



HAL
open science

Differential quantitative proteome analysis of Escherichia coli grown on acetate versus glucose

Christian Treitz, Brice Enjalbert, Jean-Charles Portais, Fabien Letisse, Andreas Tholey

► **To cite this version:**

Christian Treitz, Brice Enjalbert, Jean-Charles Portais, Fabien Letisse, Andreas Tholey. Differential quantitative proteome analysis of Escherichia coli grown on acetate versus glucose. *Proteomics*, 2016, 16 (21), pp.2742-2746. <10.1002/pmic.201600303>. <hal-01601640>

HAL Id: hal-01601640

<https://hal.science/hal-01601640v1>

Submitted on 5 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

DATASET BRIEF

Differential quantitative proteome analysis of *Escherichia coli* grown on acetate versus glucose

Christian Treitz¹, Brice Enjalbert², Jean-Charles Portais², Fabien Letisse² and Andreas Tholey¹

¹ Systematische Proteomforschung & Bioanalytik, Institut für Experimentelle Medizin, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

² LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

Relative protein abundances of *Escherichia coli* MG1655 growing exponentially on minimal medium with acetate or glucose as the sole carbon source were investigated in a quantitative shotgun proteome analysis with TMT6-plex isobaric tags. Peptides were separated by high resolution high/low pH 2D-LC, using an optimized fraction pooling scheme followed by mass spectrometric analysis. Quantitative data were acquired for 2099 proteins covering 49% of the predicted *E. coli* proteins, showing system-wide effects of growth conditions. In total, 507 proteins showed a fold change of at least 1.5 and 205 proteins changed by more than twofold. Significant differences in abundance were observed for most of the proteins in the central carbon metabolism and in proteins relevant for amino acid and protein synthesis, processing of environmental information and scavenging of a variety of alternate carbon sources. Periplasmic-binding proteins were also more abundant on acetate, especially proteins involved in scavenging extracellular resources such as sugars. All MS data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (dataset identifier PXD003863).

Received: July 19, 2016
Revised: August 26, 2016
Accepted: September 2, 2016

Keywords:

2D-LC MS / Isobaric labeling / Metabolic network / Systems biology



Additional supporting information may be found in the online version of this article at the publisher's web-site

The gram negative bacterium *Escherichia coli* is one of the best studied model organisms. This bacterium is of immense relevance for biotechnological and biomedical research. The knowledge of the influence of environmental factors, e.g. nutrients and other growth conditions is important to understand the regulation of the organism's behavior at molecular level.

Growth on acetate or glucose as carbon sources represents two distinct metabolic states for *E. coli*. Glucose is a classic substrate for biotechnological applications. This high yield carbon source promotes fast growth and is mainly metabolized through the glycolytic and TCA pathways. The growth

of *E. coli* on acetate is important for the survival of the bacterium in the gastrointestinal tract of endotherms [1]. Acetate is a byproduct of glucose-based growth and a less-efficient carbon source whose consumption requires gluconeogenesis and shunting the TCA cycle through the glyoxylate cycle [2].

The metabolic organization of *E. coli* on glucose or on acetate has been analyzed in a number of “omics” studies. Transcriptome datasets are available for growth on acetate relatively to glucose on defined minimal medium [3, 4]. Metabolome and transcriptome data were used to describe the dynamics of the molecular processes involved in glucose-acetate transition [5]. *Escherichia coli* W3110 cells, grown in the presence or absence of acetate, were previously examined by gel-based differential proteomics [6]; 37 and 17 proteins were found to be increased or decreased, respectively, in abundance in the presence of acetate. Tryptic peptides derived from *E. coli* K-12 ATCC-10798 cells grown on M9 media with either lactose, glucose, or acetate were used

Correspondence: Dr. Andreas Tholey, Systematische Proteomforschung & Bioanalytik, Institut für Experimentelle Medizin, Christian-Albrechts-Universität zu Kiel, 24105 Kiel, Germany

E-mail: a.tholey@iem.uni-kiel.de

Fax: +49 431 597 1887

Abbreviations: EF, enrichment factor; PSM, peptide spectrum matches

Colour Online: See the article online to view Figs. 1 and 2 in colour.

as reference samples for LC-MS method development [7]. However, the emphasis of this previous study was on the analytical approach, rather than the quantitative data or proteome coverage. More recently, a comprehensive label-free quantitative *E. coli* proteome dataset across 22 conditions, including minimal growth medium with glucose and acetate as carbon source, was published for *E. coli* strain BW25113, incidentally including data for the K-12 strain MG1655 on glucose but not on acetate [8].

Here, we performed a quantitative proteome analysis using an isobaric labeling/2D-LC approach described earlier [9] to provide a reference dataset showing the protein level differences of *E. coli* K-12 strain MG1655 grown on acetate versus glucose.

Escherichia coli K-12 MG1655 strain was cultured to mid-exponential phase in triplicate in M9 minimal medium containing either glucose (15 mM) or acetate (45 mM; details: Supporting Information). After proteome isolation, protein reduction/alkylation and tryptic digestion, the six samples, each labeled with one TMT reagent, were pooled and separated by 2D-LC, employing RPLC at pH 10 in the first dimension, and ion pair reverse phase LC at pH 2 in the second dimension. In the first dimension, 50 fractions were collected. To maximize usage of chromatographic space and minimize measurement time in the second dimension, a fraction pooling scheme (Supporting Information Fig. 1) was used, leading to ten pooled fractions. Second dimension separation was coupled on-line to an *LTQ Orbitrap Velos* MS (Thermo, Bremen, Germany). Experimental details of LC and MS are provided in the Supporting Information. Each of the ten pooled fractions was analyzed in four replicates, with two additional runs used to optimize acquisition parameters. The first two technical replicates were processed and, after database search, used to compile an exclusion list of *m/z* values with high confidence peptides. Corresponding *m/z* values were excluded from data-dependent MS/MS acquisition for the third and fourth replicate to reduce redundant scan events and increase protein coverage. The final dataset contained 42 LC-MS/MS raw files (ProteomeXchange Identifier PXD003863).

Recalibrated raw files were subjected to multiple database search using Proteome Discoverer Software (v1.4). MASCOT (v2.1), SequestHT, and MS Amanda search algorithms were used for database search in *E. coli* K-12 reference proteome dataset including common contaminants. Detailed search parameters are listed in Supporting Information. Percolator (v1.2) was used for error calculation combining the results of the database searches with a false discovery rate restricted to 0.01 by *q*-value [10]. Only unique peptides were used for quantification. TMT ratios of peptide spectrum matches (PSM) were median normalized. To minimize ratio compression and underestimation, reporter ion signals of PSM were corrected for signal interference and filtered for signal interference and relative precursor intensity as described previously [11] (see Supporting Information). TMT ratios of protein and peptide groups were

calculated as the median TMT ratio of the corresponding PSM.

Statistical analysis was performed with the Perseus software (www.maxquant.org). Log₂ transformed TMT ratios (acetate/glucose “A/G”) were tested against log₂ control TMT ratios using Welch’s *t*-test (ratios between biological replicates of the same treatment: A/A and G/G). Resulting *p*-values were adjusted for multiple testing by permutation-based FDR calculation. GO and KEGG annotations were assigned from UniProt or Ecocyc [12]. Perseus software was used for enrichment/depletion of GO and KEGG categories by Fisher’s exact test and 1D annotation enrichment [13] (see Supporting Information).

In 42 LC-MS runs with 722 764 MS/MS spectra 19 105 unique peptide sequences could be identified with an FDR < 1%, matching to 3123 *E. coli* proteins. Enforcing protein grouping by strictest parsimony principles filtered the identifications to 2590 *E. coli* protein groups of which 2249 were identified by multiple peptides. Relative quantification was possible for 2462 protein groups using only proteotypic (unique) peptides which were reduced to 2309 by filtering PSM to minimize ratio compression. For statistical analysis, we only considered 2099 quantified protein groups (Supporting Information Table 1) identified by multiple peptides (identifications based on an average of 11 and a median of seven peptides). More than 50% of the quantifications are supported by data from at least five unique peptides (Fig. 1A).

A noticeable difference between *E. coli* cells grown on acetate versus glucose was the overall protein content. The yield of protein extraction was much lower in cells grown on acetate versus glucose. From results for BCA protein determination and the median of TMT ratios, an overall A/G protein ratio of 0.61 was calculated, relating to equal amounts of OD = 1 culture. The combination of different volumes of the protein extracts and subsequent data normalization consequently lead to an accumulated shift of A/G ratios (Fig. 1B).

The dataset showed remarkable coverage of metabolic pathways, whereas significantly fewer proteins responsible for membrane transport or involved in other distinct cellular processes including transposition, cell adhesion, and DNA recombination were represented. In terms of localization, proteins from cytoplasm ($q < 1 \times 10^{-45}$; enrichment factor (EF) = 1.53), the cytosol ($q < 1 \times 10^{-30}$; EF = 1.86) as well as proteins located in the periplasmic space ($q < 1 \times 10^{-7}$; EF = 1.38) were enriched, whereas membrane proteins were generally depleted ($q < 1 \times 10^{-12}$; EF = 0.82; Supporting Information Table 2).

Using Welch’s *t*-test and a fold change cut-off of 1.5 for TMT ratios, we identified 507 differentially abundant proteins of which 221 proteins were lower and 286 higher abundant in acetate medium (Fig. 1C).

Relative quantification was possible for a near-complete set of the proteins associated with central carbon metabolism (Fig. 2). Proteins with increased abundance on acetate were significantly enriched in proteins of the TCA cycle

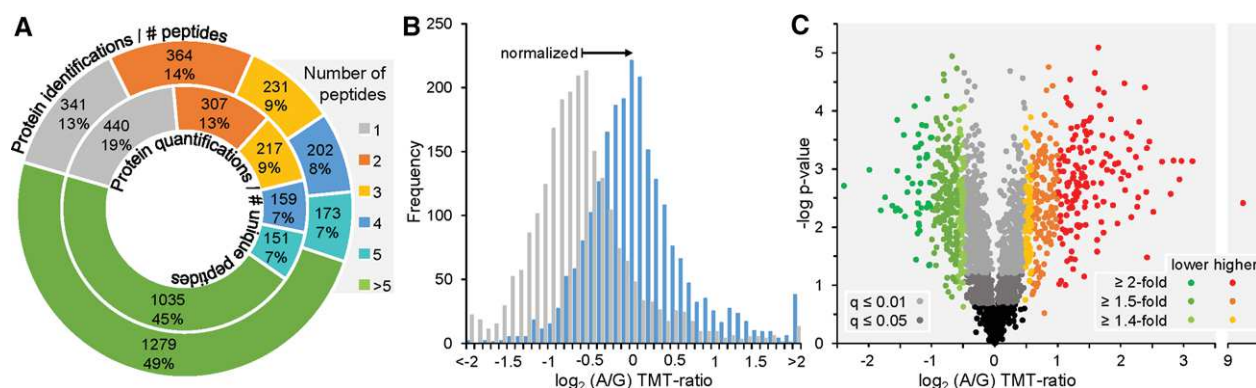


Figure 1. Data assessment. (A) Peptide evidences per protein group identification (outer) and number of unique peptides supporting protein group quantification (inner circle). (B) Data before (gray) and after normalization (blue). (C) Volcano plot showing relative abundances of the selected 2099 *E. coli* proteins between cells grown on acetate and glucose.

($q < 1 \times 10^{-5}$; EF = 4.59). As expected, the glyoxylate cycle proteins were more abundant on acetate (isocitrate lyase AceA and the two malate synthase isozymes AceB and GlcB). The isocitrate dehydrogenase kinase/phosphatase AceK regulates the node between TCA and glyoxylate cycles. While AceK showed a higher abundance on acetate (1.8-fold), a significant change of its target isocitrate dehydrogenase was detected exclusively for peptides containing the phosphorylation site (Supporting Information Fig. 2). The glycolytic proteins PfkA, Ppc, and PykF were more abundant on glucose while their gluconeogenic counterparts Fbp, PckA, and Pps were more abundant on acetate. In a previous works, the *fbp* gene was surprisingly reported with the same level of expression on acetate and glucose [4, 14]. We therefore conclude that this gene is regulated at the translational level. Finally, the acetyl-CoA synthetase protein Acs, cation/acetate symporter ActP, and inner membrane proteins YjcH, all belonging to the same operon and related to acetate utilization, were among the proteins with the highest relative (A/G) abundance.

Beside central carbon metabolism, significant changes were observed in other functional groups. Enzymes involved in fatty acid degradation were much more abundant on acetate while proteins necessary for fatty acid biosynthesis were not significantly affected (Supporting Information Fig. 3). For the translation machinery, virtually all ribosomal polypeptides (with the exception of YkgO and YkgM) were observed in lower levels on acetate ($q < 1 \times 10^{-16}$). This includes a reduction of the majority of enzymes involved in amino acid synthesis pathways, whereas amino acid degradation was reinforced (Supporting Information Fig. 4). A significant shift toward higher abundance in acetate was also observed for proteins involved in flagellar assembly ($q < 1 \times 10^{-4}$). This correlated with a previous observation stating that *E. coli* motility increase as a long-term adaptation to acetate in complex medium [15].

Periplasmic-binding proteins were also more abundant on acetate, in particular proteins involved in scavenging extracellular resources such as sugars (galactose, ribose, maltose, xy-

lose, arabinose), amino acids and peptides, vitamins and ions (Supporting Information Fig. 5). Phosphotransferase systems allowing the transport of sugars were more abundant on acetate for galactitol, cellobiose, and *N*-acetyl-D-glucosamine, while phosphotransferase systems for glucose, mannose, and fructose were less abundant.

Finally, the correlation between gene expression at the transcript and protein level was assessed by comparison with transcriptomic data [4]. Of 2099 quantified proteins in our study, data of 2085 entries could be compared (Supporting Information Fig. 6; Fig. 2 for central carbon metabolism). Proteomic and transcriptomic data showed a moderate linear relationship (Pearson's $r = 0.51$; Spearman's $r_s = 0.54$). This highlights a determinant role of transcriptional control in programs for acetate or glucose consumption. A comparison of our dataset with that recently provided by Schmidt et al. [8], using a related *E. coli* strain showed a high congruency for the majority of proteins (Supporting Information Fig. 7). However, also significant differences could be observed. For example, proteins associated with locomotion and the glyoxylate bypass showed different trends in both datasets.

To conclude, this work provides a comprehensive quantitative dataset of *E. coli* proteins on different media delivering important information to understand metabolic behavior, both in terms of organization and of regulatory processes. Our dataset compares biological situations similar to that described in other proteomics studies [8]; however, as different *E. coli* strains and different analytical platforms (label-free quantification vs. Isobaric labeling) were used, the datasets provide complementary information which can serve as a useful source for future comparative studies both in terms of technology and biology.

This work was supported by the DFG Cluster of Excellence "Inflammation at Interfaces," (CL-X) and the SFB1182 (project A1).

The authors have declared no conflict of interest.

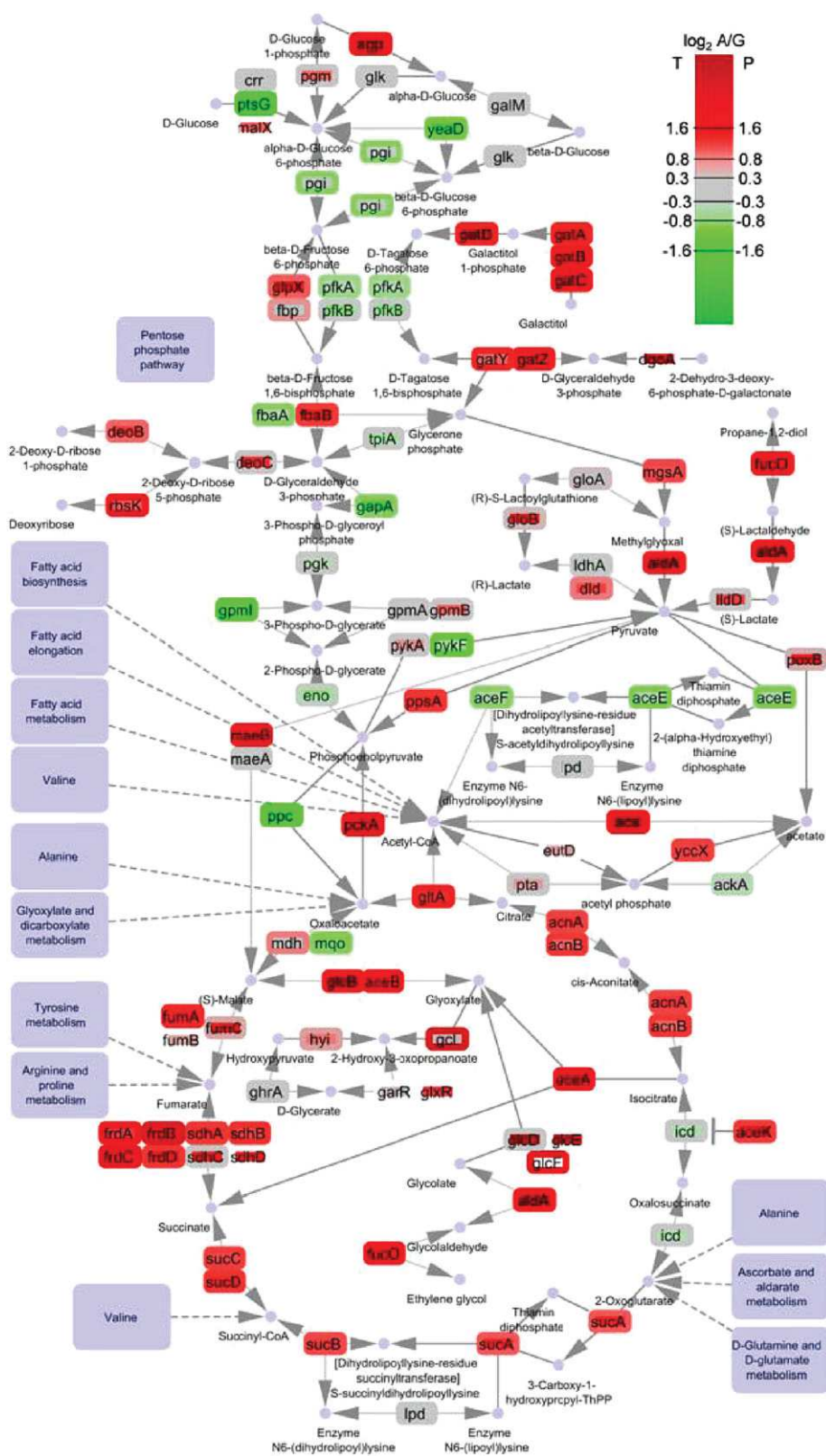


Figure 2. Central carbon metabolism. Proteins are represented by encoding genes, node colors indicate relative expression values log₂ acetate/glucose at protein (outside) and transcript level [4] inside.

References

- [1] Wei, B., Shin, S., LaPorte, D., Wolfe, A. J., Romeo, T., Global regulatory mutations in *csrA* and *rpoS* cause severe central carbon stress in *Escherichia coli* in the presence of acetate. *J. Bacteriol.* 2000, *182*, 1632–1640.
- [2] Clark, D. P., Cronan, J. E., Two-carbon compounds and fatty acids as carbon sources. *EcoSal Plus* 2005, doi: 10.1128/ecosalplus.3.4.4.
- [3] Oh, M. K., Rohlin, L., Kao, K. C., Liao, J. C., Global expression profiling of acetate-grown *Escherichia coli*. *J. Biol. Chem.* 2002, *277*, 13175–13183.
- [4] Liu, M., Durfee, T., Cabrera, J. E., Zhao, K. et al., Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*. *J. Biol. Chem.* 2005, *280*, 15921–15927.
- [5] Enjalbert, B., Letisse, F., Portais, J. C., Physiological and molecular timing of the glucose to acetate transition in *Escherichia coli*. *Metabolites* 2013, *3*, 820–837.
- [6] Kirkpatrick, C., Maurer, L. M., Oyelakin, N. E., Yoncheva, Y. N. et al., Acetate and formate stress: opposite responses in the proteome of *Escherichia coli*. *J. Bacteriol.* 2001, *183*, 6466–6477.
- [7] Silva, J. C., Denny, R., Dorschel, C., Gorenstein, M. V. et al., Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome—a sweet tale. *Mol. Cell. Proteomics* 2006, *5*, 589–607.
- [8] Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrne, E. et al., The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* 2016, *34*, 104–110.
- [9] Treitz, C., Cassidy, L., Hockendorf, A., Leippe, M., Tholey, A., Quantitative proteome analysis of *Caenorhabditis elegans* upon exposure to nematicidal *Bacillus thuringiensis*. *J. Proteomics* 2015, *113*, 337–350.
- [10] Kall, L., Canterbury, J. D., Weston, J., Noble, W. S., MacCoss, M. J., Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* 2007, *4*, 923–925.
- [11] Savitski, M. M., Mathieson, T., Zinn, N., Sweetman, G. et al., Measuring and managing ratio compression for accurate iTRAQ/TMT quantification. *J. Proteome Res.* 2013, *12*, 3586–3598.
- [12] Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A. et al., EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* 2013, *41*, D605–D612.
- [13] Cox, J., Mann, M., 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* 2012, *13*(Suppl 16), S12.
- [14] Enjalbert, B., Cocaign-Bousquet, M., Portais, J. C., Letisse, F., Acetate exposure determines the diauxic behavior of *Escherichia coli* during the glucose-acetate transition. *J. Bacteriol.* 2015, *197*, 3173–3181.
- [15] Polen, T., Rittmann, D., Wendisch, V. F., Sahn, H., DNA microarray analyses of the long-term adaptive response of *Escherichia coli* to acetate and propionate. *Appl. Environ. Microbiol.* 2003, *69*, 1759–1774.