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.

Published in:
Biochimica et Biophysica Acta. Proteins and Proteomics

DOI:
10.1016/j.bbapap.2015.05.017

Citation for published version (APA):
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,b,c

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

**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 CRP regulator CRP. Indeed, the cAMP–CRP complex modulates expression of many catabolic genes [10,13], and cAMP levels in the cells are under control of the glucose-specific part of the Phosphoenolpyruvate-dependent Phospho-Transferase system (‘glucose-PTS’), via the level of phosphorylation of Enzyme IIAGlc [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
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 21 chemostat vessels (Applikon, The Netherlands) with a working volume of 11 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 nitrito-acetic acid as a chelator, 0.17 mM Na₂SeO₃, and 20 mM glucose was used. For a reference culture, ¹⁵NH₄Cl (98 atom % ¹⁵N; Sigma Aldrich) was used as the N source, instead of NH₄Cl. 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 × 15 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.](image-url)
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 log2 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 μl/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 × 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 a concentrated stock solution and the latter mix was from Roche), 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 KH2PO4, 25% (v/v) ACN, pH 2.9 and solvent B; 10 mM KH2PO4, 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 C18 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 15N, 13N 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 14N, 15N tryptic peptide mixture, were injected as a 10 μl 0.1% (v/v) TFA aqueous solution and loaded onto the PepMap100 C18 (5-μm particle size, 100-Å pore size, 300-μm inner diameter × 5 mm length) pre-column. The peptides were eluted via an Acclaim PepMap 100 C18 (3-μm particle size, 100-Å pore size, 75-μm inner diameter × 250 mm length) analytical column (Thermo Scientific, Etten-Leur, The Netherlands) using a linear gradient from 0.1% formic acid/6% CH3CN/94% H2O (v/v) to 0.1% formic acid/40% CH3CN/60% H2O (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/Δm). 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/Δm). 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 15N labeling method, to enable MASCOT to identify both 14N and 15N 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 I/H; Integration method: Simsons; 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 old 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 <PXD0000xx>.

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 3, 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 log2 ratios of the transcripts as a function of time (compare [30]). For each biological replicate, 8 possible time profiles (i.e.: 2 × 2 × 1 × 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 $^{14}$N/$^{15}$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 $^{14}$N/$^{15}$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 transcriptomic 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 $\chi^2$ distribution with 2 degrees of freedom was calculated [31]. This is the elliptic region ($p < 0.01$) as shown in Fig. 6A. The outbound 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 µM [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 ($\mu$) 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
Post-transcriptionally regulated genes reported in Kramer et al., 2009, Valgepea et al., 2013, Baughman and Nomura, 1983, Dean et al., 1981, and De Lay and Gottesman, 1983, within an intergenic region of an operon or to their cellular functions.

Cell division and DNA replication
- minE Cell division topological specificity factor
  - III
- dnaG Chaperone protein DnaG
  - III

Translation
- rpsA DNA-directed RNA polymerase subunit α
  - IV
- rplA 50S ribosomal protein L6
  - I
- rplM 50S ribosomal protein L19
  - II
- rplP 50S ribosomal protein L13
  - I
- rpsQ 50S ribosomal protein L19
  - II
- rpsX 50S ribosomal protein L24
  - I
- rpmA 50S ribosomal protein L27
  - I
- rpmC 50S ribosomal protein L33
  - I
- rpmD 30S ribosomal protein S7
  - I
- rpmE 30S ribosomal protein S16
  - I

Central carbon metabolism
- emo Enolase
  - IV
- acnB Phosgenepyruvate carboxylase
  - II

Membrane proteins
- nuoC NADH-quinone oxidoreductase subunit C/D
  - III
- ompT Outer membrane protein X
  - I
- potD Periplasmic substrate-binding protein
  - II

ATP-binding protein
- yjjK Uncharacterized ABC transporter

Other metabolisms
- dahA PTS-dependent dihydroxyacetone kinase, dihydroxyacetone-binding subunit dahA

Post-transcriptionally regulated genes identified in this study and grouped according to their cellular functions.

<table>
<thead>
<tr>
<th>GN</th>
<th>Function</th>
<th>Quadrant</th>
<th>Slope</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>arlM</td>
<td>Shikimate kinase 1</td>
<td>I</td>
<td>0.0039</td>
<td>24.2</td>
</tr>
<tr>
<td>cyoA</td>
<td>Sulfate/thiosulfate import ATP-binding protein CysA</td>
<td>I</td>
<td>0.0175</td>
<td>80.0</td>
</tr>
<tr>
<td>cyoD</td>
<td>Sulfite reductase [NADPH] hemoprotein β-component</td>
<td>I</td>
<td>0.0075</td>
<td>108.1</td>
</tr>
<tr>
<td>cyoP</td>
<td>Thiolsulfate-binding protein</td>
<td>I</td>
<td>0.0054</td>
<td>102.9</td>
</tr>
<tr>
<td>gltA</td>
<td>Glutamine synthetase</td>
<td>I</td>
<td>0.0033</td>
<td>87.6</td>
</tr>
<tr>
<td>gatA</td>
<td>Glutamate synthase [NADPH] large chain</td>
<td>I</td>
<td>0.0099</td>
<td>47.3</td>
</tr>
<tr>
<td>merB</td>
<td>Cystathionine γ-synthase</td>
<td>I</td>
<td>0.0046</td>
<td>124.7</td>
</tr>
<tr>
<td>merC</td>
<td>Cystathionine β-isomerase</td>
<td>IV</td>
<td>-0.045</td>
<td>34.6</td>
</tr>
<tr>
<td>merD</td>
<td>M-dihomocystine-binding lipoprotein metQ</td>
<td>I</td>
<td>0.018</td>
<td>56.9</td>
</tr>
<tr>
<td>fibV</td>
<td>Cystine-binding periplasmic protein-SynR</td>
<td>I</td>
<td>0.0005</td>
<td>55.4</td>
</tr>
</tbody>
</table>

Post-transcriptionally regulated genes identified in Kramer et al., 2009, Valgepea et al., 2013, Baughman and Nomura, 1983, Dean et al., 1981, and De Lay and Gottesman, 1983, within an intergenic region of an operon or to their cellular functions.
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 replicates, 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 TuF$\alpha$ 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 who’s 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 OmpA 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 (GxdD) 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 EI\text{IA} (crr), and EI\text{ICB} (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. 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 log2 of a gene expression level. Only significant clusters were 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.](image-url)
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(AUC/55); see Supplementary Fig. S2). This linearity between the slopes and 2AUC 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 $\chi^2$ 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
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 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 $p < 0.01$ (Table 1) and they are joined by another 45 genes at the significance level of $p < 0.05$ (Supplementary Table S4).

The genes identified using this approach with the significance level of $p < 0.01$ 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, sopA, ompX, cysA, and gltB, whereas feedback regulation was observed for genes encoding the RNA polymerase α subunit (rpoA), pyridine nucleotide transhydrogenase β subunit (ptnB), lysyl-tRNA synthetase (lysS), and amino acid periplasmic-binding proteins (metQ and flu). 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 [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 RNaseIl 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 sopA. 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′→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: gpp, 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; PO2) 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 before all the post-transcriptional regulation mechanisms of the genes 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-paring 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.jbmbap.2015.05.017.

References


