Deciphering the transcriptional changes in *Escherichia coli* strains C41(DE3) and C43(DE3) that makes them a superior choice for membrane protein production.

Chaille T. Webb[1, *] *and Trevor Lithgow*[1].

[1]Infection and Immunity Program, Biomedicine Discovery Institute and Department of Microbiology, Monash University, Melbourne, Vic., Australia

Email addresses: chaille.webb@monash.edu, trevor.lithgow@monash.edu

*Correspondence: chaille.webb@monash.edu

Keywords: Transcriptome, BL21(DE3), C41(DE3), C43(DE3), Walker strains, membrane protein.

ABSTRACT

Background: The overproduction of membrane proteins for functional and structural protein analysis remains a bottleneck in the continuing quest for understanding biological systems. For recombinant membrane proteins, the Walker strains C41(DE3) and C43(DE3) are a valuable tool because they are capable of producing levels of functional protein that would otherwise be toxic to the cell. At the genome level, amongst only a handful of genetic changes, mutations in the *lac*UV5 promoter region upstream from the bacteriophage T7 RNA polymerase gene distinguish these strains from BL21(DE3) but do not inform on how the strains have adapted for superior production of recombinant membrane proteins.

Results: Comparative transcriptomic analyses revealed a moderate change in gene expression in C41(DE3) and C43(DE3) compared to their parent strain BL21(DE3) under standard growth conditions. However, under the conditions used for membrane protein production (with plasmid carriage and addition of IPTG), the differential response of C41(DE3) and C43(DE3) compared to their parent strain BL21(DE3) was striking. Over 2000 genes were differentially expressed in C41(DE3) with a two-fold change and false discover rate < 0.01 and 1700 genes differentially expressed in C43(DE3) compared to their parent strain BL21(DE3).

Conclusion: These results illuminate the cellular adaptations occurring in the Walker strains to alleviate the toxic effects that can occur during membrane protein production, whilst providing changes in metabolism pathways required for membrane protein biogenesis. The BL21(DE3) derivatives strains C41(DE3) and C43(DE3), are adept to the process of membrane biogenesis in *E. coli*, making them superior to their parent strain for the expression of membrane proteins and potentially other toxic proteins.

BACKGROUND

Early estimates that up to 30% of all proteins are integral membrane proteins (1) have been confirmed by proteo-genomic assessments (2). In most cases, solving the structures for these membrane proteins rested on an intractable problem of how to generate sufficient quantities of purified membrane protein. A major hurdle in achieving this is known to be due to cellular constraints that control membrane protein synthesis, targeting and folding (2-6). In this quest, many *E. coli* strains have been developed to enhance the cell's ability to overexpress native and non-native proteins. The BL21(DE3) strain had been engineered to expresses a bacteriophage T7 RNA polymerase (T7RNAP), to transcribe a gene of interest at high efficiency, thus producing large amounts of the corresponding protein (7, 8). In BL21(DE3), the gene encoding T7RNAP is under the control of the *lac*UV5 promoter, a stronger variant of the endogenous *lac* promoter (9). The promoter is subject to repression by LacI. When Isopropyl-β -D-thiogalactopyranoside (IPTG) is added to a cell containing this system, it causes LacI to dissociate from the *lac*UV5 promoter, resulting in the production of T7RNAP and subsequent gene expression.

Numerous studies aimed at high-level expression of membrane proteins in BL21(DE3) have failed, with the membrane protein deemed "toxic". This represents a common bottleneck during protein expression trials (10, 11). A screen for random mutations in the BL21(DE3) host that could survive high-level expression of a membrane protein (the mitochondrial oxoglutarate-malate carrier), recovered the mutant host C41(DE3) (12). One membrane protein, subunit b of bacterial F_1F_0-ATPase, was not tolerated by strain C41(DE3), so a further selection was undertaken generating strain C43(DE3) which also successfully expressed at least four membrane proteins that could otherwise not be expressed: subunit c of the F_1F_0-ATPase, an alanine-H^+ symporter, the mitochondrial ADP/ATP carrier and the mitochondrial phosphate

carrier (12). A subsequent independent evaluation suggested that in 66% of expression constructs tested, the "toxicity" of the plasmids was so high as to prevent identification of any plasmid-transformed BL21(DE3) colonies. The same test when performed in C41(DE3) or C43(DE3) strains demonstrated that all expression constructs could be recovered from transformants with varying expression levels of each membrane protein tested (13).

More than a decade after their discovery, comparative sequence analysis revealed the genetic differences between BL21(DE3) and its derivatives C41(DE3) and C43(DE3) (14). There are seven mutations in C41(DE3), and twelve in C43(DE3), compared to BL21(DE3): common to both derivatives are three single nucleotide polymorphisms (SNPs) in the *lac*UV5 promoter region of the gene encoding T7RNAP. These mutations are responsible for very low levels of T7RNAP and subsequent improvement of protein production(15, 16). Mutations in the genes *yehU* and *rbsD* are present in both C41(DE3) and C43(DE3); however, further analysis of these two genes, encoding a putative two-component sensor protein (YehU) and D-ribose pyranase (RbsD), ruled out any role for these factors in membrane protein expression (14). In addition to the common mutations, C41(DE3) also contains additional point mutations in three genes encoding inner membrane proteins (*proY*, *melB*, *ycgO*) and *yhhA* that encodes a secreted, natively-disordered protein of unknown function. Since all these changes had reverted in the C43(DE3) derivative, they were deemed to not be important for membrane protein expression. C43(DE3) contains mutations in the genes *dcuS, fur, yibJ, yjcO* and *lacI* (14). There are two copies of *lacI* on the BL21(DE3) chromosome, one next to the lac operon and the second in the DE3 region. The mutation of *lacI* in C43(DE3) was mapped to the latter. In addition, there is an IS*1* element inserted into the promoter of *cydA* and an excision of an IS*4* element restores expression of *lon,* which encodes the ATP-dependent protease Lon (14, 16). The Lon protease is associated with regulated protein degradation for the purpose of protein quality control (17,

18). Two large genomic deletions across *ccmF~ompC* and *yjiV-yjjN* were also identified in only C43(DE3) (14).

The general feature that makes a membrane protein particularly toxic to *E. coli* can be gleaned from considering the process of membrane protein biogenesis. Transmembrane proteins have amino acid compositions skewed in favour of hydrophobic residues, particularly Leu, Ile, Val, Phe and Ala (19-24) and the hydroxylated amino acid residues Ser and Tyr (21). This establishes two factors that can be rate-limiting to the translation of membrane proteins (i) the activity of the metabolic pathways that synthesize these amino acids, and (ii) aminoacyl-tRNA availability: the transcripts for membrane proteins often feature rare codons, and these are known to impact on overall membrane protein expression levels through several mechanisms impacting on mRNA stability and rates of protein synthesis (4, 6). Furthermore, high-level membrane protein accumulation depends on the availability of molecular chaperones and protein translocases that catalyze the controlled assembly of the nascent proteins into the bacterial membranes (25).

This study assessed the differential gene expression of C41(DE3) and C43(DE3) strains in comparison to their parental strain BL21(DE3), in the presence and absence of a prototypical protein expression vector with or without the inducer IPTG. The analysis details an array of changes implemented throughout the cell. We describe the trends observed and comment on their relevance to enhance membrane protein expression whilst ensuring the expression remains nontoxic to the cell.

RESULTS

Minimal transcriptional changes occur in C41(DE3) and C43(DE3) compared to their parental strain BL21(DE3).

We first wanted to investigate what changes occur in C41(DE3) and C43(DE3) compared to the parental strain BL21(DE3) in the absence of protein expression. Each strain was grown in rich media (Luria Broth, LB) to mid-log phase (OD_{600nm} = 0.6) at 37 °C. RNA was then isolated and subjected to transcriptional analysis. Each experiment was performed in biological triplicate. RNA libraries were prepared and sequenced. The analysis was performed using RNAsik (26), mapping all reads to the reference strain BL21(DE3) CP001509.3, which was recently updated (27), and the analysis performed in Degust (28). The data quality was confirmed using Degust (Table S1). Using a $\log_2$ fold change ($\log_2$FC) $\geq|1|$ and a false discovery rate (FDR) of less than 0.01 as the cut-off for significant differential expression, a total of 115 genes were differentially expressed in C41(DE3) from which 68 genes were upregulated, and 47 genes were downregulated (Supp Fig. S1a). In C43(DE3) a total of 239 genes were differentially expressed under the same conditions, where 140 genes were upregulated and 99 were downregulated (Supp Fig. S1b). Remarkably, both strains had the same proportion of genes upregulated (59%) and downregulated (41%) overall. C43(DE3) has twice as many genes differentially expressed compared to C41(DE3), however, there was a 33% overlap in upregulated genes (i.e. 52 of 156 unique genes) (Fig. 1a) and a 21% overlap in downregulated genes (i.e. 25 of 121 unique genes) demonstrating some similiarities between the two derivative strains (Fig. 1b).

Transformed strains, induced with IPTG.

Given the use of the strains for recombinant protein production, we assessed the gene expression profiles when the strains were transformed with a plasmid, pACYCDuet-1, and grown in the presence of the inducer IPTG. The pACYCDuet-1 plasmid is a derivative of the

P15A miniplasmid (29). It has a copy number of ~10, carries the *lacI* gene to provide control over gene expression and a gene that confers chloramphenicol resistance for plasmid selection. The assessment was made of the host *E. coli* strains, in the absence of any specific "gene of interest" cloned into the plasmid, so as not to mask the features of C41(DE3) and C43(DE3) that optimize them for membrane protein expression.

Cells containing the pACYCDuet-1 vector were grown in LB growth medium (with chloramphenicol) to mid-log phase at 37 °C, before the addition of IPTG. Cells were then grown for a further two hours at 37 °C. At this point, the cells were collected and RNA was isolated and subjected to transcriptional analysis. All strains grew similarly before and after the addition of IPTG (Fig. 2a). To delineate between the two different experimental parameters, the strains in this experiment were named BL21(DE3)$_{EV+IPTG}$, C41(DE3)$_{EV+IPTG}$ and C43(DE3)$_{EV+IPTG}$.

RNA libraries were prepared and sequenced, the transcripts analysed with Degust, and the data quality assessed statistically (Table S3). Significant changes were observed with a total of 2018 genes identified as differentially expressed in C41(DE3)$_{EV+IPTG}$ where 1024 genes were upregulated and 994 genes were downregulated as defined by a change in expression of $\log_2$FC ≥1 and FDR ≥0.01. In C43(DE3)$_{EV+IPTG}$ a total of 1646 genes were differentially expressed under the same conditions where 827 genes were upregulated and 819 were downregulated (Fig. 2). Comparison of the differential gene expression in C41(DE3) $_{EV+IPTG}$ and C43(DE3) $_{EV+IPTG}$ shows largely similar expression profiles in volcano plots (Fig. 2b and 2c). Venn diagrams demonstrate the majority of genes that are differentially expressed are common to both of the strains: making up 60% of differentially expressed genes in C41(DE3) (1206 of 2018 unique genes) and 73% of genes in C43(DE3) (1206 of 1646 unique genes) (Fig. 2d and

2e). Despite this, there are no similarities or overall patterns with respect to the largest fold change of differentially expressed genes between C41(DE3) and C43(DE3), (Supp. Table S4).

Metabolism pathways are significantly changed in C41(DE3)$_{EV+IPTG}$ and C43(DE)$_{EV+IPTG}$.

Many of the functional pathways appear to be very similar between C41(DE3) and C43(DE3). The differentially expressed genes in C41(DE3)$_{EV+IPTG}$ and C43(DE)$_{EV+IPTG}$ were classified according to their COG pathways (Fig. 3) and also using their KEGG annotations (Supp Fig. 2). Genes that were significantly changed in both C41(DE3)$_{EV+IPTG}$ and C43(DE)$_{EV+IPTG}$ were identified. The large majority of differentially expressed genes encode proteins involved in metabolism, particularly energy production and conversion (C) including TCA cycle genes (*sucABCD*), ATP biosynthesis (*atpAGH*) and respiration (*nuoABCEFGIJKLM*). A range of genes involved in amino acids metabolism were upregulated. These include genes in the biosynthetic pathways and metabolism of cysteine, methionine, tryptophan, tyrosine, phenylalanine, alanine, proline; amino acids utilized particularly for membrane protein biogenesis. We note that intermediates from TCA cycle also feed directly into the pathways for leucine, isoleucine and valine biosynthesis.

Other areas significantly upregulated include carbohydrate transport and metabolism (G), including melibiose transporters (*melA, melB*), trehalose/glucose metabolism (*otsA, otsB, treC*) and pyruvate metabolism (*pykA*) and inorganic ion transport and metabolism (P) inclusing taurine transport (*tauA, tauD*) and oligopeptide ABC transporters (*oppB, oppC, oppD*). Many of these genes encode membrane proteins, which we hypothesized might place demands to increased capacity in the membrane protein biogenesis pathway. In addition, activation of genes mediating transcription/translation processes (K) and cell wall/membrane/envelope biogenesis (M) are upregulated $\log_2$FC>2 with FDR<0.01 (Fig. 3). We find that in many of these pathways the same proportion of genes are being upregulated and downregulated

concomitantly. This suggests there are global changes in play within specific pathways that are unique to the derivative strains possibly affecting their response to membrane protein biogenesis (Fig. 3).

Adaptations for inner membrane biogenesis in C41(DE3) and C43(DE3)

We discovered that genes encoding several molecular chaperones and components of the membrane biogenesis pathway are transcribed at higher levels in $C41(DE3)_{EV+IPTG}$ and $C43(DE3)_{EV+IPTG}$ compared to $BL21(DE3)_{EV+IPTG}$. The schematic in Figure 4 illustrates this by showing the upregulation of genes involved in protein folding and targeting to the Sec translocon for membrane protein folding and assembly. These adaptations also explain why C41(DE3) and C43(DE3) can be used to express soluble proteins that are usually toxic (13, 30).

Table 1. Genes involved in protein folding and targeting that are upregulated in response to induction.

Gene	$C41(DE3)_{EV+IPTG}$ Log_2FC (FDR)	$C43(DE3)_{EV+IPTG}$ Log_2FC (FDR)	Role
secA	2.8($3.6x10^{-10}$)	3.5($1.2x10^{-10}$)	Chaperone and ATPase involved in the Sec protein translocation pathway
secB	2.5 ($9.3x10^{-10}$)	2.1 ($7x10^{-9}$)	Cytoplasmic chaperone involved in the Sec proteins translocation pathway
dnaK	2.7 ($2.8x10^{-7}$)	2.9 ($2.2x10^{-5}$)	Heat-shock protein Hsp70, cytoplasmic chaperone
dnaJ	1.7 ($3.0 x 10^{-5}$)	1.8 ($1.9x10^{-5}$)	Heat-shock protein Hsp40, cytoplasmic chaperone
htpG	3 1.($2.0x10^{-7}$)	3.4 ($6.4x10^{-8}$)	Heat-shock protein Hsp90, cytoplasmic chaperone
ybbN	3.0 ($2.9 x 10^{-9}$)	3.1 ($2.4x10^{-9}$)	Co-chaperone to DnaK/DnaJ/GrpE and GroEL/ES
groEL	2.8 ($1.7x10^{-8}$)	2.8 ($3.1x10^{-8}$)	Cytoplasmic chaperone
groES	4.0 ($8.0 x10^{-9}$)	3.6 ($3.2x10^{-8}$)	Cytoplasmic chaperone
hslU	4.0 ($1.0x10^{-8}$)	3.1($1.5x10^{-7}$)	Cytosolic chaperone and ATPase component of HslUV protease
hslV	4.6 ($1.5x10^{-7}$)	3.7 ($1.4x10^{-6}$)	Peptidase component of HslUV protease
clpB	2.8 ($2.2 x10^{-7}$)	2.8 ($1.6x10^{-7}$)	Cytosolic chaperone
yidC	1.2($4.3 x10^{-7}$)	0.9 ($1.8x10^{-5}$)	Inner membrane insertase

Membrane protein biogenesis relies on protein translocases: in the inner membrane (the SecYEG translocase for unfolded polypeptides or the TAT for folded proteins) and the outer membrane protein (BAM, the core β-barrel assembly machinery) and TAM (the translocation and assembly module of the β-barrel assembly machinery). Periplasmic intermediates are maintained by a series of chaperones and proteases. Table 2 shows no significant changes were seen in genes involved in the membrane-embedded components of the SecYEG machinery, although there was some significant upregulation of some of the TAT translocon components (*tatC, tatD, tatE*). The TAM components of the outer membrane protein assembly machinery were downregulated, while the BAM components remained unchanged.

Genes encoding for the two chaperones of the Sec translocation pathway, SecA and SecB, are both upregulated as are molecular chaperones located in the cytoplasm (*groEL, groES, dnaK* and *ybbN*) and periplasm (*degP, degQ* and *fkpA*). These are documented in Fig. 4B. The increased abundance of these chaperones could provide capacity to collect nascent membrane proteins prior to engagement with the membrane translocases, and to assist in the folding of the domains of the membrane proteins that protrude into the cytoplasm and periplasm.

Several genes involved in polysaccharide biosynthesis were upregulated, which is essential for building the outer leaflet of the outer membrane surface (31). Genes in the retrograde phospholipid trafficking pathway *mlaABCD* were significantly downregulated. It remains untested but possible that these changes might reorganise membrane structure to be permissive for enhanced inner membrane protein accumulation.

Table 2 Differentially expressed genes involved in membrane biogenesis.

Gene	C41(DE3)$_{EV+IPTG}$ Log_2FC (FDR)	C43(DE3)$_{EV+IPTG}$ Log_2FC (FDR)
Periplasmic chaperones and proteases		
surA	0.8 ($7.9x10^{-6}$)	0.6 ($7.3x10^{-5}$)
skp	0.3 (0.2)	0.7 (8.7x10-3)
fkpA	1.7 ($1.2x10^{-7}$)	1.4 ($1.3x10^{-6}$)
degP	1.7 ($3.0x10^{-7}$)	1.1 ($2.6x10^{-5}$)
degQ	2.2 ($8.4x10^{-7}$)	1.4 ($7.4x10^{-5}$)
ptrA	-0.1 (0.6)	-0.4 (0.04)
spy	0.3 (0.1)	0.1 (0.1)
Outer membrane biosynthesis		
bamA	0.3 (0.02)	0.2 (0.1)
bamB	-0.2 (0.2)	0 (0.8)
bamC	-0.4 ($8.9x10^{-3}$)	0 (0.8)
bamD	-0.5 ($7.3x10^{-4}$)	-0.4 ($9.9x10^{-3}$)
bamE	-0.4 (0.1)	-0.4 (0.1)
tamA	-0.7	-1.0 (7.2x $x10^{-4}$)
tamB	-1.1	-1.4 (5.6x $x10^{-3}$)
Inner membrane biosynthesis		
secY	-0.4 (0.01)	-0.2 (0.2)
secE	0.3 (0.03)	-0.3 (0)
secD	-0.5 ($1.3x10^{-3}$)	-0.2 (0.1)
secF	-0.1 (0.4)	0.5 ($2.5x10^{-3}$)
secG	-1.4 ($6.8x10^{-7}$)	-0.8 ($1.0x10^{-4}$)
yajC	0.4 (0.03)	0.4 (0.04)
tatA	-0.2 (0.09)	-0.7 ($3.8x10^{-4}$)
tatB	-0.6 ($6.5x10^{-4}$)	0.8 ($3.9x10^{-5}$)
tatC	1.3 ($4.9x10^{-6}$)	0.5 ($8.9x10^{-3}$)
tatD	0.8 ($3.0x10^{-3}$)	0 (0.9)
tatE	2.6 ($3.6x10^{-6}$)	2.3 ($1.8x10^{-5}$)
Polysaccharide biosynthesis		
lptA	2.3 (1.6x10-9)	1.8 ($1.6x10^{-8}$)
lptB	2.0 (1.2x10-7)	1.4 ($3.8x10^{-6}$)
lptC	2.0 (6.4x10-9)	1.3 ($4.5x10^{-6}$)
lptD	0.6 (5.3x10-5)	-0.1 (0.5)
lptE	-0.6(7.7x10-5)	-0.5 ($1.2x10^{-3}$)
Phospholipid trafficking		
mlaA	-0.9($9.2x10^{-5}$)	-0.2 (0.3)
mlaB	0.2(0.4)	-0.3 (0.2)
mlaC	-0.3(0.01)	-0.8 ($4.9x10^{-6}$)
mlaD	0.1(0.3)	-0.4 ($2.5x10^{-3}$)
mlaE	-1.7($7.7x10^{-8}$)	-2.0 ($3.3x10^{-9}$)
mlaF	-1.6($2.2x10^{-8}$)	-1.9 ($9.4x10^{-9}$)

Significant DEGs are highlighted in grey.

315 ***Functionally unknown, unassigned and uncharacterized***

316 Many of the genes that were differentially expressed are categorised as "functionally unknown"
317 (S) using the COG annotation (Fig. 3), or as "unassigned" using the KEGG annotators (Supp
318 Fig. 2). Since the COG annotations have not been updated for several years, we were interested
319 in confirming the number of genes with still unknown function. The genes were compared to a
320 recently published Y-ome: an updated list of every uncharacterized gene in *E. coli* K-12
321 MG1655 (32). From the total of 1024 genes upregulated in C41(DE3)$_{EV+IPTG}$, 226 of these
322 were assigned to group [S], the "uncharacterized" COG identifier (Supp Table S6). From these
323 226 genes, 138 remain uncharacterized evidenced by their presence in the *E. coli* MG1655 Y-
324 ome list of uncharacterized genes. This accounts for 14-18% of all differentially expressed
325 genes in C41(DE3)$_{EV+IPTG}$ and C43(DE3) $_{EV+IPTG}$ (Supp Table S6). . Further characterisation of
326 these various genes will aid in the overall understanding of cellular responses to enhance
327 membrane protein expression.

328

329 ***How do the genetic mutations in C41(DE3) and C43(DE3) affect their transcriptome profile?***

330 The genomes of C41(DE3) and C43(DE3) are published confirming the known mutations
331 present in the T7RNAP and also identifying several other changes (14). We were interested in
332 determining if these mutations affected the expression of their corresponding genes. Both
333 strains contain mutations in the *lacUV5* promoter region of the T7RNAP that revert it back to a
334 weaker form. A downregulation of T7RNAP is observed in all strains compared to their
335 respective BL21(DE3) controls (Supp. Table S6). As discussed earlier the *rbsD* IS*3* excision
336 causes upregulation of the rbs operon.

337

338 Genomic sequencing of C41(DE3) identified four unique changes not passed onto C43(DE3)
339 (14); Supp. Table S5). Of the three genes containing a single amino acid change, there is a

significant upregulation of *melB* and *yhhA* in our analysis in both C41(DE3)$_{EV+IPTG}$ and C43(FDE3)$_{EV+IPTG}$ in comparison to BL21(DE3)$_{EV+IPTG}$ but no significant change in *ycgO* expression (Supp. Table S5). MelB is a sugar transporter of melibiose coupled with cation exchange (33) that has been shown to be affected by membrane composition of the inner, thus changes may merely reflect a cellular response to an altered membrane environment (34). YhhA is an uncharacterised protein that contains a signal sequence suggesting it localizes to the cell envelope; however, nothing more has been reported about this gene.

C43(DE3) contains mutations in the genes *dcuS, fur, cydA, yibJ, yjcO, lon* and *lacI* (14). The majority of mutations do not invoke any significant changes in their gene expression compared to the relevant BL21(DE3) controls. The gene encoding the ATP-dependent protease Lon, is significantly upregulated due to the excision of an IS*4* element that restores expression of *lon* (Supp. Table S5) (14, 16). The Lon protease is associated with regulated protein degradation for the purpose of protein quality control (17, 18). A point mutation in the lac repressor, *lacI* present in the DE3 region of C43(DE3) results in the downregulation of *lacI* expression in C43(DE3) compared to BL21(DE3) albeit not significant according to our set parameters (log_2FC -1.3, FDR 0.03; Supp Fig. S5). This mutation in C43(DE3) has previously been suggested to be less responsive to its inducer allolactose (13) and subsequently results in superior repression of the lac operon. In the presence of the vector pACYCDuet-1 and induction with IPTG, the *lacI* expression observed is masked by the contribution of the plasmid encoded *lacI* expression.

DISCUSSION

The expression of membrane proteins in bacteria remains a popular strategy in the quest to obtain large amounts of stable and folded recombinant protein for use in structural and

functional studies. The C41(DE3) and C43(DE3) strains remain an initial port of call for the expression of membrane or toxic proteins owing to the anecdotal and published success of these strains over the past few decades since their generation.

The selection process for C41(DE3) and C43(DE3)

The C41(DE3) and C43(DE3) strains have been used to express membrane proteins from prokaryote and eukaryote sources. The genetic changes that occurred in the generation of these strains are limited to a handful of genes particularly involved in transcription and translation, changes which were not informative to the mechanism for increased membrane protein production. We hypothesise that during selection for C41(DE3) and C43(DE3), stress responses triggered by demanding production of the mitochondrial membrane protein has selected for changes in the basal expression parameters of the cell, including features that tune the strains transcriptional program to give better outcomes after addition of IPTG. In this study, we have explored the global changes in transcriptomes by comparing the changes induced in the transcriptome of C41(DE3) and C43(DE3) with those in the parent strain BL21(DE3).

The majority of cellular adaptations occur under conditions for the induction of protein expression. As predicted, this did not require the inclusion of a “protein of interest”. Upon addition of IPTG, C41(DE3) and C43(DE3) strains activate multiple pathways/operons associated with protein production to a greater extent than the parental BL21(DE3). In particular, we found a significant increase in the genes encoding molecular chaperones and proteases, as well as factors required for translocation into or across the inner membrane.

Re-tooling membrane protein biogenesis in C41(DE3) and C43(DE3)

Proteins destined for the inner membrane are typically targeted to the Sec translocon, either co-translationally or post-translationally in an unfolded state assisted by molecular chaperones, for vectorial integration into the inner membrane (35). YidC is another integral membrane protein that catalyzes the integration of membrane proteins into the inner membrane (36). Populations of YidC thereby engage with the ribosome, with the SecYEG-SecDF-YajC complex to enhance membrane protein insertion (37-39). Our results showed the strains C41(DE3) and to a lesser extent C43(DE3), contained a moderate increase in the transcription of *yidC,* that would contribute to an increased capacity to integrate inner membrane proteins (40). Notably, the protein expressed to generate C43(DE3), subunit b form F_1-F_O ATPase, is not a substrate of YidC, thereby not requiring it during inner membrane integration (41). Genes like *yidC* are networked in transcriptional circuitry such that depletion of YidC in *E. coli* causes a change in the expression of ~250 genes of various functions, including energy metabolism; metabolite transport, protein folding and quality control; translation and transcription, and were reported to overlap genes regulated through the CpxAR-mediated stress pathway and phage-shock response (42).

How are these transcriptional networks activated?

The cell envelope stress response in *E. coli* is initiated during membrane protein production via two pathways: the two-component CpxAR system and the σE response. The σE/σ24 transcriptional program is encoded by the gene *rpoE* (43, 44). In our study there was no upregulation of *rpoE* nor are the major negative regulators *rseA, rseB* or *rseC,* however, several genes regulated by the cell envelope stress response directly related to membrane protein folding are upregulated; such as the periplasmic proteases *degP, degQ, degS* and *fkpA* (45). The same scenario occurs in the CpxAR system, that also enacts a transcriptional response to various stresses including membrane-protein defects (46). From our results some the stress-

414 inducible operons (*secA* and *araF*) were upregulated, as were genes encoding chaperones
415 generally under the CpxAR control, including the protease/chaperone *degP*, the disulfide oxidase
416 *dsbA* and *yccA*.
417 In a global assessment of the CpxAR system, De Wulf et al (2002) identified 100 target
418 operons, including the *pps*, *aroF/aroK*, *rpoE/rseABC* and *secA* operons (47). The study also
419 showed that the signal transduction pathway coordinating the Cpx response interacts in
420 unexpected ways with several other transcriptional control circuits.
421
422 Recombinant protein expression is another stress that triggers deployment of alternative sigma
423 factors (48). For example, expression of a fusion protein consisting of the periplasmic maltose-
424 binding protein and beta-galactosidase (MalE-LacZ) blocks the export of other proteins
425 destined for secretion via the Sec pathway, and results in induction of *secA, groEL* and *dnaK*
426 (49), and deletion of *secB* also triggers induction of *dnaK, groEL, htpG, clpB* (F84.1), *grpE* and
427 *groES* (50). Many of these genes are coordinated by the σ^H (σ^{32}) transcriptional program or
428 heat-shock response, regulated by the *rpoH* gene (43, 51-53), We saw many of these genes
429 upregulated in C41(DE3) and C43(DE3) (as outlined in Table 1, Fig. 5).
430
431 ***The distinctions between C41(DE3) and C43(DE3)***
432 Genetically and transcriptionally the C41(DE3) and C43(DE3) strains are largely similar. By
433 way of comparison, we assessed the unique genes that were differentially expressed in
434 C43(DE3). Only 150 genes unique to C43(DE3) were differentially expressed (Fig. 1). When
435 looking at the gene distribution according to known COG pathways, many genes clustered
436 within energy production and conversion, inorganic ion transport and metabolism and
437 carbohydrate transport and metabolism. A further 15% of genes/ORFs have no annotated
438 function. Notably, C43(DE3) was isolated due to its ability to express subunit b of the F_1-F_0

ATPase and most importantly, assemble it into the inner membrane. The other candidate proteins tested for expression by Miroux and colleagues were all expressed in inclusion bodies (12). We postulate that the additional changes implemented by C43(DE3) may have contributed to the already present upregulation of functional pathways seen in C41(DE3) to enhance inner membrane protein biogenesis.

Concluding remarks

When starting a new protein expression project, researchers often apply anecdotal preferences when choosing an expression strain of choice. One key aspect that may be overlooked is that by selecting an inducible vector that contains its own copy of *lacI*, researchers will mask some of those advantageous traits generated in C43(DE3) as identified in our present study. In order to take full advantage of C43(DE3), one would need to use an inducible vector without *lacI* or alternatively utilise the C41(DE3) strain.

In addition to C41(DE3) and C43(DE3), other strains are gaining attention. For tighter control over the switch to induction of protein expression, Lemo21(DE3) is another BL21(DE3) derivative that expresses a T7RNAP inhibitor under the control of a titratable rhamnose promoter (16, 54). Two other BL21(DE3) derivatives directly from BL21(DE3) are C44(DE3) and C45(DE3) (55). These bacterial hosts offer unique features to improve membrane protein expression including tight repression of gene expression at 37 °C, a tunable expression with increased IPTG and continuous protein production throughout the exponential and stationary phases of growth. Sequencing of the C44(DE3) and C45(DE3) genomes identified the mutations responsible for these extensive cellular changes. In the future it will be interesting to explore the gene regulation that has evolved in these strains and compare this with C41(DE3) and C43(DE3), particularly via the Y-ome components. As we have shown here, understanding

464 the breadth of phenotypic and functional changes that occurs from very limited genetic
465 differences will only aid researchers more in their construction of an efficient and selective tool
466 kit for their protein production strategies.
467
468 **Figure Legends**
469 **Figure 1. Transcriptomic changes observed in C41(DE3) and C43(DE3) compared to**
470 **their parent strain BL21(DE3).** Venn diagrams illustrating the differentially expressed genes
471 (DEGs) that are **a** upregulated and **b** downregulated in C41(DE3) and C43(DE3), compared to
472 gene expression in their parent strain BL21(DE3). Differential expression is defined by a
473 change in expression with a $\log_2$FC ≥1 and FDR ≤0.01. **c** and **d** Differential gene expression
474 organised by COG classification in C41(DE3) (blue) and C43(DE3) (red) respectively.
475
476 **Figure 2. Transcriptomic changes observed in C41(DE3)$_{EV+IPTG}$ and C43(DE3)$_{EV+IPTG}$**
477 **compared to their parent strain BL21(DE3)$_{EV+IPTG}$.** **a** Growth curves of C41(DE3) (■),
478 C43(DE3) (▲) and BL21(DE3)(●) grown in LB at 37°C and induced with 0.2 mM IPTG (blue
479 dotted line) and grown for a further 3 hours. Volcano plots showing the differential expression
480 in **b** C41(DE3)$_{EV+IPTG}$ and **c** C43(DE3)$_{EV+IPTG}$ ($\log_2$FC ≥1, FDR ≤0.01). Venn diagrams
481 illustrating the differentially expressed genes (DEGs) that are **d** upregulated and **e**
482 downregulated in C41(DE3)$_{EV+IPTG}$ and C43(DE3)$_{EV+IPTG}$, compared to their parent strain
483 BL21(DE3)$_{EV+IPTG}$.
484
485 **Figure 3. Changes in the transcriptome of C41(DE3)$_{EV+IPTG}$ and C43(DE3)$_{EV+IPTG}$.** **a** and **b**
486 Differential gene expression organised by COG classification in C41(DE3)$_{EV+IPTG}$ (blue) and
487 C43(DE3)$_{EV+IPTG}$ (green) respectively.
488

Figure 4. C41(DE3) and C43(DE3) are ready for protein expression with upregulated genes involved in membrane protein biogenesis. Membrane protein precursor proteins will be targeted to the inner membrane Sec translocon for membrane insertion and folding. The cell can activate many molecular chaperones and protease inhibitors to ensure this pathway remains efficient. This schematic diagram of membrane protein biogenesis illustrating the pathways involved in membrane protein biogenesis. Proteins in shades of red highlight those genes upregulated in C41(DE3) $_{EV+IPTG}$ and C43(DE3)$_{EV+IPTG}$ compared to their expression in BL21(DE3) $_{EV+IPTG}$ (refer to Table 1).

Figure 5. Unique changes in the transcriptome of C41(DE3)$_{EV+IPTG}$ and C43(DE3)$_{EV+IPTG}$. Differential gene expression organised by KEGG classification of genes unique to C41(DE3)$_{EV+IPTG}$ that are **a** upregulated and **b** down regulated and in C43(DE3)$_{EV+IPTG}$ that are **c** upregulated and **d** downregulated.

Methods

Bacterial strains, plasmids, and culture conditions

Overexpress* C41(DE3) and C43(DE3) strains were purchased from Lucigen (cat#60452-1), the BL21(DE3) strain was purchased from Novagen (cat# 70235-3). Triplicate samples were grown overnight at 37 °C and used to inoculate 25mL Luria broth (LB) cultures and grown to mid-log phase (0.4-0.6). For the second round of RNAseq samples, the three strains were transformed with the vector pACYCDuet-1 (Novagen) and selected on LB agar plates containing 34ug/mL chloramphenicol and grown overnight at 37°C. Overnight cultures supplemented with chloramphenicol were grown in triplicate from three independent colonies and grown at 37°C. These were used to inoculate 25mL LB cultures supplemented with chloramphenicol and grown to mid-log phase (0.4-0.6). For RNA extraction, ~1- 1.5mL of

cultures were mixed with 2 times volume of Bacteria Protect Reagent (QIAGEN), vortexed and incubated at room temperature for 5 min. Cells were pelleted (4,600 rpm, 10 min, RT) and then stored at -20°C for RNA extraction.

RNA extraction

Total RNA was purified using RNeasy Kit (QIAGEN Protocol 4 and Protocol 7 with on-column DNase treatment). Briefly, Lysis buffer was added to the cells, followed by the RLT buffer and ethanol. Lysate was loaded onto the RNeasy spin column and centrifuged to bind material. The column was washed with buffer RW1. DnaseI stock solution was added to the column and incubated for 15min at RT. Additional RW1 buffer was added and then centrifuged again. The column was washed with RPE buffer and then the RNA eluted with RNAse-free water. RNA concentration and quality were assessed by A_{260}/A_{280} readings by Nanodrop and presence of degradation assessed on an agarose gel. Further analysis of RNA using Fluorimetric quantitation by Invitrogen Qubit with the Invitrogen Quant-iT dsDNA HS Assay Kit and CE integrity analysis by Agilent Fragment Analyzer (FA) using Agilent HS RNA kit was performed by Micromon.

Transcriptomics and analysis

All cDNA libraries and sequencing was performed by Micromon however two different platforms were used. The first samples [BL21(DE3), C41(DE3), C43(DE3)] cDNA libraries were prepared using the Epicentre ScriptSeq Complete (bacteria) V2 library construction chemistry with 2500ng of input RNA and were then sequenced on an Illumina NextSeq500, SBS V2 chemistry, single-end 75b reads, 1 sequencing lane (high-output), with 0.85pM loading concentration according to the manufacturer's instructions. The second set of sample [$BL21(DE3)_{EV+IPTG}$, $C41(DE3)_{EV+IPTG}$, $C43(DE3)_{EV+IPTG}$] cDNA libraries were also prepared

using the Epicentre ScriptSeq Complete (bacteria) V2 library construction chemistry with 5000ng of input RNA and were then sequenced on a MGITech MGISEQ2000-RS, DNBSEq chemistry V2, with a PE100 Customized V1 sequencing kit, paired-end 100b reads, 1 sequencing lane (FCL) and prepared according to the manufacturer's instructions.

The raw fastq files were analyzed using the RNAsik pipeline (26) where the bwa mem aligner (56) was used to align reads to the BL21(DE3) reference genome CP001509.3, the reference GFF and FASTA files were downloaded from the RefSeq database. Reads were quantified with featureCounts (57) producing the raw gene count matrix and various quality control metrics, all summarised in a MultiQC report (58). The gene count matrix was analysed with Degust (28), a web tool which performs differential expression analysis using limma voom normalisation (59) producing counts per million (CPM) library size normalisation and trimmed mean of M values (TMM) normalisation (60) for RNA composition, and also several quality plots such as classical multidimensional scaling [MDS] and MA plots. Differentially expressed genes were defined as those showing a >2-fold change in expression ($\log_2$ expression ratio ≥ 1) with a false-discovery rate (FDR) of ≤ 0.01. Of note, in sample $BL21(DE3)_{EV+IPTG}$ _3 a large proportion of the sequenced data was mapped to the vector pACYCDuet-1 leaving ~30% of data available (Supp. Table 5.), however, since all replicates behave well on the MDS plot, the differential expression was taken as meaningful and reliable.

Genes were mapped using EggNOGmapper (61, 62) to multiple identifiers including COG and KEGG and GO annotations. Those genes classed as the COG identifier [S] or unannotated were further investigated for function using the recently published Y-ome list (32).

Declarations

Consent for publication Not applicable

Competing interests The authors declare that they have no conflict of interest.

Funding This work was funded through the NHMRC Program in Cellular Microbiology.

Acknowledgments The RNAseq experiments were performed by Micromon at Monash University. The RNAseq data analysis was performed by Kirill Tsyganov from the Monash Bioinformatics Platform, Monash University. We thank them for the analysis and also helpful discussions. We thank Dr Christopher Stubenrauch, Dr Rhys Dunstan, Dr Deeana Deveson and Rebecca Bamert for critical reading of this manuscript.

Author's contributions CTW and TL developed the project design. CTW performed all experiments, analysed the data, wrote the manuscript. TL contributed to the editing and revising of the manuscript. All authors have read and approved the final manuscript.

Access to data The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (63) and are accessible through GEO Series accession number GSE153028 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE153028) and GSE153029 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE153029).

Supplementary Data

Figure S1 Volcano Plots showing the distribution of differential expression in **a** C41(DE3) and **b** C43(DE3).

Figure S2 Distribution of genes upregulated using their KEGG identifiers for A) $C41(DE)_{EV+IPTG}$ and B) $C43(DE3)_{EV+IPTG}$. In $C41(DE)_{EV+IPTG}$, 82% of genes (827 total genes) contained a KEGG identifier and 82% of $C43(DE)_{EV+IPTG}$ genes (1024 total genes).

Table S1 General statistics of transcriptomic analysis of BL21(DE3), C41(DE3) and C43(DE3) strains.

589 **Table S2** Genes with the highest fold change in differential expression of genes (DEGs) in
590 C41(DE3) and C43(DE3).

591 **Table S3** General statistics of transcriptomic analysis of BL21(DE3)$_{EV+IPTG}$, C41(DE3)$_{EV+IPTG}$
592 and C43(DE3)$_{EV+IPTG}$ strains.

593 **Table S4** Genes with the highest fold change in differential expression of genes (DEGs) in
594 C41(DE3)$_{EV+IPTG}$ and C43(DE3)$_{EV+IPTG}$.

595 **Table S5** Differential expression of genes that contain genetic mutations in the BL21(DE3)
596 derivative strains C41(DE3) and C43(DE3).

597 **Table S6** Number of DEGs that are uncharacterized in C41(DE3) and C43(DE3).

598

599 **Group1 RNAseq Data.xlsx** RNAseq analysis using Degust of C41(DE3) and C43(DE3)
600 strains compared to BL21(DE3) with a $\log_2FC \geq 1$ and FDR ≤ 0.01 cutoff.

601 **Group2 RNAseq Data.xlsx** RNAseq analysis using Degust of C41(DE3)$_{EV+IPTG}$ and C43(DE3)
602 $_{EV+IPTG}$ strains compared to BL21(DE3)$_{EV+IPTG}$ with a $\log_2FC \geq 1$ and FDR ≤ 0.01 cutoff.

603

604 **References**

605

606 1. Wallin E, von Heijne G. Genome-wide analysis of integral membrane proteins from
607 eubacterial, archaean, and eukaryotic organisms. Protein Sci. 1998;7(4):1029-38.
608 2. Bill RM, Henderson PJ, Iwata S, Kunji ER, Michel H, Neutze R, et al. Overcoming barriers
609 to membrane protein structure determination. Nat Biotechnol. 2011;29(4):335-40.
610 3. Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene
611 expression in Escherichia coli. Science. 2009;324(5924):255-8.
612 4. Norholm MH, Toddo S, Virkki MT, Light S, von Heijne G, Daley DO. Improved
613 production of membrane proteins in Escherichia coli by selective codon substitutions. FEBS
614 Lett. 2013;587(15):2352-8.
615 5. Sonoda Y, Cameron A, Newstead S, Omote H, Moriyama Y, Kasahara M, et al. Tricks of
616 the trade used to accelerate high-resolution structure determination of membrane proteins.
617 FEBS Lett. 2010;584(12):2539-47.
618 6. Norholm MH, Light S, Virkki MT, Elofsson A, von Heijne G, Daley DO. Manipulating the
619 genetic code for membrane protein production: what have we learnt so far? Biochim Biophys
620 Acta. 2012;1818(4):1091-6.

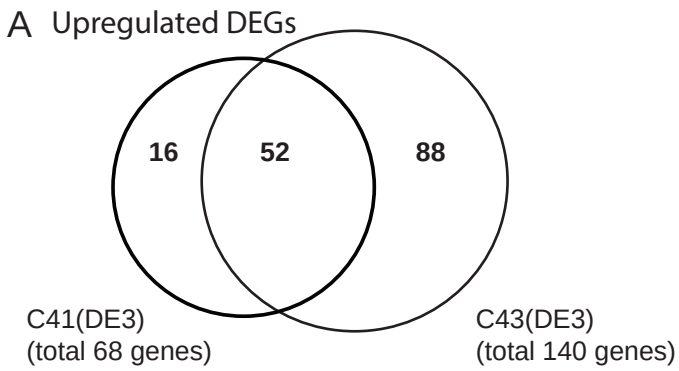
7. Iost I, Guillerez J, Dreyfus M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. J Bacteriol. 1992;174(2):619-22.
8. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. 1990;185:60-89.
9. Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol. 1986;189(1):113-30.
10. Wagner S, Baars L, Ytterberg AJ, Klussmeier A, Wagner CS, Nord O, et al. Consequences of membrane protein overexpression in Escherichia coli. Mol Cell Proteomics. 2007;6(9):1527-50.
11. Zhang Z, Kuipers G, Niemiec L, Baumgarten T, Slotboom DJ, de Gier JW, et al. High-level production of membrane proteins in E. coli BL21(DE3) by omitting the inducer IPTG. Microb Cell Fact. 2015;14:142.
12. Miroux B, Walker JE. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J Mol Biol. 1996;260(3):289-98.
13. Dumon-Seignovert L, Cariot G, Vuillard L. The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). Protein Expr Purif. 2004;37(1):203-6.
14. Kwon SK, Kim SK, Lee DH, Kim JF. Comparative genomics and experimental evolution of Escherichia coli BL21(DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction. Sci Rep. 2015;5:16076.
15. Schlegel S, Genevaux P, de Gier JW. De-convoluting the Genetic Adaptations of E. coli C41(DE3) in Real Time Reveals How Alleviating Protein Production Stress Improves Yields. Cell Rep. 2015;10(10):1758-66.
16. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, et al. Tuning Escherichia coli for membrane protein overexpression. Proc Natl Acad Sci U S A. 2008;105(38):14371-6.
17. Phillips TA, VanBogelen RA, Neidhardt FC. lon gene product of Escherichia coli is a heat-shock protein. J Bacteriol. 1984;159(1):283-7.
18. Van Melderen L, Aertsen A. Regulation and quality control by Lon-dependent proteolysis. Res Microbiol. 2009;160(9):645-51.
19. Baeza-Delgado C, Marti-Renom MA, Mingarro I. Structure-based statistical analysis of transmembrane helices. Eur Biophys J. 2013;42(2-3):199-207.
20. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, et al. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature. 2005;433(7024):377-81.
21. Prilusky J, Bibi E. Studying membrane proteins through the eyes of the genetic code revealed a strong uracil bias in their coding mRNAs. Proc Natl Acad Sci U S A. 2009;106(16):6662-6.
22. Saidijam M, Azizpour S, Patching SG. Comprehensive analysis of the numbers, lengths and amino acid compositions of transmembrane helices in prokaryotic, eukaryotic and viral integral membrane proteins of high-resolution structure. J Biomol Struct Dyn. 2018;36(2):443-64.
23. Senes A, Gerstein M, Engelman DM. Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. J Mol Biol. 2000;296(3):921-36.

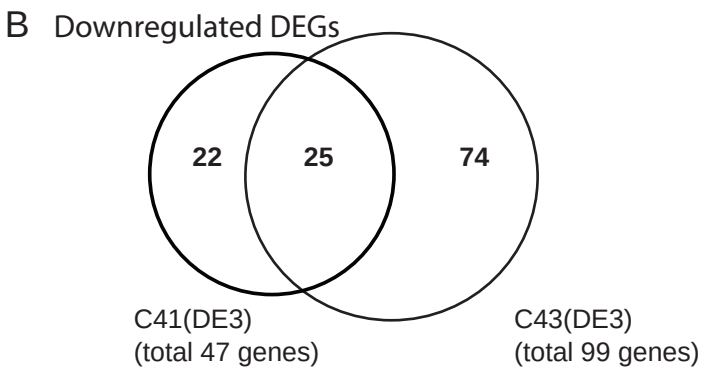
24. Ulmschneider MB, Sansom MS, Di Nola A. Properties of integral membrane protein structures: derivation of an implicit membrane potential. Proteins. 2005;59(2):252-65.
25. Schlegel S, Hjelm A, Baumgarten T, Vikstrom D, de Gier JW. Bacterial-based membrane protein production. Biochim Biophys Acta. 2014;1843(8):1739-49.
26. Tsyganov Kirill AJP, Stuart Kenneth Archer, and David Powell. RNAsik: A Pipeline for Complete and Reproducible RNA-seq Analysis That Runs Anywhere with Speed and Ease. Journal of Open Source Software. 2018;3.
27. Kim S, Jeong H, Kim EY, Kim JF, Lee SY, Yoon SH. Genomic and transcriptomic landscape of Escherichia coli BL21(DE3). Nucleic Acids Res. 2017;45(9):5285-93.
28. Powell D, R., Perry, Andrew, Milton, Michael. . "Degust: Interactive Rna-Seq Analysis. Zenodo. 2019.
29. Chang AC, Cohen SN. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J Bacteriol. 1978;134(3):1141-56.
30. Lebendiker M, Danieli T. Production of prone-to-aggregate proteins. FEBS Lett. 2014;588(2):236-46.
31. Woodward L, Naismith JH. Bacterial polysaccharide synthesis and export. Curr Opin Struct Biol. 2016;40:81-8.
32. Ghatak S, King ZA, Sastry A, Palsson BO. The y-ome defines the 35% of Escherichia coli genes that lack experimental evidence of function. Nucleic Acids Res. 2019;47(5):2446-54.
33. Ethayathulla AS, Yousef MS, Amin A, Leblanc G, Kaback HR, Guan L. Structure-based mechanism for Na(+)/melibiose symport by MelB. Nat Commun. 2014;5:3009.
34. Hariharan P, Tikhonova E, Medeiros-Silva J, Jeucken A, Bogdanov MV, Dowhan W, et al. Structural and functional characterization of protein-lipid interactions of the Salmonella typhimurium melibiose transporter MelB. BMC Biol. 2018;16(1):85.
35. Luirink J, Yu Z, Wagner S, de Gier JW. Biogenesis of inner membrane proteins in Escherichia coli. Biochim Biophys Acta. 2012;1817(6):965-76.
36. Dalbey RE, Kuhn A, Zhu L, Kiefer D. The membrane insertase YidC. Biochim Biophys Acta. 2014;1843(8):1489-96.
37. Berger I, Jiang Q, Schulze RJ, Collinson I, Schaffitzel C. Multiprotein Complex Production in E. coli: The SecYEG-SecDFYajC-YidC Holotranslocon. Methods Mol Biol. 2017;1586:279-90.
38. Kedrov A, Wickles S, Crevenna AH, van der Sluis EO, Buschauer R, Berninghausen O, et al. Structural Dynamics of the YidC:Ribosome Complex during Membrane Protein Biogenesis. Cell Rep. 2016;17(11):2943-54.
39. Stenberg F, Chovanec P, Maslen SL, Robinson CV, Ilag LL, von Heijne G, et al. Protein complexes of the Escherichia coli cell envelope. J Biol Chem. 2005;280(41):34409-19.
40. Nannenga BL, Baneyx F. Reprogramming chaperone pathways to improve membrane protein expression in Escherichia coli. Protein Sci. 2011;20(8):1411-20.
41. van der Laan M, Urbanus ML, Ten Hagen-Jongman CM, Nouwen N, Oudega B, Harms N, et al. A conserved function of YidC in the biogenesis of respiratory chain complexes. Proc Natl Acad Sci U S A. 2003;100(10):5801-6.
42. Wang P, Kuhn A, Dalbey RE. Global change of gene expression and cell physiology in YidC-depleted Escherichia coli. J Bacteriol. 2010;192(8):2193-209.
43. Erickson JW, Gross CA. Identification of the sigma E subunit of Escherichia coli RNA polymerase: a second alternate sigma factor involved in high-temperature gene expression. Genes Dev. 1989;3(9):1462-71.

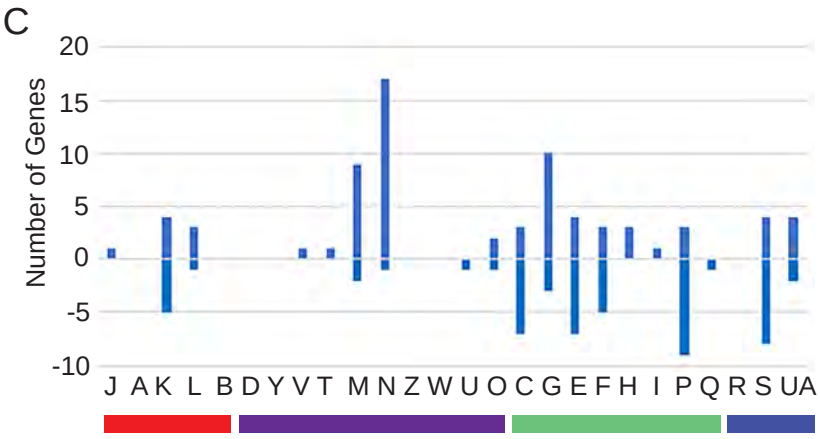
44. Raina S, Missiakas D, Georgopoulos C. The rpoE gene encoding the sigma E (sigma 24) heat shock sigma factor of Escherichia coli. EMBO J. 1995;14(5):1043-55.
45. Rhodius VA, Suh WC, Nonaka G, West J, Gross CA. Conserved and variable functions of the sigmaE stress response in related genomes. PLoS Biol. 2006;4(1):e2.
46. Raivio TL, Leblanc SK, Price NL. The Escherichia coli Cpx envelope stress response regulates genes of diverse function that impact antibiotic resistance and membrane integrity. J Bacteriol. 2013;195(12):2755-67.
47. De Wulf P, McGuire AM, Liu X, Lin EC. Genome-wide profiling of promoter recognition by the two-component response regulator CpxR-P in Escherichia coli. J Biol Chem. 2002;277(29):26652-61.
48. Gamer J, Multhaup G, Tomoyasu T, McCarty JS, Rudiger S, Schonfeld HJ, et al. A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the Escherichia coli heat shock transcription factor sigma32. EMBO J. 1996;15(3):607-17.
49. Ito K, Akiyama Y, Yura T, Shiba K. Diverse effects of the MalE-LacZ hybrid protein on Escherichia coli cell physiology. J Bacteriol. 1986;167(1):201-4.
50. Wild J, Walter WA, Gross CA, Altman E. Accumulation of secretory protein precursors in Escherichia coli induces the heat shock response. J Bacteriol. 1993;175(13):3992-7.
51. Zhao K, Liu M, Burgess RR. The global transcriptional response of Escherichia coli to induced sigma 32 protein involves sigma 32 regulon activation followed by inactivation and degradation of sigma 32 in vivo. J Biol Chem. 2005;280(18):17758-68.
52. Nagai H, Yano R, Erickson JW, Yura T. Transcriptional regulation of the heat shock regulatory gene rpoH in Escherichia coli: involvement of a novel catabolite-sensitive promoter. J Bacteriol. 1990;172(5):2710-5.
53. Wang QP, Kaguni JM. A novel sigma factor is involved in expression of the rpoH gene of Escherichia coli. J Bacteriol. 1989;171(8):4248-53.
54. Schlegel S, Lofblom J, Lee C, Hjelm A, Klepsch M, Strous M, et al. Optimizing membrane protein overexpression in the Escherichia coli strain Lemo21(DE3). J Mol Biol. 2012;423(4):648-59.
55. Angius F, Ilioaia O, Amrani A, Suisse A, Rosset L, Legrand A, et al. A novel regulation mechanism of the T7 RNA polymerase based expression system improves overproduction and folding of membrane proteins. Sci Rep-Uk. 2018;8.
56. Li H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with Bwa-Mem. http://arxivorg/abs/13033997. 2013.
57. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.
58. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-8.
59. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29.
60. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3):R25.
61. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol. 2017;34(8):2115-22.
62. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309-D14.

760 63. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and
761 hybridization array data repository. Nucleic Acids Res. 2002;30(1):207-10.
762

Figure 1







INFORMATION STORAGE AND PROCESSING
[J] Translation, ribosomal structure and biogenesis
[A] RNA processing and modification
[K] Transcription
[L] Replication, recombination and repair
[B] Chromatin structure and dynamics

CELLULAR PROCESSES AND SIGNALING
[D] Cell cycle control, cell division, chromosome partitioning
[Y] Nuclear structure
[V] Defense mechanisms
[T] Signal transduction mechanisms
[M] Cell wall/membrane/envelope biogenesis
[N] Cell motility
[Z] Cytoskeleton
[W] Extracellular structures
[U] Intracellular trafficking, secretion, and vesicular transport
[O] Posttranslational modification, protein turnover, chaperones

METABOLISM
[C] Energy production and conversion
[G] Carbohydrate transport and metabolism
[E] Amino acid transport and metabolism
[F] Nucleotide transport and metabolism
[H] Coenzyme transport and metabolism
[I] Lipid transport and metabolism
[P] Inorganic ion transport and metabolism
[Q] Secondary metabolites biosynthesis, transport and catabolism

POORLY CHARACTERIZED
[R] General function prediction only
[S] Function unknown
unassigned

Figure 2

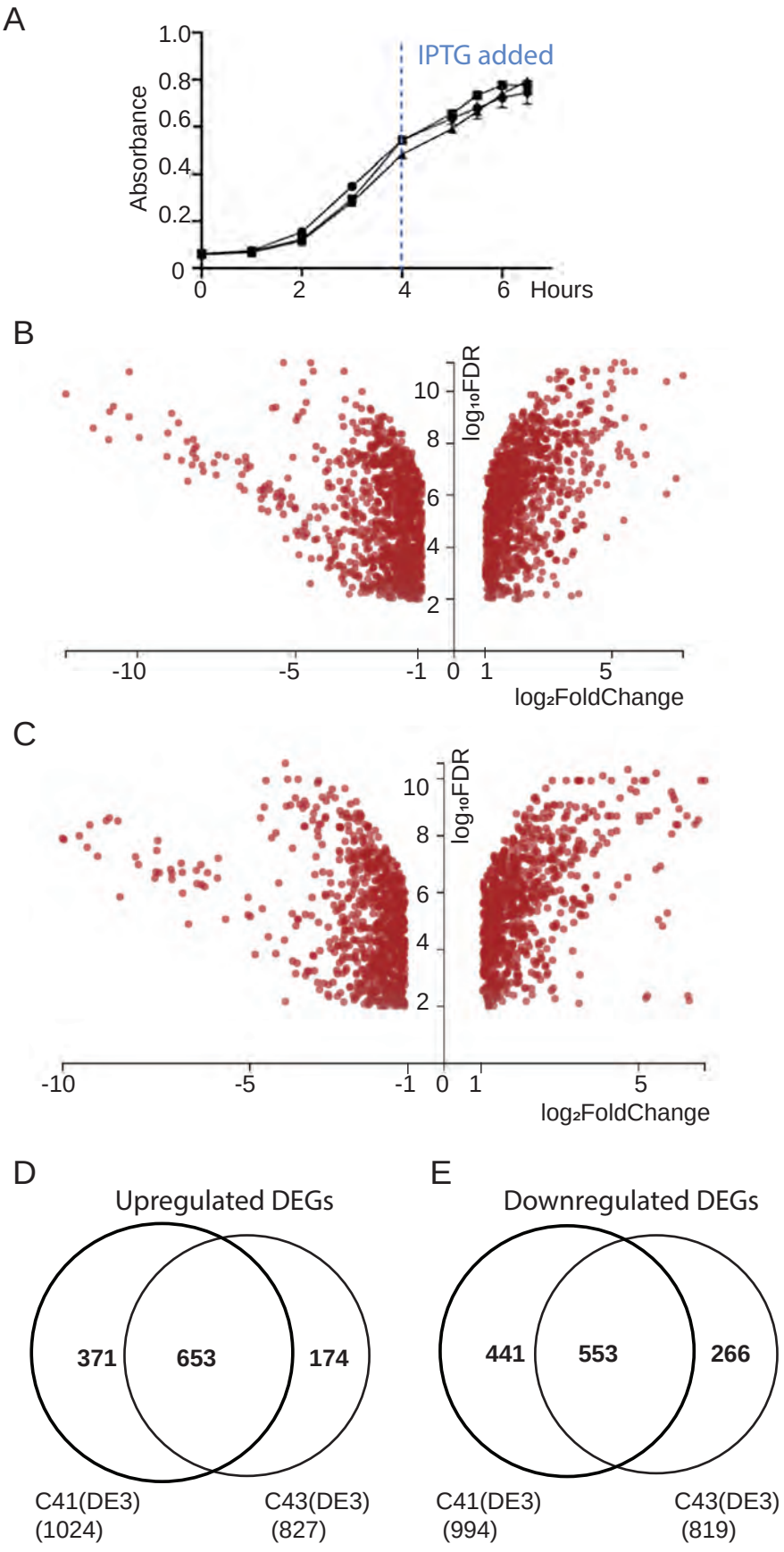
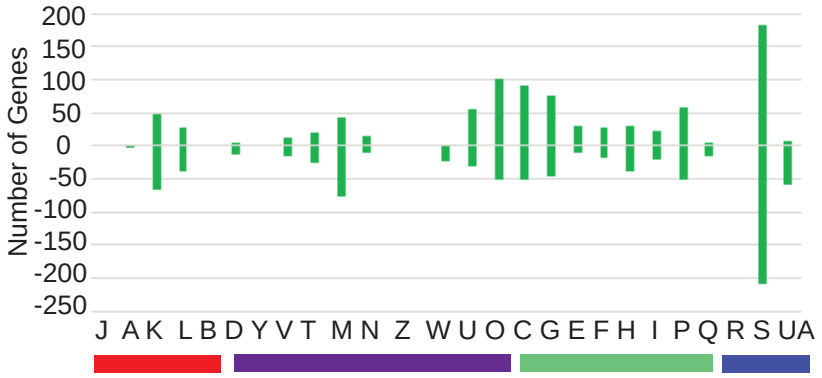

A
IPTG added
Absorbance
Hours
B
log₁₀FDR
log₂FoldChange
C
log₁₀FDR
log₂FoldChange
D
Upregulated DEGs
371
653
174
C41(DE3)
(1024)
C43(DE3)
(827)
E
Downregulated DEGs
441
553
266
C41(DE3)
(994)
C43(DE3)
(819)

Figure 3

A

B



INFORMATION STORAGE AND PROCESSING
[J] Translation, ribosomal structure and biogenesis
[A] RNA processing and modification
[K] Transcription
[L] Replication, recombination and repair
[B] Chromatin structure and dynamics

CELLULAR PROCESSES AND SIGNALING
[D] Cell cycle control, cell division, chromosome partitioning
[Y] Nuclear structure
[V] Defense mechanisms
[T] Signal transduction mechanisms
[M] Cell wall/membrane/envelope biogenesis
[N] Cell motility
[Z] Cytoskeleton
[W] Extracellular structures
[U] Intracellular trafficking, secretion, and vesicular transport
[O] Posttranslational modification, protein turnover, chaperones

METABOLISM
[C] Energy production and conversion
[G] Carbohydrate transport and metabolism
[E] Amino acid transport and metabolism
[F] Nucleotide transport and metabolism
[H] Coenzyme transport and metabolism
[I] Lipid transport and metabolism
[P] Inorganic ion transport and metabolism
[Q] Secondary metabolites biosynthesis, transport and catabolism

POORLY CHARACTERIZED
[R] General function prediction only
[S] Function unknown
unassigned

Figure 4.

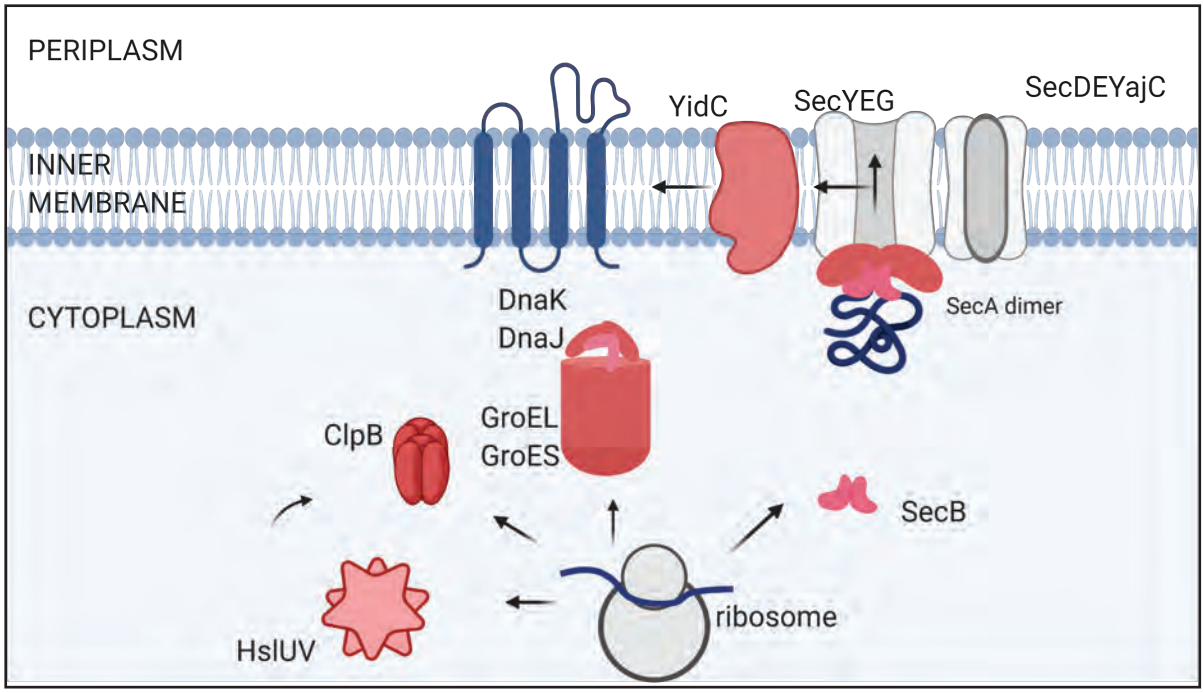
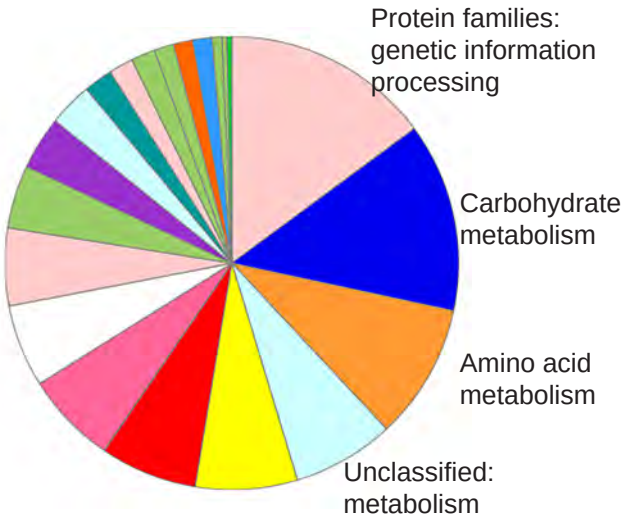

PERIPLASM
INNER
MEMBRANE
CYTOPLASM
YidC
SecYEG
SecDEYajC
SecA dimer
DnaK
DnaJ
GroEL
GroES
ClpB
SecB
ribosome
HslUV

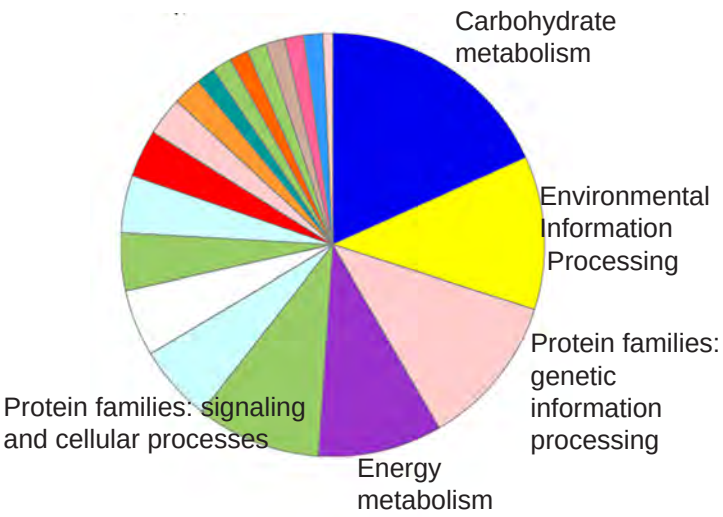
Figure 5.

A



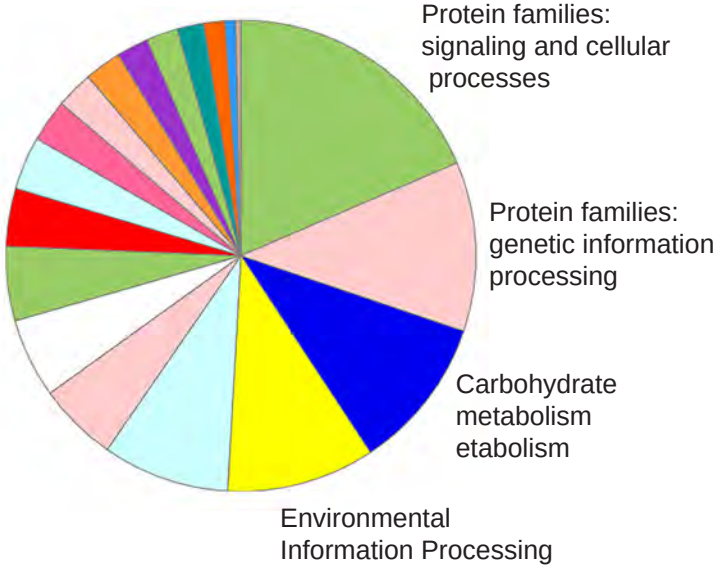
C41(DE3)EV+IPTG
upregulated DEGs
(289 of 371)

C



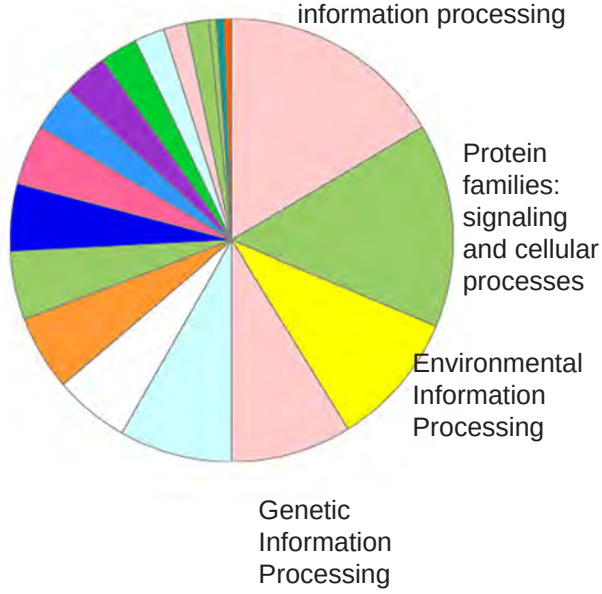
C43(DE3)EV+IPTG
upregulated DEGs
(137 of 174)

B



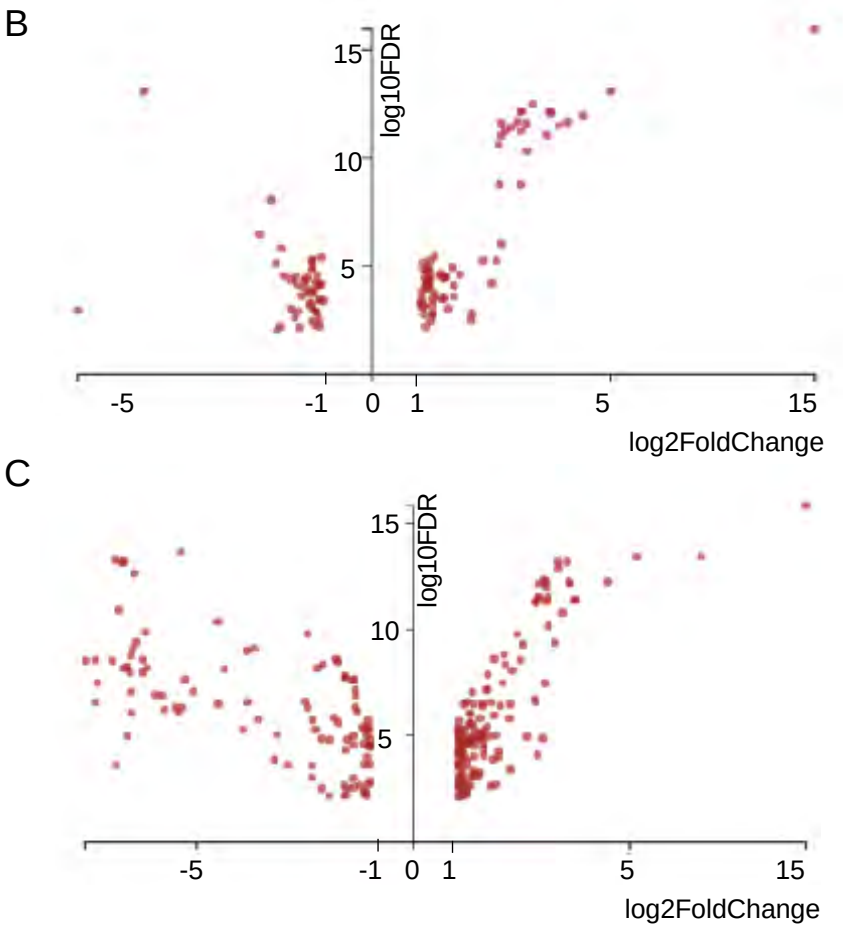
C41(DE3)EV+IPTG
downregulated DEGs
(275 of 441)

D



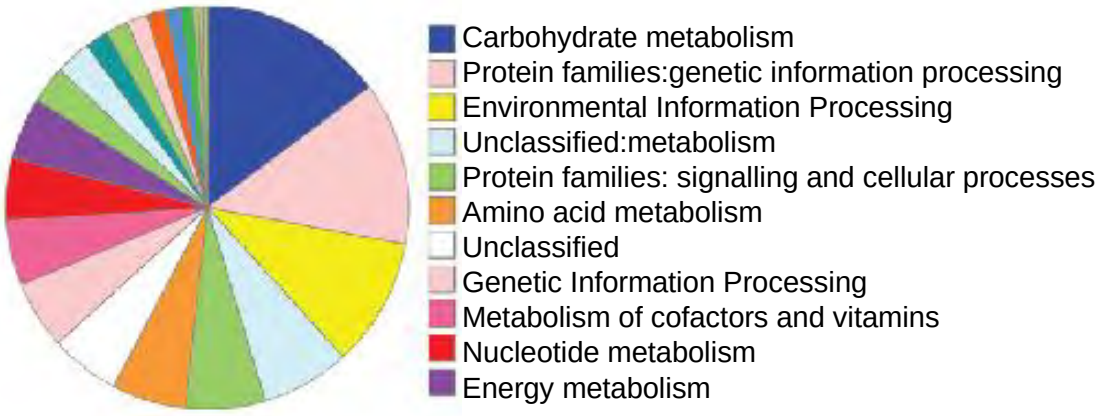
C43DE3)EV+IPTG
downregulated DEGs
(186 of 266)

Supplementary Figure S1 Volcano Plots showing the distribution of differential expression in **a** C41(DE3) and **b** C43(DE3).



Supplementary Figure S2 Distribution of genes upregulated using their KEGG identifiers for A) C41(DE)$_{EV+IPTG}$ and B) C43(DE3)$_{EV+IPTG}$. In C41(DE)$_{EV+IPTG}$, 82% of genes (827 total genes) contained a KEGG identifier and 82% of C43(DE)$_{EV+IPTG}$ genes (1024 total genes).

A



B

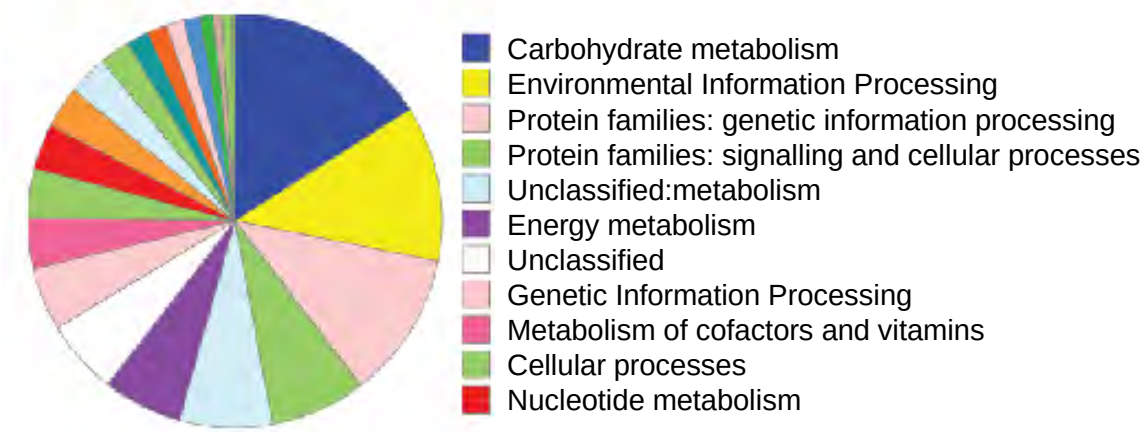


Table S1 General statistics of transcriptomic analysis of BL21(DE3), C41(DE3) and C43(DE3) strains.

Sample	M Assigned	% Assigned	M Reads Mapped	% Mapped	% Aligned
BL21(DE3)_1	11.4	81.4	28.0	99.6	100
BL21(DE3)_2	196.4	80.8	484.1	99.6	100
BL21(DE3)_3	4.2	23.9	35.2	99.6	100
C41(DE3)_1	14.3	77.1	36.8	99.4	99
C41(DE3)_2	12.8	77.6	32.8	99.5	99
C41(DE3)_3	11.9	77.4	30.6	99.5	100
C43(DE3)_1	9.9	78.0	25.2	99.6	100
C43(DE3)_2	12.1	79.5	30.1	99.6	100
C43(DE3)_3	19.0	79.5	47.5	99.6	100

Table S2 Genes with the highest fold change in differential expression of genes (DEGs) in C41(DE3) and C43(DE3).

	C41(DE3)			C43(DE3)			
Gene ID	**Name**	**log2FC**	**FDR**	**Gene ID**	**Name**	**log2FC**	**FDR**
Upregulated				**Upregulated**			
ACT45429.1	*rbsA*	9.3	1.06E-16	ACT45429.1	*rbsA*	9.08	1.46E-16
ACT45430.1	*rbsC*	5	7.88E-14	ACT45197.1	*yhjX*	6.66	3.61E-14
ACT45431.1	*rbsB*	4.44	1.06E-12	ACT45430.1	*rbsC*	5.18	3.61E-14
ACT42964.1	*flgB*	4.1	2.09E-12	ACT45431.1	*rbsB*	4.5	5.45E-13
ACT42965.1	*flgC*	3.93	3.12E-12	ACT42964.1	*flgB*	3.75	4.11E-12
ACT42963.1	*flgA*	3.76	8.49E-13	ACT42965.1	*flgC*	3.73	4.17E-12
ACT42966.1	*flgD*	3.72	6.91E-13	ACT42963.1	*flgA*	3.62	7.27E-13
ACT42962.1	*flgM*	3.67	8.93E-12	ACT42966.1	*flgD*	3.6	5.45E-13
ACT42967.1	*flgE*	3.38	3.03E-13	ACT42290.1	*lon*	3.56	6.51E-14
ACT43703.2	*flhE*	3.26	4.79E-11	ACT42962.1	*flgM*	3.48	1.60E-11
ACT42961.1	*flgN*	3.24	2.37E-12	ACT42967.1	*flgE*	3.35	1.37E-13
ACT42969.1	*flgG*	3.16	6.91E-13	ACT45177.1	*dctA*	3.35	6.51E-14
ACT42970.1	*flgH*	3.14	5.31E-12	ACT43705.1	*flhB*	3.28	4.27E-10
ACT43705.1	*flhB*	3.13	1.60E-09	ACT43703.2	*flhE*	3.13	6.77E-11
Downregulated				**Downregulated**			
ACT45427.1	*insK-6*	-6.13	1.16E-03	ACT46012.1	*hpaR*	-7.59	2.99E-09
ACT42586.1	*T7 RNAP*	-4.76	7.88E-14	ACT43962.1	*alkB*	-7.37	2.67E-07
ACT45583.1	*rhaB*	-2.34	3.49E-07	ACT46009.1	*hpaD*	-7.36	2.73E-09
ACT42050.1	*fhuA*	-2.11	8.91E-09	ACT46011.1	*hpaG*	-7.32	3.37E-08
ACT44523.1	*hyuA*	-1.99	7.60E-06	ACT46013.1	*tsr*	-6.98	2.99E-09
ACT45258.1	*NA*	-1.98	9.54E-03	ACT45997.1	*hsdR*	-6.91	5.32E-14
ACT42860.1	*yccF*	-1.9	6.69E-03	ACT45427.1	*insK-6*	-6.9	2.52E-04
ACT45671.1	*zraP*	-1.88	1.57E-06	ACT46000.1	*yjiX*	-6.82	1.26E-11
ACT45582.2	*rhaA*	-1.82	2.87E-05	ACT45999.1	*yjiA*	-6.75	6.51E-14
ACT45581.1	*rhaD*	-1.68	4.57E-05	ACT46001.1	*yjiY*	-6.71	6.51E-14
ACT45980.1	*yjiH*	-1.68	1.01E-03	APW29184.1	*yojO*	-6.71	6.75E-09
ACT45580.1	*rhaM*	-1.6	2.56E-03	ACT46004.1	*hpaA*	-6.64	5.62E-09
ACT44529.1	*ssnA*	-1.58	3.02E-05	ACT43963.1	*ada*	-6.62	1.02E-05
ACT45155.1	*slp*	-1.56	7.42E-05	ACT45991.1	*mcrC*	-6.57	1.09E-08

Table S3. General statistics of transcriptomic analysis of BL21(DE3)$_{EV+IPTG}$, C41(DE3)$_{EV+IPTG}$ and C43(DE3)$_{EV+IPTG}$ strains.

Sample	M Assigned	% Assigned	M Reads Mapped	% Mapped	% Aligned
BL21(DE3)$_{EV+IPTG}$ _1	1.4	13.2	2.8	26.5	27
BL21(DE3)$_{EV+IPTG}$ _2	1.4	12.4	2.9	25.5	25
BL21(DE3)$_{EV+IPTG}$ _3	7.5	2.8	2.8	18.0	18
C41(DE3)$_{EV+IPTG}$ _1	6.4	72.1	8.3	93.6	94
C41(DE3)$_{EV+IPTG}$ _2	6.4	72.4	8.2	93.4	93
C41(DE3)$_{EV+IPTG}$ _3	5.2	70.3	6.9	92.8	93
C43(DE3)$_{EV+IPTG}$ _1	6.1	66.7	7.9	85.8	86
C43(DE3)$_{EV+IPTG}$ _2	5.8	68.2	7.4	87	87
C43(DE3)$_{EV+IPTG}$ _3	5.8	68.3	7.4	86	87

Table S4. Genes with the highest fold change in differential expression of genes (DEGs) in C41(DE3)$_{EV+IPTG}$ and C43(DE3)$_{EV+IPTG}$.

	C41(DE3)$_{EV+IPTG}$			C43(DE3)$_{EV+IPTG}$			
Gene ID	**Name**	**log$_2$FC**	**FDR**	**Gene ID**	**Name**	**log$_2$FC**	**FDR**
Upregulated				**Upregulated**			
ACT45430	*rbsC*	7.25	2.39E-11	ACT45430	*rbsC*	6.70	1.16E-10
ACT45161	*gadE*	7.03	2.29E-07	ACT45429	*rbsA*	6.60	1.16E-10
ACT45429	*rbsA*	6.73	4.09E-11	ACT42966	*flgD*	6.53	2.78E-09
ACT45768	*yjdN*	6.73	8.97E-07	ACT43845	*gatB*	6.46	3.95E-09
ACT44788	*tdcB*	6.08	2.73E-08	ACT44795	*garL*	6.30	6.24E-03
ACT45155	*slp*	6.00	2.35E-09	ACT45386	*tnaC*	6.27	4.19E-03
ACT45781	*melB*	5.71	1.64E-11	ACT42967	*flgE*	6.24	2.03E-09
ACT45433	*rbsR*	5.69	6.54E-08	ACT42969	*flgG*	6.04	1.15E-09
ACT44630	*ECD_0282*	5.59	3.31E-08	ACT42965	*flgC*	6.01	4.22E-09
ACT46001	*yjiY*	5.47	2.06E-08	ACT42968	*flgF*	5.96	4.06E-09
ACT45432	*rbsK*	5.45	1.64E-11	ACT42964	*flgB*	5.93	5.49E-10
ACT44631	*ECD_0282*	5.40	1.03E-09	ACT43843	*gatD*	5.89	4.09E-10
ACT44026	*nuoK*	5.37	8.89E-06	ACT42971	*flgI*	5.85	8.74E-09
ACT43983	*nrdA*	5.36	7.63E-12	ACT43844	*gatC*	5.85	2.94E-10
Downeregulated				**Downeregulated**			
ACT43859	*ECD_0203*	-12.27	1.25E-10	ACT43959	*eco*	-9.81	1.24E-08
ACT43857	*ECD_0203*	-11.40	2.58E-09	ACT43956	*napA*	-9.77	1.38E-08
ACT43860	*ECD_0203*	-10.92	7.25E-09	ACT43960	*mqo*	-9.37	1.00E-08
ACT43837	*yegQ*	-10.88	5.92E-10	ACT45998	*mrr*	-9.17	2.56E-08
ACT43858	*ECD_0203*	-10.77	3.74E-10	ACT45996	*hsdM*	-9.11	4.22E-09
ACT43822	*dcd*	-10.27	9.56E-10	ACT43952	*napC*	-8.91	5.27E-08
ACT43848	*gatY*	-10.26	1.71E-11	ACT45997	*hsdR*	-8.69	2.99E-09
ACT43821	*asmA*	-9.95	5.90E-09	ACT45427	*insK-6*	-8.60	2.37E-09
ACT43799	*wcaN*	-9.07	3.22E-08	ACT43954	*napH*	-8.39	3.05E-09
ACT43827	*yegI*	-9.04	3.96E-09	ACT45992	*mcrB*	-8.33	1.41E-06
ACT43847	*gatZ*	-8.92	6.93E-10	ACT46013	*tsr*	-7.93	5.52E-08
ACT45427	*insK-6*	-8.71	1.61E-09	ACT46004	*hpaA*	-7.86	1.04E-07
ACT43825	*alkA*	-8.61	4.54E-08	ACT43951	*ccmA*	-7.43	3.12E-07
ACT43824	*yegE*	-8.43	8.33E-09	ACT43958	*napF*	-7.41	1.88E-07

Table S5 Differential expression of genes that contain genetic mutations in the BL21(DE3) derivative strains C41(DE3) and C43(DE3).

Gene name	Gene ID	Genetic Mutation	C41(DE3)		C43(DE3)		C41(DE3)$_{EV+IPTG}$		C43(DE3)$_{EV+IPTG}$	
			log$_2$FC	FDR	log$_2$FC	FDR	log$_2$FC	FDR	log$_2$FC	FDR
Present in C41(DE3) only										
proY	ACT42249.1	Synonymous SNP	-0.1	7.00E-01	0	9.00E-01	-1.3	1.80E-07	-1.5	8.54E-07
melB	ACT45781.1	Non synonymous SNP	0.2	5.00E-01	0.2	2.60E-01	5.7	1.60E-11	3.1	2.06E-09
ycgO /cvrA	ACT43058.1	Non synonymous SNP	-0.2	4.00E-01	0.2	3.00E-01	-0.75	7.90E-04	-0.01	9.50E-01
yhhA	ACT45099.1	Non synonymous SNP	-0.3	9.00E-02	0.1	5.00E-01	3.3	1.80E-09	2.2	5.60E-09
Present in C43(DE3) only										
dcuS	ACT45785.1	Frameshift	0.2	5.00E-01	2.3	8.60E-09	-2	4.70E-06	-0.3	1.80E-01
fur	ACT42516.1	Val insertion	-0.3	1.00E-01	-0.1	4.00E-01	0.6	8.20E-03	0.97	1.90E-05
lacI	ACT42585.1	Non synonymous SNP	0.3	-6.00E-01	-1.2	3.00E-02	2.3	7.30E-04	2.6	1.60E-04
lon	ACT42290.1	Activation of Lon	0.1	6.00E-01	3.6	6.50E-14	-0.4	1.50E-01	4.4	4.20E-09
yibJ / rhsJ	ACT45249.1	Synonymous SNP	-0.5	6.00E-02	-0.1	8.00E-01	-0.8	2.00E-02	-1.3	8.90E-04
yjcO	ACT45740.1	Non synonymous SNP	-0.4	4.00E-02	-0.5	8.90E-03	0.9	1.50E-04	0.2	2.40E-01
cydA	ACT42570.1	IS1 insertion in promoter	-0.2	3.20E-01	-0.9	8.90E-06	0.3	2.00E-02	0.7	7.40E-05
Present in both strains										
I	ACT42586.1	mutations in the lac*UV5* promoter	-4.8	7.90E-14	-1.2	3.00E-02	-4.5	7.60E-12	-4.1	2.90E-11
rbsD		IS3 excision								
yehU	ACT43879.1	Non-synonymous SNP	0.5	6.40E-03	0.2	3.00E-01	-0.94	1.90E-05	-0.4	1.00E-02

**Differential expression in comparison to BL21(DE3). # Differential expression compared toBL21(DE3)$_{EV+IPTG}$. 'np' represents genes that were not present in this analysis.*

Table S6. Number of DEGs that are uncharacterized in C41(DE3) and C43(DE3).

	C41(DE3)$_{EV+IPTG}$ upregulated	**C41(DE3)$_{EV+IPTG}$ downregulated**	**C43(DE3)$_{EV+IPTG}$ upregulated**	**C43(DE3)$_{EV+IPTG}$ downregulated**
Total DEGs	1024	994	827	819
Genes with unknown function*	226	294	192	267
Genes matched to *E. coli* MG1655	183	222	171	172
E. coli MG1655 Y-ome	138	180	131	136
% of genes with unknown function with respect to total DEGs	14%	18%	16%	17%

**Genes of unknown function as identified by the COG identifier 'S' and unannotated.*