



UvA-DARE (Digital Academic Repository)

Time-series analysis of the transcriptome and proteome of *Escherichia coli* upon glucose repression

Borirak, O.; Rolfe, M.D.; de Koning, L.J.; Hoefsloot, H.C.J.; Bekker, M.; Dekker, H.L.; Roseboom, W.; Green, J.; de Koster, C.G.; Hellingwerf, K.J.

DOI

[10.1016/j.bbapap.2015.05.017](https://doi.org/10.1016/j.bbapap.2015.05.017)

Publication date

2015

Document Version

Final published version

Published in

Biochimica et Biophysica Acta. Proteins and Proteomics

License

CC BY

[Link to publication](#)

Citation for published version (APA):

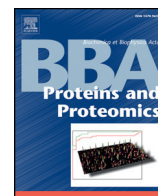
Borirak, O., Rolfe, M. D., de Koning, L. J., Hoefsloot, H. C. J., Bekker, M., Dekker, H. L., Roseboom, W., Green, J., de Koster, C. G., & Hellingwerf, K. J. (2015). Time-series analysis of the transcriptome and proteome of *Escherichia coli* upon glucose repression. *Biochimica et Biophysica Acta. Proteins and Proteomics*, 1854(10, Part A), 1269-1279.
<https://doi.org/10.1016/j.bbapap.2015.05.017>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.
UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)



Time-series analysis of the transcriptome and proteome of *Escherichia coli* upon glucose repression



Orawan Borirak^a, Matthew D. Rolfe^b, Leo J. de Koning^c, Huub C.J. Hoefsloot^d, Martijn Bekker^a, Henk L. Dekker^c, Winfried Roseboom^c, Jeffrey Green^b, Chris G. de Koster^c, Klaas J. Hellingwerf^{a,*}

^a Molecular Microbial Physiology, Swammerdam Institute for Life Sciences, University of Amsterdam, The Netherlands

^b Krebs Institute, Molecular Biology and Biotechnology, University of Sheffield, United Kingdom

^c Mass Spectrometry of Biomacromolecules, Swammerdam Institute for Life Sciences, University of Amsterdam, The Netherlands

^d Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, The Netherlands

ARTICLE INFO

Article history:

Received 2 March 2015

Received in revised form 12 May 2015

Accepted 28 May 2015

Available online 4 June 2015

Keywords:

Carbon catabolite repression

Time-series analysis

Transcriptomics analysis

Proteomics analysis

Post-transcriptional control

Small regulatory RNA

ABSTRACT

Time-series transcript- and protein-profiles were measured upon initiation of carbon catabolite repression in *Escherichia coli*, in order to investigate the extent of post-transcriptional control in this prototypical response. A glucose-limited chemostat culture was used as the CCR-free reference condition. Stopping the pump and simultaneously adding a pulse of glucose, that saturated the cells for at least 1 h, was used to initiate the glucose response. Samples were collected and subjected to quantitative time-series analysis of both the transcriptome (using microarray analysis) and the proteome (through a combination of ¹⁵N-metabolic labeling and mass spectrometry). Changes in the transcriptome and corresponding proteome were analyzed using statistical procedures designed specifically for time-series data. By comparison of the two sets of data, a total of 96 genes were identified that are post-transcriptionally regulated. This gene list provides candidates for future in-depth investigation of the molecular mechanisms involved in post-transcriptional regulation during carbon catabolite repression in *E. coli*, like the involvement of small RNAs.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Escherichia coli is the best studied Gram-negative bacterial species to date [1]. This makes it the ideal prokaryote in which to study physiological adaptation, and the involvement of post-transcriptional regulation therein. The availability of omics-analysis techniques has, particularly in bacteria, opened up the possibility of analyzing biological function with a 'systems approach' [2]. A simplifying assumption, however, that is often made in systems analyses is that the level of expression of a certain protein is proportional to the abundance of the corresponding mRNA. However, many studies did not find good correlation between protein- and mRNA abundance (for review see: [3]), which suggests that in systems analysis, regulation of gene expression at the post-transcriptional level must be taken into account. Indeed several examples of this type of regulation have recently been documented [4–8]. For this reason here we investigate the extent of post-transcriptional

regulation in a physiological response that is based on a global (i.e. genome-wide) alteration of the level of gene expression. For this, we have selected carbon catabolite repression (CCR) in *E. coli* [9,10] as our model system.

The mechanisms underlying the ability of *E. coli* to grow on a wide range of carbon sources has already been studied for decades, through both physiological and genetic studies [11,12]. *E. coli*'s preferred use of glucose is brought about by CCR [13]: if in a batch culture of this organism the majority of the available glucose has been catabolized, metabolism is reprogrammed to prepare the organism for use of alternative carbon sources [14]. When glucose becomes available again, the cells undergo another major transition, i.e. CCR that involves both the cellular transcriptional- and metabolic networks. Regulation of the CCR is complex and controlled at multiple levels. It is assumed to be predominantly regulated at the level of transcription via the 'alarmone' cAMP, which in turn forms a complex with the global transcriptional regulator CRP. Indeed, the cAMP–CRP complex modulates expression of many catabolic genes [10,13], and cAMP levels in the cell are under control of the glucose-specific part of the Phosphoenol-pyruvate-dependent Phospho-Transferase system ('glucose-PTS'), via the level of phosphorylation of Enzyme IIA^{Glc} [10]. 'Systems' studies of CCR generally focus on the transcriptional regulation of gene expression [11,12, 15–17], whereas the involvement of post-transcriptional regulation in

Abbreviations: AUC, area under the curve; CCR, carbon catabolite repression; ED, Entner–Doudoroff; FDR, False Discovery Rate; MudPIT, Multidimensional Protein Identification Technology; PTS, Phosphotransferase system; SCXC, Strong Cation Exchange Chromatography; SLP, Substrate-level phosphorylation; STEM, Short Time-series Expression Miner.

* Corresponding author.

E-mail address: K.J.Hellingwerf@uva.nl (K.J. Hellingwerf).

CCR has not even been investigated under steady state conditions, let alone dynamically.

Because of the 'global' nature of the CCR response, we considered it plausible that post-transcriptional regulation would constitute a significant part of it. This expectation is strengthened by the results of several recent studies of other regulation mechanisms, in which it was shown that transcript levels often poorly correlate with the corresponding protein levels [18–21]. Furthermore, a detailed time-series proteomics analysis of carbon catabolite repression, combined with transcript profiling analysis, has not yet been reported. Data derived from time-series experiments provide a richer source of information than a single time point measurement. Interpretation of cellular responses with the latter approach usually raises the question of whether the most informative time point was selected for optimal data analysis. Moreover, time-series data tend to reduce measurement noise and thus increase the accuracy of the conclusions.

Here we present the results of a set of experiments that enabled us to quantify a genome-wide time-series of both the transcript- and the protein-levels in *E. coli* cells, subjected to a change from glucose-limiting conditions, i.e. CCR-free, to glucose-excess conditions, i.e. with CCR. This could be achieved via stopping the pump of a chemostat, with simultaneous addition of a glucose pulse that saturates the cells for a period of at least 1 h (see Fig. 1 for the experimental design). Physiological and molecular genetic evidence that such a glucose pulse indeed activates CCR has recently been described elsewhere [22]. The aim of the current experiments was to investigate the significance of post-transcriptional control in CCR in *E. coli*, by comparing dynamic alterations of the transcriptome, with those of the corresponding proteome. The results of such measurements were subjected to statistical analyses, specifically designed for time-series data. The 'Area Under the Curve' (AUC), representing the relative change in RNA level of a specific gene, was used to calculate the change in the amount of the transcript during the whole time-series, while a simple linear regression

was used to determine the change in the amount of the corresponding protein per unit of time. Using a genome-wide comparative analysis between the transcript- and the corresponding protein-levels, we have identified 96 genes that are regulated post-transcriptionally, 51 of them with a significance level of <0.01 , and another 45 genes at the significance level of <0.05 .

The discovery of this extensive involvement of post-transcriptional regulation in CCR provides a starting point for the more detailed understanding of this physiological response. Mechanisms that may contribute to this added layer of regulation are briefly discussed.

2. Materials and methods

2.1. Bacterial strain and growth conditions

E. coli MG1655 was grown under glucose-limited conditions in 2 l chemostat vessels (Applikon, The Netherlands) with a working volume of 1 l at a dilution rate of 0.2 h^{-1} . Culture conditions and medium composition were selected as previously described [22]. Briefly, a minimal medium [23] supplemented with 20 mM nitrilo-acetic acid as a chelator, $0.17 \mu\text{M Na}_2\text{SeO}_3$, and 20 mM glucose was used. For a reference culture, $^{15}\text{NH}_4\text{Cl}$ (98 atom % ^{15}N ; Sigma Aldrich) was used as the N source, instead of NH_4Cl . Temperature was controlled at 37°C and pH was maintained at 6.9 ± 0.1 by titrating with sterile 4 M NaOH. The culture was aerated with 0.5 l/min water-saturated air and agitated with a propeller at 600 rpm. Pre-cultures were grown in the same medium, except that 100 mM sodium phosphate buffer was used to increase buffering capacity. After the chemostat culture reached steady state, 50 mM glucose (final concentration) was added to initiate the CCR response. Simultaneously, the medium feed was stopped and cells were harvested for further analyses by conventional rapid sampling (i.e. using a slight overpressure) at 5, 15, 30, and 60 min after the glucose pulse. Cells harvested from the steady state of the chemostat were used as a reference (i.e. time = 0) sample.

2.2. Transcriptomic analysis

2.2.1. RNA sample preparation

RNA was isolated using the RNeasy mini kit (Qiagen) as described previously [22]. The oligonucleotide microarrays (design ID 029412) [24] used in this study were obtained from Agilent Technologies (Stockport, UK). Each microarray slide ($8 \times 15 \text{ k}$ format) was based on the Agilent *E. coli* catalogue microarray (G4813A-020097), which covered 4287 *E. coli* K-12 MG1655 genes and was supplemented by an additional 311 probes designed using eArray (Agilent Technologies) for recently identified genes, re-annotated genes and small non-coding RNAs.

2.2.2. Microarray analysis

Isolated RNA was directly converted to fluorescently labeled cDNA as described elsewhere [25]. The cDNA produced from RNA samples at 5, 15, 30 and 60 min after perturbation of the chemostat with a glucose pulse were all hybridized with cDNA produced from the steady-state sample (i.e. $t = 0$), which was used as the reference. Two RNA samples (from biological replicates) were obtained for each time point, and these were hybridized twice as dye-swaps (i.e. technical replicates) and thus provide four replicates in total. These choices provide sufficient data for robust statistical filtering. Quantification of the cDNA samples, microarray hybridization, and washing and scanning of the arrays, were carried out as described in the Fairplay III labeling kit (Agilent Technologies, 252009, Version 1.1). Scanning was performed with a high resolution microarray scanner (Agilent Technologies). GeneSpring GX v7.3 (Agilent Technologies) was used for data normalization and data analysis. The transcriptomic data have been deposited in ArrayExpress (Accession number E-MTAB-2398).

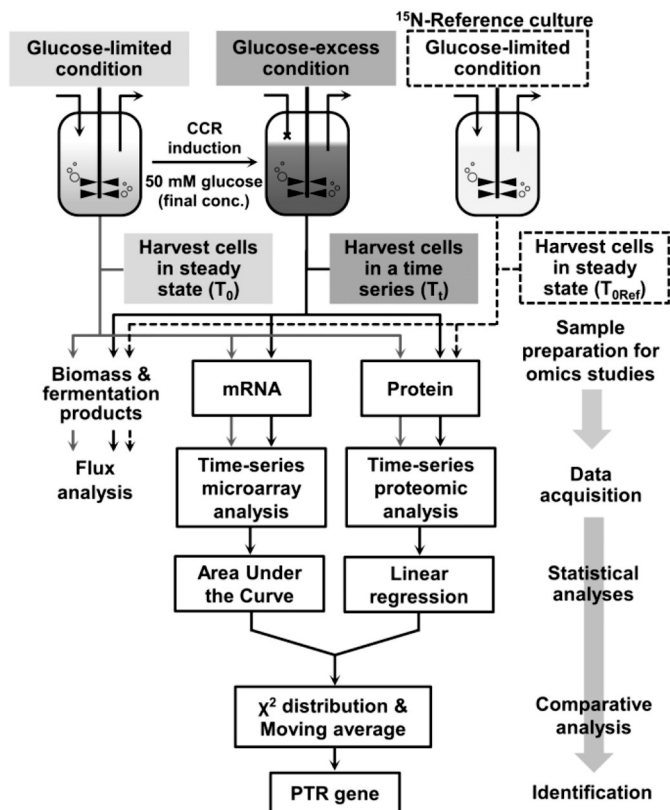


Fig. 1. Experimental design and identification strategy for post-transcriptionally regulated (PTR) genes used in this study.

2.2.3. Short Time-series Expression Miner (STEM) analysis

The STEM analysis tool [26] that is integrated with Gene Ontology (GO) enrichment analysis, was used to cluster genes that show a similar temporal expression pattern by using normalized \log_2 ratios of the transcriptomic data. A total of 50 possible temporal gene expression profiles (out of the 81 possible) were computed. Genes were then assigned to the best-fitting profile using the STEM clustering algorithm. The significance level of each profile was calculated based on the ratio of the number of assigned genes to that profile, versus the number of expected genes to a profile using Permutation test, and corrected by Bonferroni correction as previously described [27].

2.3. Proteomic analysis

2.3.1. Protein sample preparation

Cells (approximately 20 ml) were harvested from the chemostat by conventional rapid sampling (i.e. using a slight overpressure) directly into an ice-cold tube with a small volume of 50 $\mu\text{g}/\text{ml}$ chloramphenicol and a 1/10 dilution of a Complete Protease Cocktail inhibitors mix (both from a concentrated stock solution and the latter mix was from Roche), and then centrifuged at $4000 \times g$ for 5 min at 4 °C. Cell pellets were immediately frozen with liquid nitrogen and subsequently lyophilized and stored at 80 °C until use. Sampled cells were mixed with cells from the ^{15}N -reference culture at a 1:1 ratio based on OD_{600} , and then suspended in an extraction buffer that consisted of 6 M urea, 0.5 mM EDTA, 2% (w/v) SDS, and complete protease inhibitors cocktail mixture (Roche) in 100 mM NH_4HCO_3 lysis buffer, after which the mixture was sonicated. The protein concentration was measured in all samples with the bicinchoninic acid (BCA) assay (BioRad). Next, 200 μg protein was subjected to trypsin digestion, using the gel-assisted digestion method as previously described [28]. The resulting peptide mixture was then lyophilized after extraction from the gel.

2.3.2. Strong Cation Exchange Chromatography

Peptides were resuspended in 0.1% (v/v) trifluoroacetic acid (TFA) plus 50% (v/v) acetonitrile (ACN), and then loaded onto a PolySULPHOETHYLAspartamide™ column (2.1 mm ID, 10 cm length) on an Ultimate HPLC system, connected to a fraction collector (LC Packings, Amsterdam, The Netherlands). Elution (flow rate: 0.1 ml/min) was performed using solvent A; 10 mM KH_2PO_4 , 25% (v/v) ACN, pH 2.9 and solvent B; 10 mM KH_2PO_4 , 500 mM KCl and 25% (v/v) ACN, pH 2.9. A stepwise gradient was used of 2%, 4%, 6%, 8%, 10%, 20%, 50%, and 100% of B. The program was run for 120 min, in which the step-gradient started after 40 min and lasted for 10 min at each step. The elution was monitored via absorbance measurements at 214 nm. Accordingly, 8 separate fractions were collected. Then, these samples were lyophilized and stored at -80 °C. Before being analyzed by mass spectrometry, the samples were re-suspended in 0.1% (v/v) TFA plus 3% (v/v) ACN. Fractions eluted from 6 to 10% (v/v) of solvent B were combined, as well as those collected from 20 to 100% of solvent B. Therefore, a total of 4 fractions was generated, and subsequently desalted with a C_{18} reversed phase tip (Varian).

2.3.3. LC-FT-MS/MS data acquisition, data processing and relative protein quantification

For 3 biological replicates, the proteomes of the cells harvested at steady state (i.e. at $t = 0$ min), and of the cells harvested at $t = 5, 15, 30$ and 60 min after induction of CCR, were analyzed with mass spectrometry. The LC-FT-MS/MS data of each of the 4 SCX fractions of the ^{14}N , ^{15}N isotopic tryptic peptide mixture of these proteomes were acquired with an ApexUltra Fourier transform ion cyclotron resonance mass spectrometer (Bruker Daltonic, Bremen, Germany) equipped with a 7 T magnet and a nano-electrospray Apollo II DualSource™ coupled to an Ultimate 3000 (Dionex, Sunnyvale, CA, USA) HPLC system. The 60 samples, each containing 400 ng of the ^{14}N , ^{15}N tryptic peptide mixture, were injected as a 10 μl 0.1% (v/v) TFA aqueous

solution and loaded onto the PepMap100 C_{18} (5- μm particle size, 100-Å pore size, 300- μm inner diameter \times 5 mm length) pre-column. The peptides were eluted via an Acclaim PepMap 100 C_{18} (3- μm particle size, 100-Å pore size, 75- μm inner diameter \times 250 mm length) analytical column (Thermo Scientific, Etten-Leur, The Netherlands) using a linear gradient from 0.1% formic acid/6% $\text{CH}_3\text{CN}/94\%$ H_2O (v/v) to 0.1% formic acid/40% $\text{CH}_3\text{CN}/60\%$ H_2O (v/v) over a period of 120 min at a flow rate of 300 nl/min. Data-dependent Q-selected peptide ions were fragmented in the hexapole collision cell at an Argon pressure of 6×10^{-6} mbar (measured at the ion gauge) and the fragment ions were detected in the FTICR cell at a resolution of up to 60,000 (m/ Δ m). Instrument mass calibration was better than 1 ppm over an m/z range of 250 to 1500. The MS/MS rate was about 2 Hz. This yielded more than 9000 MS/MS spectra over the 120 min LC-MS/MS chromatogram.

Raw FT-MS/MS data of the 4 SCX peptide fractions were processed as multi-file (MudPIT) with the MASCOT DISTILLER program, version 2.4.3.1 (64 bits), MDRO 2.4.3.0 (MATRIX science, London, UK), including the Search toolbox and the Quantification toolbox. Peak-picking for both MS and MS/MS spectra was optimized for a mass resolution of up to 60,000 (m/ Δ m). Peaks were fitted to a simulated isotope distribution with a correlation threshold of 0.7, and with a minimum signal-to-noise ratio of 2. The processed data were searched with the MASCOT server program 2.3.02 against the complete *E. coli* K12 proteome database from the UniProt consortium (release: June, 2012; 4271 entries in total) with the redundancy removed with DBtoolkit [29]. The database was complemented with its corresponding decoy data base for statistical analyses of peptide false discovery rate (FDR). Trypsin was used as the enzyme and 1 missed cleavage was allowed. Carbamidomethylation of cysteine was used as a fixed modification and oxidation of methionine as a variable modification. In addition to the search for tryptic peptides, semi-tryptic peptides were allowed in order to monitor selectivity of digestion. The peptide mass tolerance was set to 5 ppm and the peptide fragment mass tolerance was set to 0.01 Da. The quantification method was set to the metabolic ^{15}N labeling method, to enable MASCOT to identify both ^{14}N and ^{15}N peptides. The MASCOT MudPIT peptide identification score was set to a cut-off of 20. At this cut-off, and based on the number of assigned decoy peptide sequences, a peptide false discovery rate (FDR) of ~2% for all analyses was obtained. Using the quantification toolbox, the isotopic ratio for all identified proteins was determined as weighted average of the isotopic ratios of the corresponding light over heavy peptides. Selected critical settings were: require bold red: on, significance threshold: 0.05; Protocol type: precursor; Correction: Element 15 N; Value 99.4; Report ratio L/H; Integration method: Simons; Integration source: survey; Allow elution time shift: on; Elution time delta: 20 s; Std Err. Threshold: 0.15, Correlation Threshold (Isotopic distribution fit): 0.98; XIC threshold: 0.1; All charge states: on; Max XIC width: 200 s; Threshold type: at least homology; Peptide threshold value: 0.05; unique pepseq: on. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [30] with the dataset identifier <PXD000xxx>.

2.4. Statistical analysis

2.4.1. Transcriptome – Area Under the Curve (AUC)

The transcriptomic data consist of 2 biological replicates of samples taken at 0 (i.e. the steady state), and 5, 15, 30, and 60 min after the glucose pulse, with technical duplicate measurements at all-time points (except at 30 min, when only a single sample was available). To characterize a change in the amount of transcript over time, we determined Area Under the Curve (AUC) of normalized \log_2 ratios of the transcripts as a function of time (compare [30]). For each biological replicate, 8 possible time profiles (i.e.: $2 \times 2 \times 1 \times 2$) can be constructed per gene. The AUC values of the 8 time profiles were calculated using the trapz function of MATLAB. These values were averaged and next the

average AUC value of the two biological replicates was calculated to obtain the average value. This final AUC value represents the relative change in the amount of a transcript during the whole time-series. The normalized unlogged ratio of the transcript at each time point and the calculated AUC values are provided in the Supplementary Table S1.

2.4.2. Proteome — linear regression analysis

The normalized $^{14}\text{N}/^{15}\text{N}$ isotopic ratio for all proteins and for all sampling points is listed in the Supplementary Table S2. They represent the relative abundance level of a protein. The number of available data points for any protein varies between 1 and 16, as the data were generated from 3 biological replicates and 5 time points, plus 1 additional technical duplicate. Any change in the abundance of a protein is governed by the balance between its rate of production (via transcription and translation) and its rate of degradation. To estimate relative changes in protein concentration, the proteins for which at least 6 separate data points were available, the abundance was analyzed with linear regression, using the MATLAB function `regress`. Time was used as the explanatory variable and the normalized protein $^{14}\text{N}/^{15}\text{N}$ isotopic ratio as the dependent variable. The resulting regression coefficient (i.e. the slope) represents the change of the relative amount of the protein per unit of time, which can also be interpreted as 'a net production (with positive slope) or net degradation (i.e. when the slope is negative) rate'. A rate is considered to be significant if the value zero (i.e. the null hypothesis) is outside the 95% confidence interval of the calculated slope. The calculated rates of change in relative protein abundance of 557 proteins, with the corresponding *p*-values, are provided in Supplementary Table S3.

2.4.3. Integrated analysis of transcriptomic and proteomic data

2.4.3.1. Calculation of the confidence region for the first null-hypothesis that transcript and corresponding protein level do not change. To be able to calculate confidence regions for this null hypothesis, first the statistical properties of the measurement distribution under this null hypothesis must be derived. For the transcript level all the duplicate measurements were used to obtain a standard deviation per batch. Under the null hypothesis all the \log_2 ratios, the values used in the analysis, are zero. So in order to obtain the statistical distribution of the AUC values the artificial transcriptomic data were drawn from a normal distribution with zero mean and the standard deviation from the batch under consideration. Then the same procedure as described previously was used to calculate the AUC for the artificial transcriptomics data. This was done 1000 times. The obtained AUC values were fitted with a normal distribution again using the MATLAB function `normfit`, and the mean and standard deviation were calculated.

For the proteome a normal distribution is fitted to the measurements at $t = 0$ for the three batches. From this fit the means and standard deviations of the relative changes in the level of a specific protein are determined using the `normfit` function of MATLAB. Under the null hypothesis that no changes in protein level occur during the experiment, the measurements at times 5, 15, 30 and 60 min come from the same distribution as the measurements from $t = 0$. In an artificial data file every value from the original data is replaced by a value from a normal distribution with the appropriate mean and standard deviation. This means that if the value was from batch 2 also the mean and standard deviation from batch 2 was used. If there was no value in the original data this was also the case in the artificial data. Then for an artificial data file the slopes were calculated as described above. This was repeated for 1000 artificial data files. The calculated slopes were then used to fit a normal distribution and both its mean and standard deviation were determined with `normfit`.

With this we have the distribution of the artificial data under the null-hypothesis for both the AUC values and the slopes. Then to get a confidence region where neither the transcript — nor the protein level did change significantly, the χ^2 distribution with 2 degrees of freedom

was calculated [31]. This is the elliptic region ($p < 0.01$) as shown in Fig. 6A. The outboard regions of this ellipse were then used as the significance threshold for the transcript level (AUC). This means that any gene having AUC values > 16.75 or < -16.75 was considered as significantly changed in its transcript level (with $p < 0.01$). The corresponding values for the 'slope' are: < -0.004 and > 0.004 .

2.4.3.2. Use of a 'moving average' to identify genes with a disproportionate change in the level of its mRNA and the corresponding protein. For each quantifiable protein the value for the slope (i.e. from the time-series proteomics analysis and that represents the relative change in its abundance) was averaged over 13 ($= n$) values, i.e. the value for the slope of the protein itself, plus the slope of the 6 proteins with the closest lower AUC value and those with the 6 nearest higher AUC values. If this averaged slope falls outside the 99% confidence interval (as determined by linear regression) for the protein under consideration, the corresponding gene is considered to be subject to post-transcriptional regulation (Table 1 and Supplementary Table S4). The same analysis was carried out with $n = 11$ and $n = 15$ (results not shown). This resulted in essentially the same list of genes subject to PTR.

3. Results

3.1. Dynamic analysis of the physiological characteristics of *E. coli* cells upon glucose repression, induced in cells growing in a steady-state chemostat culture

To investigate the contribution of post-transcriptional control in *E. coli* upon initiation of the CCR response, we first created a reference condition where no CCR is present by using a glucose-limited chemostat culture. Under these conditions no residual glucose could be detected in the chemostat cultures, which means that the glucose concentration was lower than $50 \mu\text{M}$ [22]. This culture was then pulsed with an excess of glucose to initiate the CCR, simultaneously with stopping the pump of the chemostat. The de-repressed nature of the cells at steady state, and the switch to the glucose-repressed state that we aimed to achieve were experimentally validated as described elsewhere [22]. Briefly, growth and physiological characteristics of the glucose-repressed cells were monitored through biomass- and fermentation-product measurements, and additionally via a time-series measurement of the expression level of selected genes. The time-series gene expression analysis was performed with RT-PCR. The results showed that there was no CCR in *E. coli* cells growing in a glucose-limited chemostat, as such cells had the ability to immediately consume alternative carbon sources (confirmed via addition of selected sugars to the culture), while two selected genes, i.e. *ptsG* (encoding a PTS-glucose transporter enzyme IIBC), and *crp* (encoding the global transcriptional regulator CRP), were shown to be repressed at least up to 60 min after the glucose pulse. The decreased expression level of these genes confirms that the initiation of glucose repression by the pulse of glucose was successful.

The growth rate (μ) of the *E. coli* culture after the glucose pulse increased from 0.2 up to 0.5 h^{-1} after 60 min, and slightly increased thereafter up to 90 min. Sugar and organic acid measurements showed that the glucose concentration was in excess at all times during this 60 min time window (glucose concentration 38 mM 60 min after a 50 mM pulse [22]) and that acetate was the only fermentation product observed up to 120 min after the glucose pulse. From such a pulsed chemostat, cells were harvested at steady state ($t = 0$) and in a time-series at 5, 15, 30, and 60 min after the glucose pulse, to be used for further analyses of both the transcriptome and the proteome (Fig. 1).

3.2. Quantitative transcriptomic analysis of Carbon Catabolite Repression

Time-series transcriptomic data were measured using microarrays containing 4057 protein encoding genes, 152 pseudo-genes, and 78 small non-coding RNAs and other RNA elements. Transcript levels at

Table 1
List of post-transcriptionally regulated genes identified in this study and grouped according to their cellular functions.

GN	Function	Quadrant	Slope	AUC
Amino acid biosynthesis				
<i>aroK</i> ^f	Shikimate kinase 1	I	0.0039	24.2
<i>cysA</i> ^f	Sulfate/thiosulfate import ATP-binding protein CysA	I	0.0175	80.0
<i>cysI</i> ^f	Sulfite reductase [NADPH] hemoprotein β -component	I	0.0075	108.1
<i>cysP</i> ^a	Thiosulfate-binding protein	I	0.0054	102.9
<i>glnA</i> ^f	Glutamine synthetase	I	0.0033	87.6
<i>gltB</i> ^{f,b}	Glutamate synthase [NADPH] large chain	I	0.0099	47.3
<i>metB</i> ^f	Cystathionine γ -synthase	I	0.0046	124.7
<i>metC</i>	Cystathionine β -lyase metC	IV	-0.0045	34.6
<i>metQ</i>	D-methionine-binding lipoprotein metQ	I	0.0018	56.9
<i>fitY</i>	Cystine-binding periplasmic protein-SymR	I	0.0005	55.4
Cell division and DNA replication				
<i>minE</i>	Cell division topological specificity factor	III	-0.0051	-18.3
<i>nlpI</i>	Lipoprotein NlpI	IV	-0.0020	23.8
<i>dnaj</i>	Chaperone protein Dnaj	III	-0.0051	-6.3
Transcription				
<i>rpoA</i>	DNA-directed RNA polymerase subunit α	I	0.0001	51.8
<i>sspA</i>	Stringent starvation protein A	IV	-0.0011	45.4
<i>stpA</i>	DNA-binding protein stpA	I	0.0260	83.3
<i>yifE</i>	UPF0438 protein yifE	I	0.0034	17.1
Translation				
<i>hfq</i>	Protein hfq	IV	-0.0028	17.8
<i>hpf</i>	Ribosome hibernation promoting factor	IV	-0.0080	11.5
<i>lysS</i>	Lysine-tRNA ligase	I	0.0008	44.4
<i>pheT</i>	Phenylalanine-tRNA ligase β -subunit	II	0.0003	-46.3
<i>srnB</i> ^f	ATP-dependent RNA helicase SrmB-IsrB	I	-0.0032	39.8
<i>tig</i>	Trigger factor	IV	0.0036	30.7
<i>typA</i>	GTP-binding protein typA/BipA	I	0.0081	45.2
<i>ygfZ</i> ^f	tRNA-modifying protein ygfZ	II	0.0055	-36.4
Ribosomal proteins				
<i>rplA</i> ^c	50S ribosomal protein L1	I	0.0027	60.0
<i>rplF</i>	50S ribosomal protein L6	I	0.0036	72.6
<i>rplI</i>	50S ribosomal protein L9	I	0.0049	40.9
<i>rplM</i>	50S ribosomal protein L13	I	0.0025	60.4
<i>rplS</i>	50S ribosomal protein L19	II	0.0042	-17.1
<i>rplX</i>	50S ribosomal protein L24	I	0.0034	75.5
<i>rpmA</i>	50S ribosomal protein L27	I	0.0037	36.5
<i>rpmG</i> ^f	50S ribosomal protein L33	I	0.0058	50.3
<i>rpsC</i> ^d	30S ribosomal protein S7	I	0.0043	29.7
<i>rpsP</i> ^b	30S ribosomal protein S16	I	0.0038	18.4
Central carbon metabolism				
<i>eno</i>	Enolase	IV	-0.0001	48.6
<i>ppc</i> ^b	Phosphoenolpyruvate carboxylase	II	0.0012	-19.7
<i>talB</i>	Transaldolase B	IV	-0.0020	47.1
Membrane proteins				
<i>nuoC</i> ^f	NADH-quinone oxidoreductase subunit C/D	III	-0.0039	-79.7
<i>ompX</i> ^e	Outer membrane protein X	I	0.0255	103.7
<i>potD</i>	Periplasmic substrate-binding protein subunit of Putrescine/spermidine ABC transporter	IV	-0.0014	35.1
<i>yjjK</i> ^f	Uncharacterized ABC transporter ATP-binding protein YjjK	IV	-0.0010	42.8
Other metabolisms				
<i>dhaK</i> ^f	PTS-dependent dihydroxyacetone kinase, dihydroxyacetone-binding subunit dhaK	IV	-0.0070	20.7
<i>pntB</i>	NAD(P) transhydrogenase subunit beta	I	0.0004	51.3
<i>recA</i> ^b	Protein RecA	IV	-0.0022	38.2
<i>suhB</i>	Inositol-1-monomophosphatase	I	0.0228	42.4
<i>pepB</i> ^f	Peptidase B	IV	-0.0039	27.3
<i>rnr</i> ^f	Ribonuclease R	IV	-0.0029	39.4
<i>pdxB</i>	Erythronate-4-phosphate dehydrogenase	IV	-0.0016	73.6
<i>yeeX</i>	UPF0265 protein yeeX	IV	-0.0035	16.4
<i>yfgD</i> ^f	Uncharacterized protein YfgD	IV	-0.0007	53.5

Post-transcriptionally regulated genes reported in ^aKramer et al., 2009, ^bValgepea et al., 2013, ^cBaughman and Nomura, 1983, ^dDean et al., 1981, and ^eDe Lay and Gottesman, 2009. ^fGene containing a Repetitive Extragenic Palindrome sequence (REP element) within an intergenic region of an operon or 3'UTR of transcription unit. Underlined genes are additionally identified when a set of 556 proteins having at least 6 separate data points were used instead of a set of proteins having 8 separate data points (490 proteins). Quadrant refers to Fig. 6B.

$t = 5, 15, 30,$ and 60 min after the glucose pulse were normalized with the transcript level at $t = 0$. The average coefficient of variation among replicates at 5, 15, 30, and 60 min was 20%, 21%, 16%, and 31%, respectively. Significant changes in the amount of transcript over the observed time after the glucose pulse were determined using the 'Area Under the Curve' (AUC) approach (see **Materials and methods**: "Statistical analysis"). The AUC value represents the average change in the relative amount of a transcript present during the whole time-series, and is shown in Supplementary Table S1, along with the normalized unlogged ratio of each transcript at each time point. Besides that, the Short Time-series Expression Miner (STEM) analysis tool [26] that is integrated with Gene Ontology (GO) enrichment analysis, was used to cluster genes that show a similar temporal expression pattern using normalized \log_2 ratios. In total, 10 significant temporal gene expression profiles were clustered ($p < 0.01$) as shown in Fig. 2.

As expected, when glucose is introduced into the glucose-limited chemostat culture, genes that express proteins involved in carbohydrate metabolism, including carbohydrate transport (GO:0008643) and carbohydrate catabolic process (GO:0016052), are down-regulated. Interestingly expression of these same genes recovered after 30 min, as can be seen in Profile 0, while genes involved in cellular biosynthetic processes and cellular growth are up-regulated (e.g. DNA metabolic processes (GO:0006259), ribosomal proteins (GO:0003735), and nucleotide metabolic processes (GO:0009165)), as revealed by Profile 43, 41, and 40, respectively. Genes encoding proteins involved in RNA binding (GO:0003723) or in translation (GO:0006412) are enriched in Profile 38. Profile 49 represents the expression pattern of genes encoding proteins functioning in sulfur compound transport (GO:0072348). Phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS; GO:0009401) encoding genes, including *fruB*, *gatB*, *gatC*, *manZ*, *srlB*, and *srlE*, are clustered in Profile 19. In contrast, genes encoding carbohydrate transport (GO:0008643), especially ATP-binding cassette (ABC) transporters, including *malE*, *malF*, *malG*, *malK*, *araF*, and *araG*, follow the expression pattern of Profile 9. Profile 8 shows the expression of a set of genes that are down-regulated rapidly from the beginning onwards and remain so up to at least 15 min after initiation of the glucose pulse. Profile 8 is enriched in genes expressing proteins involved in cellular respiration (GO:0045333) and includes the NDH-I complex and some genes participating in the TCA cycle, i.e. *acnA*, *fumA*, *sucC*, and *sucD*, while Profile 10 displays the expression pattern of genes encoding proteins involved in aerobic respiration (GO:0009060), e.g. *acnB*, *gltA*, *sdhA*, *sdhB*, *sdhC*, and *sdhD*, that are rapidly down-regulated at first, but recover after 15 min after the glucose pulse. In contrast to most other genes involved in (cellular) respiration, *ndh* (encoding NADH: ubiquinone oxidoreductase II (NDH-II)) is up-regulated in a pattern that follows the expression pattern of Profile 38 (for more detail: see the STEM analysis files in Supplementary Data S1).

3.3. Quantitative proteomic analysis of Carbon Catabolite Repression

In parallel with the analysis of the transcriptome, quantitative time-series measurements of the proteome were carried out by using a stable-isotope labeling technique and LC-FTMS. Reference cells from a glucose-limited culture grown on $^{15}\text{NH}_4^+$ were harvested at steady state ($T_{0\text{ref}}$) and then mixed equally (based on OD_{600}) with cells derived from a ^{14}N culture (see Fig. 1). Then, these mixed samples (i.e. $t_0/t_{0\text{ref}}$, $t_5/t_{0\text{ref}}$, $t_{15}/t_{0\text{ref}}$, $t_{30}/t_{0\text{ref}}$, and $t_{60}/t_{0\text{ref}}$) were individually processed and analyzed as described in Materials and Methods: "Statistical analysis". Three independent experiments with the ^{14}N cultures were performed, resulting in a total of 873 quantified proteins. Errors in the 1:1 mixing of the ^{14}N cell cultures, with the ^{15}N reference cultures, were corrected by normalizing each dataset for all time points. This was done by setting the $^{14}\text{N}/^{15}\text{N}$ isotopic ratio for the TufA protein to 1. TufA has previously been used as the internal standard for corrections of protein injection between technical replicates and also for variation in protein loading after growth on different carbon sources [32]. After normalization, a

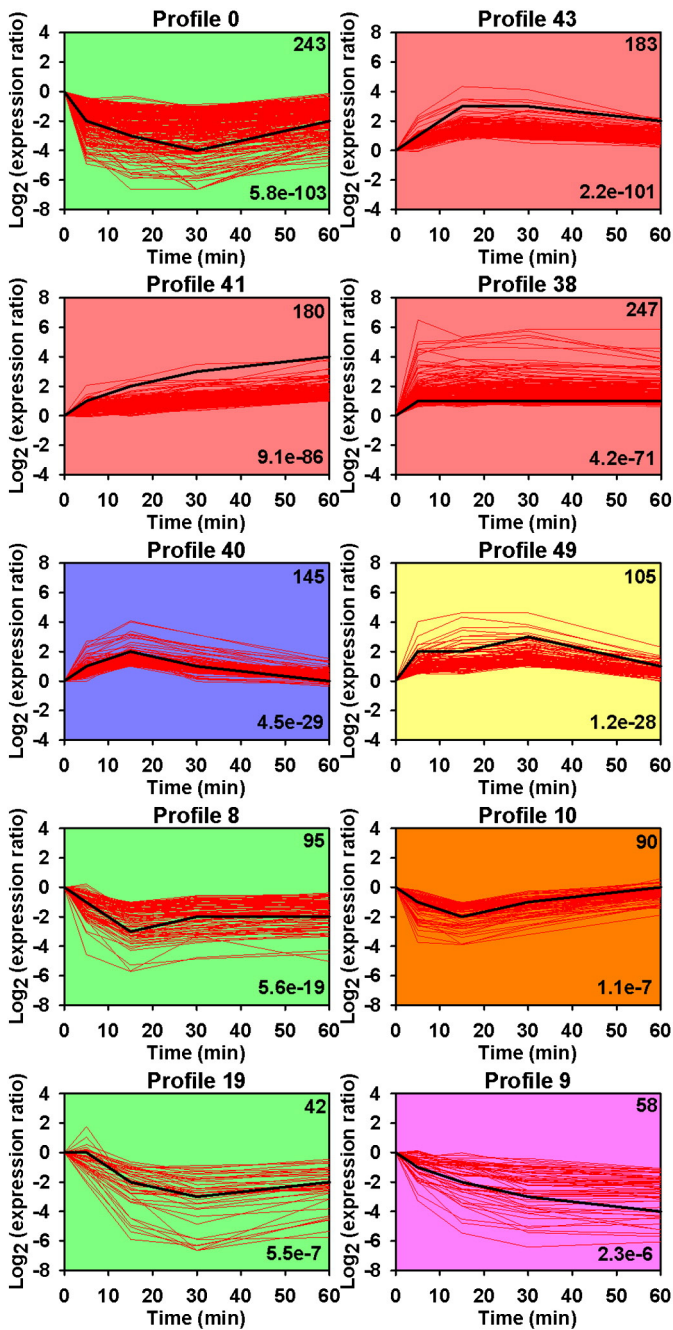


Fig. 2. Time-series transcriptomic analysis of *E. coli* upon glucose repression. Genes were clustered using Short Time-series Expression Miner (STEM) analysis tool with the abscissa representing the time scale of 5, 15, 30, and 60 min after the glucose pulse, and the ordinate the \log_2 of a gene expression level. Only significant profiles are shown here and arranged according to the significance level. The significance level (p -value) of the profile was calculated based on the number of genes assigned (right top) versus the number of genes expected and is indicated at the right-bottom. Black lines represent temporal gene expression model of the profile itself while red lines are genes that were assigned to the profile. The significant profiles which are similar to each other are grouped as a cluster of profiles, and are given the same color.

normal distribution of the protein $^{14}\text{N}/^{15}\text{N}$ isotopic ratios at $t = 0$ of around 1 (with $R^2 = 0.99$) is obtained, as shown in Fig. 3. Standard deviations in the protein $^{14}\text{N}/^{15}\text{N}$ isotopic ratios, both before and after normalization of the three biological replicas, are within 10%, revealing the accuracy of the protein quantification. As an alternative, normalization of the protein $^{14}\text{N}/^{15}\text{N}$ isotopic ratios in each data set was completed using their median value of the $^{14}\text{N}/^{15}\text{N}$ isotopic ratio [33]. As shown in Supplementary Table S5, there is no significant difference in the

results between normalization on the median values, and on the values of the TufA isotopic ratios. Nearly 80% of the quantified proteins were detected in at least two biological replicates, an observation which attests to the excellent reproducibility of the experiments (Supplementary Fig. S1A). The resulting normalized protein $^{14}\text{N}/^{15}\text{N}$ isotopic ratios for all time points are listed in the Supplementary Table S2.

The abundance of a total of 557 proteins as a function of time after the glucose pulse was then subjected to linear regression analysis (see Statistical analysis). Approximately 60% of the proteins analyzed showed a slope significantly different from zero ($p < 0.05$). Of those, 115 (20%), and 228 (40%) proteins were significantly induced and repressed by CCR, respectively. The calculated changes in protein abundance (slopes) of 557 proteins, with the corresponding p -value and the change in transcript level (i.e. AUC value), are provided in Supplementary Table S3.

3.4. Response of the proteome upon Carbon Catabolite Repression

Consistent with the immediate upshift in growth rate upon addition of the pulse of glucose [22] and with the transcriptomics results (see above), the time series analysis of *E. coli*'s proteome revealed that the majority of the proteins whose level is up-regulated, are ribosomal proteins (40%), and proteins involved in nucleotide- or amino acid biosynthesis (19% and 18%, respectively; Fig. 4A). Also expressions of proteins involved in scavenging nucleotides and amino acids, such as the uracil permease (UraA), and the periplasmic oligopeptide-binding protein OppA were increased. Presumably, increasing amounts of iron-sulfur cluster containing proteins are required in this upshift of growth rate, because the levels of the sulfate transporter (i.e. CysA, and CysP) and of glutaredoxin-4 (GrxD) were remarkably up-regulated too.

A particularly intriguing member of the group of proteins whose level is significantly upregulated by the glucose pulse is StpA, a multifunctional protein that is homologous to H-NS [34]. StpA has a role in the regulation of glucose catabolism, as it represses the *bgl* operon [35, 36]. Furthermore, it has recently been shown to display RNA chaperone activity, both *in vitro* and *in vivo* [37,38] and is, together with H-NS, responsible for the "high expression level of essential and growth-associated genes and low levels of stress-related and horizontally acquired genes" [39,40].

A group of enzymes involved in central carbon metabolism, which includes glycolysis, the TCA cycle, the glyoxylate shunt and the pentose phosphate pathway, are the most important down-regulated proteins by CCR, in good agreement with the transcriptome data, as shown in Fig. 4B. Also the alternative sugar transporters galactitol permease and mannose permease are sharply down-regulated, as well as the general PTS components EI (*ptsI*) and HPr (*ptsH*), and the glucose-specific PTS components EIIA (*crr*), and EIICB (*ptsG*), which is also in agreement with the results of the transcript analyses reported here and with the ~ 3.5 -fold decrease in expression from a *ptsG-lacZ* fusion in nitrogen-limited chemostat cultures (mM residual glucose) compared to glucose-limited chemostat cultures (μM residual glucose) [41]. Similarly, ATP synthase and several components of the respiratory chain (i.e. the NDH-I complex and the cytochrome *bd-I* terminal oxidase) are also down-regulated, while the expression level of NDH-II (which has a lower H^+/e^- stoichiometry than NDH-I [42,43]) increased. Also the latter observation is consistent with the microarray results.

The slope derived from the linear regression model not only indicates the *direction of change* in protein abundance, but also reveals the net change in the rate of production (i.e. the rate of production minus the rate of degradation) for each protein. These changes in rate differ considerably between individual proteins and between protein categories. The most pronounced differences, after grouping of the members of the proteome of *E. coli* in the MultiFun [44] categories, are shown in Fig. 5. Among the 3 most up-regulated groups, the highest production rate observed (i.e. for the nucleotide biosynthesis group) is 12-fold higher than observed for the lowest (i.e. the ribosomal protein group).

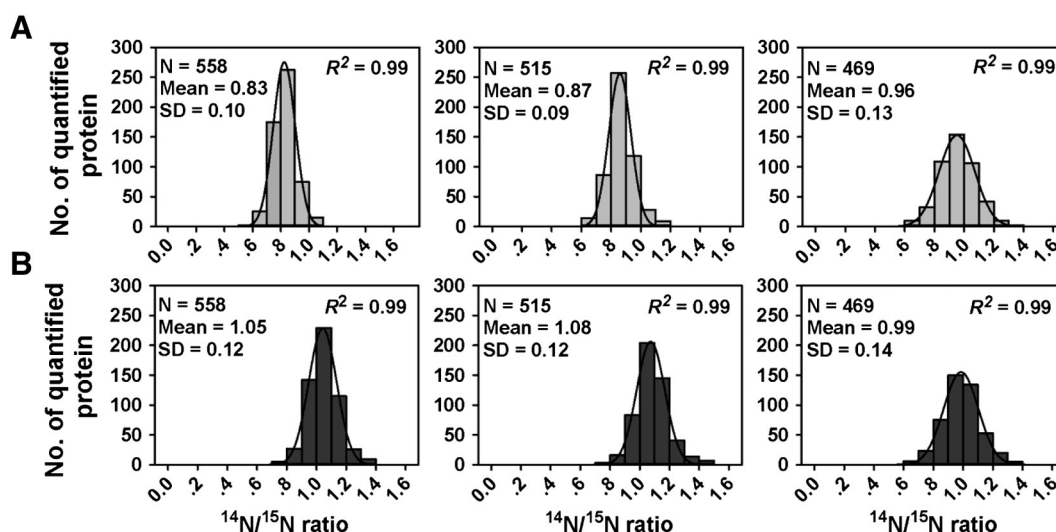


Fig. 3. Normalization of protein ratios at steady state, $t = 0$, against TufA. Ordinate represents number of quantified proteins while abscissa indicates the protein $^{14}\text{N}/^{15}\text{N}$ isotopic ratio. (A) Normal distribution of the protein ratios before normalization from three independent experiments. (B) Normal distribution of the normalized protein ratios. Number of quantified protein, mean value of the protein ratios, standard deviation of the protein ratios, and R^2 are indicated.

The standard deviation of the production rates in the amino acid biosynthesis group and in the nucleotide biosynthesis group is 63% and 41%, respectively. In contrast, the production rates of the ribosomal proteins show less variation: 26%. These results suggest that the production of the ribosomal proteins is strictly controlled and mutually synchronized.

In contrast to the differences observed for the (MultiFun categories of the) up-regulated proteins, the rates of decreased abundance of the down-regulated proteins is relatively constant (Fig. 5). Decreased abundance will be due to a combination of ‘dilution’ of the protein because decreased relative rate of synthesis, plus active proteolytic degradation. The fact that the limiting rate seems to converge to a value close to the growth rate after the glucose pulse, may suggest that the former contribution may be dominant. Nevertheless, the results show that during the glucose response protein expression in *E. coli* is predominantly controlled at the synthesis level, considering that most of the down-regulated proteins appear to be degraded gradually and passively at a similar rate.

3.5. Time-series analyses of Transcriptome vs. Proteome

To identify genes whose products, in addition to CCR control, may be subject to post-transcriptional regulation, we have applied a new type of statistical analysis of the genome-wide omics data that relates the relative change in a transcript level with the change in expression of the corresponding protein that it brings about. In Fig. 6A, this data has been plotted with increasing AUC values (i.e. change in relative mRNA abundance) as the explanatory variable and the relative change of the corresponding protein (i.e. the slope) as the dependent variable. In agreement with the first-order approximation, i.e. that the relative

change in mRNA abundance is proportional to the relative change in protein abundance, the data points in this plot seem to be related exponentially. This is further confirmed by plotting the slopes against AUC^* ($= 2^{(\text{AUC}/55)}$; see Supplementary Fig. S2). This linearity between the slopes and 2^{AUC} is expected as changes in both mRNA and protein levels are expressed relative to level prior to the perturbation of the cells with the pulse of glucose. Quantitative analysis of the linearity of this fit reveals that 36% of the covariance in the relative change of the abundance of the *E. coli* proteome in the glucose response is defined by changing transcript levels (Supplementary Fig. S2). The slope of this line holds information on the average efficiency of translation of the genes affected by glucose repression; however, a molecular interpretation of its numerical value is prevented by the fact that alteration of protein levels have not yet come to equilibrium in the time window available.

To identify genes that are subject to post-transcriptional regulation a method nevertheless was selected that is independent of the mathematical relation between AUC and slope. First the region in the plot was identified in which neither a change in the relative transcript abundance, nor in that of the corresponding protein level, can be considered as significantly changed. Such a region is defined by a χ^2 distribution with two degrees of freedom (see red and blue ellipse in Fig. 6A for $p < 0.01$ and $p < 0.05$, respectively). All genes from within this ellipse were then excluded from further analysis. In the absence of further gene specific regulatory mechanisms, genes with similarly altered relative transcription level (i.e. AUC values) are predicted to also have similarly altered relative protein abundances. Neighboring genes on the AUC axis in Fig. 6A, for which this first-order hypothesis holds, are therefore expected to also have similar slope values (i.e. proportionally altered production rates/abundance). If their slope values are not

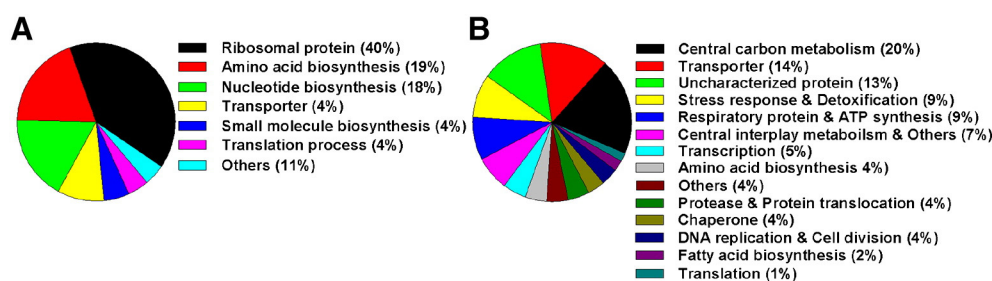


Fig. 4. Quantitative time-series proteomic analysis of *E. coli* upon a glucose repression. Proteins with significantly changed expression level ($p < 0.05$) were grouped based on their cellular functions. (A) Significantly up-regulated protein categories. (B) Significantly down-regulated protein categories.

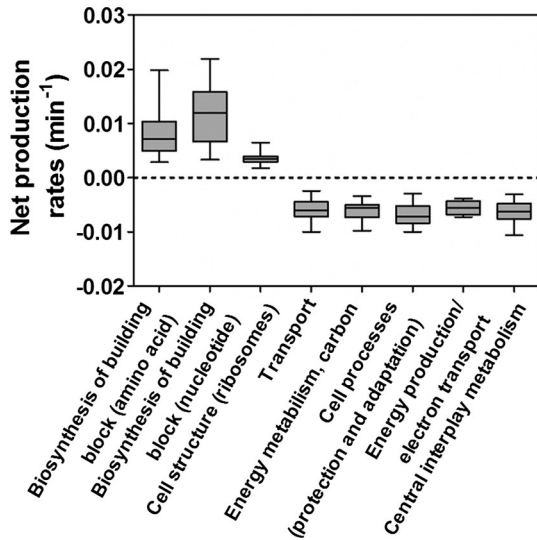


Fig. 5. Relative net production rates of the abundant proteins, grouped based on the MultiFun categories. Box plots show the relative net synthesis rates of the three most up-regulated protein categories, and the five most down-regulated protein categories. The horizontal bars represent the median, the first and the third quartile, while whiskers represent minimum and maximum values.

similar, then the level of expression of that protein must be subject to post-transcriptional regulation. For genes of this latter class there must be a post-transcriptional mechanism that influences the relative amount of produced protein by other means than solely via the relative

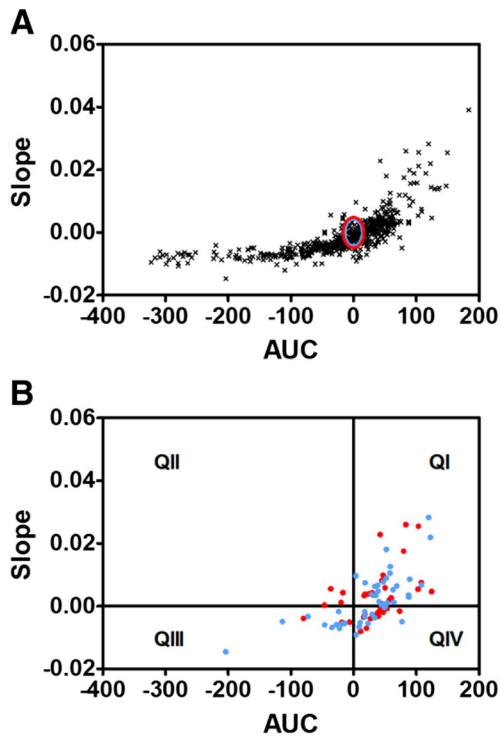


Fig. 6. Comparison of transcript level, expressed as Area Under the curve (AUC), with the rate of change of the corresponding protein abundance, expressed as the slope of the amount of the respective protein, against time upon a glucose repression. (A) The area in which neither transcript- nor protein abundance is changed significantly, at $p < 0.01$ (red ellipse) and at $p < 0.05$ (blue ellipse). Genes from these areas are excluded from further analysis. (B) Genes for which the correlation between transcript level and protein abundance differs significantly from the average with $p < 0.01$ (red dot), and $p < 0.05$ (blue dot). These are the genes that are identified as being post-transcriptionally regulated. QI to QIV refer to a Cartesian coordinate.

mRNA concentration. In order to identify all genes for which this conclusion about post-transcriptional regulation holds, we used the moving average approach as described in Materials and Methods: Statistical analysis: Genes having a correlation between the change in their relative transcript abundance and the change in the level of the corresponding protein that is statistically significantly different from the rest, are thereby identified as post-transcriptionally regulated (PTR) genes. Using this analysis, a total of 51 genes has been identified at the significance level of < 0.01 (Table 1) and they are joined by another 45 genes at the significance level of < 0.05 (Supplementary Table S4).

The genes identified using this approach with the significance level of < 0.01 are shown in Fig. 6B; red dot, and are listed in Table 1. Of these 51 genes, eight genes have already been reported in the literature to be regulated at the post-transcriptional level [45–49]. The 51 genes can be classified into 4 groups, corresponding to the four quadrants of Fig. 6C. The first two groups (Quadrants I and IV) contain genes transcriptionally activated by the glucose pulse. The abundance of their transcript increases (i.e. positive AUC); however, the production of the corresponding proteins is constrained by additional post-transcriptional control(s). The post-transcriptional regulation of genes from these quadrants can be further sub-divided into three possible control mechanisms, i.e. (i) increase in the translation rate, (ii) feedback regulation/ or inhibition of the translation process and (iii) altered mRNA stability. Enhanced translation rates were observed (Table 1) for genes in Quadrant I, like *suhB*, *stpA*, *ompX*, *cysA*, and *gltB*, whereas feedback regulation was observed for genes encoding the RNA polymerase α subunit (*rpoA*), pyridine nucleotide transhydrogenase β subunit (*pntB*), lysyl-tRNA synthetase (*lysS*), and amino acid periplasmic-binding proteins (*metQ* and *fliY*). As these genes exhibit significant increases in transcript level, but show little change in protein abundance. The increase in translation rate of *ompX* is possibly due to the fact that the small RNAs, CyaR and MicA, that inhibit translation of *ompX* [5,49], were significantly down-regulated (see Supplementary Table S1). Inada and Nakamura, (1996) proposed that the expression of *suhB*, encoding inositol monophosphatase, is auto-regulated via its own translation product, by negatively modulating mRNA stability. But the underlying mechanism that controls *suhB* mRNA decay is unclear and whether or not inositol monophosphatase directly or indirectly modifies the activity of RNaseIII is not known [50]. However, in the specific transition elicited by addition of a pulse of glucose, it is clear that the activation of *suhB* mRNA degradation was abolished.

Genes identified in Quadrant IV are possibly regulated by the inhibition of translation and/or the rate of mRNA turnover. Genes that may well be inhibited at the level of translation are *eno*, *pdxB*, and *sspA*. Notably ~35% of the genes (out of a total of 17) in this quadrant contain a Repetitive Extragenic Palindrome sequence (REP) element, within an intergenic region of a polycistronic mRNA or in the 3'UTR of a monocistronic transcript. The REP element has been reported to extend the half-life of the upstream mRNA by protecting the 3' end of the transcript from attack by a 3' \rightarrow 5' exonuclease [51,52]

In contrast to the first group of identified PTR genes (i.e. those in Quadrants I and IV), the group in Quadrant II represents genes of which transcription was repressed (i.e. negative AUC) by glucose, while the corresponding protein increased in abundance. Possible mechanisms underlying this type of regulation could be: (i) a strongly increased rate of translation or (ii) an extension of the half-life of the protein. Several genes in this quadrant have previously been shown to be regulated post-transcriptionally by changes in growth rate [45]. Examples are: *ggp*, *gltB*, and *rpsP*.

The abundance of proteins encoded by genes present in Quadrant III might be regulated by increased proteolysis. The rates of degradation of MinE and DnaJ are remarkably higher than those of the rest of the genes within this quadrant, considering the relatively small rate of change in their transcript level. Considering that these proteins are involved in cell division and DNA replication, respectively, tight regulation of their abundance is to be expected.

4. Discussion

Catabolite/glucose repression has already been studied for many decades, but a full understanding of this regulation mechanism is still not available [53]. Significantly, transcriptional regulation through the global regulator CRP and the concentration of cAMP only, cannot fully explain all the re-programming in *E. coli* to adjust metabolism to the availability of its preferred carbon source. In this study, the de-repressed CCR state of *E. coli* cells was established using the chemostat culturing technique, in combination with the induction of CCR by addition of a saturating pulse of glucose, as described elsewhere [22]. The fact that cells cultured in a chemostat under glucose limitation are in the carbon catabolite de-repressed state was known from previous work [54]. This chemostat-based approach of eliciting glucose repression has advantages over previous genome-wide studies of the CCR response: (i) No multiple comparisons between wild-type and mutant cells are needed, (ii) the chemostat provides a well-controlled time-independent environment, whereas batch cultures are often falsely assumed to represent a (quasi) steady state, because in batch cultures bacteria like *E. coli* experience significant variation of both their chemical- (e.g. pH; P_{O_2}) and their physical environment [55].

As shown in the Results, after switching from glucose-limited to glucose-excess conditions, expression of more than 50% of the genes of the *E. coli* MG1655 genome was changed significantly. This observed high percentage of genes with a significant change in expression level is a result of: (i) The accuracy we could achieve in the multiple measurements in the time-series analysis, and (ii) the procedure selected to initiate the CCR. Regarding accuracy: Via selection of time-series measurements, genes that show a slight but consistent increase or decrease in expression as a function of time, are distinguished more accurately from background noise than in probing at a single time-point, which has resulted in the high proportion of statistically significantly altered gene expression levels detected. Regarding the second point, the procedure that we selected for initiation of glucose repression, this causes, in parallel to the CCR response, also an increase in growth rate of the cells [22]. Therefore, a subset of the PTR genes that we have identified in this study may in fact respond to increased growth rate, rather than to glucose de/repression specifically, like e.g. some of those in Quadrant II (see also Results section).

In spite of the duplicate samples taken for the mRNA quantification, and an assay in triplicate for the proteomics samples, the relative error ranges of the protein slopes are larger than those for the AUC in the transcriptome assay, as can be deduced from the shape of the ellipses in Fig. 6A. Further reduction of these error ranges in this time-series experiment could be obtained by more frequent sampling or via the analysis of a larger number of parallel samples. The first of these options, however, would require larger chemostat vessels.

The limit value of the negative slopes (implying decrease relative protein abundance) against the AUC values (Fig. 6A) suggests that a limit value of the ordinate is observed at AUC values < -150 (see left-hand part of Fig. 6A). The change in protein abundance of these genes is due to a combination of degradation and ‘dilution’ relative to newly synthesized protein for continued growth. The limit value of the decrease of relative protein abundance after a glucose pulse is 0.008 min^{-1} , which is very close to the maximal growth rate of the cells after the glucose pulse. This suggests that dilution is a more important mode of decreased protein levels than active proteolysis and supports the idea that change in protein abundance during CCR is regulated primarily via synthesis (Figs. 5 and 6A) which is consistent with the characteristics of other physiological transitions reported previously [20,45,56,57].

Decreased expression of selected genes encoding proteins that contribute to cellular respiration was expected [58]: Faster growth requires more ribosomes for anabolism, like e.g. nucleotide-synthesizing enzymes. Therefore, at higher growth rates cells rely more on ATP production through pathways with a higher net thermodynamic driving force,

e.g. ATP-uncoupled pathways, which can catalyze with higher molecular turnover numbers [58]. The occurrence of this shift to the use of lower efficiency/higher turnover pathways is confirmed by the observation that key glycolytic genes/enzymes, especially *pfkA*/*PfkA* and *pykA*/*PykA*, are down-regulated significantly (see Supplementary Tables S1 and S4). This will allow for more expression of components involved in anabolism that then will allow faster growth [59].

In contrast to this general expectation, after ~30 min glucose-repressed *E. coli* cells exhibit a catabolic efficiency that is higher than under glucose-limited growth conditions [22], possibly because the higher glucose concentrations drive an increased flux through the still incomplete suppression of the high-efficiency catabolic pathways. Furthermore, down-regulation of the central carbon- and energy-metabolism is unlikely to be driven by the general stress response, as the master regulator of this response (*rpoS*, [60]), was down-regulated significantly in parallel with oxidative-, osmotic-, starvation-, and DNA damage-stress response genes (see Supplementary Table S1). The importance of control of CCR, beyond the level of transcription, has recently also been emphasized by Geiselmann, de Jong and others [53,61].

Significantly, recent study of Beisel and Storz (2011) has proposed a model in which the sRNA Spot 42 forms a coherent feed-forward loop with CRP and helps to regulate the expression of catabolic genes at the level of translation and/or the mRNA stability [62]. More than one hundred sRNAs have been predicted computationally to exist in *E. coli*. Of those, only 80 sRNAs have been experimentally validated [63]. To date, many sRNA targeting programs are available, e.g. CopraRNA-sRNA targeting [64], RNAPredator [65], and targetRNA [66]. However, identifying genes that are regulated by sRNA remains challenging as the sRNA and mRNA base-pairing sequences are usually very short (7–10 base pairs) [6], which may result in many false-positive predictions. The approach we have applied in this study does not allow one to distinguish between mechanisms that a cell could use to reduce transcript (nor protein) levels, i.e. whether it is due to increased mRNA degradation or due to the inhibition of transcription. Nevertheless, the mRNA molecules in *E. coli* usually have a half-life of between 3 and 8 min [67]. Therefore, genes exhibiting a very rapid decrease in their transcript level during the first 5 min after the glucose pulse, as those clustered in Profiles 0 and 8 (Fig. 2), may be subject to one of the selective mRNA-degradation mechanisms. Other candidate mechanisms are the involvement of a riboswitch [68] or of ribosome stalling [69]. Obviously, much more detailed experimental validation is necessary before all the post-transcriptional regulation mechanisms of the genes identified in this study will have been elucidated.

Transparency Document

The Transparency document associated with this article can be found, in the online version.

Acknowledgements

The authors would like to thank Dr. Jeroen van der Steen for the omics data integration. The work of O.B. was funded through a grant from the Higher Education Commission of Thailand.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.bbapap.2015.05.017>.

References

- [1] F.C. Neidhardt, I.R. Curtiss, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, et al., *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, ASM Press, Washington, D.C., 1996
- [2] H. Kitano, Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology, *Curr. Genet.* 41 (2002) 1–10.

- [3] T. Maier, M. Guell, L. Serrano, Correlation of mRNA and protein in complex biological samples, *FEBS Lett.* 583 (2009) 3966–3973.
- [4] C.S. Wadler, C.K. Vanderpool, A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide, *Proc. Natl. Acad. Sci. U. S. A.* 104 (2007) 20454–20459.
- [5] J. Johansen, M. Eriksen, B. Kallipolitis, P. Valentin-Hansen, Down-regulation of outer membrane proteins by noncoding RNAs: unraveling the cAMP-CRP- and sigmaE-dependent CyaR-ompX regulatory case, *J. Mol. Biol.* 383 (2008) 1–9.
- [6] J.B. Rice, C.K. Vanderpool, The small RNA SgrS controls sugar-phosphate accumulation by regulating multiple PTS genes, *Nucleic Acids Res.* 39 (2011) 3806–3819.
- [7] C.L. Beisel, G. Storz, Discriminating tastes: physiological contributions of the Hfq-binding small RNA Spot 42 to catabolite repression, *RNA Biol.* 8 (2011) 766–770.
- [8] G. Desnoyers, E. Masse, Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq, *Genes Dev.* 26 (2012) 726–739.
- [9] M. Liu, T. Durfee, J.E. Cabrera, K. Zhao, D.J. Jin, F.R. Blattner, Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*, *J. Biol. Chem.* 280 (2005) 15921–15927.
- [10] J. Deutscher, C. Francke, P.W. Postma, How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria, *Microbiol. Mol. Biol. Rev.* MMBR 70 (2006) 939–1031.
- [11] M.K. Oh, L. Rohlin, K.C. Kao, J.C. Liao, Global expression profiling of acetate-grown *Escherichia coli*, *J. Biol. Chem.* 277 (2002) 13175–13183.
- [12] G. Gosset, Z. Zhang, S. Nayyar, W.A. Cuevas, M.H. Saier Jr., Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*, *J. Bacteriol.* 186 (2004) 3516–3524.
- [13] P.W. Postma, J.W. Lengeler, G.R. Jacobson, Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria, *Microbiol. Rev.* 57 (1993) 543–594.
- [14] B. Gorke, J. Stulke, Carbon catabolite repression in bacteria: many ways to make the most out of nutrients, *Nat. Rev.* 6 (2008) 613–624.
- [15] H. Tao, C. Bausch, C. Richmond, F.R. Blattner, T. Conway, Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media, *J. Bacteriol.* 181 (1999) 6425–6440.
- [16] R.M. Gutierrez-Rios, J.A. Freyre-Gonzalez, O. Resendis, J. Collado-Vides, M. Saier, G. Gosset, Identification of regulatory network topological units coordinating the genome-wide transcriptional response to glucose in *Escherichia coli*, *BMC Microbiol.* 7 (2007) 53.
- [17] R. Khankal, J.W. Chin, D. Ghosh, P.C. Cirino, Transcriptional effects of CRP* expression in *Escherichia coli*, *J. Biol. Eng.* 3 (2009) (3–1611–3–13).
- [18] Y. Taniguchi, P.J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, et al., Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells, *Science* 329 (2010) 533–538.
- [19] P. Lu, C. Vogel, R. Wang, X. Yao, E.M. Marcotte, Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation, *Nat. Biotechnol.* 25 (2007) 117–124.
- [20] G. Kramer, R.R. Sprenger, M.A. Nessen, W. Roseboom, D. Speijer, L. de Jong, et al., Proteome-wide alterations in *Escherichia coli* translation rates upon anaerobiosis, *Mol. Cell. Proteomics* MCP 9 (2010) 2508–2516.
- [21] L. Nie, G. Wu, W. Zhang, Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: a multiple regression to identify sources of variations, *Biochem. Biophys. Res. Commun.* 339 (2006) 603–610.
- [22] O. Borirak, M. Bekker, K.J. Hellingwerf, Molecular physiology of the dynamic regulation of carbon catabolite repression in *Escherichia coli*, *Microbiology* 160 (2014) 1214–1223.
- [23] C.G.T. Evans, D. Herbert, D.W. Tempest, Chapter XIII The Continuous Cultivation of Micro-organisms: 2. Construction of a Chemostat, in: *Methods Microbiol.* Academic Press 1970, pp. 277–327.
- [24] S. McLean, R. Begg, H.E. Jesse, B.E. Mann, G. Sanguinetti, R.K. Poole, Analysis of the bacterial response to Ru(CO)₃C(Glycinate) (CORM-3) and the inactivated compound identifies the role played by the ruthenium compound and reveals sulfur-containing species as a major target of CORM-3 action, *Antioxid. Redox Signal.* 19 (2013) 1999–2012.
- [25] M.D. Rolfe, A. Ter Beek, A.I. Graham, E.W. Trotter, H.M. Asif, G. Sanguinetti, et al., Transcript profiling and inference of *Escherichia coli* K-12 ArcA activity across the range of physiologically relevant oxygen concentrations, *J. Biol. Chem.* 286 (2011) 10147–10154.
- [26] J. Ernst, Z. Bar-Joseph, STEM: a tool for the analysis of short time series gene expression data, *BMC Bioinforma.* 7 (2006) 191.
- [27] J. Ernst, G.J. Nau, Z. Bar-Joseph, Clustering short time series gene expression data, *Bioinforma. Oxf. Engl.* 21 (Suppl. 1) (2005) i159–i168.
- [28] C.L. Han, C.W. Chien, W.C. Chen, Y.R. Chen, C.P. Wu, H. Li, et al., A multiplexed quantitative strategy for membrane proteomics: opportunities for mining therapeutic targets for autosomal dominant polycystic kidney disease, *Mol. Cell. Proteomics* MCP 7 (2008) 1983–1997.
- [29] L. Martens, J. Vandekerckhove, K. Gevaert, DBTtoolkit: processing protein databases for peptide-centric proteomics, *Bioinforma. Oxf. Engl.* 21 (2005) 3584–3585.
- [30] M.V. Lee, S.E. Topper, S.L. Hubler, J. Hose, C.D. Wenger, J.J. Coon, et al., A dynamic model of proteome changes reveals new roles for transcript alteration in yeast, *Mol. Syst. Biol.* 7 (2011) 514.
- [31] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, Wiley 1984, pp. 72–74.
- [32] J.C. Silva, R. Denny, C. Dorschel, M.V. Gorenstein, G.Z. Li, K. Richardson, et al., Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: a sweet tale, *Mol. Cell. Proteomics* MCP 5 (2006) 589–607.
- [33] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (2008) 1367–1372.
- [34] A. Free, R.M. Williams, C.J. Dorman, The StpA protein functions as a molecular adapter to mediate repression of the bgl operon by truncated H-NS in *Escherichia coli*, *J. Bacteriol.* 180 (1998) 994–997.
- [35] A. Free, M.E. Porter, P. Deighan, C.J. Dorman, Requirement for the molecular adapter function of StpA at the *Escherichia coli* bgl promoter depends upon the level of truncated H-NS protein, *Mol. Microbiol.* 42 (2001) 903–917.
- [36] T. Wolf, W. Janzen, C. Blum, K. Schnez, Differential dependence of StpA on H-NS in autoregulation of *stpA* and in regulation of *bgl*, *J. Bacteriol.* 188 (2006) 6728–6738.
- [37] R. Grossberger, O. Mayer, C. Waldsich, K. Semrad, S. Urschitz, R. Schroeder, Influence of RNA structural stability on the RNA chaperone activity of the *Escherichia coli* protein StpA, *Nucleic Acids Res.* 33 (2005) 2280–2289.
- [38] L. Rajkowitz, R. Schroeder, Dissecting RNA chaperone activity, *RNA* 13 (2007) 2053–2060.
- [39] A. Zhang, S. Rimsky, M.E. Reaban, H. Buc, M. Belfort, *Escherichia coli* protein analogs StpA and H-NS: regulatory loops, similar and disparate effects on nucleic acid dynamics, *EMBO J.* 15 (1996) 1340–1349.
- [40] R. Srinivasan, V.F. Scolari, M.C. Lagomarsino, A.S.N. Seshasayee, The genome-scale interplay amongst xenogene silencing, stress response and chromosome architecture in *Escherichia coli*, *Nucleic Acids Res.* 43 (2015) 295–308.
- [41] S. Seeto, L. Notley-McRobb, T. Ferenci, The multifactorial influences of RpoS, Mlc and cAMP on ptsG expression under glucose-limited and anaerobic conditions, *Res. Microbiol.* 155 (2004) 211–215.
- [42] K. Matsushita, T. Ohnishi, H.R. Kaback, NADH-ubiquinone oxidoreductases of the *Escherichia coli* aerobic respiratory chain, *Biochemistry (Mosc)* 26 (1987) 7732–7737.
- [43] P. Sharma, M.J. Teixeira de Mattos, K.J. Hellingwerf, M. Bekker, On the function of the various quinone species in *Escherichia coli*, *FEBS J.* 279 (2012) 3364–3373.
- [44] M.H. Serres, M. Riley, MultiFun, a multifunctional classification scheme for *Escherichia coli* K-12 gene products, *Microb. Comp. Genomics* 5 (2000) 205–222.
- [45] K. Valgepea, K. Adamberg, A. Seiman, R. Vilu, *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins, *Mol. Biosyst.* 9 (2013) 2344–2358.
- [46] G. Kramer, R.R. Sprenger, J. Back, H.L. Dekker, M.A. Nessen, J.H. van Maarseveen, et al., Identification and quantitation of newly synthesized proteins in *Escherichia coli* by enrichment of azidohomoalanine-labeled peptides with diagonal chromatography, *Mol. Cell. Proteomics* MCP 8 (2009) 1599–1611.
- [47] G. Baughman, M. Nomura, Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in *Escherichia coli*, *Cell* 34 (1983) 979–988.
- [48] D. Dean, J.L. Yates, M. Nomura, Identification of ribosomal protein S7 as a repressor of translation within the *str* operon of *E. coli*, *Cell* 24 (1981) 413–419.
- [49] N. De Lay, S. Gottesman, The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior, *J. Bacteriol.* 191 (2009) 461–476.
- [50] T. Inada, Y. Nakamura, Autogenous control of the *subB* gene expression of *Escherichia coli*, *Biochimie* 78 (1996) 209–212.
- [51] S.F. Newbury, N.H. Smith, E.C. Robinson, I.D. Hiles, C.F. Higgins, Stabilization of translationally active mRNA by prokaryotic REP sequences, *Cell* 48 (1987) 297–310.
- [52] S.F. Newbury, N.H. Smith, C.F. Higgins, Differential mRNA stability controls relative gene expression within a polycistronic operon, *Cell* 51 (1987) 1131–1143.
- [53] A. Kremling, J. Geiselmann, D. Ropers, H. de Jong, Understanding carbon catabolite repression in *Escherichia coli* using quantitative models, *Trends Microbiol.* 23 (2015) 99–109.
- [54] U. Lendenmann, T. Egli, Is *Escherichia coli* growing in glucose-limited chemostat culture able to utilize other sugars without lag? *Microbiol. Read. Engl.* 141 (Pt 1) (1995) 71–78.
- [55] M. Bekker, G. Kramer, A.F. Hartog, M.J. Wagner, C.G. de Koster, K.J. Hellingwerf, et al., Changes in the redox state and composition of the quinone pool of *Escherichia coli* during aerobic batch-culture growth, *Microbiol. Read. Engl.* 153 (2007) 1974–1980.
- [56] A. Boysen, J. Moller-Jensen, B. Kallipolitis, P. Valentin-Hansen, M. Overgaard, Translational regulation of gene expression by an anaerobically induced small non-coding RNA in *Escherichia coli*, *J. Biol. Chem.* 285 (2010) 10690–10702.
- [57] R. Shalgi, J.A. Hurt, I. Krykbaeva, M. Taipale, S. Lindquist, C.B. Burge, Widespread regulation of translation by elongation pausing in heat shock, *Mol. Cell* 49 (2013) 439–452.
- [58] A. Flamholz, E. Noor, A. Bar-Even, W. Liebermeister, R. Milo, Glycolytic strategy as a tradeoff between energy yield and protein cost, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 10039–10044.
- [59] D. Molenar, R. van Berlo, D. de Ridder, B. Teusink, Shifts in growth strategies reflect tradeoffs in cellular economics, *Mol. Syst. Biol.* 5 (2009) 323.
- [60] H. Weber, T. Polen, J. Heuveling, V.F. Wendisch, R. Hengge, Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity, *J. Bacteriol.* 187 (2005) 1591–1603.
- [61] S. Berthoumioux, H. de Jong, G. Baptist, C. Pinel, C. Ranquet, D. Ropers, et al., Shared control of gene expression in bacteria by transcription factors and global physiology of the cell, *Mol. Syst. Biol.* 9 (2013) 634.
- [62] C.L. Beisel, G. Storz, The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*, *Mol. Cell* 41 (2011) 286–297.
- [63] R. Raghavan, E.A. Groisman, H. Ochman, Genome-wide detection of novel regulatory RNAs in *E. coli*, *Genome Res.* 21 (2011) 1487–1497.
- [64] P.R. Wright, A.S. Richter, K. Papenfort, M. Mann, J. Vogel, W.R. Hess, et al., Comparative genomics boosts target prediction for bacterial small RNAs, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) E3487–E3496.

- [65] F. Eggenhofer, H. Tafer, P.F. Stadler, I.L. Hofacker, RNAPredator: fast accessibility-based prediction of sRNA targets, *Nucleic Acids Res.* 39 (2011) W149–W154.
- [66] B. Tjaden, TargetRNA: a tool for predicting targets of small RNA action in bacteria, *Nucleic Acids Res.* 36 (2008) W109–W113.
- [67] J.A. Bernstein, A.B. Khodursky, P.H. Lin, S. Lin-Chao, S.N. Cohen, Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 9697–9702.
- [68] E. Nudler, A.S. Mironov, The riboswitch control of bacterial metabolism, *Trends Biochem. Sci.* 29 (2004) 11–17.
- [69] K. Ito, S. Chiba, K. Pogliano, Divergent stalling sequences sense and control cellular physiology, *Biochem. Biophys. Res. Commun.* 393 (2010) 1–5.