605.
20. Albe KR, Butler MH, Wright BE. 1990. Cellular concentrations of enzymes and their substrates. *J Theor Biol* **143**: 163-95.
21. Milo R, Jorgensen P, Moran U, Weber G, et al. 2010. BioNumbers—the database of key numbers in molecular and cell biology. *Nucl Acids Res* **38**: D750-53.
22. Brocchieri L, Karlin S. 2005. Protein length in eukaryotic and prokaryotic proteomes. *Nucl Acids Res* **33**: 3390-400.
23. Zhang J. 2000. Protein-length distributions for the three domains of life. *Trends Genet* **16**: 107-9.
24. Zimmerman SB, Trach SO. 1991. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J Mol Biol* **222**: 599-620.
25. Ellis RJ. 2001. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol* **11**: 114-9.
26. Wang Z, Shen W, Kotler DP, Heshka S, et al. 2003. Total body protein: a new cellular level mass and distribution prediction model. *Am J Clin Nutr* **78**: 979-84.
27. Dennis PP, Bremer H. 1974. Macromolecular composition during steady-state growth of *Escherichia coli* B-r. *J Bacteriol* **119**: 270-81.
28. Inoye S, Takeishi K, Lee N, DeMartini M, et al. 1976. Lipoprotein from the outer membrane of *Escherichia coli*: purification, paracrystallization, and some properties of its free form. *J Bacteriol* **127**: 555-63.
29. Luciani AM, Rosi A, Matarrese P, Arancia G, et al. 2001. Changes in cell volume and internal sodium concentration in HeLa cells during exponential growth and following lonidamine treatment. *Eur J Cell Biol* **80**: 187-95.
30. Lu P, Vogel C, Wang R, Yao X, et al. 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **25**: 117-24.
31. Ghaemmghami S, Huh W-K, Bower K, Howson RW, et al. 2003. Global analysis of protein expression in yeast. *Nature* **425**: 737-41.
32. Taniguchi Y, Choi PJ, Li G, Chen H, et al. 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**: 533-8.
33. Maass S, Sievers S, Zühlke D, Kuzinski J, et al. 2011. Efficient, global-scale quantification of absolute protein amounts by integration of targeted mass spectrometry and two-dimensional gel-based proteomics. *Anal Chem* **83**: 2677-84.
34. Futcher B, Latter GI, Monardo P, McLaughlin CS, et al. 1999. A sampling of the yeast proteome. *Mol Cell Biol* **19**: 7357-68.
35. Von der Haar T. 2008. A quantitative estimation of the global translational activity in logarithmically growing yeast cells. *BMC Syst Biol* **2**: 87.
36. Maier T, Schmidt A, Güell M, Kühner S, et al. 2011. Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol Syst Biol* **7**: 511.
37. Beck M, Malmström JA, Lange V, Schmidt A, et al. 2009. Visual proteomics of the human pathogen *Leptospira interrogans*. *Nat Methods* **6**: 817-23.