Published in final edited form as:

Nature. 2009 August 6; 460(7256): 762–765. doi:10.1038/nature08184.

Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*

Johan Malmström[1,*], **Martin Beck**[1,*], **Alexander Schmidt**[1,4], **Vinzenz Lange**[1,4], **Eric W. Deutsch**[2], and **Ruedi Aebersold**[1,2,3,4,5]

[1] Institute of Molecular Systems Biology, ETH Zurich (Swiss Federal Institute of Technology), Wolfgang Pauli-Str. 16, CH-8093 Zurich, Switzerland [2] Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103-8904, USA [3] Faculty of Science, University of Zurich, Switzerland [4] Competence Center for Systems Physiology and Metabolic Diseases, Zurich, Switzerland

Abstract

Mass spectrometry based methods for relative proteome quantification have broadly impacted life science research. However, important research directions, particularly those involving mathematical modeling and simulation of biological processes, also critically depend on absolutely quantitative data, i.e. knowledge of the concentration of the expressed proteins as a function of cellular state. Until now, absolute protein concentration measurements of a significant fraction of the proteome (73%) have only been derived from genetically altered *S. cerevisiae* cells 1, a technique that is not directly portable from yeast to other species. In this study we developed and applied a mass spectrometry based strategy to determine the absolute quantity i.e. the average number of protein copies per cell in a cell population, for a significant fraction of the proteome in genetically unperturbed cells. Applying the technology to the human pathogen *Leptospira interrogans*, a spirochete responsible for Leptospirosis 4, we generated an absolute protein abundance scale for 83% of the mass spectrometry detectable proteome, from cells at different states. Taking advantage of the unique cellular dimensions of *L. interrogans*, we used cryo electron tomography (cryoET) morphological measurements to verify at the single cell level the average absolute abundance values of selected proteins determined by mass spectrometry on a population of cells. As the strategy is relatively fast and applicable to any cell type we expect that it will become a cornerstone of quantitative biology and systems biology.

The developed strategy combines three mass spectrometry based proteomic methods, absolute quantification using isotope labeled reference peptides 2, label free quantification

[5] correspondences should be addressed to aebersold@imsb.biol.ethz.ch.
[*] contributed equally.

Supplementary Information: Supplementary Information is linked to the online version of the paper at www.nature.com/nature

Author information: The mass spectrometry data including spectra and the identified peptides and proteins have been deposited into a PeptideAtlas instance found at www.peptideatlas.org. Reprints and permissions information is available at www.nature.com/reprints. The authors have declared no competing financial interest. Correspondence and request for materials should be addressed to R.A (aebersold@imsb.biol.ethz.ch).

and high throughput proteome sequencing by liquid chromatography tandem mass spectrometry 3. In the first step, we used isoelectric focusing by off-gel electrophoresis to fractionate tryptic digests of whole cell protein extracts 5 and high performance LC-MALDI and LC-ESI tandem mass spectrometry using directed precursor ion selection 6 to identify the peptides contained in the respective fractions. Under the selected growth conditions we could identify 2221 proteins, corresponding to 61% of the ORFs predicted from the *L. interrogans* genome 4. From more than 90 LC-MS/MS runs more than 410000 fragment ion spectra (MS/MS spectra) were acquired of which 145703 were assigned to 18303 unique peptides at a peptide false discovery rate (FDR) of less than 1% 7 and the identified peptides and proteins were assembled into a PeptideAtlas instance as previously described www.peptideatlas.org 8 (see supplement for more information).

In the second step, we selected from the PeptideAtlas, 32 peptides corresponding to 19 proteins at different abundance levels determined by the number of matched MS/MS spectra acquired in step 1 above (spectral counts). The absolute abundance levels for the 19 proteins was determined using SRM and heavy stable isotope labeled reference peptides 2 (Table S1). Knowing the number of cells used to generate the sample and the amount of the heavy labeled peptides added, the copy number for the selected proteins could be calculated. The cellular abundance of these anchor proteins ranged from 40 to 15.000 copies per cell (Table S1).

In the third step we used extracted precursor ion intensities for peptides derived from LC-MS maps acquired from trypsinized cell lysates. We then calculated the total protein ion intensity for the respective proteins by using the median intensities from the three most intense peptides matching to a specific protein 9 10 11. The 19 anchor proteins with SRM determined copy numbers then served as calibration points for translating the relative abundance measurements based on extracted peptide precursor ion intensities 3 and spectral counting into absolute abundance measurements. Absolute protein abundance estimates were thereby obtained for 769 proteins using extracted ion intensities and the absolute abundance of an additional 1095 proteins was estimated by means of spectral counting 12 (Table S2). The number of proteins with estimated absolute protein abundance corresponds to 51 % of the ORF's predicted from the *L. interrogans* genome. At a logarithmic scale the extracted ion intensities correlate well with the absolute protein abundance (Figure 1A). The accuracy of the absolute abundance measurement was determined by bootstrapping the extracted protein ion intensities against the SRM values. This statistical analysis allows validating different sets of reference peptides independently, by randomly removing a fraction of the dataset, rebuilding the linear model and estimating the protein concentration of the initially removed data points. Since the real value of these data points is known from the SRM measurements the average error can be estimated by multiple sampling events. The bootstrapping provided an estimated average error rate of 1.8 fold for the extracted ion intensities and ∼3 fold for the spectral counting (Figure 1B, S4). In summary, by using this three step approach relying on three complementary mass spectrometry methods, we generated a proteome map of absolute protein concentrations, for 51% of the ORF's predicted from the *L. interrogans* genome, corresponding to 83% of the proteome observable by deep proteome mapping, with an average error rate of less than 3-fold.

To assess the accuracy of the absolute protein abundance values generated by mass spectrometry, we applied cryo electron tomography (cryoET) as an orthogonal and independent method. The extraordinarily thin cross section of *L. interrogans* cells (100-180 nm) makes them an ideal specimen for cryo electron tomography (cryoET) measurements 13. We assessed the accuracy of the mass spectrometry derived absolute abundance values *in vivo* by benchmarking against distinct morphological features with known subunit composition and structure, using a library of cryo electron tomograms, covering subvolumes of more than 40 individual *L. interrogans* cells. The selected structures used to benchmark the absolute protein abundance estimations have been studied in depth both in vitro and in vivo using biochemical and structural biology techniques and their structure and composition are known. The following features were analyzed:

i. *Flagellar length: L. interrogans* cells contain 2 periplasmic flagella that emanate from flagellar motors at both poles (Figure 1C, red) and protrude towards the middle of the cell. The flagella are obvious features in *L. interrogans* cryo electron tomograms (Figure 1 and S3, blue). FlaB1 the major core component and the most abundant flagellar protein was estimated by mass spectrometry at 12.000 copies per cell. We determined that each cell contained 2000 copies of flagellar protein FlaB2, 300 copies of FlaB3 and 3500 copies of FlaB4. Both flagella combined thus contain an estimated average of 17800 copies of FlaB proteins organized into 11 protofilaments 14 with an inter subunit spacing of ∼52 Å. From these data we calculated an average length of both flagella to 8.4 μm. This value correlates well with the measured flagellar length, based on the following considerations. The average cell length is 11.5 μm (determined from low magnification projection images, data not shown) and the combined length of both flagella is somewhat lower than the average cell length, as is apparent from the tomograms as shown in Fig. S3.

ii. *Flagellar motor:* The structure and protein composition of bacterial flagellar motors has been previously determined in various species 15 16. The flagellar motor is an obvious structure in tomograms of *L. interrogans* (Figure 1, red features). One of its components, FliF, has been shown to occur in 26 copies per motor in *Salmonella typhimurium* 15,16. From the acquired tomograms we conclude that each *L. interrogans* cell has one motor at each pole and that FliF, assuming evolutionary conservation of the subunit composition, is expected to occur in 52 copies per cell. For this low abundant protein the average FliF copy number measured by mass spectrometry was 43, a value that is in agreement with the expected value within the estimated accuracy of the method.

iii. *Methyl-accepting protein receptors (MCPs):* MCP's function in metabolite sensing 17, forming arrays that are straight forward to discern by cryoET 18 and localize to the proximity of the flagellar motor in *L. interrogans* (Figure 1C, green features). A receptor unit cell contains 6 molecules of MCPs that occupy an area of ∼50 nm^2 19. By quantitative mass spectrometry we found 6000 MCPs per cell which translates into an estimated area of about 50 000 nm^2 per cell occupied by the receptors. This estimate is in agreement with the area occupied by receptor arrays

in the tomograms which was determined to occupy 40 000 nm^2 (200 nm * 100 nm per pole).

iv. *Total cellular protein:* Based on the proteome wide absolute abundance data and the cell volume measurements by cryoET we determined the total cellular protein concentration to ∼250 mg/ml by summing all protein copies and dividing by the volume. A total cellular protein concentration of ∼250 mg/ml is in agreement with earlier studies in *E.Coli* 20.

In summary, based on the validation of the quantitative mass spectrometry data with an orthogonal method we can confirm the estimated accuracy of the determined absolute abundance protein scale. The dynamic range of absolute abundance scale spans minimally 3 orders of magnitude from 40000 copies per cell for protein LipL32 to single digit protein copies for low abundant proteins (Table S2). The mass spectrometric approach described here is generic, while the validation of the data by cryoET is dependent on the cellular dimensions and therefore not generic.

After the original proteome map of a bacterial species has been established, absolute protein quantification in repeat measurements of differentially perturbed cells is possible and straight forward. We investigated the quantitative changes in the *L. interrogans* proteome caused by treatment with Ciprofloxacin, an antibiotic that inhibits DNA-gyrase 21. *L. interrogans* cells were counted and harvested at three different time points of treatment, 3h, 24 h and 48h, in five independent biological experiments, the proteins were extracted and digested followed by mass spectrometric analysis. On average around 1000 proteins per state could be identified and quantified with absolute protein concentrations. Within the set of observed proteins more than 200 proteins changed their protein abundance more than two fold (Table S2).

In contrast to typical quantitative proteomics investigations where relative abundance levels are obtained that support the comparison of protein abundance between the same protein across samples, this approach allows comparisons between different proteins across samples. This is exemplified in Figure 2, where the identified proteins were grouped into biological functions using GeneOntology (GO) 22. By taking into account the protein copy numbers determined in this study and the protein length, the fraction of the cellular protein synthesis budget associated with a particular cellular function could be calculated. The data indicate a significant discrepancy between the number of ORFs and protein molecules within a certain GO group. More than 40% of the ORF's (30% of the identified proteins) in *L. interrogans* are hypothetical, and associated with unknown biological processes. By summing up the total copy number of all proteins belonging to this group and multiplying with the protein length of these proteins and comparing it to the total number of cellular proteins and their protein length, we estimate that the hypothetical proteins constitute only 12.7% of the total cellular protein synthesis budget (Figure 2, blue). This indicates that the hypothetical proteins generally are low abundant. *L. interrogans* cells invest a large fraction of their cellular protein synthesis budget on the processes protein synthesis and folding, electron transport, cell motility and especially the proteins of the external encapsulating structure. In contrast, chemotaxis, even though it involves a large number of genes, has only a moderate impact on the protein synthesis budget. It is noteworthy that the cell invests a

large proportion of the total protein synthesis budget to maintain a relatively small group of proteins at a very high cellular concentration as in case of the proteins of the external encapsulating structure (Figure 2, bright green), suggesting that the functions carried out by these proteins are critical for the cell (for more information see Supplement material).

The ability to detect the absolute concentration of a significant fraction of the proteome allowed us to also examine how the proteome as a whole compensated for the changes of expression of specific proteins. As an example, upon exposure to Ciprofloxacin the cells reacted to the DNA-topo-isomeric stress by a fifteen-fold up-regulation of recombinase A (recA), a measurement that is consistent with literature values 23 24. The increase in recA was accompanied by an enormous increase of 15 hypothetical proteins that comprised approximately 20% of the entire proteome after Ciprofloxacin treatment (Figure 2). Interestingly, this large redistribution of the proteome did not significantly change the total cellular protein concentration. Therefore, the large increase in the abundance of the group of previously hypothetical proteins upon ciprofloxacin exposure was compensated by a slight reduction of other high abundant protein classes, such as the ribosomal proteins and proteins involved in nucleotide binding (Figure 2, red). This indicates that in *L. interrogans*, the cells strive to maintain a certain total number of protein components, i.e a constant cellular proteome concentration.

In this study we describe an integrated mass spectrometric technique for the determination of the absolute concentration of proteins representing a significant fraction of the proteome of genetically unperturbed microbial cells. The technique was applied to the proteome of *L. interrogans*, a microbial species with a sequenced genome containing 3658 predicted ORF's 4,25. Out of the 2221 identified proteins, 1864 proteins, representing 51 % of the predicted proteome, were provided with estimated copy per cell numbers. The cellular protein concentrations estimated by the technique were assessed by bootstrapping and by the orthogonal method, cryoET tomography which allowed us to detect quantify and localize specific protein complexes in single *L. interrogans* cells in near life state. The cryoET tomography measurements were possible on whole *L. interrogans cells* due to the extraordinarily high cytoplasmic contrast and the unusual dimensions of the cells. The validated method was applied to study the reorganization of the *L. interrogans* proteome upon exposure of the antibiotics Ciprofloxacin. The data indicate that the cells react by expressing massive amounts of a small number of normally unexpressed proteins of unknown function while keeping the total cellular protein constant. The described technique is fast, efficient and can be applied to various biological systems of low and medium complexity in future studies.

Methods Summary

Please see the supplementary material for a detailed methods description. In brief, *Leptospira interrogans serovar Copenhageni* cells of the strain Fiocruz L1-130 were cultivated as previously described 26 and perturbed for 24h with 5 μg/ml Ciprofloxacin (antibiotic treatment). For establishing the proteome map, peptides from a trypsinized full cell lysates were separated by off-gel electrophoresis using twice the 3-10 *pI* range and once the 3-7 range. After iso-electric focusing the fractions were subjected to LC-MS/MS as

previously described 5,27. MS/MS spectra were searched using SEQUEST against the predicted proteome from *Leptospira interrogans serovar Copenhageni str*, complete genome (NC_005823, NC_005824). The data were integrated into PeptideAtlas as previously described 8 and are available for browsing and downloading at http://www.peptideatlas.org/. To determine the absolute protein quantities, the cells were counted and harvested through centrifugation. Upon cell lysis the proteins were digested with trypsin followed by C18 reversed-phase peptide clean up. Peptide quantification was determined via SRM as previously described 28, based on heavy labeled reference peptides. For determination of MS1 intensities, samples were analyzed with an hybrid LTQ-FT-ICR mass spectrometer operated as previously described 6 and MS1-based peak extraction and alignment into MasterMaps was done using SuperHirn 3. Cells from stimulated or non-stimulated cultures, respectively, were pipetted onto Quantifoil R2/1, 200 mesh copper grids (Plano), rapidly plunge frozen into liquid ethane29 and then introduced into a Technai F20 cryo electron microscope. Tomograms were acquired and reconstructed as previously described 30.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project has been funded in part by ETH Zurich, the Swiss National Science Foundation (grant # 31000-10767), Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health (contract No. N01-HV-28179), by SystemsX.ch the Swiss initiative for systems biology, in part by the PROSPECTS (proteomics in time and space) European network of excellence, and with funds from the ERC project “Proteomics V3.0” for R.A. J.A.M. was supported by a fellowship from the Swedish society for medical research (SSMF), M.B. was supported by a long-term fellowship of the European Molecular Biology Organization and a Marie Curie fellowship of the European Commission, A.S. and V.L. were supported by the Competence Center for Systems Physiology and Metabolic Diseases. We thank O. Medalia and the electron microscopy facility of ETH Zurich (EMEZ) for support and David A. Haake for critical reading of the manuscript.

References

1. Ghaemmaghami S, et al. Global analysis of protein expression in yeast. Nature. 2003; 425(6959): 737. [PubMed: 14562106]
2. Gerber SA, et al. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A. 2003; 100(12):6940. [PubMed: 12771378]
3. Mueller LN, et al. SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling. Proteomics. 2007; 7(19):3470. [PubMed: 17726677]
4. Ren SX, et al. Unique physiological and pathogenic features of Leptospira interrogans revealed by whole-genome sequencing. Nature. 2003; 422(6934):888. [PubMed: 12712204]
5. Malmstrom J, et al. Optimized peptide separation and identification for mass spectrometry based proteomics via free-flow electrophoresis. Journal of Proteome Research. 2006; 5(9):2241. [PubMed: 16944936]
6. Schmidt A, et al. An Integrated, Directed Mass Spectrometric Approach for In-depth Characterization of Complex Peptide Mixtures. Mol Cell Proteomics. 2008; 7(11):2138. [PubMed: 18511481]
7. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Analytical chemistry. 2002; 74(20):5383. [PubMed: 12403597]

8. Desiere F, et al. The PeptideAtlas project. Nucleic acids research. 2006; 34(Database issue):D655. [PubMed: 16381952]
9. Silva JC, et al. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics. 2006; 5(1):144. [PubMed: 16219938]
10. Vogel C, Marcotte EM. Calculating absolute and relative protein abundance from mass spectrometry-based protein expression data. Nat Protoc. 2008; 3(9):1444. [PubMed: 18772871]
11. Ishihama Y, et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics. 2005; 4(9):1265. [PubMed: 15958392]
12. Lu P, et al. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol. 2007; 25(1):117. [PubMed: 17187058]
13. Lucic V, Forster F, Baumeister W. Structural studies by electron tomography: from cells to molecules. Annual review of biochemistry. 2005; 74:833.
14. Samatey FA, et al. Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. Nature. 2001; 410(6826):331. [PubMed: 11268201]
15. Jones CJ, Macnab RM, Okino H, Aizawa S. Stoichiometric analysis of the flagellar hook-(basal-body) complex of Salmonella typhimurium. Journal of molecular biology. 1990; 212(2):377. [PubMed: 2181149]
16. Sosinsky GE, et al. Mass determination and estimation of subunit stoichiometry of the bacterial hook-basal body flagellar complex of Salmonella typhimurium by scanning transmission electron microscopy. Proc Natl Acad Sci U S A. 1992; 89(11):4801. [PubMed: 1594581]
17. Charon NW, Goldstein SF. Genetics of motility and chemotaxis of a fascinating group of bacteria: the spirochetes. Annu Rev Genet. 2002; 36:47. [PubMed: 12429686]
18. Briegel A, et al. Location and architecture of the Caulobacter crescentus chemoreceptor array. Molecular microbiology. 2008; 69(1):30. [PubMed: 18363791]
19. Briegel, Ariane, et al. Location and architecture of the Caulobacter crescentus chemoreceptor array. 2008 in press.
20. Elowitz MB, et al. Protein mobility in the cytoplasm of Escherichia coli. J Bacteriol. 1999; 181(1): 197. [PubMed: 9864330]
21. Shalit I, Barnea A, Shahar A. Efficacy of ciprofloxacin against Leptospira interrogans serogroup icterohaemorrhagiae. Antimicrobial agents and chemotherapy. 1989; 33(5):788. [PubMed: 2751290]
22. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000; 25(1):25. [PubMed: 10802651]
23. Dwyer DJ, Kohanski MA, Hayete B, Collins JJ. Gyrase inhibitors induce an oxidative damage cellular death pathway in Escherichia coli. Molecular systems biology. 2007; 3:91. [PubMed: 17353933]
24. Reckinger AR, Jeong KS, Khodursky AB, Hiasa H. RecA can stimulate the relaxation activity of topoisomerase I: Molecular basis of topoisomerase-mediated genome-wide transcriptional responses in Escherichia coli. Nucleic acids research. 2007; 35(1):79. [PubMed: 17151069]
25. Nascimento AL, et al. Comparative genomics of two Leptospira interrogans serovars reveals novel insights into physiology and pathogenesis. J Bacteriol. 2004; 186(7):2164. [PubMed: 15028702]
26. Haake DA, et al. Changes in the surface of Leptospira interrogans serovar grippotyphosa during in vitro cultivation. Infect Immun. 1991; 59(3):1131. [PubMed: 1997416]
27. Yi EC, Lee H, Aebersold R, Goodlett DR. A microcapillary trap cartridge-microcapillary high-performance liquid chromatography electrospray ionization emitter device capable of peptide tandem mass spectrometry at the attomole level on an ion trap mass spectrometer with automated routine operation. Rapid Communications in Mass Spectrometry. 2003; 17(18):2093. [PubMed: 12955739]
28. Lange V, et al. Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring. Mol Cell Proteomics. 2008
29. Dubochet J, et al. Cryo-electron microscopy of vitrified specimens. Q Rev Biophys. 1988; 21(2): 129. [PubMed: 3043536]

30. Beck M, et al. Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. Nature. 2007; 449(7162):611. [PubMed: 17851530]

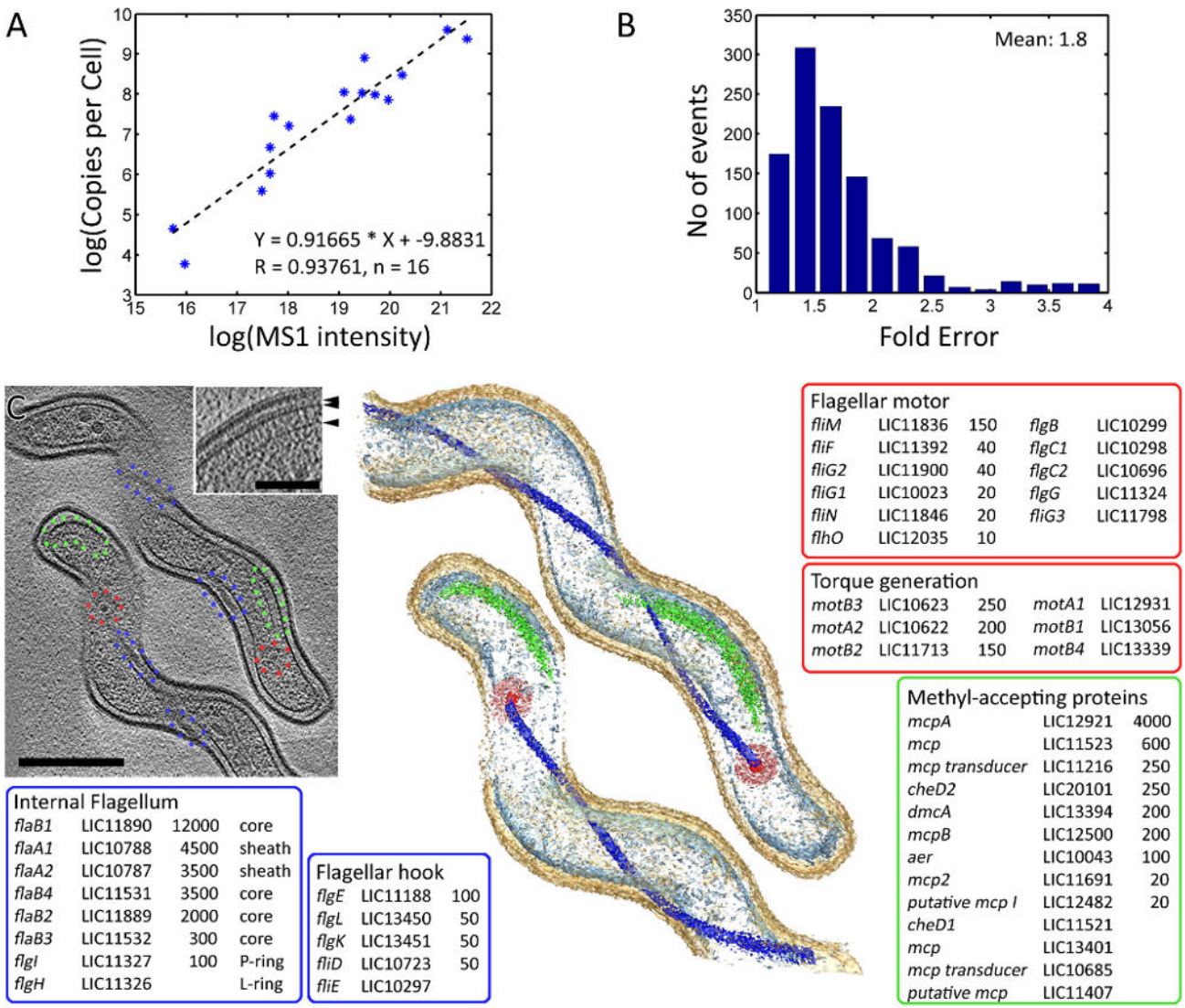

Figure 1. Large scale determination of cellular protein concentrations

A) Natural logarithm of extracted precursor ion intensities plotted against the natural logarithm of copies/cell for 16 proteins quantified by SRM. B) Distribution of error rates determined by bootstrapping. C) Slice through tomographic reconstruction, substructures are marked color-coded as described below (scale bar 200 nm). The inset shows a close-up of methyl-accepting proteins (scale bar 100 nm). The boxes display the gene products making up the different components of methyl-accepting proteins (green), periplasmic flagella (dark blue), the flagellar stator (transparent red) and rotor (dark red).

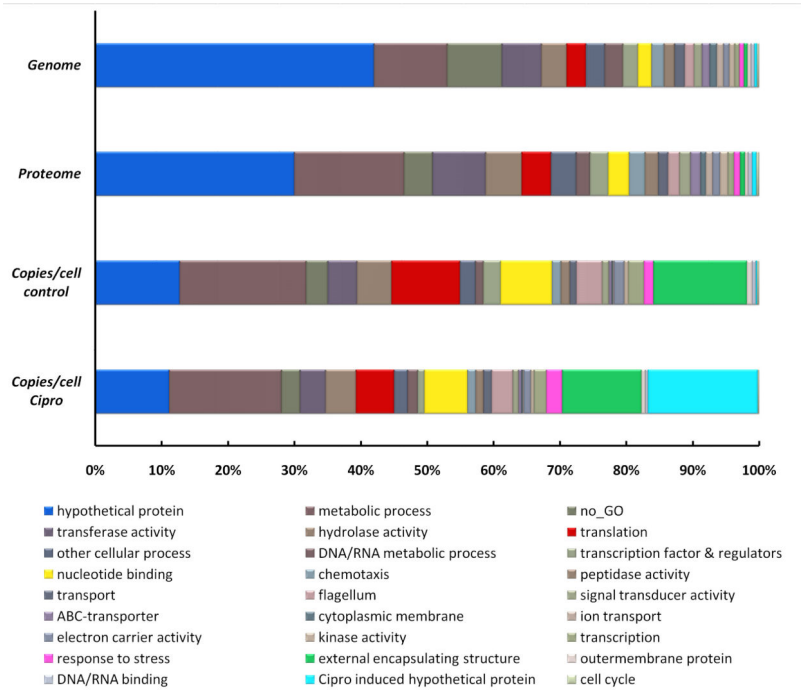

Figure 2. Abundance levels of selected protein groups (by GeneOntology)

The number of genes, number of identified proteins and copies per cell for two cellular states (control and ciprofloxacin treated). Please note that hypothetical proteins are underrepresented in the copies per cell calculation as compared to gene numbers, while members of the protein folding, encapsulating structure and electron transport group are largely overrepresented.