*Scientific paper*

# Gene Expression Profiling of Recombinant Protein Producing *E. coli* at Suboptimal Growth Temperature

**Mitja Mahnič,[1,*] Špela Baebler,[3] Andrej Blejec,[3] Špela Jalen,[4] Kristina Gruden,[3] Viktor Menart and Simona Jevševar[2]**

[1] *Development center Slovenia, Lek Pharmaceuticals d.d., a Sandoz company, Verovškova 57, 1526 Ljubljana, Slovenia*

[2] *Sandoz Biopharmaceuticals Mengeš, Lek Pharmaceuticals d.d., a Sandoz company, Kolodvorska 27, 1234 Mengeš, Slovenia*

[3] *Department of Biotechnology and Systems Biology, National Institute of Biology, Večna pot 111, 1000 Ljubljana, Slovenia*

[4] *Laboratory for Biosynthesis and Biotransformation, National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia*

** Corresponding author: E-mail: mitja.mahnic@sandoz.com; Phone: + 386 1 5803275*

*Received: 07-03-2011*

***Dedicated to inspiring mentor dr. Viktor Menart.***

## Abstract

Recent studies have revealed that at lower cultivation temperatures (25 °C) much higher percentage of correctly folded recombinant hG-CSF protein can be extracted from inclusion bodies. Hence, the goal of our research was to investigate mechanisms determining characteristics of non-classical inclusion bodies production using gene expression profiling, focusing on proteases and chaperones gene expression. Statistical analysis of microarray data showed prominent changes in energy metabolism, in metabolism of amino acids and nucleotides, as well as in biosynthesis of cofactors and secondary metabolites if the culture was grown below its optimal temperature. Moreover, 24 differentially expressed up to now known genes classified among proteases, chaperones and other heat or stress related genes. Among chaperones UspE and among proteases YaeL and YeaZ might play an important role in accumulation of correctly folded recombinant proteins. Membrane localized protease *yaeL* gene was found to have higher activity at 25 °C and is thus potentially functionally related to the more efficient recombinant protein production at lower temperatures. The results of this study represent advance in the understanding of recombinant protein production in *E. coli*. Genes potentially influencing production of recombinant protein at lower growth temperature represent basis for further research towards improvement of *E. coli* production strains as well as fermentation process.

**Keywords:** Recombinant protein production, non-classical inclusion bodies, *E. coli*, expression microarrays, YaeL protease, GroEL chaperone

## 1. Introduction

The *„art nouveau"* of the modern biopharmaceutical industry is getting towards understanding of the mechanisms underlying recombinant protein production in different organisms. The obtained knowledge should further contribute to improvements of fermentation processes and thus gain economical advantages for the industry.

*Escherichia coli* is the most widely used recombinant protein producing organism due to its ease of cultivation and fast production rate. Protein misfolding is a common event during bacterial over-expression of recombinant genes.[1] Incorrectly folded or misfolded proteins can appear as a result of cell exposure to the environmental stress, such as elevated temperatures and over-expression of recombinant genes. The resulting misfolded proteins may be degraded by proteases, fold by chaperones, or aggregated and sequestered as inclusion bodies (IBs).[2] Hence, a common limitation of recombinant protein production in bacteria is the formation of insoluble protein aggregates known as IBs.[3] It has been believed for a long time that IB proteins are biologically inactive and therefore

undesired in bioprocesses.[4] The potential of chaperones in assisting folding of misfolded proteins has been investigated from several aspects.[5] On the other hand, it has already been reported that functional proteins could be easily extracted from IBs using non denaturing mild detergents and polar solvents provided that cultures were grown at lower temperatures.[3,4,6–9] Such IBs, termed „non-classical" inclusion bodies (ncIBs) by Jevševar et al. (2005),[7] are defined by containing large amount of correctly folded protein precursor produced in *E. coli* at lower temperature (around 25 °C). Compared to classical IBs they are characterized by higher fragility and solubility, irreversible contraction at acidic pH and most importantly, by a high amount of correctly folded target protein or its precursor.[6]

One of the most important recombinant proteins in the field of modern oncology is human granulocyte colony stimulating factor (hG-CSF) protein. Due to its regulatory role in the growth, differentiation, survival, and activation of neutrophils and their precursors, hG-CSF is central to neutrophil-based immune defenses[1]. Four types of hG-CSF are clinically available: a glycosylated form (lenograstim) produced in CHO cells, an N-terminal replaced nonglycosylated form of granulocyte colony-stimulating factor (nartograstim),[10] and nonglycosylated form (filgrastim), both produced by using the expression in *E. coli.*[1] In addition to aforementioned forms long acting form of filgrastim – PEGfilgrastim, a modified PEGylated filgrastim enabling less frequent administration has been available since 2002.

As shown before[7] cultivation temperature was the most important variable affecting properties of hG-CSF IBs and thus its efficient production. Therefore the goal of our research was to investigate mechanisms determining characteristics of ncIBs production by comparing physiology of recombinant *E. coli* [BL21 (DE3)] at three different temperatures (T = 25 °C (suboptimal), 37 °C (optimal) and 42 °C (heat shock), respectively) using gene expression profiling approach. As formation of various proteases and chaperones under different temperature conditions was previously reported[6–8,11] we have inspected behavior of these genes in more detail.

# 2. Experimental

## 2. 1. Cultures and Plasmids

In this study the recombinant *E. coli* strain BL21 (DE3) (Novagen), carrying expression plasmid pET3a without hG-CSF insert (control strain) or with hG-CSF insert [Fopt5] (production strain) was used. hG-CSF insert ([Fopt5]) was prepared as described in[7].

## 2. 2. Culture Conditions

Bacterial inoculum of the production and control strain was prepared in a shake flask culture and grown overnight at 25 °C and at 160 rpm in the LBPG/amp100 medium[7]. After reaching optical density of $OD_{600nm} \approx 4$ the inoculum was transferred to the GYSP medium and immediately induced with IPTG. The cultures were then incubated in shake flasks at 160 rpm and at three different temperatures (i.e. T = 25 °C, 37 °C and 42 °C, respectively) until the appropriate culture´s optical density (OD), indicating the transition from the exponent to the stationary phase was reached ($OD_{600nm} \approx 10$ for the culture grown at 25 °C and $OD_{600\,nm} \approx 4$ for the cultures grown at 37 °C or 42 °C).[7] At that point cultures were stabilized with RNA protect Bacteria Reagent (Qiagen), aliquoted, centrifuged and the bacterial pellet was stored at – 80 °C for further RNA and protein expression analysis. The cultivation experiment was repeated 3 times, thus yielding 18 samples altogether.

## 2. 3. Isolation of Total RNA and DNase Treatment

RNA isolation and DNase treatment was performed as described by Petek et al. (2010),[12] except for substituting lysostaphin with lysocyme (500 mg/ml) in the cell lysis step. RNA quality, quantity and integrity were checked by NanoDrop (NanoDrop Technologies, USA), gel electrophoresis and Bioanalyzer (Agilent Technologies).

## 2. 4. Microarray Hybridization

Purified RNA (approximately 30 μg) was used for the cDNA synthesis and direct labeling (Superscript II, Invitrogen). Luciferase control mRNA (1 ng/μg; Promega) and 3 μg of random primers were added to each RNA sample. This was followed by 10 minutes of incubation at 70 °C and immediate chilling on ice. cDNA synthesis was carried out using SuperScript II reverse transcriptase (Invitrogen) according to manufacturer's instructions. Synthesized cDNA was purified using MinElute PCR purification Kit (Qiagen). The concentration of cDNA, efficiency of dye (Cy5) integration and integrity of labeled cDNA were checked by NanoDrop and gel electrophoresis. Pre-designed oligonucleotide microarrays (CustomArray™ 12K Microarray, Combimatrix Corporation) containing 12,000 features arrays of complete *E. coli* genome (4,200 genes, positive and negative control sequences) were used. Labeled cDNA was hybridized to the arrays according to the protocol recommended by CombiMatrix except for using 2X formamide based hybridization buffer (Genisphere) containing Salmon testis DNA (1 μg / μl, Sigma) and shorter hybridization time (1h).

## 2. 5. Microarray Imaging and Data Analysis

After hybridization semiconductor microarray surfaces were covered by imaging solution and were scanned using a fluorescence LS200 scanner (TECAN).[13] Combi-

matrix Microarray Images Software was used for image analysis and quality control. Further data analysis was performed in R software environment for statistical computing and graphics (http://www.r-project.org/). Bioconductor´s packages *affy, limma and KEGGsoap* were used for quality control, preprocessing, statistical significance testing[14] and annotation. The data was normalized using the quantile normalization. Intensities of the factory-built in control probes were compared to information on Combimatrix FAQ internet site to confirm the validity of the chosen preprocessing approach. Differentially expressed transcripts were functionally analysed according to Gene Ontology (GO) using GSEA[15]. Normalized data was further analyzed for significance using the linear models with different contrast settings and empirical Bayes ($p<0.05$). To minimize the possibility of false positive results, all $\log_2$ values of gene expression ratios between –0.5 and 0.5 were considered not relevant and were excluded from the data interpretation[16]. The final results were expressed as $\log_2$ values (logFC) of the ratios between the mean expression in sample groups. Moreover, the dataset was defined by GeneID identification tags for easier access to different knowledge databases. Differentially expressed genes (DE) were visualized in the EcoCyc database (http://www.ecocyc.org/expression.html) that allows representation of the obtained results on the metabolic and signaling pathways.

The microarray data have been deposited in NCBI´s Gene expression Omnibus and are accessible through GEO Series accession number GSE25561 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25561).

## 2. 6. Real-Time PCR

Six DE genes *yeaZ*, *ydcP*, *groEL*, *ecpD*, *torD* and *uspD (yiiT)* were selected for real-time PCR analysis based on TaqMan® MGB™ technology[17]. 16S rRNA was used as the reference gene. Gene specific sequences were chosen for assay design using NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The assays were designed by Applied Biosystems, primer and probe sequences are listed in Table 1. Total RNA (approximately 3 µg) was reverse transcribed using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to manufactures instructions.

Real-time PCR reaction were set up as described in Petek et al. (2010)[12] in ABI PRISM 7900 HT Fast Sequence Detection System (Applied Biosystems) using 5 µl reactions and standard cycling parameters. Data quality control and analysis was performed as described in Petek et al. (2010).[12] For the purpose of comparison of qPCR and microarray data, the results were expressed as the $\log_2$ of the ratio between relative gene expressions at two different temperatures for the control and production strain and as the ratio between relative gene expressions in control and production strain grown at the same temperature. Statistical significance of differences in gene expression was calculated using the same model as in analysis of microarray data.[13]

## 2. 7. Western Blotting

SDS-PAGE was performed using a 4–12% NuPAGE$^R$ Novex Bis-Tris gel (Invitrogen) according to manufacturer´s protocols. Prior to electrophoresis all samples were resuspended in 10 mM TRIS/HCl (pH = 8.0) buffer and diluted according to their final OD to obtain similar sample loads. Samples were further treated by addition of NuPAGE$^R$ LDS sample buffer, denatured for 10 minutes at T = 70 °C and applied to gel. Electrophoresis was performed at 200 V, 125 mA for 40 minutes at room temperature. Proteins that were separated with SDS-PAGE were afterwards transferred onto the Nitrocellulose membrane by using iBlot™ Dry blotting System (Invitrogen). Immunodetection was made by primary antibody (GroEL antibody mouse Monoclonal, IgG1, Antibodies-online GmbH) followed by secondary antibodies (Anti-mouse IgG – HRP, Sigma). Colorimetric detection was achieved with addition of detection solution (mixture of solution A (15 mg of 4-chloro-1-naftol dissolved in 5 ml of Methanol) and solution B (15 ml of $H_2O_2$ added to 25 ml of TBS (pH = 7.5). At the end the membrane was imaged and obtained images were further analyzed by ImageJ program (Image Processing and Analysis in Java, http://rsbweb.nih.gov/ij/) used for optical density measurements. We semi-quantified all proteins from electrophoresis gel images and GroEL from western blotting membranes images and determine the relative GroEL content in different samples.

**Table 1:** Primer and probe sequences used for real-time PCR

| Gene | Forward primer | Reverse primer | Probe |
|---|---|---|---|
| *yeaZ* | CGCTGATATTCGGCCCAGTAAA | GTGCTGGCAGCCATTGAC | CTTCGCCCATTCGCG |
| *ydcP* | GATATTGGCGCGTTCGATTCG | GAGATGATCTTTCGCCACTTTCAAT | CAGGCCGATAAATTT |
| *groEL* | GCAACTCTGGTTGTTAACACCAT | TGCAGCATAGCTTTACGACGAT | AACCGCAGCGACTTT |
| *ecpD* | GAACACGCTCTCTCTGTCTTTAGG | CAAACGTGGGCAAACAATCAAATT | ACAGCCAGCACCTCAC |
| *torD* | ACAGGACGAGCAAGAGATTAAACG | CGTTGAAATTGCCGCTGGTTT | CCCTGCCTCAACTAAC |
| *uspD* | CGGATAACAAGCTGTATAAACTGACGAA | GCATTTCTCCGCGTTCAATACG | TCGGCCATTGAATATT |
| 16S rRNA | GGAGTACGGCCGCAAGGT | CATGCTCCACCGCTTGTG | AAAACTCAAATGAATTGACG |

# 3. Results

It is known that cultivation at T = 25 °C yields high amounts of correctly folded hG-CSF protein within the IBs[7]. 98% of the expressed hG-CSF was in the form of IBs and only 2% was produced in the cytoplasm.[8] In these conditions, recombinant hG-CSF protein accumulated to yield 35% to 40% of the total proteins of *E. coli*, almost 50% of hG-CSF extracted from the IBs showed biological activity.[7,8] In order to investigate the underlying mechanisms of this phenomenon the experimental design was set to profile *E. coli* gene expression at different cultivation temperatures. Therefore, control (recombinant strain carrying empty expression plasmid) and production strain (recombinant strain carrying expression plasmid containing hG-CSF) were incubated at three different temperatures, 25 °C, 37 °C and 42 °C, respectively. The cultures were sampled just before transition into stationary state, when the cells were still fully viable and already under stress of recombinant protein production. At this point maximal plasmid copy number was determined as well as maximal productivity was achieved – accumulation level of hG-CSF reached plateau.[7,18] After initial data analysis special attention was given to the temperature dependent changes in expression levels of proteases, chaperones and other heat or stress related genes.

## 3. 1. Overview of Identified Gene Expression Differences

Linear models were set to identify DE genes in comparisons of the following group pairs: control strain cultures grown at 37 °C compared to 25 °C (C37_25), control strain cultures grown at 42 °C compared to 25 °C (C42_25), production strain cultures grown at 37 °C compared to 25 °C (P37_25), production strain cultures grown at 42 °C compared to 25 °C (P42_25) as well as comparison of control and production strain cultures grown at 25 °C (P_C25), 37 °C (P_C37), 42 °C (P_C42). 282 DE genes showed statistically significant differences in expression levels if comparing control strain cultures grown at T = 37 °C and T = 25 °C and 226 DE genes if comparing control strain cultures grown at T = 42 °C and T = 25 °C. In production strain 28 and 34 DE genes showed changes in expression levels at T = 37 °C and T = 42 °C compared to T = 25 °C, respectively. Comparison in gene expression profiles of control and production strain grown at the same cultivation temperature identified on average 113 DE genes; i.e. from 74 to 161 DE genes at all studied temperatures. Additionally, two-way analysis of variance was performed to test for the effects of recombinant protein production (P_C) and growth temperature as well as interaction of both factors (Supporting information 1). The percentage of DE genes identified was similarly up to about 7% (i.e. up to about 305 DE genes) as in pair-wise comparisons (Fig.1). When comparing control

**Table 2:** Functional analysis of changes in gene expression using GSEA. GO terms identified as significantly enriched ($p < 0.05$) are presented. Up and down regulated genes were analysed separately. P-production strain, C-control strain, 37_25, 42_37, 42_25-comparison of DE genes identified due to different growth temperature.

| | Comparison | | Comparison | |
|---|---|---|---|---|
| | C37_25 | | P37_25 | |
| | Up-regulated | Down-regulated | Up-regulated | Down-regulated |
| GO terms classes | 0005198 Structural molecule activity | 0003676 Nucleic acid binding | 0006144 Purine base metabolic process | 0003723 RNA binding |
| | 0005576 Extracellular region | 0003887 DNA-directed DNA polymerase activity | 0009086 Methionine biosynthetic process | 0003735 Structural constituent of ribosome |
| | 0006096 Glycolysis | | | 0005840 Ribosome |
| | 0008198 Ferrous iron binding | 0008080 N-acetyltransferase activity | 0043190 ATP-binding cassette (ABC) transporter complex | 0006412 Translation |
| | 0009245 Lipid A biosynthetic process | 0019843 rRNA binding | | 0006935 Chemotaxis |
| | 0010181 FMN binding | | | 0009103 Lipopolysaccharide biosynthetic process |
| | | | | 0009279 Cell outer membrane |
| | | | | 0015288 Porin activity |
| | | | | 0015453 Oxidoreduction-driven active transmembrane transporter activity |
| | | | | 0019843 rRNA binding |
| | | | | 0030529 Ribonucleoprotein complex |
| | | | | 0030964 NADH dehydrogenase complex |
| | | | | 0046930 Pore complex |
| | | | | 0048038 Quinone binding |

| | C42_25 | | P42_25 | |
|---|---|---|---|---|
| | **Up-regulated** | **Down-regulated** | **Up-regulated** | **Down-regulated** |
| **GO terms classes** | 0005624 Membrane fraction<br>0005829 Cytosol<br>0006096 Glycolysis<br>0006457 Protein folding<br>0006950 Response to stress<br>0009408 Response to heat<br>0042802 Identical protein binding<br>0045454 Cell redox homeostasis<br>0051082 Unfolded protein binding<br>0051301 Cell division | 0008152 Metabolic process | 0000156 Two-component response regulator activity<br>0000160 Two-component signal transduction system (phosphorelay)<br>0004812 Aminoacyl-trna ligase activity<br>0006457 Protein folding<br>0008233 Peptidase activity<br>0009408 Response to heat<br>0042803 Protein homodimerization activity<br>0045454 Cell redox homeostasis<br>0051082 Unfolded protein binding | 0000049 tRNA binding<br>0003723 RNA binding<br>0003735 Structural constituent of ribosome<br>0003924 GTPase activity<br>0005525 GTP binding<br>0005840 Ribosome<br>0006412 Translation<br>0009060 Aerobic respiration<br>0009279 Cell outer membrane<br>0015288 Porin activity<br>0015453 Oxidoreduction-driven active transmembrane transporter activity<br>0016597 Amino acid binding<br>0019843 rRNA binding<br>0030529 Ribonucleoprotein complex<br>0046930 Pore complex |
| | **C42_37** | | **P42_37** | |
| | **Up-regulated** | **Down-regulated** | **Up-regulated** | **Down-regulated** |
| **GO terms classes** | 0003677 DNA binding<br>0004518 Nuclease activity<br>0005506 Iron ion binding<br>0006260 DNA replication<br>0006457 Protein folding<br>0006865 Amino acid transport<br>0030170 Pyridoxal phosphate binding<br>0042802 Identical protein binding<br>0051082 Unfolded protein binding<br>0055114 Oxidation-reduction process | 0000287 Magnesium ion binding<br>0004519 Endonuclease activity<br>0005198 Structural molecule activity<br>0005215 Transporter activity<br>0005529 Sugar binding<br>0005975 Carbohydrate metabolic process<br>0006310 DNA recombination<br>0006811 Ion transport<br>0008033 tRNA processing<br>0008152 Metabolic process<br>0008198 Ferrous iron binding<br>0008237 Metallopeptidase activity<br>0008565 Protein transporter activity<br>0008643 Carbohydrate transport<br>0009279 Cell outer membrane<br>0009306 Protein secretion<br>0009425 Bacterial-type flagellum basal body | 0005576 Extracellular region<br>0005624 Membrane fraction<br>0006071 Glycerol metabolic process<br>0006457 Protein folding<br>0006935 Chemotaxis<br>0007049 Cell cycle<br>0008033 tRNA processing<br>0008198 Ferrous iron binding<br>0008270 Zinc ion binding<br>0008565 Protein transporter activity<br>0009103 Lipopolysaccharide biosynthetic process<br>0009289 Pilus<br>0009306 Protein secretion<br>0030145 Manganese ion binding<br>0045454 Cell redox homeostasis<br>0051082 Unfolded protein binding | 0000299 Integral to membrane of membrane fraction<br>0003676 Nucleic acid binding<br>0005840 Ribosome<br>0006412 Translation<br>0006633 Fatty acid biosynthetic process<br>0006814 Sodium ion transport<br>0008652 Cellular amino acid biosynthetic process<br>0009086 Methionine biosynthetic process<br>0016779 Nucleotidyltransferase activity<br>0019843 rRNA binding<br>0020037 Heme binding<br>0030529 Ribonucleoprotein complex<br>0050660 Flavin adenine dinucleotide binding<br>0051539 4 iron, 4 sulfur cluster binding |

strain cultures grown at T = 37 °C and T = 42 °C to cultures grown at T = 25 °C the ratio between up-regulated and down-regulated genes was on average 50 : 50 (35 : 31 DE genes, respectively) as it was true also for the gene expression profiles of the production strain (data not shown). When comparing control and production strain cultures grown at the same temperature the ratio between up-regulated and down-regulated genes was on average 65 : 35 in favor of down-regulated at T = 25 °C (7 : 3 DE genes, respectively) and T = 42 °C (11 : 9 DE genes, respectively), respectively and 55 : 45 (7 : 6 DE genes, respectively) in favor of up-regulated at T = 37 °C (Supporting information 2).

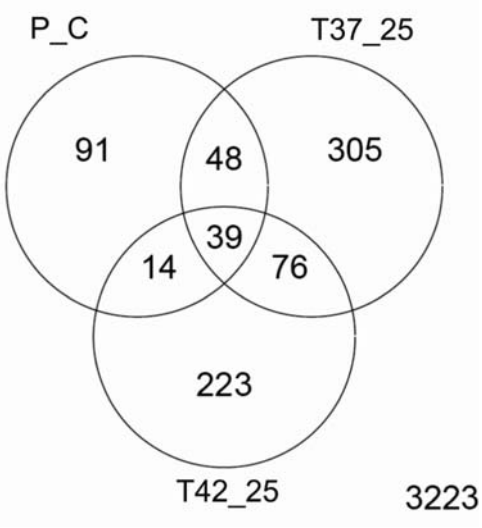



**Figure 1:** Venn diagram of differentially expressed genes identified by analysis of variance. P_C: strain dependent changes in gene expression, T37_25, T42_25 – temperature dependent changes in gene expression. p-value cut-off 0.05.

For easier biological interpretation functional analysis of gene expression changes was performed using GSEA (Table 2). To gain a better overview of the studied processes DE genes were visualized in EcoCyc Omics viewer which is based on its own functional ontology system (Supporting information 3). Prominent changes in energy metabolism, in metabolism of amino acids and nucleotides, as well as in biosynthesis of cofactors and secondary metabolites if the culture was grown below its optimal temperature were observed. Some DE genes are annotated as transporters. Also some signaling pathways related genes were identified as DE if comparing strains grown at different temperatures. If comparing production to control strain grown at suboptimal temperature genes identified as DE are related to carbohydrates biosynthesis, amino acids biosynthesis and in biosynthesis of cofactors, while at higher growth temperatures (T = 37 °C and T = 42 °C) more general changes in energy metabolism, metabolism of amino acids and secondary products were detected (Supporting information 2).

## 3. 2. Gene Expression Analysis of Chaperones, Proteases and Other Stress Related Genes

According to EcoCyc and KEGG databases and literature survey,[1,19,20] 131 of *E. coli´*s genes are coding for molecular chaperones, proteases, sigma factors and other heat or stress related proteins.

**Table 3:** Temperature dependent changes in gene expression of chaperones, proteases and other heat shock and stress induced proteins in recombinant E. coli BL21 (DE3) strain: Genes identified as DE and having negative logFC (log2) ratio in particular comparison are indicated as 'down', while DE genes with positive logFC ratio are indicated as 'up' regulated. Other classification is aligned with EcoCyc classification of genes. Comparisons of gene expression of production strain grown at different temperatures (P42_25, P37_25) and control strain grown at different temperatures (C42_25, C37_25) are presented. Functional classification of proteases was performed according to EcoCyc (http://www.ecocyc.org/expression.html).

| EcoCyc ID | Gene (ecoid) | Protein function | Comparison | | | |
|---|---|---|---|---|---|---|
| | | | P37_25 | P42_25 | C37_25 | C42_25 |
| **Chaperones and other stress induced proteins** | | | | | | |
| EG10599 | *groEL* | chaperone | | up | | up |
| EG10241 | *dnaK* | chaperone | | | | up |
| EG10240 | *dnaJ* | chaperone | | up | | |
| EG11534 | *ibpA* | chaperone | | | | up |
| G6463 | *clpS* | chaperone | | | down | |
| EG12055 | *ccmE* | chaperone | | | | down |
| EG11973 | *ecpD* | chaperone | | | down | |
| G378 | *fliJ* | predicted chaperone | | | down | |
| EG11976 | *fliQ* | flagellar biosythesis | | | down | down |
| G7039 | *fliY* | cystine binding protein | | | | down |
| EG12195 | *torD* | chaperone | | | down | |
| EG11877 | *uspD* | stress protein | | | | up |
| EG11246 | *uspE* | stress protein | | | up | |
| G7743 | *hslR* | heat shock protein | | | | down |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Cytoplasm proteases** | | | | | | |
| G6991 | *yeaZ* | Hypothetical protease | | | down | |
| **Cytoplasmic membrane proteases** | | | | | | |
| EG10397 | *glpG* | intramembrane serine protease GlpG | | | up | |
| EG10956 | *sohB* | Possible Ser protease SohB | | | down | down |
| EG12436 | *yaeL* | RseP zinc protease | down | down | | |
| **Periplasm proteases** | | | | | | |
| EG11893 | *dacD* | PBP-6B, D-alanyl-D-alanine carboxypeptidase | | | | down |
| EG10463 | *degP* | Serine protese Do | | | | up |
| G6746 | *ydcP* | Putative protease YdcP | | up | | |

**Table 4:** Comparison of gene expression of proteases, chaperones and stress induced proteins in production and control BL21 (DE3) strain at different temperatures: Genes identified as DE and having negative logFC ratio in particular comparison are indicated as 'down', while DE genes with positive logFC ratio are indicated as 'up' regulated. Other classification is aligned with EcoCyc classification of genes. Comparisons of production versus control strain at 25 °C (P_C25), 37 °C (P_C37) and 42 °C (P_C42) are presented. Gene names and their classification is aligned with EcoCyc (http://www.ecocyc.org/expression.html).

| **EcoCyc ID** | **Gene (ecoid)** | **Protein function** | **Comparison** | | |
|---|---|---|---|---|---|
| | | | **P_C25** | **P_C37** | **P_C42** |
| **Sigma factors and regulators** | | | | | |
| EG10897 | *rpoH* ($\sigma^{32}$) | sigma factor controlling the heat shock response | | | down |
| EG12121 | *rssB* | sigma regulator | | | down |
| **Chaperones and stress induced proteins** | | | | | |
| G7039 | *fliY* | cystine binding protein | | | up |
| G6357 | *hscC* | Chaperone | | | up |
| EG11877 | *uspD* | stress protein | | | down |
| EG11246 | *uspE* | stress protein | up | | up |
| **Cytoplasm proteases** | | | | | |
| G6991 | *yeaZ* | Hypothetical protease | down | | |
| **Cytoplasmic membrane proteases** | | | | | |
| EG10956 | *sohB* | Possible Ser protease SohB | down | | |
| EG12436 | *yaeL* | RseP zinc protease | up | | |
| **Periplasm proteases** | | | | | |
| EG11893 | *dacD* | PBP-6B, D-alanyl-D-alanine carboxypeptidase | | | up |

Expression of these genes in our growth conditions was inspected in more detail. All together 24 genes showed significant changes in their expression levels in any of the comparisons (Table 3, 4). Most distinctive are gene expression profiles of cultures grown at 42 °C as most of DE genes were identified in comparisons C42_25, P42_25 and P_C42. As the heat shock response is well studied in *E.coli* we have compared our results to already published data, but our focus was set more to the differences in gene expression related to suboptimal growth temperature (P37_25, C37_25 and P_C25). Few genes coding for chaperones (*clpS*, *ecpD*, *fliJ*, *torD*, *uspE*) and a *glpG* gene coding for cytoplasmic membrane protease were regulated specifically due to change of temperature from 25 °C to 37 °C (and not to 42 °C). Interestingly, *uspE* gene was found to be the only chaperone gene and *yaeL* the only protease gene with higher expression in production strain compared to control strain when grown at 25 °C (Table 4). Similarly, *yeaZ* gene coding for putative cytoplasmic protease and *sohB* gene coding for putative periplasmic protease were found to be down-regulated in production strain when grown at 25 °C (Table 4).

## 3. 3. Verification of Microarray Results by Real-time PCR

Quantitative real-time PCR (qPCR) was used to verify microarray results. Among genes identified as DE in at least one of the comparisons six were chosen for further analysis: four genes coding for chaperones (*groEL*, *ecpD*, *torD*, *uspD (yiit)*) and two genes coding for proteases (*yeaZ*, *ydcP*). Gene expression changes obtained by microarrays were in all comparisons confirmed by qPCR, except for two (Table 5).

## 3. 4. Temperature Dependent Accumulation of GroEL Chaperone

Information on gene expression was for GroEL chaperone complemented also on the level of protein accumulation.

**Table 5:** Comparison of gene expression measurements by microarrays and quantitative real-time PCR: Difference in expression is given as logFC. Comparisons of production versus control strain at 25 °C (P_C25) and 42 °C (P_C42) are presented as well as comparisons of production strain grown at different temperatures (P42_25) and control strain grown at different temperatures (C42_25, C37_25). M – microarray analysis results, qPCR – quantitative real-time PCR results. NS – nonsignificant difference in expression as detected in microarray analysis.

| Gene | Method | P42_25 | C37_25 | C42_25 | P_C25 | P_C42 |
|---|---|---|---|---|---|---|
| *groEL* | M | 0.7 | NS | 1.7 | NS | NS |
| | qPCR | 1.6 | –0.7 | 1.1 | | |
| *ecpD* | M | NS | –1.2 | NS | NS | NS |
| | qPCR | | –0.7 | | | |
| *torD* | M | NS | –1.2 | NS | NS | NS |
| | qPCR | | –1.0 | | | |
| *uspD* | M | 1.1 | NS | 2.5 | NS | –1.4 |
| | qPCR | 1.6 | | 1.0 | | |
| *yeaZ* | M | NS | –2.2 | NS | –1.6 | NS |
| | qPCR | | –0.9 | | –1.4 | |
| *ydcP* | M | 1.6 | NS | NS | NS | NS |
| | qPCR | 0.9 | | | | |

The results showed (Figure 2) that GroEL protein content is increasing with increasing temperature in both cultures which is in line with the results of microarray hybridizations and qPCR. However, expression levels of GroEL are higher at T = 37 °C than at T = 42 °C in production strain. Statistically significant differences were obtained for comparison of protein content at 42 °C and 25 °C in control strain ($p < 0.05$) and for comparison of control and production strain at 42 °C ($p < 0.05$).

# 4. Discussion

Up to now many studies of *E. coli* transcriptome analysis changes during heat shock and/or recombinant protein production have been reported.[21–23] However, not many studies focused on transcriptome changes in *E. coli* that occur during recombinant protein production at lower temperatures (i.e. at T = 25 °C) in particular with regard to formation of non classical IBs. Although in conditions investigated, we have observed limited changes in gene expression due to recombinant protein production in *E. coli*, our study suggests that some chaperones and proteases play an important role in the efficient production of the recombinant protein hG-CSF at suboptimal growth temperature. Interestingly, very limited changes in gene expression due to recombinant protein production in *E. coli* were observed (in contrast to control strain) in metabolic and regulatory pathways irrespective of the temperature growth condition (data not shown). Such small differences might indicate that the overexpression of recombinant protein could be more important stress factor for the recombinant protein producing cells which is able to „diminish" the temperature associated changes on gene expression.

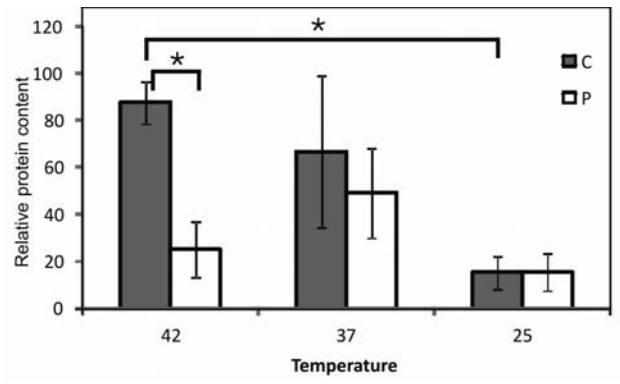



**Figure 2:** Recombinant *E. coli* GroEL content at different growth temperatures. Western blot analysis was performed for GroEL content in control strain (grey columns; (C)) and production strain (white columns; (P)). Bars represent standard error of biological replicate measurements. Statistically significant differences are marked with * ($p < 0.05$). Original gel and membrane images can be found in Supporting information 4.

## 4. 1. Role of Chaperones in Production of Active Recombinant Protein hG-CSF at Lower Temperature

The role of chaperones is to assist in proper folding of both native and recombinant cell proteins[24]. For example, it was shown that overproduction of GroES and GroEL or DnaK and DnaJ should prevent protein aggregation.[25,26] We have observed increased expression levels of *groEL* (Table 3) at T = 42 °C in comparison to T = 25 °C in both, control and production strain and only in one strain type for *dnaK* and *dnaJ* which is in agreement with the previously published results[27,28] and should assist in folding of proteins in heat stress conditions. Under stress *E. coli* needs chaperones first for its own survival and only afterwards for production of recombinant proteins. Thus, at elevated temperatures or overproduction of recombinant proteins chaperones might become limiting.[24] In our study, the levels of GroEL were found to be lower in production strain compared to control (Fig. 2), indicating even further lack of this protein in producing strain. The increased expression level in comparison between production and control strain grown at T = 25 °C was observed for *uspE* gene (Table 4). The observed increased levels of *uspE* genes might therefore only confirm that they are induced due to a variety of stresses (e.g. heat shock or recombinant protein production)[29] while the exact mechanism of their function remains unknown.

Flagellum-specific chaperones (e.g. FliJ, FliQ and FliY), the type III cytoplasmic chaperone family members, assist in folding and export of flagellar proteins.[30] *fliY*, coding for FliY, a periplasmic[31] cystine-binding protein[32] participates in regulation of class III transcription through regulation $\sigma^F$ (*rpoF*, $\sigma^{28}$) activity.[33] *rpoF* is a minor sigma factor responsible for initiation of transcription

of a number of genes involved in motility.[34] Higher activity of *fliY* gene (a significant decrease in expression level at T = 42 °C in comparison to T = 25 °C in the control strain, Table 3) at suboptimal growth temperature, activates *rpoF* whose expression levels remain unchanged (see Supporting information). Similarly, *fliQ* and *fliJ* coding for integral membrane components of the flagellar export apparatus[30,34] showed increased expression levels at T = 25 °C in the control strain (Table 3). Direct implication of the increased activity of flagellum specific chaperones on efficiency of hG-CSF protein expression and folding on the basis of currently known functions of Fli proteins cannot be taken. These results suggest that *E. coli* flagellum-specific chaperones might have some so far unidentified functions that could contribute to improved folding of recombinant proteins.

### 4. 2. Hypothetical Protease YeaZ is Repressed in Recombinant *E.coli* Producing hG-CSF When Grown at 25 °C

At elevated temperatures cytoplasm proteases and cytoplasmic membrane proteases are in majority down-regulated and periplasm proteases up-regulated (Table 3). The only cytoplasm protease that has higher gene activity at lower growth temperature (in the control strain) and could thus contribute to more efficient recombinant protein production is YeaZ. Potential role of periplasm located proteases in structure of IBs is even more obscure. Interestingly however, the expression level of hypothetical protease *yeaZ* gene was decreased at 25 °C in production strain if compared to control strain (Table 4). Although not much data is available on function of this protein, it is known that together with two essential proteins YjeE and YgjD it forms an interaction network whose cellular role remains unknown.[35] Nevertheless, YeaZ could be an important factor in protein interactions that lead to different accumulation levels of recombinant protein and presents interesting area for future research.

### 4. 3. Membrane Localized Protease YaeL is Induced in Recombinant *E.coli* at Suboptimal Temperature

It is known that the function of YaeL is to cleave RseA in the cytoplasmic or intramembrane region[36] and to exhibit proteolytic activity toward RpoE and RpoH.[37] Continual degradation of RseA by YaeL is needed to provide the cell with sufficient free RpoE to support cell viability.[38] According to our results *yaeL* is upregulated at suboptimal temperature in production strain if compared to both temperatures (Table 3). This suggests that cell viability is increased at suboptimal temperature, hence YaeL could be an important factor in protein interactions that lead to different accumulation levels of recombinant proteins and presents interesting area for future research.

## 5. Conclusions

*Escherichia coli* adaptation to internal and external signals is an extremely important mechanism and must be tightly regulated in order to assure survival of the organism. Recombinant protein production is representing a stress for bacterial cells as is also change in the growth temperature. An interesting and so far not understood phenomenon is the increased content of correctly folded recombinant protein in IBs of *E. coli* grown at 25 °C.

Our hypothesis was that chaperones and/or proteases might have an important role in this process. The lowered activity of a certain group of proteases should render the newly formed molecules of hG-CSF less vulnerable to proteolytic hydrolysis. On the other hand, some chaperones could be identified that contribute to proper folding of protein in IBs at suboptimal growth temperature. We have found several chaperones (Table 3) that were regulated in our experiment, mostly however had increased levels detected in cultures grown at 42 °C and not at 25 °C. The only chaperone with higher gene activity in production strain grown at 25 °C was UspE (Table 4). No other experimental data is currently available that would link these proteins to folding of recombinant proteins. Among proteases, cytoplasmic proteases are of main interest since IBs are mostly found in the bacterial cytoplasm.[24] Gene expression of cytoplasm protease *yeaZ* was found to be decreased at 25 °C in production strain whereas gene expression of cytoplasmic membrane protease *yaeL* was found to be increased at 25 °C in production strain (Table 4). Thus *yaeL* and *yeaZ* genes are candidates for improved recombinant protein production at lower growth temperature. Taken altogether, gene expression profiling provided additional information on mechanisms of recombinant protein production in *E. coli* at suboptimal growth temperatures. Further research is needed to directly confirm the potential role of candidate genes in this process.

## 6. Acknowledgements

The authors would like to acknowledge Neža Turnšek and Katja Barle for help with sample preparations and qPCR analysis, Karmen Čerkić and Maruša Hafner for help with cultivation of *E. coli* and Mateja Skok for Western blotting analysis. The work was partially conducted at the Center for Functional Genomics and Bio-Chips, Ljubljana, Slovenia and was also supported by Sandoz, Biopharmaceuticals, Lek Pharmaceuticals d.d. and Slovenian research agency programs P1-0340 and P4-0165. This work is dedicated to inspiring mentor

dr. Viktor Menart who has unexpectedly passed away during the study.

## 7. References


1. L. S. A. Vanz, G. Renard, S. M. Palma, J. M. Chies, S. L. Dalmora, L. A. Basso, D. S. Santos, *Microb. Cell Fact.* **2008**, *7*, 13.
2. M. Miot, J-M. Betton, *Microb. Cell Fact.* **2004**, *3*, 4.
3. S. Ventura, A. Villaverde, *Trends Biotechnol.* **2006**, *24,* 179–85.
4. E. García-Fruitós, N. González-Montalbán, M. Morell, A. Vera, R. M. Ferraz, A. Arís, S. Ventura, A. Villaverde, *Microb. Cell Fact.* **2005**, *4*, 27.
5. J. Tydemers, A. Mogk, B. Bukau, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 777–788.
6. Š. Peternel, M. Bele, V. Gaberc-Porekar, V. Menart, *Microb. Cell Fact.* **2006**, *5*, P23.
7. S. Jevševar, V. Gaberc-Porekar, I. Fonda, B. Podobnik, J. Grdadolnik, V. Menart, *Biotechnol. Prog.* **2005**, *21*, 632–639.
8. Š. Peternel, J. Grdadolnik, V. Gaberc-Porekar, R. Komel, *Microb. Cell Fact.* **2008**, *7*, 34.
9. M. Martínez-Alonso, N. González-Montalbán, E. García-Fruitós, A. Villaverde, *Microb. Cell Fact.* **2009**, *8*, 4.
10. K. Maruyama, K. Tsuji, R. Tanaka, K. Yamada, Y. Kodera, T. Nakahata, *Bone Marrow Transplant.* **1998**, *22*, 313–320.
11. S. Jevševar, J. Palčič, Š. Jalen, A. Pavko, *Acta Chim. Slov.* **2007**, *54*, 360–365.
12. M. Petek, Š. Baebler, D. Kuzman, A. Rotter, Z. Podlesek, K. Gruden, M. Ravnikar, U. Urleb, *BMC Microbiol.* **2010**, *10*, 159.
13. Š. Baebler, H. Krečič-Stres, A. Rotter, P. Kogovšek, K. Cankar, E.J. Kok, K. Gruden, M. Kovač, J. Žel, M. Pompe-Novak, M. Ravnikar, *Mol. Plant Pathol.* **2009**, *10*, 263–75.
14. G. K. Smyth, *Stat. Appl. Genet. Mol. Biol.* **2004**, *3*, Article 3.
15. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S.L. Pomeroy, T. R. Golub, E.S. Lander, J. P. Mesirov, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102(43)*, 15278–9.
16. M. Pompe-Novak, K. Gruden, Š. Baebler, H. Krečič-Stres, M. Kovač, M. Jongsma, M. Ravnikar, *Physiol. Mol. Plant Pathol.* **2006**, *67*, 237–247.
17. T. Nolan, R. E. Hands, S. A. Bustin, *Nat. Protoc.* **2006**, *1*, 1559–1582.
18. M. Škulj., V. Okršlar, Š. Jalen, S. Jevševar, P. Slanc, B. Štrukelj, V. Menart, *Microb. Cell Fact.* **2008**, *7*, 6.
19. M. Arifuzzaman, M. Maeda, A. Itoh, K. Nishikata, C. Takita, R. Saito, T. Ara, K. Nakahigashi, H-C. Huang, A. Hirai, K. Tsuzuki, S. Nakamura, M. Altaf-Ul-Amin, T. Oshima, T. Baba, N. Yamamoto, T. Kawamura, T. Ioka-Nakamichi, M. Kitagawa, M. Tomita, S. Kanaya, C. Wada, H. Mori, *Genome Res.* **2006**, *16*, 686–691.
20. S. Gottesman, *Annu. Rev. Genet.* **1996**, *30*, 465–506.
21. S. W. Harcum, F. T. Haddadin, *J. Ind. Microbiol. Biotechnol.* **2006**, *33*, 801–814.
22. M. D. Carruthers, C. Minion, *FEMS Microbiol. Lett.* **2009**, *295*, 96–102.
23. M. Gadgil, V. Kapur, W-S. Hu, *Biotechnol. Prog.* **2005**, *21*, 689–699.
24. A. Villaverde, M. M. Carrió, *Biotechnol. Lett.* **2003**, *25*, 1385–1395.
25. A. Gragerov, E. Nudler, N. Komissarova, G. A. Gaitanaris, M. E. Gottesman, V. Nikiforov, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 10341–10344.
26. F. U. Hartl, *Nature* **1996**, *381*, 571–579.
27. M. Y. Sherman, A. L. Goldberg, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8648–8652.
28. J. Chatellier, F. Hill, P. A. Lund, A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9861–9866.
29. N. Gustavsson, A. Diez, T. Nyström, *Mol. Microbiol.* **2002**, *43*, 107–17.
30. T. Minamino, K. Namba, *J. Mol. Microbiol. Biotechnol.* **2004**, *7*, 5–17.
31. D. S. Mytelka, M. J. Chamberlin, *J. Bacteriol.* **1996**, *178*, 24–34.
32. J. D. Butler, S. W. Levin, A. Fachiano, L. Miele, A. B. Mukheriee, *Life Sci.* **1993**; *52*, 1209–15.
33. T. Minamino, R. M. Macnab, *J. Bacteriol.* **1999**, *181*, 1388–94.
34. X. Mao, Y. Zhang, Y. Xu, *PLoS One*, **2011**, *6–7*, e22556.
35. J. I. Handford, B. Ize, G. Buchanan, G. P. Butland, J. Greenblat, A. Emili, T. Palmer, *J. Bacteriol.* **2009**, *191*, 4701–4.
36. K. Kanehara, K. Ito, Y. Akiyama, Genes Dev. 2002, 16, 2147–2155.
37. C. Dartigalongue, H. Loferer, S. Raina, EMBO J., 2001, 20, 5908–5918.
38. S. E. Ades, Curr. Opin. Microbiol., 2004, 7, 157–162.

## Povzetek

V nedavnih raziskavah je bilo dokazano, da se v inkluzijskih telesih pri nižjih temperaturah gojenja (25 °C) izloča večji delež pravilno zvitega rekombinantnega proteina hG-CSF. Naša raziskava je bila zato usmerjena v preučevanje in razlago procesov, ki sodelujejo pri tvorbi »neklasičnih« inkluzijskih teles predvsem s pomočjo preučevanja genskega izražanja proteaz in šaperonov. V primeru gojenja kultur pri suboptimalni temperaturi smo s statistično analizo oligonukleotidnih mikromrež dobili vidne rezultate sprememb v izražanju genov vpletenih v energetski metabolizem, v metabolizem aminokislin in nukleotidov ter kofaktorjev in sekundarnih metabolitov. Izmed do zdaj znanih genov, ki nosijo zapis za šaperone, proteaze in ostale s stresom povezane proteine jih je bilo kar 24 statistično značilno diferencialno izraženih. Tako so UspE med šaperoni in YaeL ter YeaZ med proteazami najbolj pomembni kandidati s potencialno vlogo pri akumulaciji aktivnih rekombinantnih proteinov. Ugotovili smo tudi, da ima pri nižjih temperaturah gen *yaeL*, ki nosi zapis za izražanje v membrani lokalizirane proteaze YaeL večjo aktivnost kar ga funkcijsko potencialno povezuje z bolj učinkovitim izražanjem rekombinantnih proteinov. Rezultati raziskave zato predstavljajo napredek v razumevanju produkcije rekombinantih proteinov bakterije *E. coli*. Prav ti geni, ki pri nižji temperaturi gojenja potencialno vplivajo na produkcijo rekombinantnih proteinov predstavljajo osnovo za nadaljnje raziskave pri izboljšavi produkcijskih sevov bakterije *E. coli* kot tudi pri izboljšavi fermentacijskih procesov.

# Supporting information

## Supporting information 1

Results of statistical testing for differentially expressed genes. Two-way analysis of variance was performed with the cut-off p-value 0.05. DE genes due to recombinant protein production (P_C), growth temperature (37_25, 42_25) and interaction of both factors (interaction) are listed. Each worksheet within the Excel file (ANOVA analysis) consists of different columns describing numerical identification of a gene (geneid), EcoCyc identification number (ecocyc), gene name (ecoid), definition and class of the protein, statistical significance (P. value),…. and normalized data using quantile normalization at different temperatures (e.g. N-25-BL21-10-1 stands for normalized data of strain of *E. coli* control strain grown at T = 25 °C and harvested at OD = 10).

## Supporting information 2

**Table 1:** Temperature dependent changes in energy metabolism, in metabolism of amino acids, and of nucleotides, in biosynthesis of cofactors and secondary metabolites of recombinant BL21 (DE3) strain: Comparisons of gene expression of control strain grown at different temperatures (C42_25, C37_25)- Genes are identified as DE and having negative logFC ratio in particular comparison are indicated as 'down', while DE genes with positive logFC ratio are indicated as 'up' regulated. Gene names and their classification is aligned with EcoCyc (http://www.ecocyc.org/expression.html).

| Gene (ecoid) | Protein function Cellular function | Pathways | Comparison C37_25 | C42_25 |
|---|---|---|---|---|
| **C1 compounds utilization and assimilation** | | | | |
| *frmA* | glutathione-dependent formaldehyde dehydrogenase | formaldehyde oxidation II (gluthathione dependent) | down | down |
| **alcohols degradation** | | | | |
| *glpB* | glycerol-3-phosphate dehydrogenase (anaerobic) | glycerol degradation I | up | up |
| *glpD* | glycerol 3-phosphate dehydrogenase (aerobic) | glycerol degradation I | up | up |
| **carbohydrates biosynthesis** | | | | |
| *eno* | enolase | gluconeogenesis I | up | |
| *ybhA* | pyridoxal phosphatase/fructose 1,6-bisphosphatase | gluconeogenesis I | down | |
| *galM* | galactose-1-epimerase | galactose degradation I (Leloir pathway) | up | |
| *rbsD* | ribose pyranase | ribose degradation | down | |
| *mdh* | malate dehydrogenase | gluconeogenesis I | | up |
| *pgi* | phosphoglucose isomerase | gluconeogenesis I | | up |
| *mngB* | α-mannosidase | 2-O-a-mannosyl-D-glycerate degradation | down | down |
| *araB* | L-ribulokinase | L-arabinose degradation I | down | down |
| **fatty acid and lipids biosynthesis** | | | | |
| *lpxP* | palmitoleoyl acyltransferase | superpathway of KDO2-lipid A biosythesis | | down |
| *waaA* | KDO transferase | lipid-A-precursor biosynthesis | up | |
| *cfa* | cyclopropane fatty acyl phospholipid synthase | cyclopropane fatty acid (CFA) biosynthesis | up | |
| *waaA* | KDO transferase | superpathway of KDO2-lipid A biosythesis | up | |
| **Fermentation** | | | | |
| *acnB* | 2-methylisocitrate dehydratase | 2-methylcitrate cycle I | down | down |
| *acnB* | bifunctional aconitate hydratase 2 | mixed acid fermentation | down | down |
| *mdh* | malate dehydrogenase | mixed acid fermentation | | up |
| *pykF* | pyruvate kinase I | mixed acid fermentation | up | |
| *ldhA* | D-lactate dehydrogenase | mixed acid fermentation | down | |
| *frdB* | fumarate reductase iron-sulfur protein | mixed acid fermentation | down | |
| **Glycolysis** | | | | |
| *pgi* | phosphoglucose isomerase | glyoxylate bypass and TCA | | up |
| *mdh* | malate dehydrogenase | glyoxylate bypass and TCA | | down |
| *sucC* | succinyl-CoA synthetase (β subunit) | glyoxylate bypass and TCA | | down |
| *acnB* | bifunctional aconitate hydratase 2 | glyoxylate bypass and TCA | | down |
| *ybhA* | pyridoxal phosphatase/fructose 1,6-bisphosphatase | glycolysis I | down | |

| Gene (ecoid) | Protein function Cellular function | Pathways | Comparison C37_25 | C42_25 |
|---|---|---|---|---|
| **Glycolysis** | | | | |
| *pfkB* | 6-phosphofructokinase II | glycolysis I -Etner-Doudoroff | up | |
| *eno* | enolase | glycolysis I Etner-Doudoroff | up | |
| *pykF* | pyruvate kinase I | glycolysis I Etner-Doudoroff | up | |
| *ybhA* | pyridoxal phosphatase/fructose 1,6-bisphosphatase | glycolysis I-TCA | down | |
| *aceB* | | glycolysis I-TCA | down | |
| *acnB* | bifunctional aconitate hydratase 2 | glycolysis I-TCA | down | |
| *pfkA* | malate synthase A | glycolysis I-TCA | up | |
| *eno* | enolase | glycolysis I-TCA | up | |
| *pykF* | pyruvate kinase I | glycolysis I-TCA | up | |
| **inorganic nutrients metabolism** | | | | |
| *ssuD* | FMNH2-dependent alkanesulfonate monooxygenase | two-component alkanesulfonate monooxygenase | down | down |
| *cysH* | 3'-phospho-adenylylsulfate reductase | sulfate reduction I (assimilatory) | | up |
| *cysJ* | sulfite reductase (flavoprotein subunit complex) | sulfate reduction I (assimilatory) | | up |
| **respiration** | | | | |
| *acnB* | bifunctional aconitate hydratase 2 | anaerobic respiration | down | down |
| *mdh* | malate dehydrogenase | anaerobic respiration | | up |
| *pykF* | pyruvate kinase I | anaerobic respiration | up | |
| *eno* | enolase | anaerobic respiration | up | |
| *frdB* | fumarate reductase iron-sulfur protein | anaerobic respiration | down | |
| **pentose phosphate pathways** | | | | |
| *tktA* | transketolase I | pentose phosphate pathway | down | |
| *rpe* | ribulose phosphate 3-epimerase | pentose phosphate pathway | up | |
| **superpathway of glycolysis and Entner Doudoroff** | | | | |
| *pgi* | phosphoglucose isomerase | glycolysis I – Enter-Doudoroff | | up |
| **cell structures biosynthesis** | | | | |
| *pgi* | phosphoglucose isomerase | colanic acid building blocks biosythesis | | up |
| *rffG* | dTDP-glucose 4,6-dehydratase 2 | enterobacterial common antigen biosythesis | down | down |
| *glf* | UDP-galactopyranose mutase | dTDP-L-rhamnose biosythesis I | up | |
| *rfbC* | dTDP-4-dehydrorhamnose 3,5-epimerase | dTDP-L-rhamnose biosythesis I | down | |
| **amine and polyamines biosynthesis** | | | | |
| *argI* | ornithine carbamoyltransferase | aminopropylcadaverine biosynthesis | | down |
| *nagB* | glucosamine-6-phosphate deaminase | N-acetylglucosamine degradation | | up |
| **amino acids biosynthesis** | | | | |
| *ydiB* | quinate/shikimate dehydrogenase | tryptophan biosythesis | down | |
| *trpC* | phosphoribosylanthranilate isomerase / indole-3-glycerol phosphate synthase | tryptophan biosythesis | down | |
| *trpB* | tryptophan synthase, β subunit dimer | tryptophan biosythesis | up | |
| *ilvI* | catalytic subunit of acetolactate synthase III | leucine biosynthesis | down | |
| *ilvI* | catalytic subunit of acetolactate synthase III | isoleucine biosynthesis (from threonine) | down | |
| *ansB* | L-asparagine aminohydrolase II | glutamate degradation II | down | |
| *ilvI* | catalytic subunit of acetolactate synthase III | isoleucine biosynthesis I (from threonine) | down | |
| *speE* | spermidine synthase | aminopropylcadaverine biosynthesis | down | |
| *gcvT* | aminomethyltransferase | glycine cleavage complex | down | |
| *dapE* | N-succinyl-L-diaminopimelate desuccinylase | lysine biosythesis I | down | down |
| *proA* | glutamate-5-semialdehyde dehydrogenase | proline biosythesis I | | down |
| *iscS* | cysteine desulfurase | alanine biosythesis I | | up |
| *cysK* | cysteine synthase A | cysteine biosythesis I | | up |
| **nucleosides and nucleotides biosynthesis** | | | | |
| *pyrE* | orotate phosphoribosyltransferase | pyrimidine ribonucleotides interconversion (de novo biosythesis) | up | up |
| *deoD* | purine nucleoside phosphorylase deoD-type | salvage pathways of adenine, hypoxanthine and their nucleosides | | up |
| *deoD* | purine nucleoside phosphorylase deoD-type | salvage pathways of guanine, xanthine and their nucleosides | | up |
| *dcd* | dCTP deaminase | pyrimidine deoxyribonucleotides de novo biosythesis | down | down |

| Gene (ecoid) | Protein function<br>Cellular function | Pathways | Comparison<br>C37_25 | <br>C42_25 |
|---|---|---|---|---|
| **nucleosides and nucleotides biosynthesis** | | | | |
| *trpC* | indole-3-glycerol phosphate synthase / phosphoribosylanthranilate isomerase | tryptophan biosythesis | | down |
| *cdd* | cytidine deaminase | salvage pathways of pyrimidine ribonucleotides | up | |
| *udp* | uridine phosphorylase | salvage pathways of pyrimidine ribonucleotides | up | |
| **nucleosides and nucleotides degradation and recycling** | | | | |
| *deoD* | purine nucleoside phosphorylase deoD-type | purine deoxyribonucleosides degradation | | up |
| *deoD* | purine nucleoside phosphorylase deoD-type | degradation of purine ribonucleosides | | up |
| *cdd* | cytidine deaminase | pyrimidine deoxyribonucleosides degradation | up | |
| *cdd* | cytidine deaminase | degradation of pyrimidine ribonucleoside | up | |
| *udp* | uridine phosphorylase | degradation of pyrimidine ribonucleoside | up | |
| **secondary metabolites degradation** | | | | |
| *idnD* | L-idonate 5-dehydrogenase | L-idonate degradation | | down |
| *idnO* | 5-keto-D-gluconate 5-reductase | L-idonate degradation | up | up |
| *garL* | α-dehydro-β-deoxy-D-glucarate aldolase | D-glucarate and D-galactarate degradation | up | |
| *pfkB* | 6-phosphofructokinase II | galactitol degradation | up | |
| *gatZ* | Subunit of tagatose-1,6-bisphosphate aldolase 2 | galactitol degradation | up | |
| *gatY* | Subunit of tagatose-1,6-bisphosphate aldolase 2 | galactitol degradation | up | |
| **cofactors, prosthetic groups, electron carriers biosynthesis** | | | | |
| *cysG* | uroporphyrin III C-methyltransferase | siroheme biosythesis | | up |
| *cobC* | predicted α-ribazole-5'-P phosphatase | adenosylcobalamin salvage from cobinamide I | down | down |
| *thiC* | thiamin biosynthesis protein ThiC | thiamin biosynthesis I | down | down |
| *thiG* | thiazole synthase | thiamin biosynthesis I | down | down |
| *gshB* | glutathione synthetase | glutathione biosythesis | | up |
| *ribE* | 6,7-dimethyl-8-ribityllumazine synthase | flavin biosythesis I (bacteria) | | up |
| *ubiH* | 2-octaprenyl-6-methoxyphenol hydroxylase | ubiquinone-8 biosythesis (prokaryotic) | | down |
| *btuR* | cobinamide adenosyltransferase / cobalamin adenosyltransferase | adenosylcobalamin salvage from cobinamide I | up | |
| *cobU* | cobinamide-P guanylyltransferase / cobinamide kinase | adenosylcobalamin salvage from cobinamide I | up | |
| *thiM* | hydroxyethylthiazole kinase | thiamin biosynthesis I | up | |
| *nadK* | NAD kinase | NAD phosphorylation and dephosphorylation | down | |
| *pdxH* | pyridoxine 5'-phosphate oxidase / pyridoxamine 5'-phosphate oxidase | pyridoxal 5´-phosphate biosythesis and salvage | up | |
| *folE* | GTP cyclohydrolase I | tetrahydrofolate biosythesis | up | |
| *folB* | dihydroneopterin aldolase | tetrahydrofolate biosythesis | up | |
| *gcvT* | Aminomethyltransferase | formylTHF biosythesis I | down | |

**Table 2:** Functional analysis of differences in gene expression observed between production and control BL21 (DE3) strain grown at different temperatures: Genes identified as DE and having negative logFC ratio in particular comparison are indicated as 'down', while DE genes with positive logFC ratio are indicated as 'up' regulated. Other classification is aligned with EcoCyc classification of genes. Comparisons of production versus control strain at 25 °C (P_C25), 37 °C (P_C37) and 42 °C (P_C42) are presented. Gene names and their classification is aligned with EcoCyc (http://www.ecocyc.org/expression.html).

| Gene (ecoid) | Protein function | Pathways | Comparison | | |
|---|---|---|---|---|---|
| | **Energy metabolism** | | **P_C25** | **P_C37** | **P_C42** |
| **inorganic nutrients metabolism** | | | | | |
| *ssuD* | FMNH2-dependent alkanesulfonate monooxygenase | two-component alkanesulfonate monooxygenase | down | down | |
| **carbohydrates biosynthesis** | | | | | |
| *mngB* | α-mannosidase | 2-O-a-mannosyl-D-glycerate degradation | down | | |
| *araB* | L-ribulokinase | L-arabinose degradation I | down | | |
| *galM* | galactose-1-epimerase | galactose degradation I (Leloir pathway) | up | | |
| *mdh* | malate dehydrogenase | gluconeogenesis I | | | down |
| *pgi* | phosphoglucose isomerase | gluconeogenesis I | | | down |

| Gene (ecoid) | Protein function | Pathways | Comparison | | |
|---|---|---|---|---|---|
| | **Energy metabolism** | | **P_C25** | **P_C37** | **P_C42** |
| **Fermentation** | | | | | |
| *acnB* | 2-methylisocitrate dehydratase | 2-methylcitrate cycle I | | up | |
| *mdh* | malate dehydrogenase | mixed acid fermentation | | | down |
| **pentose phosphate pathways** | | | | | |
| *rpe* | ribulose phosphate 3-epimerase | pentose phosphate pathway | | down | |
| *tktA* | transketolase I | pentose phosphate pathway | | up | |
| *talA* | transaldolase A | pentose phosphate pathway (oxidative branch) | | | down |
| *gnd* | 6-phosphogluconate dehydrogenase (decarboxylating) | pentose phosphate pathway (oxidative branch) | | | up |
| **fatty acid and lipids biosynthesis** | | | | | |
| *cfa* | cyclopropane fatty acyl phospholipid synthase | cyclopropane fatty acid (CFA) biosynthesis | | down | |
| **Respiration** | | | | | |
| *mdh* | malate dehydrogenase | anaerobic respiration | | | down |
| *acnB* | 2-methylisocitrate dehydratase | anaerobic respiration | | up | |
| **Glycolysis** | | | | | |
| *mdh* | malate dehydrogenase | glyoxylate bypass and TCA | | | down |
| *pgi* | phosphoglucose isomerase | glyoxylate bypass and TCA | | | down |
| *acnB* | 2-methylisocitrate dehydratase | glyoxylate bypass and TCA | | up | |
| **superpathway of glycolysis and Entner Doudoroff** | | | | | |
| *pgi* | phosphoglucose isomerase | glycolysis I – Enter-Doudoroff | | | down |
| **carbohydrates degradation** | | | | | |
| *rpib* | allose-6-phosphate isomerase / ribose-5-phosphate isomerase B | D-allose degradation | | | up |
| **amino acids degradation** | | | | | |
| *cysK* | cysteine synthase A | L-cysteine degradation II | | | down |
| | **Metabolism of amino acids** | | | | |
| **amino acids biosynthesis** | | | | | |
| *ydiB* | quinate/shikimate dehydrogenase | tryptophan biosythesis | down | | down |
| *trpC* | phosphoribosylanthranilate isomerase / indole-3-glycerol phosphate synthase | tryptophan biosythesis | down | | |
| *dapE* | N-succinyl-L-diaminopimelate desuccinylase | lysine biosythesis I | | up | |
| *gcvT* | Aminomethyltransferase | glycine cleavage complex | | | down |
| *cysK* | cysteine synthase A | cysteine biosythesis I | | | down |
| *proA* | glutamate-5-semialdehyde dehydrogenase | proline biosythesis I | | | up |
| *iscS* | cysteine desulfurase | alanine biosythesis I | | | down |
| *dapA* | dihydrodipicolinate synthase | aspartate biosynthesis | | | down |
| **amino acids degradation** | | | | | |
| *cysK* | cysteine synthase A | L-cysteine degradation II | | | down |
| **cell structures biosynthesis** | | | | | |
| *pgi* | phosphoglucose isomerase | colanic acid building blocks biosythesis | | | down |
| *glf* | UDP-galactopyranose mutase | *O*-antigen building blocks biosynthesis *(E. coli)* | | | up |
| | **Biosynthesis of cofactors and secondary metabolites** | | | | |
| **secondary metabolites degradation** | | | | | |
| *gatY* | Subunit of tagatose-1,6-bisphosphate aldolase 2 | galactitol degradation | up | down | |
| *gatZ* | Subunit of tagatose-1,6-bisphosphate aldolase 2 | galactitol degradation | | down | |
| *idnD* | L-idonate 5-dehydrogenase | L-idonate degradation | | up | |
| *yjjN* | predicted L-galactonate oxidoreductase | L-galactonate degradation | | | up |
| *uxaB* | altronate oxidoreductase | D-galacturonate degradation I | | | up |

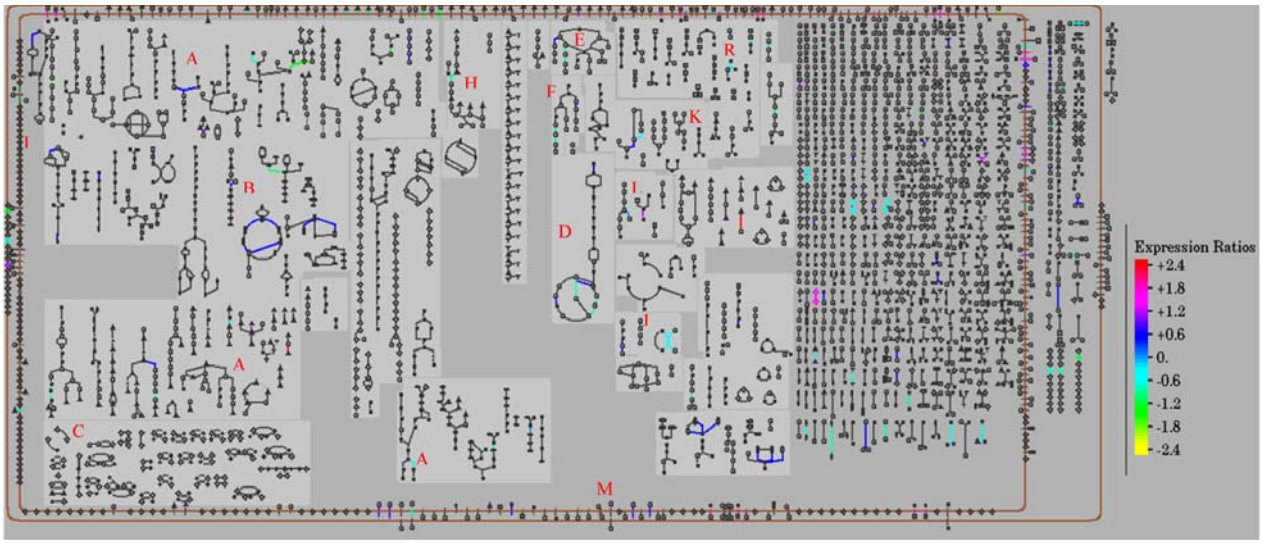
| Gene (ecoid) | Protein function | Pathways | Comparison | | |
|---|---|---|---|---|---|
| | **Energy metabolism** | | **P_C25** | **P_C37** | **P_C42** |
| **cofactors, prosthetic groups, electron carriers biosynthesis** | | | | | |
| *cobU* | cobinamide-P guanylyltransferase / cobinamide kinase | adenosylcobalamin salvage from cobinamide I | up | | |
| *thiC* | thiamin biosynthesis protein ThiC | thiamin biosynthesis I | down | | |
| *cobC* | predicted α-ribazole-5'-P phosphatase | adenosylcobalamin salvage from cobinamide I | | up | |
| *thiG* | thiazole synthase | thiamin biosynthesis I | down | | |
| *gcvT* | aminomethyltransferase | formylTHF biosythesis I | | | down |
| *frmA* | glutathione-dependent formaldehyde dehydrogenase | formaldehyde oxidation II (gluthathione dependent) (gluthathione dependent) | | up | |
| *pdxH* | pyridoxine 5'-phosphate oxidase / pyridoxamine 5'-phosphate oxidase | pyridoxal 5´-phosphate biosythesis and salvage | | down | |
| *cysG* | uroporphyrin III C-methyltransferase | siroheme biosythesis | | | down |
| *btuR* | cobinamide adenosyltransferase / cobalamin adenosyltransferase | adenosylcobalamin salvage from cobinamide I | | | up |
| *cobs* | cobalamin 5'-phosphate synthase / cobalamin synthase | adenosylcobalamin salvage from cobinamide I | | | up |
| *thiM* | hydroxyethylthiazole kinase | thiamin biosynthesis I | | | up |
| *gshB* | glutathione synthetase | glutathione biosythesis | | | down |
| *ribE* | 6,7-dimethyl-8-ribityllumazine synthase | flavin biosythesis I (bacteria) | | | down |
| *ribF* | bifunctional riboflavin kinase / FMN adenylyltransferase | | | | up |

## Supporting information 3

**Figure 1:** Overview of changes in gene expression of recombinant *E. coli* grown at different temperatures DE gene sets from comparison A) between 37 °C and 25 °C, and B) between 42 °C and 25 °C in control strain are presented using EcoCyc Omics Viewer [http://biocyc.org/ECOLI/overview-expression-map]. Lines in the diagram correspond to reactions; nodes correspond to cell components (□ – carbohydrates, △ – amino acids, ◇ – proteins, ▽ – cofactors, T – tRNA, ○ – other). Biosynthetic pathways are positioned in the left of the cytoplasm, degradative pathways in the right, signaling pathways in the bottom-left. Reactions not assigned to any pathway are in the far right of the cytoplasm. In the inner and outer membranes are transporters (with arrows) and other membrane proteins. In the periplasm are periplasmic enzymes and other periplasmic proteins. LogFC ratios of DE genes are presented coloured according to the scale on the image. Each figure (A and B) is marked with red letters to show cellular function. A = cofactors, prosthetic groups, electron carriers biosynthesis; B = fatty acids and lipids biosynthesis; C = signal transduction pathways; D = glycolysis; E = fermentation; F = respiration; G = pentose phosphate pathways; H = amines and polyamines biosynthesis; I = superpathway of glycolysis and Entner-Doudoroff; J = inorganic nutrients metabolism; K = secondary metabolites degradation; L= alcohols degradation; M= transporters

A)

B)

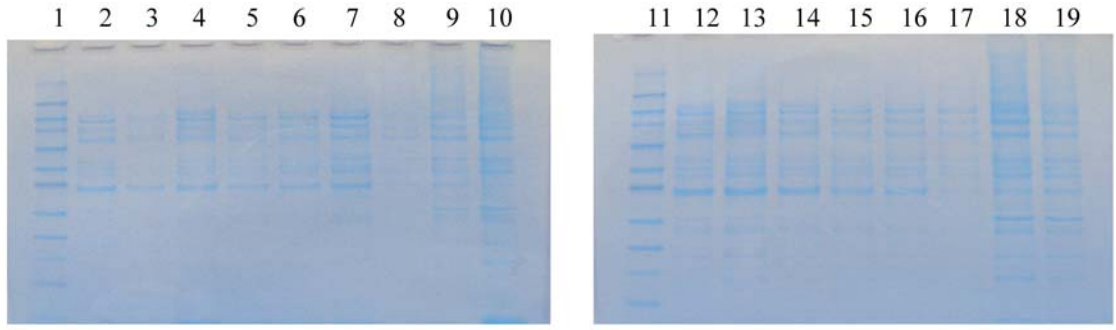



## Supporting information 4

Original images supporting figure 2 of the manuscript

A) SDS – PAGE of *E. coli* proteins and B) Western blott analysis for GroEL: 1) Standard, 2) P42-1, 3) P42-2, 4) P42-3, 5) P37-1, 6) P37-2, 7) P37-3, 8) P25-1, 9) P25-2, 10) P25-3 (A) 11) Standard, 12) C42-1, 13) C42-2, 14) C42-3, 15) C37-1, 16) C37-2, 17) C37-3, 18) C25-1, 19) C25-2

A)




B)