Published in final edited form as:
Biotechnol Bioeng. 2014 May ; 111(5): 980–999. doi:10.1002/bit.25169.

Dynamic Transcriptional Response of *Escherichia coli* to Inclusion Body Formation

Faraz Baig[1], **Lawrence P. Fernando**[b,§], **Mary Alice Salazar**[b,§], **Rhonda R. Powell**[c], **Terri F. Bruce**[c], and **Sarah W. Harcum**[*,a]

[a]Department of Bioengineering, Clemson University, 301 Rhodes Research Center, Clemson, SC 29634, United States

[b]Department of Chemistry, Clemson University, 219 Hunter Laboratories, Clemson, SC 29634, United States

[c]Department of Biological Sciences, Clemson University, 132 Long Hall, Clemson, SC 29634, United States

Abstract

Escherichia coli is used intensively for recombinant protein production, but one key challenge with recombinant *E. coli* is the tendency of recombinant proteins to misfold and aggregate into insoluble inclusion bodies (IBs). IBs contain high concentrations of inactive recombinant protein that require recovery steps to salvage a functional recombinant protein. Currently, no universally effective method exists to prevent IB formation in recombinant *E. coli*. In this study, DNA microarrays were used to compare the *E. coli* gene expression response dynamics to soluble and insoluble recombinant protein production. As expected and previously reported, the classical heat-shock genes had increased expression due to IB formation, including protein folding chaperones and proteases. Gene expression levels for protein synthesis-related and energy-synthesis pathways were also increased. Many transmembrane transporter and corresponding catabolic pathways genes had decreased expression for substrates not present in the culture medium. Additionally, putative genes represented over one-third of the genes identified to have significant expression changes due to IB formation, indicating many important cellular responses to IB formation still need to be characterized. Interestingly, cells grown in 3% ethanol had significantly reduced gene expression responses due to IB formation. Taken together, these results indicate that IB formation is complex, stimulates the heat-shock response, increases protein and energy synthesis needs, and streamlines transport and catabolic processes, while ethanol diminished all of these responses.

Keywords

Protein aggregates; Inclusion bodies; DNA microarrays; stress; ethanol

Introduction

Escherichia coli is one of the most intensively used organisms for recombinant protein production. It can grow rapidly on inexpensive media and is easily modified genetically (Swartz 2001). However, *E. coli* also has a tendency to misfold recombinant proteins, forming insoluble inclusion body (IB) aggregates in the cell (Baneyx 1999; Baneyx and Mujacic 2004; Basu et al. 2011; Swartz 2001). IBs are dense refractile particles that contain

[*]Corresponding author. Tel.: + 1-864-656-6865; harcum@clemson.edu(S.W. Harcum).
[§]These authors contributed equally

mostly the recombinant protein (Allen et al. 1992; Carrio and Villaverde 2002; Ventura and Villaverde 2006; Villaverde and Carrio 2003); however, proteases have been isolated from IBs (Jordan and Harcum 2002) and heat-shock proteins have been identified as associated with IBs (Carrio and Villaverde 2005). Additionally, IBs have amyoid-like structure, where some IB-embedded proteins retain biological activity (Garcia-Fruitos et al. 2007b; Gatti-Lafranconi et al. 2011; Peternel et al. 2007; Sabate et al. 2010). Unlike early notions that IBs were inert, recent work has demonstrated that IBs are dynamic entities within the cell that migrate to the cell poles, fuse, and dissolve as the cells grow (Rokney et al. 2009). The high degree of purity, biological activity, and consistent structures has initiated research to use IBs as drug-delivery devices (Garcia-Fruitos et al. 2012; Garcia-Fruitos and Villaverde 2010; Liovic et al. 2012; Peternel and Komel 2010; Rodriguez-Carmona and Villaverde 2010)

In order to obtain a biologically active protein from IBs, most often additional time-consuming and low-yield purification steps are required (Basu et al. 2011; Hoffmann and Rinas 2001); however, recent progress has been made to develop less time-consuming process with higher yields (Peternel 2013; Peternel and Komel 2010; Peternel and Komel 2011; Porowinska et al. 2012; Singh et al. 2012). In parallel to purification improvements, many culture and cloning methods have been developed to reduce or control IB accumulation. These methods include reduced culture temperatures, reduced gene expression rates, adjusted codon usage, protein engineering, co-expression of molecular chaperones, and heat-stimulation of chaperones (Chen et al. 2002; Garcia-Fruitos et al. 2005; Hoffmann et al. 2004; Ignatova et al. 2000; Jevsevar et al. 2005; Martinez-Alonso et al. 2010; Pan et al. 2003; Petersson et al. 2004; Schlieker et al. 2002; Strandberg and Enfors 1991; Striedner et al. 2003; Villaverde and Carrio 2003). Despite all these characterization studies, it is not yet possible to predict *a priori* the solubility of a recombinant protein with greater than 90% accuracy (Agostini et al. 2012; Diaz et al. 2010; Magnan et al. 2009; Smialowski et al. 2012; Smialowski et al. 2007).

DNA microarray data have been used to determine coordinated regulation patterns, regulatory circuits, and signal transduction systems in *E. coli* (Cheung et al. 2003; Choi et al. 2003; Conway and Schoolnik 2003; Duerrschmid et al. 2008; Gill et al. 2001; Haddadin and Harcum 2005; Harcum and Haddadin 2006; Lee and Lee 2005; Mahnic et al. 2012; Marisch et al. 2013; Nahku et al. 2010; Oh and Liao 2000; Oh et al. 2002; Richmond et al. 1999; Rohlin et al. 2002; Selinger et al. 2003; Wendisch et al. 2001; Yoon et al. 2003). With respect to IBs, two past studies examined the *E. coli* transcriptome responses to IBs (Lesley et al. 2002; Smith 2007); however, these studies only examined the transcriptome after significant amounts of IBs accumulated. Since IBs form over time after induction, examining the dynamic change in gene expression may lead to a better understanding of the cascade of transcriptional events that lead to IBs.

The objective of this study was to determine the dynamic transcriptional response of *E. coli* to IB formation. DNA microarrays were used to characterize gene expression changes due to IB formation. The gene expression changes due to IB formation were directly compared to gene expression changes due to soluble recombinant protein production. Since the addition of ethanol has been shown to increase the solubility of IB-prone proteins, the effects of ethanol on the gene expression response to IB formation were also examined.

Materials and Methods

Bacterial Strain and Plasmids

E. coli MG1655 were obtained from the American Type Culture Collection (ATCC). The plasmid pTVP1GFP (gift from A. Villaverde) encodes the VP1 capsid of foot-and-mouth

disease (Liu et al. 2006) fused to green fluorescent protein (GFP) (Garcia-Fruitos et al. 2007a). The pGFPCAT plasmid was constructed from the pTrcHis-GFP_{UV}/CAT plasmid (gift from WE Bentley) (Cha et al. 2000), where the GFP_{UV} was replaced with the GFP from the pTVP1GFP plasmid; however, the GFP protein is located on the amino-terminus of the VP1 protein and on the carboxyl-terminus of the CAT protein (Salazar et al. 2013). *E. coli* MG1655 were transformed with either the pTVP1GFP or pGFPCAT plasmid. The plasmids are very similar with common lineages, including pBR322 origins, *lacI* expression, ampicillin-resistance, and isopropyl β-D-1-thiogalactopyranoside (IPTG) inducibility through a *trc* promoter. Additionally, the GFPCAT and VP1GFP fusion proteins have similar sizes, 519 and 451 amino acid residues, respectively (58.88 and 50.42 kDa) (Cha et al. 2000; Garcia-Fruitos et al. 2007b).

Culture Conditions

E. coli MG1655 pTVP1GFP and pGFPCAT were cultured in a minimal medium described previously (Korz et al. 1995). Frozen stock (1 mL, stored at −80°C) were thawed and added to the minimal medium containing 40 μg/mL ampicillin (Hyclone). Cells were grown overnight in a shaker incubator (C24, New Brunswick Scientific, Inc.) at 37°C and 250 rpm to approximately 2.5 OD. ODs were obtained at 600 nm with a spectrophotometer (Spectronic 20 Genesys), where 1 OD is equivalent to 0.50 g dry cell weight per L. Samples were diluted with deionized water to obtain absorbance readings in the linear range (0 to 0.25 OD). The overnight cultures were used to inoculate the experimental flasks, which contained 120 mL medium in a 500 mL shake flasks for an initial cell density of 0.05 OD. These cultures were placed in a water bath shaker at 37°C and 200 rpm (C76, New Brunswick Scientific). All samples were taken without removing the flasks from the water bath or stopping the aeration due to shaking.

All cultures were induced in the mid-exponential phase (OD of 0.5) with 1 mM IPTG. Parallel control cultures were not induced (uninduced). Samples for the DNA microarrays were collected just prior to induction (time 0) and 5, 20, 40, and 60 minutes post-induction for the induced cultures; samples at 60 minutes were also collected for the uninduced cultures. Ethanol-treated VP1GFP cultures (induced and uninduced conditions) were run as additional controls; ethanol was added when the cultures reached 0.25 OD to a final concentration of 3%. Samples at time 0 and 60 minutes were collected for the ethanol-treated cultures (both induced and uninduced conditions). All samples for DNA microarray analysis were immediately stabilized in RNAprotect Bacteria Reagent (Qiagen, Inc.) and processed as per manual instructions. Briefly, the RNAprotect solution was removed by centrifugation (14,500 x g, 10 minutes, Hermle Labnet Z383K centrifuge), and the cell pellets were stored in −80˚C until used for RNA isolation. All culture conditions were conducted in a minimum of biological triplicates (*i.e.*, three biological replicates).

Protein Production Analysis

VP1GFP protein production was measured by the fluorescence signal for *E. coli* pTVP1GFP cultured in minimal medium and induced with 1 mM IPTG. Samples were harvested and immediately assayed with the Influx Cell Sorter flow cytometer (BD, Inc.) with a 488 nm Argon excitation laser and a 530/40 nm emission filter. Fluorescence levels from 100,000 cells were averaged to obtain the fluorescent intensity of the sample. CAT activity was quantified using the kinetic assay described by Rodriguez and Tait (Rodriguez and Tait 1983), and adapted to a 96-well plate format (Sharma et al. 2007a). Additionally, CAT biological activity was confirmed by growing induced *E. coli* pGFPCAT cells on minimal medium agar plates containing 0.61 mM chloramphenicol.

Protein Localization

Widefield fluorescent microscopy of unstained and immunolabeled cells—*E. coli* pTVP1GFP and pGFPCAT samples, as well as ethanol-treated *E. coli* pTVP1GFP samples, were harvested 60-minutes post-induction from the induced and the parallel uninduced cultures. The harvested cells were fixed in 4% paraformaldehyde overnight at 4°C. The fixed cells were centrifuged and washed twice with 20 mM glycine/phosphate buffered saline (PBS). The washed cells were incubated in 20 mM glycine/PBS for 10 minutes to block background fluorescence caused by paraformaldehyde. For images of VP1GFP, an aliquot of cells was removed, washed twice in PBS, resuspended in PBS:Glycerol (1:1), and mounted on slides. Since GFP did not exhibit fluorescence in cells expressing GFPCAT, immunofluorescence staining was performed as described previously (Welter et al. 2002), where the uninduced and induced *E. coli* pTVP1GFP were used as negative controls for the anti-CAT antibody and induced *E. coli* pPROExCAT were used as the positive controls (Sharma et al. 2007b) (data not shown). Briefly, *E. coli* cells were permeabilized using 0.05% Triton-X-100 in PBS for 20 minutes at room temperature. Permeablized cells were washed once with PBS, then incubated in blocking solution (3% bovine serum albumin (BSA)/10% goat serum/PBS) for 20 minutes at room temperature (with rotation). The blocked cells were washed once with PBS, then incubated with rabbit anti-CAT antibody (Sigma) for 1 hour at room temperature (with rotation) at a dilution of 1:250 in 1% BSA/PBS. Then, the *E. coli* cells were washed twice in 1% BSA/PBS, followed by incubation in Alexa Fluor 594 goat anti-rabbit secondary antibody (Life Technologies, Carlsbad, CA) diluted at 1:1000 in BSA/PBS for two hours at room temperature with rotation in the dark. Finally, cells were washed two times in 1% BSA/PBS, followed by one wash in PBS, resuspended in PBS:Glycerol (1:1), and mounted on slides. All traditional widefield fluorescence and differential interference contrast (DIC) microscopy images of the unstained and anti-CAT antibody probed cells were collected using a Nikon Ti Eclipse (Nikon Instruments, Melville, NY) microscope and a 60X oil immersion TIRF objective (NA = 1.49). For all images presented, a 1.5X booster lens was also utilized. Image processing was performed using Nikon NIS Elements AR Software, Version 3.2.

Super-resolution microscopy—For super-resolution images of GFP localization, *E. coli* pTVP1GFP (treated without or with ethanol) were harvested 60-minutes post-induction, and an aliquot of *E. coli* was removed and washed two times in PBS. Cells were resuspended in 100 mM β-mercaptoethylamine (Sigma) in PBS (pH 7.4), and mounted on depression slides. A Leica Widefield Super-resolution Ground State Depletion (GSD) System (Leica Microsystems, Buffalo Grove, IL), equipped with a 100X oil immersion objective (NA = 1.47) was used for image capture. For acquisition of super-resolution images, samples were pumped using 100% of 488 nm laser power until the frame rate reached 0.2, and images were acquired using 50% of the 488 nm laser power. Image processing and analysis was performed using LAS-AF Software (Leica).

RNA isolation and Characterization

Total RNA isolation was performed using RNeasy Mini or Midi Kits (Qiagen, Inc.). A Nanodrop spectrophotometer (ND 1000 from Thermo Scientific, Inc.) was used to quantify RNA. The Agilent 2100 Bioanalyzer with Expert Software (Version B.02.07.SI532) with Prokaryote Total RNA series II assay settings was used to obtain RNA integrity numbers (RIN) with the RNA 6000 Nanochip Kit. Total RNA was used to synthesize the first strand cDNA using the Superscript First-Strand Synthesis System for RT-PCR (Invitrogen, Inc) as per the Nimblegen instruction manual (Version 3.2). The RNA 6000 Nanochip Kit was also used to quantify mRNA (using the mRNA protocol) after second strand synthesis.

DNA microarrays

Custom *Escherichia coli* DNA microarrays (12 arrays per slide x 135K probes per array) with probes (45–60mer, 10 probes per target, 3 copies of each probe on array) for 4,281 *E. coli* genes and probes for mGFP, VP1, ampicillin resistance gene (Amp^r), and CAT were prepared by Roche NimbleGen. The DNA microarrays were processed at Florida State University's NimbleGen Certified Microarray Facility in Tallahassee, Florida. NimbleGen's NimbleScan software normalizes the gene expression levels with a quantile normalization method in order to reduce obscuring variation between samples. The software uses a Robust Multichip Average (RMA) algorithm to generate Calls files (_RMA.calls) that contain normalized average gene expression values. The probe sequences and raw gene expression data have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus (Accession Number: GSE47732).

The DNA microarray data was imported into ArrayStar®, a gene expression analysis software program (DNASTAR, Madison, WI) from the RMA.call files. Technical replicate gene expression levels were scaled using the "global averaging" data transformation. An ANOVA test ($p \leq 0.10$) was conducted on the gene expression values for all of the culture conditions analyzed using ArrayStar®. A total of 14 sets of biological triplicates and one set of six biological replicates (time 0 of the ethanol-treated VP1GFP cultures) were analyzed.

Gene annotations were obtained from the ASAP database of the University of Wisconsin-Madison on January 5, 2013 for *Escherichia coli* MG1655 Version m56 (Glasner et al. 2003). For genes annotated with unknown products or function (*i.e.*, genes labeled predicted, putative, or conserved), EcoCyc (Version 16.5) was used (accessed January 7, 2013) to provide additional information about the gene product or function.

Statistical Analysis

Tukey's W post-hoc testing ($p \leq 0.05$) and regression analysis ($p \leq 0.05$) were subsequently applied to identify genes with significant differences in gene expression between these conditions using Minitab 16 (State College, PA). Statistical analysis of growth rates was conducted using the generalized linear model (GLM) method ($p \leq 0.05$) with the JMP 10 software (SAS Institute Inc.). One- way ANOVA analysis was used for fluorescent intensity comparisons ($p \leq 0.05$), including the IB and smaller features size and number comparisons.

Results and Discussion

Cell Growth and Protein Production

The overall objective of this study was to characterize the dynamic gene expression variability in *E. coli* due to insoluble and soluble protein production. The pTVP1GFP and pGFPCAT plasmids and the VP1GFP and GFPCAT proteins were selected to minimize differences between the culture conditions, except for protein solubility (Salazar et al. 2013). *E. coli* MG1655 pTVP1GFP and pGFPCAT were cultured in synchronized shake flasks to produce either the mostly insoluble VP1GFP or the soluble GFPCAT proteins, respectively. One set of triplicate VP1GFP and GFPCAT cultures was induced in the mid-exponential phase with 1 mM IPTG, while uninduced cultures were run in parallel. Additionally, VP1GFP cultures treated with ethanol were examined for both induced and uninduced conditions. Samples taken for DNA microarray analyses were VP1GFP and GFPCAT induced cultures 5, 20, 40, and 60 minutes post-induction, VP1GFP and GFPCAT uninduced cultures at 0 and 60 minutes relative to the induced cultures, VP1GFP ethanol-treated uninduced cultures at 0 and 60 minutes, and VP1GFP ethanol-treated induced cultures at 60 minutes post-induction; in all, 15 conditions with at least three biological replicates were analyzed by DNA microarrays (48 total arrays). The cell density profiles for

the triplicate cultures are shown in Figure 1A. The addition of IPTG did not change the observed growth rate for any of the cultures, as shown in Figure 1A. The lines shown with the cell density measurements represent exponential growth rates of 0.55 h^{-1} and 0.45 h^{-1} for the VP1GFP/GFPCAT and the ethanol-treated VP1GFP cultures, respectively. For the ethanol-treated VP1GFP cultures, the growth rate was lower due to the ethanol addition; however, the IPTG addition did not further alter the growth rate. The observed decrease in the growth rate for the cultures after 2-hours post-induction is due to oxygen limitations as the cell numbers increase. There was no significant difference ($p \leq 0.05$) in the growth rates at any point for the uninduced and the induced paired cultures.

Flow cytometry was used to confirm and quantify VP1GFP protein production via GFP fluorescence due to IPTG-induction for both the untreated and the ethanol-treated VP1GFP cultures. As shown in Figure 1B, fluorescent intensity per cell increased linearly ($p \leq 0.05$) post-induction for both untreated and ethanol-treated VP1GFP cultures; un-averaged triplicate culture data is shown for each condition. The fluorescent intensity for the ethanol-treated induced cells was significantly higher than the untreated induced cells ($p \leq 0.05$). The fold increase in fluorescent intensity for the induced VP1GFP cells relative to the uninduced VP1GFP cells was observed to be approximately 8-fold 60-minutes post-induction (Figure 1B), while the fold increase in fluorescent intensity for the ethanol-treated induced VP1GFP cells relative to the ethanol-treated uninduced VP1GFP cells was approximately 14-fold 60-minutes post-induction. At 3.5-hours post-induction, the induced VP1GFP cells had approximately 40-fold higher fluorescent intensity than the uninduced VP1GFP cells, and ethanol-treated induced VP1GFP cells had 70-fold higher fluorescent intensity compared to the ethanol-treated uninduced VP1GFP cells. Previous researchers have suggested the soluble protein fractions increase due to ethanol additions, which might explain the higher fluorescence of the ethanol-treated induced VP1GFP cultures (Thomas and Baneyx 1997). Additionally, García-Fruitós et al. (2007) reported higher fluorescence intensity for soluble VP1GFP compared to insoluble VP1GFP. Since the addition of 3% ethanol to purified VP1GFP protein does not change the fluorescent intensity (data not shown), these data indicate that either more VP1GFP was being expressed or a higher fraction of soluble VP1GFP was being expressed in the cells treated with ethanol.

Fluorescence microscopy confirmed the localization of the VP1GFP protein in the form of IBs in the end regions of the cells (shown in green), and these regions were associated with very high signal intensities that appear after induction (Figure 2A) and agree with previous reports that IBs accumulate in the poles of the cell (Garcia-Fruitos et al. 2007b; Rokney et al. 2009). The uninduced cells did not have these bright features (data not shown). Furthermore, super-resolution microscopy (Leica) was used to resolve the large VP1GFP IBs and smaller high-intensity features for both the induced VP1GFP and ethanol-treated induced VP1GFP cultures (Figure 2C and 2D). Super-resolution microscopy identifies fluorescence events at a given location; the more events that are recorded in a location, the more white pixels appear in the image. Therefore, the white regions in the *E. coli* images represent higher concentrations of GFP as compared to the red regions in the image. As measured using the super-resolution images, the inclusion bodies for both culture conditions were 285 ± 15 nm (standard error) in diameter, where the smaller features (foci) were 65 ± 5 nm (standard error). The size and number of the IBs and features were not different between the induced and ethanol-treated induced cultures ($p > 0.05$). Interestingly, *E. coli* producing granulocyte colony-stimulating factor (G-CSF) imaged using transmission electron microscope observed 670 nm IB structures in the cell poles, but no pre-IB or smaller IB-like structures (Peternel et al. 2007). *E. coli* cells producing a mutant lambda CI repressor-GFP fusion protein were observed to have smaller fluorescent foci that moved to the cell poles; however, they were unable to resolve or size the smaller foci due to the limited microscope resolution (Rokney et al. 2009). *E. coli* producing human growth hormone (hGH) and

asparaginase as IBs were observe to have different sizes (800 nm vs 200 nm, respectively) and formation dynamics (Upadhyay et al. 2012). Thus, the exact nature of an IB entity is dependent on the recombinant protein and host interactions.

In contrast, the overall fluorescence intensity per cell was significantly different between the induced and ethanol-treated induced *E. coli* pTVP1GFP cultures ($p \leq 0.05$) as measured using traditional widefield fluorescence images taken with the Nikon system [A representative portion of the induced *E. coli* pTVP1GFP image is shown in Figure 2A, while the ethanol-treated induced *E. coli* pTVP1GFP image is not shown]. The induced *E. coli* pTVP1GFP cells had fluorescence intensities of 7750 ± 600 (standard error) relative fluorescence units, whereas the ethanol-treated induced *E. coli* pTVP1GFP cells had fluorescence intensities of 11,700 ± 630 (standard error) relative fluorescence units. The increased fluorescence intensities as measured by traditional widefield fluorescence microscopy are consistent with the flow cytometer data, which indicate a 50% increase due to the ethanol-treatment; these results are consistent with the premise that ethanol-treated *E. coli* pTVP1GFP cells produced a higher percentage of soluble recombinant protein relative to untreated *E. coli* pTVP1GFP cells. Further studies are needed to determine if the smaller bright features observed in this work with the super-resolution microscopy for the induced cells are pre-IBs that merge to form larger IBs, a high concentration of soluble protein, or smaller IBs that were previously unobserved due to resolution limitations.

GFPCAT protein production was confirmed and quantified via the CAT enzyme assay (Rodriguez and Tait 1983). Figure 1C shows the CAT enzyme assay results for both induced and uninduced GFPCAT cultures. The specific CAT activity of the uninduced cultures was approximately 34 U/mg (as expected) (Sharma et al. 2007b). The induced culture had a linear increase in specific CAT activity for the first 2 hours post-induction ($p \leq 0.05$), where the fold induction was approximately 9-fold 60-minutes post-induction. At 4-hours post-induction, the specific CAT activity of the induced GFPCAT culture was approximately 23-fold higher than that of the uninduced cultures, which is very consistent with CAT activity observed for other *trc* promoter-controlled CAT protein production systems (Cha et al. 2000; Haddadin and Harcum 2005; Harcum et al. 1992; Sharma et al. 2007a). Additionally, induced GFPCAT cultures displayed significant growth on 0.61 mM chloramphenicol minimal media plates, confirming biological activity. Further, immunostaining of the *E. coli* pGFPCAT with anti-CAT confirmed that no specific localization of the GFPCAT occurred within the cells (Figure 2B, shown in red). These CAT protein production characteristics indicate that the GFP fusion did not appear to significantly alter CAT activity. Unfortunately, the GFPCAT fusion protein does not fluoresce as much as the VP1GFP fusion, which means direct fluorescent comparison were not meaningful. The low fluorescent intensity of GFPCAT is probably due to the GFP location-the carboxyl-terminus of the CAT protein versus the amino-terminus of the VP1 protein - as these and several other factors have been reported to disrupt GFP fluorescence (Phillips 2001; Prescott et al. 1999; Zimmer 2002).

Gene expression levels for the GFP, CAT and VP1 genes on the DNA microarray were amongst the most highly expressed genes observed on the DNA microarray with intensity values in the top 0.2% of all genes analyzed. The gene expression levels for these three recombinant proteins and the ampillicin resistance gene were not significantly different between the uninduced and induced cultures; except CAT gene expression was absent from the VP1GFP cultures and VP1 gene expression was absent from the GFPCAT cultures. These gene expression data indicate that the DNA microarray signals for the plasmid-encoded genes were saturated, even for the uninduced cultures and thus could not be used for accurate quantification. However, the CAT activity fold change for GFPCAT and the GFP fluorescence fold change for the VP1GFP protein due to induction indicate that the

GFPCAT and VP1GFP protein production levels were similar, thus imposing similar amino acid utilization metabolic burdens (Bentley et al. 1991; Flores et al. 2004; Glick 1995).

Gene Expression Analysis

An ANOVA analysis ($p \leq 0.10$) of the DNA microarray data for all culture conditions and time points identified 961 genes with significant differences in expression levels between at least two conditions across the fifteen sample conditions. Tukey's W post-hoc pairwise comparisons were used to identify genes that had differential expression levels between all Time 0 and Time 60 pairings. Results of the pairwise comparisons are shown below in Table 1. Many of the pairwise comparisons were not biologically meaningful, such as the comparisons between GFPCAT uninduced cultures and ethanol-treated VP1GFP induced cultures. These types of comparisons are indicated with an "X" following the number of differentially expressed genes. Additionally, some of the pairwise comparisons had more than one condition difference (*i.e*., confounding effects). For example, the induced VP1GFP 60-minute cultures compared to the ethanol-treated uninduced VP1GFP cultures. These comparisons are indicated with a "C" following the number of differentially expressed genes. Four of these comparisons captured the differentially expressed genes that potentially could be attributed to the solubility state of the recombinant protein. These four comparisons included: 1) 0-minute to 60-minute for the induced VP1GFP cultures; 2) uninduced to induced for the VP1GFP 60-minute cultures; 3) VP1GFP to GFPCAT for the induced 60-minute cultures; and 4) ethanol-treated to untreated for the induced VP1GFP 60-minute cultures. Bold-faced numbers are used to identify these four comparisons in Table 1. Additionally, since the *E. coli* response to insoluble recombinant protein has previously been identified to elicit a partial heat-shock response (Lesley et al. 2002; Smith 2007), the number of heat-shock genes identified as differentially expressed for these four comparisons are also noted in Table 1. These four solubility-sensitive comparisons identified a total of 241 genes, which included 13 heat-shock response genes (*clpB*, *dnaJK*, *ftsH*, *gapA*, *groLS*, *hslUV*, *htpGX*, *lon*, and *rlmE*).

Since a key objective of this study was to identify the dynamic transcriptional events related to IB formation, regression analysis ($p \leq 0.05$) was used to identify genes with time-dependent behavior. The 0-, 5-, 20-, 40- and 60-minute samples were analyzed for both the VP1GFP and GFPCAT cultures, where the 0-minute samples were taken just prior to the IPTG addition. The regression analysis identified 33 genes with time-dependent behavior for the VP1GFP cultures and 92 genes with time-dependent behavior for the GFPCAT cultures. The uninduced VP1GFP and GFPCAT and the uninduced and ethanol-treated induced cultures did not have a sufficient number of time points to perform regression analyses; *i.e*., only 0-and 60-minute samples were taken. In order to identify a comprehensive set of genes that were affected by recombinant protein solubility, a union of the genes identified by Tukey's W post-hoc testing for the four solubility-sensitive pairwise comparisons and the genes identified by regression analyses for the VP1GFP and GFPCAT cultures was compiled. This union identified a total of 318 differentially expressed genes. All of the 318 differentially expressed genes are listed alphabetically by "Gene" name with fold changes for the four Tukey comparisons ($p \leq 0.05$) and with the slope direction (positive or negative) for genes with significant regression results ($p \leq 0.05$) in Table S.1 (Supporting Information).

Heat-Shock Response Genes

In wild-type *E. coli*, there are 35 classical heat-shock genes that are known to be up-regulated in response to elevated culture temperatures (*clpABPX*, *dnaJK*, *gapA*, *groLS*, *grpE*, *hflB*, *hscA*, *hslJRUV*, *htgA*, *htpGX*, *htrABCE*, *ibpAB*, *ldhA*, *lon*, *lysU*, *pspA*, *rfaD*, *rlmE*, *rpoDEH*, *yrfI*) (Gross 1996; Richmond et al. 1999). Many heat-shock genes have also

been observed to have increased gene expression levels due to soluble recombinant protein production (*htgA* and *htrCE*) (Haddadin and Harcum 2005) and misfolded recombinant protein production (*clpBP*, *dnaJK*, *groLS*, *grpE*, *hslRUV*, *htpGX*, *ibpAB*, *lon*, *rlmE*, *rpoD*, *yrfI*) (Lesley et al. 2002; Smith 2007). The ANOVA analysis ($p \leq 0.10$) identified 18 heat-shock genes as significantly regulated, and the union of the Tukey's W and regression analyses included 14 of these 18 heat-shock genes.

All 18 significantly regulated classical heat-shock genes (ANOVA $p \leq 0.10$) are listed in Table 2 with statistically significant fold changes and regression slope directions, as well as fold changes observed by Lesley *et al*. (2002) and Smith (2002) due to IB formation. The most striking difference is the response magnitudes observed compared to both of the previous studies. Specifically, the fold changes observed by Lesley *et al*. (2002) and Smith (2007) for the heat-shock genes were between 2- to 40-fold higher due to insoluble recombinant protein production; however, in the current study, fold changes were between 1.5 to 2.2 for these same genes. Several factors likely contributed to these differences: 1) In the current study, minimal medium was used to slow the growth rates to 0.55 h^{-1}, whereas LB medium was used by Lesley *et al*. (2002) and Smith (2007). Growth rates in LB are typically 2.0 h^{-1}. 2) Lesley *et al*. (2002) used 6x His-tags on all proteins, whereas in this study, only GFPCAT had a 6x His-tag. 3) Smith (2007) used *E. coli* BL21, whereas in the current study and in Lesley *et al*. (2002), *E. coli* K-12 strains were used. And, 4) Both Lesley *et al*. (2002) and Smith (2007) used strong T7-based promoters, whereas in this study, the weaker *trc* promoter was used. The significantly higher growth rates in LB medium compared to a minimal medium would allow for more cell doublings within 60 minutes, and thus, the turnover rates for mRNA species would be higher. Additionally, the T7 promoter would result in higher recombinant protein content, potentially causing more severe cellular stress. Despite these magnitude differences for the heat-shock response genes, the current study identified the same heat-shock genes as Lesley et al. (2002) and Smith (2007). Thus, the dynamic behavior obtained by the current study for heat-shock and other differentially expressed genes will have smaller magnitudes than previous work. Additionally, dynamic gene expression behavior has not been previously reported (Lesley et al. 2002; Smith 2007).

In order to better understand the time-dynamic behavior of the heat-shock genes, gene expression profiles for each of the 14 significantly affected heat-shock genes was examined for the VP1GFP and GFPCAT cultures (Figure 3). The gene expression profiles for *tig*, the trigger factor were also included. The *tig* gene encodes for one of the three major protein-folding chaperones in *E. coli*, where *dnaJK* and *groLS* encode the other two. The heat-shock genes identified by the regression analyses with positive slopes for VP1GFP are shown in the first two rows (*clpB*, *dnaK*, *gapA*, *groLS*, and *htpG*). The gene expression profiles for these six genes indicate that these genes responded early to the synthesis of VP1GFP. Responses of the other heat-shock genes (*clpP*, *dnaJ*, *ftsH*, *hslUV*, *htpX*, *lon*, and *rlmE*) were detected later, between 40 and 60 minutes post-induction. Specifically, the folding chaperone genes responded immediately to the synthesis of VP1GFP with linear increases in gene expression. In contrast, responses of the protease genes were detected later, after the insoluble VP1GFP protein had begun to accumulate within the cells. The different y-axes scales used in Figure 3 highlight that the chaperone genes, on the whole, had a much greater changes in gene expression due to VP1GFP production than the protease genes.

The DnaK-DnaJ-GrpE chaperone system, also known as the Hsp70 chaperone system, provides a substrate-binding pocket that allows for binding of misfolded protein domains. This chaperone system is ATP-controlled. Specifically, when ATP is bound, misfolded proteins can bind. When ADP is bound, the substrate-binding pocket is closed and the chaperone folds the misfolded protein. Usually, ATP hydrolysis is the rate-limiting step,

where DnaK provides the primary ATPase activity (Bukau and Horwich 1998; Keseler et al. 2013). Interestingly, the *dnaK* gene expression response appears more rapid than the *dnaJ* gene expression response and had a larger fold change for both the induced and the ethanol-treated induced VP1GFP cultures. The *grpE* gene was not identified to have statistically significant behavior. One would expect *dnaK* to respond to misfolded proteins prior to increased levels of *dnaJ* and *grpE* protein production levels due to the fundamental role DnaK plays in hydrolyzing ATP to power the DnaK-DnaJ-GrpE chaperone system.

In addition to the heat-shock genes, 30 additional genes regulated by the heat-shock transcription factor *rpoH*, or σ32, are included in the 318 differentially expressed genes (*aaeB*, *dcuR*, *entD*, *hcaR*, *kdgR*, *lexA*, *lrp*, *lspA*, *lysR*, *mlrA*, *nadR*, *nhaR*, *nor*, *phoQ*, *proB*, *rcnR*, *rhaR*, *rsmI*, *sdiA*, *yaiO*, *ybeX*, *ybeZ*, *ydeO*, *yeeJ*, *yfiR*, *yghQ*, *yjfP*, *yjfZ*, *yqeG*, and *yqjF*). The *rpoH* gene was not significantly changed due to VP1GFP protein production, which is consistent with previous observations: recombinant protein overproduction and insolubility cause similar, but not identical, responses to heat stress (Harcum and Haddadin 2006; Lesley et al. 2002; Smith 2007). Moreover, the recombinant cultures in the current study were not actually exposed to elevated temperatures, the stimulus for the *rpoH* gene.

Classical Stringent Response Genes

The classical stringent response is a stress response associated with nutritional limitations. Specifically, a stringent response is initiated by a high uncharged-tRNA to aminoacylated-tRNA ratio, caused by intracellular amino acid limitations (Cashel et al. 1996). Under a stringent response, 25 genes are up-regulated and 41 genes are down-regulated, in addition to down-regulation of the ribosome genes (Cashel et al. 1996). Under high levels of recombinant protein production, stringent-like responses have been observed (Haddadin and Harcum 2005; Rozkov et al. 2000), where Haddadin and Harcum (2005) used high IPTG-levels (5 mM) to purposely “stress” the cultures. For the VP1GFP and GFPCAT induced cultures, only 10 stringent response genes had higher levels compared to the uninduced cultures (*clpB*, *deoA*, *dnaK*, *gdhA*, *groEL*, *hisG*, *htpG*, and *thrC*S); however, all but the *thrCS* genes, represent genes also associated with the heat-shock response. None of the genes normally down-regulated during a stringent response were identified. Additionally, none of the ribosome genes were observed to be lower for the induced cultures as would be expected during a stringent response. Thus, there was no pronounced stringent response elicited by the low levels of recombinant protein being produced by these two plasmids at these IPTG-levels.

SOS Response Genes

The SOS response is elicited to repair DNA damage (Walker 1996). Ethanol is a known stressor for *E. coli* that causes numerous membrane lipid compositional changes (Dombek and Ingram 1984). During a SOS response, 29 different genes have been observed to be regulated mainly controlled by *recA* and *lexA*. Only one SOS gene (*lexA*) was identified to be different across the conditions investigated. The *lexA* gene expression was 0.9-fold lower for the VP1GFP cultures 60-min post-induction compared to the parallel control uninduced VP1GFP cultures at the same time. The fold change is consistent with a SOS response; however, no other SOS associated genes were found to have differential levels across the investigated conditions. Thus, if the SOS response was elicited at 60-min post-induction due to IB formation, the SOS response had not had time to develop and regulate other genes. The lack of an SOS response is consistent with previous observations for IPTG-induced *E. coli* cultures (Haddadin and Harcum 2005; Lesley et al. 2002; Smith 2007).

Gene Ontologies

In order to classify gene functions and coordinated behaviors in response to protein solubility state, the 318 genes with differential gene expression were grouped by common functionalities using gene ontology (GO) terms from the ASAP database and EcoCyc. Figure 4 shows the gene classifications, with the number of genes with a particular GO term indicated. Most genes had several GO term entries. In those cases, the dominant function of the encoded protein was used for classification such that no gene is represented twice in the diagram (Figure 4). For example, *lon* is listed with 13 GO terms, ranging from "response to stress" to "DNA-binding." The protein product of *lon* is the protease La, which is a well-characterized protease in *E. coli* that degrades abnormal proteins (Fischer and Glockshuber 1994.; Keseler et al. 2013). Thus, *lon* was grouped under Proteolysis. Table S.2 (Supporting Information) lists all genes grouped by classification, then alphabetically. For the 14 differentially expressed heat-shock genes, eight genes were grouped with Protein Folding (*clpB*, *dnaJK*, *groLS*, *hslUV*, and *htpG*), four genes were grouped with Proteolysis (*clpP*, *ftsH*, *htpX*, and *lon*), one gene was grouped with Energy Metabolism (*gapA*, a glyceraldehyde-3-phosphate dehydrogenase involved in glycolysis), and one gene was grouped with RNA Methylation (*rlmE*, a 23S rRNA methyltransferase).

For the 318 genes with differential gene expression and known function, the largest fraction of genes was classified within Metabolic Process (152 genes), as shown in Figure 4. The metabolic process group contained the following sub-classifications (with number of genes indicated in parentheses): Biosynthesis (38 genes), RNA Metabolism (33 genes), Catabolism (26 genes), Energy Metabolism (23 genes), Protein Metabolism (16 genes), and DNA Metabolism (5 genes); and eleven metabolic process genes did not align with these sub-classifications (*aphA*, *deoA*, *frc*, *glpK*, *guaD*, *hyuA*, *pntA*, *pps*, *sufS*, *yicI*, and *yihQ*). The next largest first-level classification was putative genes (111 genes), representing 35% of the differentially expressed genes. Since only approximately 14% of the *E. coli* genome is considered to be of unknown function (Riley et al. 2006), this high proportion of putative genes indicates that *E. coli* responds to insoluble recombinant proteins using many genes that have not been well-studied. The remaining first-level classifications were Transmembrane Transport (33 genes) and Transcription Regulation (17 genes). Of the 318 differentially expressed genes, five genes did not align with these classifications (*asr*, *fimFG*, *tsr*, and *yeeJ*), and were mainly cell wall components. These gene classifications were subsequently used to identify coordinated behavior within cellular functionalities due to the protein solubility state (*i.e.*, sensitive to IB formation).

Protein Metabolism Genes

Protein metabolism genes encode for chaperones that assist with protein synthesis and proteases that degrade misfolded or unneeded proteins. Within the protein metabolism group, nine protein folding and seven proteolysis genes were identified. The nine protein folding genes include eight classical heat-shock genes (*clpB*, *dnaJK*, *groLS*, *hslUV*, and *htpG*), plus *tig*. All of the protein-folding genes increased in gene expression due to VP1GFP production, including *tig*. The time profiles for *tig* are shown in Figure 3 with the classical heat-shock genes; although *tig* is not a classical heat-shock protein, it is a chaperone. The observed behavior of the *tig* gene, however, was more similar to the heat-shock protease genes (*clpP*, *ftsH*, *htpX*, and *lon*) as opposed to the heat-shock folding chaperone genes (*i.e.*, *tig* gene expression did not increase immediately). The seven proteolysis genes identified included four classical heat-shock genes (*clpP*, *ftsH*, *htpX*, and *lon*), plus *ompX* and *pepBN*. The *ompX* and *pepBN* gene expression profiles are very similar to the heat-shock protease genes.

Protein Synthesis-Related Gene Classifications—Many of the identified differentially expressed genes encode for proteins involved in protein synthesis, including genes involved with ribosomal subunits (*rplACEJLNQR*, *rpmDIJ*, and *rpsAFGHLMNPU*), aminoacyl-tRNA synthetases (*asnS*, *leuS*, *pheT*, *thrS*, and *tyrS*), and amino acid synthesis (*aroAG*, *dadX*, *gdhA*, *hisGH*, *lrp*, *metH*, *proB*, *thrC*, and *trpCD*). The most coordinated group of genes was the ribosomal subunit genes. Twenty ribosomal subunit genes with differential gene expression were identified, including both 30S ribosomal subunit genes (*rpsAFGHLMNPU*) and 50S ribosomal subunit genes (*rplACEJLNQR* and *rpmDIJ*). The dynamic profiles for these 20 genes were very similar to each other. To highlight this behavioral similarity, the gene expression profiles for all twenty ribosomal subunit genes are shown in the Supporting Information (Figure S.1). Due to the highly similar profiles, only the average ribosomal subunit gene profiles are shown in Figure 5A. The remaining 35 ribosomal subunit genes that did not meet statistical significance; however, were also observed to have similar gene expression profiles to the 20 differentially expressed ribosomal subunit genes. The gene expression profiles for the 35 statistically insignificant ribosomal subunit genes are shown in the Supporting Information (FigureS.2). The ribosomal subunit gene expression levels increased due to recombinant protein production: soluble GFPCAT protein production resulted in a linear increase in gene expression; VP1GFP protein production initially resulted in a similar linear increase, but between 40 and 60 minutes post-induction, a dramatic increase was observed; and VP1GFP protein production in ethanol-treated cultures was not significant. Interestingly, the ribosomal subunit gene expression profiles for the uninduced VP1GFP, the uninduced GFPCAT, and the ethanol-treated induced VP1GFP cultures were similar, showing no significant increase in gene expression levels, indicating that the ethanol addition dampened the ribosomal subunit response to VP1GFP protein production. Past studies have indicated that the amounts of ribosomal proteins and other protein synthesis machinery components can decrease in high-productivity recombinant protein systems, presumably due to aminoacyl-tRNA deprivation (Dong et al. 1995; Gallant 1979; Rinas et al. 2007). The seemingly contradictory results of this study may indicate that cells grown in minimal medium and analyzed predominantly in the early-phases of recombinant protein production (≤ 1 hour post-induction) had not yet encountered a significant shortage of amino acids or charged tRNAs. Interestingly, the ethanol treatment decreased the ribosomal subunit response, yet resulted in higher VP1GFP protein production levels, and decreased cell growth rates. The protein production and growth rate characteristics are consistent with past studies (Thomas and Baneyx 1997). Thus, it appears that the ethanol altering response may be a significant factor in the increased protein production levels.

Ribosomal Subunit Genes—The ribosomal subunit gene expression profiles show a clear decrease in expression between the uninduced time 0-minute and the 5-minute post-induction samples for the VP1GFP cultures. However, the seemingly high ribosomal subunit expression for the uninduced time 0-minute VP1GFP cultures can be attributed to one biological sample (replicate B). An ANOVA test with Bonferroni multiple testing corrections were used on the three biological replicates of VP1GFP; however, the number of differentially expressed genes was not greater than the false positive rate. There was not enough statistical evidence to warrant exclusion of any VP1GFP time 0-minute replicates, and replicate B was utilized for all data analysis; however, caution was exercised in attributing global behavior characteristics based on this time point.

Amino Acid Synthesis Genes—Amino acid synthesis genes (*aroAG*, *dadX*, *gdhA*, *hisGH*, *lrp*, *metH*, *proB*, *thrC*, and *trpCD*) encode for enzymes involved in the pathways used to generate amino acids. The gene expression profiles for these amino acid synthesis genes (averaged gene expression) are shown in Figure 5B. The amino acid synthesis gene

expression levels increased due to recombinant protein production for both the soluble and insoluble proteins, but were increased to a greater extent by VP1GFP protein production, yet this response was not observed due to VP1GFP protein production for the ethanol-treated cultures.

The increased expression levels of certain amino acid synthesis genes could potentially be attributed to the amino acid compositions of VP1GFP and GFPCAT. Relative to an average *E. coli* protein, VP1GFP and GFPCAT both contain higher percentages of phenylalanine. The phenylalanine content is 3% of amino acid residues for the average *E. coli* protein vs. 4% and 7% of total amino acid residues for VP1GFP and GFPCAT, respectively. Other amino acids with significantly different content include: Histidine (1% vs. 4% and 5%) and threonine (5% vs. 11% and 7%). In contrast, VP1GFP contains a lower percentage of tryptophan (1% vs. 0.4%), while GFPCAT contains a slightly higher percentage of tryptophan (1.0% vs. 1.3%) (Neidhardt and Umbarger 1996). Four amino acid synthesis genes related to phenylalanine, histidine, and threonine were observed to have increased gene expression levels in both the induced VP1GFP and GFPCAT cultures. These four genes are substrate-controlled (*aroG*, *hisGH*, and *thrC*): the *aroG* gene is inhibited by the TyrR-phenylalanine DNA-binding transcriptional repressor, the *hisGH* genes are inhibited by charged L-histidyl-tRNA species, and the *thrC* gene is inhibited by charged L-threonyl-tRNA species. High production levels of VP1GFP and GFPCAT likely caused phenylalanaine, threonine, and histidine levels to decrease, which then indirectly increased the gene expression levels of *aroG*, *hisGH*, and *thrC* genes by decreasing respective inhibitors.

Aminoacyl-tRNA Synthetase Genes—Aminoacyl-tRNA synthetase genes encode for the enzymes responsible for covalently binding amino acids to tRNA molecules for protein synthesis (*i.e.*, aminoacyl-tRNA charging). For the aminoacyl-tRNA synthetase genes identified in this study as differentially expressed (*asnS*, *leuS*, *pheT*, *thrS*, and *tyrS*), the most pronounced increase in expression was observed for the induced VP1GFP 60-minute cultures post-induction. Gene expression profiles for aminoacyl-tRNA synthetase genes (averaged gene expression) are shown in Figure 5C. Lesley *et al*. (2002) and Smith (2007) did not report coordinated changes for amino acid synthesis or aminoacyl-tRNA synthetase genes in response to insoluble recombinant protein, most likely because these previous studies cultured *E. coli* in LB medium, which provides high levels extracellular of amino acids (Lesley et al. 2002; Moriya et al. 2007; Smith 2007). In contrast, growth in minimal medium requires the cells to synthesize all amino acids from glucose and ammonium. In this study, the observed increase in gene expression levels for the ribosomal subunit, amino acid synthesis, and aminoacyl-tRNA synthetase genes indicates the cells were adapting to increased protein synthesis needs due to induction for both VP1GFP and GFPCAT proteins. Interestingly, the ethanol-treated induced VP1GFP cultures did not have significantly increase ribosomal subunit, amino acid synthesis, and aminoacyl-tRNA synthetase gene expression levels, yet produced more VP1GFP protein. The ribosomal subunit, amino acid synthesis, and aminoacyl-tRNA synthetase gene responses indicate that these responses are solubility-related, beyond the protein production and amino acid imbalance response, and these responses are diminished by ethanol.

Transmembrane Transport Genes

Transmembrane transport genes encode for proteins that control cellular import and export. In the current study, 33 of the transmembrane transport genes were identified as having differential expression. Within these 33 transmembrane transport genes, 13 carbohydrate transport genes (*actP*, *bglF*, *dgoT*, *fucP*, *glpF*, *gntT*, *lamB*, *malF*, *opgH*, *ugpAB*, *uhpT*, and *ulaA*) and eight amino acid transport genes (*dppBC*, *dtpB*, *eamB*, *gltS*, *gspDL*, and *tdcC*)

were identified, as well as 12 transmembrane transport genes that were not further sub-classified (*aaeB*, *acrBF*, *blc*, *btuE*, *cysP*, *lspA*, *modB*, *ompF*, *pstS*, *purP*, and *tsgA*). Twenty-three genes had significantly decreased expression due to VP1GFP production; eleven carbohydrate transport genes (*actP*, *bglF*, *dgoT*, *fucP*, *glpF*, *gntT*, *lamB*, *malF*, *ugpB*, *uhpT*, and *ulaA*), six amino acid transport genes (*dtpB*, *eamB*, *gltS*, *gspDL*, and *tdcC*), and six unclassified transport genes (*aaeB*, *acrF*, *blc*, *modB*, *purP*, and *tsgA*). These effects were not observed for the uninduced and ethanol-treated induced VP1GFP and the uninduced and induced GFPCAT cultures. The gene expression profiles for these 23 transmembrane transport genes (averaged gene expression) are shown in Figure 6A. These gene expression profiles also highlight the difficulties of observing and detecting genes with statically significant decreased gene expression levels. Specifically, if there is no mechanism to degrade a particular mRNA sequence, then the cell must grow to dilute the mRNA. The observed doubling time for these cultures in minimal media was 1.26 hours, which means after 60-minutes, the cells have not yet doubled following the induction. Thus, if the expression of a particular mRNA was completely stopped due to induction and not targeted for degradation, it would only decrease approximately 0.58-fold in 60-minutes due to growth dilution. The transmembrane genes were observed to decrease 0.88-fold due to induction of VP1GFP, indicating that gene expression did not stop for the transmembrane transport genes, but was slowed considerably.

Lesley *et al*. (2002) and Smith (2007) noted the decreased expression of several membrane transporters due to the presence of insoluble recombinant protein, including various sugar and metal ion transporters (*chaA*, *fecB*, *feoA*, *fruA*, *glpF*, *lamB*, *rbsC*, *setA*, and *ychM*) (Lesley et al. 2002; Smith 2007). This study also identified decreased gene expression of several transporters previously not associated with IBs (*aaeB*, *acrFP*, *bglF*, *blc*, *dgoT*, *dtpB*, *eamB*, *fucP*, *gltS*, *gntT*, *gspDL*, *malF*, *modB*, *purP*, *tdcC*, *tsgA*, *ugpB*, *uhpT*, and *ulaA*). Interestingly, for the transmembrane transporters that had decreased expression due to the VP1GFP protein production, 13 of these genes are directly regulated by the cyclic AMP receptor protein, Crp; including 10 of the 11 carbohydrate transport genes (*aaeB*, *actP*, *bglF*, *fucP*, *glpF*, *gntT*, *lamB*, *malF*, *modB*, *tdcC*, *ugpB*, *uhpT*, and *ulaA*). The Crp protein is known to regulate catabolite repression (Görke and Stülke 2008; Gosset et al. 2004). The Crp-controlled genes that had decreased gene expression facilitate entry of specific substrates, such as *actP* (glycolate), *bglF (*methyl-β-D-glucoside-6-phosphate and arbutin-6-phosphate), *dgoT* (galactonate), *fucP* (fucose), *glpF* (glycerol), *gntT* (gluconate), *lamB/malF* (β-maltose), and *uhpT* (hexose phosphate) (Van der Rest et al. 2000); however, in this study, none of these carbohydrates were provided in the medium, as glucose was the sole carbon source. Thus, the decreased gene expression of these Crp-controlled transporters may indicate that VP1GFP protein production shifts metabolic efforts towards glucose by decreasing expression of unnecessary transporter proteins.

Catabolism Genes

Catabolism reactions in cells control the breakdown and recycling of cellular building blocks. The catabolism classification mainly identified carbohydrate and amino acid degradation genes with differential expression. For carbohydrate degradation, 12 of the 13 genes identified decreased due to VP1GFP protein production *(bglB, dgoDK, glcD, gudD, maoC, mtlD, rhaBM, treF, uxuB*, and *yiaS*). Additionally, three amino acid degradation genes (*dtd*, *tdh*, and *tnaA*) and four unclassified catabolism genes (*caiB*, *chiA*, *cpdB*, and *nudE*) decreased due to VP1GFP protein production. The gene expression profiles for these 19 genes (averaged gene expression) are shown in Figure 6B. These catabolism genes had gene expression profiles similar to the transmembrane transport genes. In contrast, these responses were not observed in the ethanol-treated VP1GFP or GFPCAT cultures.

For the 19 catabolism pathway genes that had decreased gene expression due to VP1GFP protein production, nine genes are directly regulated by Crp (*bglB*, *caiB*, *cpdB*, *maoC*, *mtlD*, *rhaB*, *tnaA*, *uxuB*, and *yiaS*). Taken in conjunction with the coordinated decrease in gene expression for Crp-controlled transmembrane transport genes, these results indicate that catabolite repression may have occurred due to VP1GFP protein production. Interestingly, several of the transmembrane transporter genes with decreased gene expression act on the same target substrates as many of the catabolic genes with decreased gene expression due to VP1GFP protein production. For example, the *dgoT* gene encodes the D-galactonate transporter and the *dgoDK* genes encode enzymes that degrade D-galactonate (Karp et al. 2010). Table 3 lists these catabolic and transmembrane transporter genes with common target substrates. These catabolic and transmembrane transporter genes were unaffected by GFPCAT protein production and unaffected by VP1GFP protein production in the ethanol-treated cultures. Lesley (2002) and Smith (2007) did not report coordinated gene expression for the catabolism genes and transmembrane transporter genes (Lesley et al. 2002; Smith 2007). These results indicate that cells respond to the onset of insoluble recombinant protein production through a coordinated down-regulation of transporter and catabolic genes, where ethanol disrupts this response, possibly by interfering with Crp-functions.

Cofactor Synthesis Genes

Cofactors are compounds required by a protein to function properly and often modulate enzyme activity. Within the cofactor synthesis gene classification, nine genes (*dxs*, *folC*, *hemL*, *ispEFG*, and *thiCMS*) had increased gene expression levels due to VP1GFP protein production, whereas these effects were not observed for the uninduced or induced GFPCAT and the uninduced or ethanol-treated induced VP1GFP cultures. The gene expression profiles for these cofactor synthesis genes (averaged gene expression) are shown in the Supporting Information (FigureS.3). The *dxs* and *ispEFG* genes encode enzymes in the methylerythritol phosphate pathway, which primarily synthesizes membrane glycolipids, peptidoglycans, and quinols. The *thiCMS* genes encode enzymes involved in thiamine diphosphate synthesis, an essential cofactor for both pyruvate dehydrogenase and transketolase.; pyruvate dehydrogenase decarboxylates pyruvate to acetyl-CoA, and transketolase catabolizes sugars in the pentose phosphate pathway, such as ribose-5-phosphate and erythrose-4-phosphate (Keseler et al. 2013). Interestingly, the genes encoding for pyruvate dehydrogenase (*aceEF* and *lpd*) and transketolase (*tktA*) were not identified as statistically significant (See FigureS.4 in the Supporting Information), but were observed to have gene expression profiles similar the differentially expressed cofactor synthesis genes. Taken in conjunction with the changes in the transmembrane transporter gene expression levels, the increased gene expression of the *dxs* and *ispEFG* genes may indicate that VP1GFP protein production induces a physical change in the cell membrane, which required repair to the cell membrane. The increased gene expression of the *thiCMS* genes may indicate that cells needed to increase energy production due to VP1GFP protein production (Keseler et al. 2013).

Energy Metabolism Genes

Energy metabolism genes encode for enzymes that catalyze ATP, NADH, or quinols synthesis. In this study, 23 energy metabolism genes were identified with significant differential expression across the culture conditions. Six TCA cycle genes had increased expression levels due to both VP1GFP and GFPCAT protein production (*acnB*, *gltA*, *sdhAB*, and *sucBC*), where the observed increase was higher due to VP1GFP protein production. The ethanol treatment significantly reduced the gene expression responses due to VP1GFP protein production. The gene expression profiles for these six TCA cycle genes (averaged gene expression) are shown in Figure 7. Additionally, nine of the 12 remaining TCA cycle genes (*acnA*, *fumA*, *icd*, *lpd*, *mdh*, *sdhCD*, and *sucAD*) did not meet statistical significance,

yet had gene expression profiles similar to the six identified TCA cycle genes (See Figure S. 5 in Supporting Information). The coordinated increase in gene expression levels observed for these TCA cycle genes supports previous observations that recombinant protein synthesis is an energy intensive process (Glick 1995; Tseng 1997; Tseng et al. 2001) and that insoluble VP1GFP protein production results in a greater need for energy, which is disrupted by the addition of ethanol.

ATP Synthase Genes—In addition to the TCA cycle genes, two ATP synthase genes (*atpFI*) were observed to have differential gene expression due to GFPCAT and VP1GFP protein production. The *atpF* gene expression profile increased due to both GFPCAT and VP1GFP protein production similar to that observed for the ribosomal subunit genes. The *atpF* profiles are included in FigureS.6 in the Supporting Information. The *atpF* gene encodes for the subunit b of the ATP synthase F_0 sector, which is critical to proton translocation. There are seven additional genes that encode for subunits of the F_1F_0 ATP synthase complex (*atpABCDEGH*). The gene expression profiles of these seven genes were similar to *atpF*; however, these seven genes did not meet statistical significance (shown in FigureS.6 in Supporting Information). Interestingly, the *atpI* gene is considered to be unessential for the function of the ATP synthase complex (Keseler et al. 2013) and was observed to decrease due to GFPCAT and VP1GFP protein production. The observed increases of the ATP synthase genes are consistent with increased energy needs associated with recombinant protein production.

Other Energy Metabolism Genes—Of the remaining energy metabolism genes, eight genes involved in the electron transport chain had increased gene expression levels due to VP1GFP protein production (*appC*, *dld*, *hyaDE*, *napD*, *narW*, and *nuoCG*). These eight genes encode for components of the NADH oxidoreductase, cytochrome oxidase, nitrate reductase, hydrogenase 1, and lactate dehydrogenase complexes (Keseler et al. 2013). The gene expression profiles for these eight genes are shown in the Supporting Information (FigureS.7). The electron transport chain is a key energy generating pathway, where electron transfer is coupled with transmembrane proton translocation and results in a proton gradient to produce ATP via ATP synthase (Weber and Senior 2003). Lesley (2002) and Smith (2007) did not observe changes in gene expression for electron transport genes. Most likely, the multiple carbon sources present in LB medium mitigated this response (Lesley et al. 2002; Smith 2007). The observed increased gene expression response for electron transport genes due to VP1GFP protein production are consistent with the higher metabolic burden associated with recombinant protein production (Bentley et al. 1991; Glick 1995) and further, with the energy-intensive processes needed to refold or degrade misfolded proteins (Bukau and Horwich 1998).

Transcription Regulation Genes

Transcription regulation genes control transcription rates by encoding proteins that directly bind to promoter regions. In this study, 17 transcription regulation genes were identified with significantly different gene expression. Nine transcription regulation genes were observed to have decreased gene expression levels in response to VP1GFP protein production in the induced VP1GFP cultures (*atoS*, *dcuR*, *hcaR*, *lexA*, *lysR*, *nadR*, *rcnR*, *rfaH*, and *rhaR*), and one gene (*sdiA*) was observed to have decreased gene expression levels in the induced VP1GFP cultures relative to the ethanol-treated induced VP1GFP cultures. The gene expression profiles for these ten genes are shown in the Supporting Information (FigureS.8). The gene expression profiles for these transcription regulation genes are similar to the transmembrane transport genes (Figure 6A). The *atoS*, *hcaR*, *nadR*, and *rhaR* genes all control catabolic pathways, including short-chain fatty acids (*atoS*), hydrocinnamic acid (*hcaR*), NADP (*nadR*), and rhamnose (*rhaR*) degradation. The decreased gene expression

for these regulator genes collaborates with the observed decrease in expression levels for the catabolic pathways. The *lexA* gene is known to repress DNA repair as a part of the SOS response. Decreased *lexA* levels would result in higher DNA repair (Fernández de Henestrosa et al. 2000), indicating that VP1GFP protein production increased DNA damage; however, was disrupted by ethanol. The *sdiA* gene controls genes that promote cell division. Decreased *sdiA* gene expression levels would inhibit cell division (García-Lara et al. 1996; Keseler et al. 2013), indicating that VP1GFP protein production inhibits cell division, yet in the presence of ethanol with reduced growth rates, this response was mitigated.

Three transcription regulation genes (*kdgR*, *nusA*, and *phoQ*) had increased gene expression levels due to VP1GFP protein production, but were unaffected in the GFPCAT and ethanol-treated VP1GFP cultures as shown in the Supporting Information (FigureS.9). The *nusA* gene binds RNA polymerase to increase transcription efficiency for 5S, 16S, and 23S rRNA (Keseler et al. 2013; Quan et al. 2005), indicating that VP1GFP protein production stimulates higher transcription rates. The *phoQ* gene normally increases due to magnesium starvation (Kato et al. 1999), indicating that VP1GFP protein production increases cell magnesium requirements. It is known that there are at least two magnesium-dependent proteases in *E. coli*, ClpP (Gross 1996) and an unnamed protease (Jordan and Harcum 2002). Both of these proteases are known to have increased activity due to insoluble recombinant protein production. The *clpP* gene was observed to have increased gene expression levels due to VP1GFP protein production and has been observed to have increased gene expression levels due to insoluble recombinant protein production by both Lesley et al. (2002) and Smith (2007) (Lesley et al. 2002; Smith 2007). The unnamed magnesium-dependent protease was isolated and characterized from cells expressing high levels of an inclusion-body prone protein (Jordan and Harcum 2002). Thus, IB formation likely significantly increased the magnesium requirements of the cells; however, ethanol disrupted this response.

Effect of Ethanol on IB Stress Response

The effects of ethanol were characterized in this study, since it is well-known that ethanol can improve recombinant protein solubility (Thomas and Baneyx 1997); however, the mechanism is unknown. Gene expression level comparisons between the induced and the ethanol-treated induced VP1GFP cultures only identified nine genes with significant difference; however, comparisons between induced VP1GFP and GFPCAT cultures identified 126 genes. By examining the gene expression profiles for genes identified to be sensitive to VP1GFP protein production without ethanol, it was clear that ethanol affected gene expression response dynamics and mitigates or disrupted many of the responses associated with insoluble protein production. These results suggest that the ethanol-treatment relieves IB stress by mitigating the gene expression responses that are similar to the heat-shock response (Figure 3), and includes the genes for the ribosomal subunits (Figure 5A), aminoacyl-tRNA synthetase (Figure 5B), transmembrane transporter (Figure 6A), and catabolism (Figure 6B). Due to the lack of statistical differences observed for the gene expression levels between induced VP1GFP and ethanol-treated induced VP1GFP cultures, further research is needed to clarify the role of the ethanol treatment on the stress responses caused by IB formation.

Conclusion

The tendency for recombinant proteins to misfold in *E. coli* and form IBs represents a major roadblock for some large-scale production processes. This study characterized the dynamic transcriptional behavior of *E. coli* in the early stages of insoluble recombinant protein production to capture these early responses. As expected, the classical heat-shock response

genes were stimulated due to IB formation. Additionally, several protein-folding and protease genes not associated with the classical heat-shock response had increased gene expression levels due to IB formation. The increased levels for genes involved in protein-folding and in proteolysis indicates that the cells attempt to alleviate this stress by increasing synthesis of chaperones to assist with protein folding and by increasing synthesis of proteases to remove misfolded proteins.

It was also observed that components of the cellular protein synthesis machinery had increased gene expression due to recombinant protein production, but were more severely affected by IB formation, including the ribosomal subunit, amino acid synthetic pathways, and aminoacyl-tRNA synthetase genes. Thus, the cells respond to the early onset of IB formation by increasing cellular protein synthesis machinery. In contrast, several substrate-specific transmembrane transport and catabolism genes had decreased gene expression levels due to IB formation. Moreover, the Crp protein, that regulates most of the affected transmembrane transport and catabolism genes, was sensitive to ethanol. Confounding these observations with regards to transmembrane transporter and catabolism is the fact that these substrates were not present in the growth medium. Thus, the decreased gene expression of these genes partially served to alleviate the metabolic burden associated with unnecessary cell membrane components and enzymes. Furthermore, the response of the energy metabolism and electron transport chain genes suggests that the cells required more ATP as well as essential cofactors.

Prior to this study, it was unknown the dynamic relationship between the formation of the IBs and the changes in gene expression (mainly the known heat-shock and protease up-regulation), as all past studies only quantified gene expression changes after the IBs were well formed. These results indicate that most of the gene expression changes occur after the IBs are formed, which means the cell sense the presence of the IBs. In contrast, several of the heat-shock response genes reacted more quickly (prior to IB formation), indicating that these genes were reacting to the high rate of protein synthesis more so than the IB formation. The linear regression analysis for the soluble protein also demonstrated that some of these genes respond to increased protein synthesis rates. Even more surprising was the lack of gene expression responses to IB formation in the cells treated with ethanol, yet the amount of soluble recombinant protein appears to be slightly higher, while the IB size and number are identical. These results also demonstrated that are a high number of gene expression responses that have unknown function (putative genes). To truly understand how to mitigate the protein production bottlenecks associated with IB formation, these putative genes must be given functions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The plasmid pTVP1GFP was donated by E. García-Fruitós and A. Villaverde at the Universidad Autónoma de Barcelona. The plasmid pTrcHis-GFP$_{UV}$/CAT plasmid was donated by W. E. Bentley at the University of Maryland, College Park. The plasmid pGFPCAT was constructed by Dr. M. T. Morris at Clemson University. The authors would like to thank Dr. N. Vyavahare for the use of the Agilent Bioanalyzer 2100, Dr. M. Sehorn for his contributions to the development of the discussion section, Mr. Y. K. Gowtham for his technical assistance, Mr. A. N. Brodsky for proof-reading, and Dr. J. B. Harcum for his review of the statistical analysis. This research was supported by funding provided by the National Science Foundation (CBET 0738162) and an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health (P20GM103444). Image capture was performed in the Clemson Light Imaging Facility (Clemson University), which is supported by a grant from the Vice-President's Office for Research at Clemson. The GSD System was on loan to the Clemson Light Imaging Facility as a demonstration system, courtesy of Leica Microsystems. We wish to thank Leica Microsystems for the use of the GSD, along with Dr. Robyn Schlicher and

Mr. Tommy Dees of Leica Microsystems for training and assistance required for sample preparation and image capture. The authors have no conflict of interest to declare.

References

Agostini F, Vendruscolo M, Tartaglia GG. Sequence-Based Prediction of Protein Solubility. Journal of Molecular Biology. 2012; 421(2–3):237–241. [PubMed: 22172487]

Allen SP, Polazzi JO, Gierse JK, Easton AM. Two Novel Heat-Shock Genes Encoding Proteins Produced in Response to Heterologous Protein Expression in *Escherichia coli*. Journal of Bacteriology. 1992; 174(21):6938–6947. [PubMed: 1356969]

Baneyx F. Recombinant protein expression in *Escherichia coli*. Current Opinion in Biotechnology. 1999; 10(5):411–421. [PubMed: 10508629]

Baneyx F, Mujacic M. Recombinant protein folding and misfolding in *Escherichia coli*. Nature Biotechnology. 2004; 22(11):1399–1408.

Basu A, Li X, Leong SSJ. Refolding of proteins from inclusion bodies: rational design and recipes. Applied Microbiology and Biotechnology. 2011; 92(2):241–251. [PubMed: 21822901]

Bentley WE, Davis RH, Kompala DS. Dynamics of induced CAT expression in *E. coli*. Biotechnology and Bioengineering. 1991; 38(7):749–760. [PubMed: 18600801]

Bukau B, Horwich AL. The Hsp70 and Hsp60 chaperone machines. Cell. 1998; 92(3):351–366. [PubMed: 9476895]

Carrio M, Villaverde A. Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. Journal of Bacteriology. 2005; 187(10):3599–3601. [PubMed: 15866952]

Carrio MM, Villaverde A. Construction and deconstruction of bacterial inclusion bodies. 2002; 96(1): 3–12.

Cashel, M.; Gentry, GR.; Hernandez, VJ.; Vinella, D. The Stringent Response. In: Neidhardt, FC.; Curtiss, R.; Lin, ECC.; Low, KB.; Magasanik, B.; Reanikoff, WS.; Riley, M.; Schaechter, M.; Umbarger, HE., editors. Escherichia coli and Salmonella. Washington D. C: ASM Press; 1996. p. 1458-1496.

Cha HJ, F WC, J VJ, G R, Bentley WE. Observations of green fluorescent protein as a fusion partner in genetically engineered *Escherichia coli:* monitoring protein expression and solubility. Biotechnology and Bioengineering. 2000; 67(5):565–574. [PubMed: 10649231]

Chen WJ, Hong M, Li D, Lu SD. High-level expression of foreign genes via multiple joined operons and a new concept regarding the restricted constant of total amount of plasmid DNA per *Escherichia coli* cell. Chinese Medical Journal. 2002; 115(12):1785–1789. [PubMed: 12622924]

Cheung KJ, Badarinarayana V, Selinger DW, Janse D, Church GM. A microarray-based antibiotic screen identifies a regulatory role for supercoiling in the osmotic stress response of *Escherichia coli*. Genome Research. 2003; 13(2):206–215. [PubMed: 12566398]

Choi JH, Lee SJ, Lee SY. Enhanced production of insulin-like growth factor I fusion protein in *Escherichia coli* by coexpression of the down-regulated genes identified by transcriptome profiling. Applied and Environmental Microbiology. 2003; 69(8):4737–4742. [PubMed: 12902266]

Conway T, Schoolnik GK. Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. Molecular Microbiology. 2003; 47(4):879–889. [PubMed: 12581346]

Diaz AA, Tomba E, Lennarson R, Richard R, Bagajewicz MJ, Harrison RG. Prediction of Protein Solubility in Escherichia coli Using Logistic Regression. Biotechnology and Bioengineering. 2010; 105(2):374–383. [PubMed: 19739095]

Dombek KM, Ingram LO. EFFECTS OF ETHANOL ON THE ESCHERICHIA-COLI PLASMA-MEMBRANE. Journal of Bacteriology. 1984; 157(1):233–239. [PubMed: 6360997]

Dong H, Nilsson L, Kurland CG. Gratuitous Overexpression of Genes in *Escherichia coli* Leads to Growth Inhibition and Ribosome Destruction. Journal of Bacteriology. 1995; 177(6):1497–1504. [PubMed: 7883706]

Duerrschmid K, Reischer H, Schmidt-Heck W, Hrebicek T, Guthke R, Rizzi A, Bayer K. Monitoring of transcriptome and proteome profiles to investigate the cellular response of E. coli towards

recombinant protein expression under defined chemostat conditions. Journal of Biotechnology. 2008; 135(1):34–44. [PubMed: 18405993]

Fernández de Henestrosa AR, Ogi T, Aoyagi S, Chafin D, Hayes JJ, Ohmori H, Woodgate R. Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. Molecular Microbiology. 2000; 35(6):1560–1572. [PubMed: 10760155]

Fischer H, Glockshuber R. A point mutation within the ATP-binding site inactivates both catalytic functions of the ATP-dependent protease La (Lon) from *Escherichia coli*. FEBS Letters. 1994; 356:101–103. [PubMed: 7988699]

Flores S, de Anda-Herrera R, Gosset G, Bolivar FG. Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway. Biotechnology and Bioengineering. 2004; 87(4):485–494. [PubMed: 15286986]

Gallant JA. Stringent Control in *E. coli*. Annual Review of Genetics. 1979; 13:393–415.

Garcia-Fruitos E, Aris A, Villaverde A. Localization of functional polypeptides in bacterial inclusion bodies. Applied and Environmental Microbiology. 2007a; 73(1):289–294. [PubMed: 17085715]

Garcia-Fruitos E, Carrio MM, Aris A, Villaverde A. Biological and conformational state of soluble recombinant proteins. Febs Journal. 2005; 272:353–353. [PubMed: 15654874]

Garcia-Fruitos E, Martinez-Alonso M, Gonzalez-Montalban N, Valli M, Mattanovich D, Villaverde A. Divergent genetic control of protein solubility and conformational quality in *Escherichia coli*. Journal of Molecular Biology. 2007b; 374(1):195–205. [PubMed: 17920630]

Garcia-Fruitos E, Vazquez E, Diez-Gil C, Corchero JL, Seras-Franzoso J, Ratera I, Veciana J, Villaverde A. Bacterial inclusion bodies: making gold from waste. Trends in Biotechnology. 2012; 30(2):65–70. [PubMed: 22037492]

Garcia-Fruitos E, Villaverde A. Friendly production of bacterial inclusion bodies. Korean Journal of Chemical Engineering. 2010; 27(2):385–389.

García-Lara J, Shang LH, Rothfield LI. An Extracellular Factor Regulates Expression of *sdiA*, a Transcriptional Activator of Cell Division Genes in *Escherichia coli*. Journal of Bacteriology. 1996; 178(10):2742–2748. [PubMed: 8631660]

Gatti-Lafranconi P, Natalello A, Ami D, Doglia SM, Lotti M. Concepts and tools to exploit the potential of bacterial inclusion bodies in protein science and biotechnology. Febs Journal. 2011; 278(14):2408–2418. [PubMed: 21569207]

Gill RT, Delisa MP, Valdes JJ, Bentley WE. Genomic analysis of high cell density recombinant *Escherichia coli* fermentation and “cell conditioning” for improved recombinant protein yield. Biotechnology and Bioengineering. 2001; 72(1):85–95. [PubMed: 11084598]

Glasner JD, Liss P, Plunkett G, Darling A, Prasad T, Rusch M, Byrnes A, Gilson M, Biehl B, Blattner FR. others. ASAP, a systematic annotation package for community analysis of genomes. Nucleic Acids Research. 2003; 31(1):147–151. [PubMed: 12519969]

Glick BR. Metabolic Load and Heterologous Gene-Expression. Biotechnology Advances. 1995; 13(2): 247–261. [PubMed: 14537822]

Görke B, Stülke J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nature Reviews Microbiology. 2008; 6(8):613–624.

Gosset G, Zhang Z, Nayyar S, Cuevas WA, Saier MH. Transcriptome Analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*. Journal of Bacteriology. Journal of Bacteriology. 2004; 186(11):3516–3624. [PubMed: 15150239]

Gross, CA. Function and Regulation of the Heat Shock. In: Neidhardt, FC.; Curtiss, R.; Lin, ECC.; Low, KB.; Magasanik, B.; Reanikoff, WS.; Riley, M.; Schaechter, HE.; Umbarger, HE., editors. Escherichia coli and Salmonella. ASM Press; Washington D. C: 1996. p. 1382-1399.

Haddadin FT, Harcum SW. Transcriptome profiles for high-cell-density recombinant and wild-type *Escherichia coli*. Biotechnology and Bioengineering. 2005; 90(2):127–153. [PubMed: 15742388]

Harcum SW, Haddadin FT. Global transcriptome response of recombinant *Escherichia coli* to heat-shock and dual heat-shock recombinant protein induction. Journal of Industrial Microbiology and Biotechnology. 2006; 33(10):801–814. [PubMed: 16680459]

Harcum SW, Ramirez DM, Bentley WE. Optimal Nutrient Feed Policies for Heterologous Protein-Production. Applied Biochemistry and Biotechnology. 1992; 34–5:161–173.

Hoffmann F, Rinas U. Plasmid amplification in *Escherichia coli* after temperature upshift is impaired by induction of recombinant protein synthesis. Biotechnology Letters. 2001; 23(22):1819–1825.

Hoffmann F, van den Heuvel J, Zidek N, Rinas U. Minimizing inclusion body formation during recombinant protein production in *Escherichia coli* at bench and pilot plant scale. Enzyme And Microbial Technology. 2004; 34(3–4):235–241.

Ignatova Z, Enfors SO, Hobbie M, Taruttis S, Vogt C, Kasche V. The relative importance of intracellular proteolysis and transport on the yield of the periplasmic enzyme penicillin amidase in *Escherichia coli*. Enzyme and Microbial Technology. 2000; 26(2–4):165–170. [PubMed: 10689073]

Jevsevar S, Gaberc-Porekar V, Fonda I, Podobnik B, Grdadolnik J, Menart V. Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. Biotechnology Progress. 2005; 21(2):632–639. [PubMed: 15801811]

Jordan GL, Harcum SW. Characterization of up-regulated proteases in an industrial recombinant *Escherichia coli* fermentation. Journal of Industrial Microbiology and Biotechnology. 2002; 28(2): 74–80. [PubMed: 12074055]

Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, Kaipa P, Gilham F, Spaulding A, Popescu L. others. Pathway Tools version 13.0: Integrated Software for Pathway/ Genome Informatics and Systems Biology. Briefings in Bioinformatics. 2010; 11(40)

Kato A, Tanabe H, Utsumi R. Molecular Characterization of the PhoP-PhoQ Two-Component System in *Escherichia coli* K-12: Identification of Extracellular Mg^{2+}-Responsive Promoters. Journal of Bacteriology. 1999; 181(17):5516–5520. [PubMed: 10464230]

Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, others. EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Research. 2013; 41(D1):D605–D612. [PubMed: 23143106]

Korz DJ, Rinas U, Hellmuth K, Sanders EA, Decker WD. Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. Journal of Biotechnology. 1995; 39(1):59–65. [PubMed: 7766011]

Lee PS, Lee KH. Engineering HlyA hypersecretion in Escherichia coli based on proteomic and microarray analyses. Biotechnology and Bioengineering. 2005; 89(2):195–205. [PubMed: 15580578]

Lesley SA, Graziano J, Cho CY, Knuth MW, Klock HE. Gene expression response to misfolded protein as a screen for soluble recombinant protein. Protein Engineering. 2002; 15(2):153–160. [PubMed: 11917152]

Liovic M, Ozir M, Zavec AB, Peternel S, Komel R, Zupancic T. Inclusion bodies as potential vehicles for recombinant protein delivery into epithelial cells. Microbial Cell Factories. 2012; 11

Liu Y, Hu R, Zhang S, Zhang L, Wei X, Chen L. Expression of the Foot-and-Mouth Disease Virus VP1 protein using a replication-competent recombinant canine adenovirus type 2 elicits a humoral antibody response in a porcine model. Viral Immunology. 2006; 19(2):202–209. [PubMed: 16817763]

Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. Bioinformatics. 2009; 25(17):2200–2207. [PubMed: 19549632]

Mahnic M, Baebler S, Blejec A, Jalen S, Gruden K, Menart V, Jevsevar S. Gene Expression Profiling of Recombinant Protein Producing E. coli at Suboptimal Growth Temperature. Acta Chimica Slovenica. 2012; 59(1):59–69. [PubMed: 24061173]

Marisch K, Bayer K, Scharl T, Mairhofer J, Krempl PM, Hummel K, Razzazi-Fazeli E, Striedner G. A Comparative Analysis of Industrial Escherichia coli K-12 and B Strains in High-Glucose Batch Cultivations on Process-, Transcriptome- and Proteome Level. Plos One. 2013; 8(8)

Martinez-Alonso M, Garcia-Fruitos E, Ferrer-Miralles N, Rinas U, Villaverde A. Side effects of chaperone gene co-expression in recombinant protein production. Microbial Cell Factories. 2010; 9

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Research. 2007; 35:W182–W185. [PubMed: 17526522]

Nahku R, Valgepea K, Lahtvee PJ, Erm S, Abner K, Adamberg K, Vilu R. Specific growth rate dependent transcriptome profiling of Escherichia coli K12 MG1655 in accelerostat cultures. Journal of Biotechnology. 2010; 145(1):60–65. [PubMed: 19861135]

Neidhardt, FC.; Umbarger, HE. Chemical Composition of *Escherichia coli*. In: Neidhardt, FC.; Curtiss, R.; Lin, ECC.; Low, KB.; Magasanik, B.; Reanikoff, WS.; Riley, M.; Schaechter, M.; Umbarger, HE., editors. Escherichia coli and Salmonella. 2nd ed. Washington D. C: ASM Press; 1996. p. 13-16.

Oh M-K, Liao JC. DNA microarray detection of metabolic responses to protein overproduction in *Escherichia coli*. Metabolic Engineering. 2000; 2:201–209. [PubMed: 11056062]

Oh MK, Rohlin L, Kao KC, Liao JC. Global expression profiling of acetate-grown *Escherichia coli*. Journal of Biological Chemistry. 2002; 277(15):13175–13183. [PubMed: 11815613]

Pan KL, Hsiao HC, Weng CL, Wu MS, Chou CP. Roles of DegP in prevention of protein misfolding in the Periplasm upon overexpression of penicillin acylase in *Escherichia coli*. Journal of Bacteriology. 2003; 185(10):3020–3030. [PubMed: 12730160]

Peternel S. Bacterial cell disruption: a crucial step in protein production. New Biotechnology. 2013; 30(2):250–254. [PubMed: 21979792]

Peternel S, Gaberc-Porekar V, Komel R, Viktor M. Medium pH affects size and density of inclusion bodies resulting in different extractability of the target protein. Journal of Biotechnology. 2007; 131(2):S189–S189.

Peternel S, Komel R. Isolation of biologically active nanomaterial (inclusion bodies) from bacterial cells. Microbial Cell Factories. 2010; 9

Peternel S, Komel R. Active Protein Aggregates Produced in Escherichia coli. International Journal of Molecular Sciences. 2011; 12(11):8275–8287. [PubMed: 22174663]

Petersson L, Carrio MM, Vera A, Villaverde A. The impact of dnaKJ overexpression on recombinant protein solubility results from antagonistic effects on the control of protein quality. Biotechnology Letters. 2004; 26(7):595–601. [PubMed: 15168861]

Phillips GJ. Green fluorescent protein - a bright idea for the study of bacterial protein localization. Fems Microbiology Letters. 2001; 204(1):9–18. [PubMed: 11682170]

Porowinska D, Marszalek E, Wardecka P, Komoszynski M. In vitro renaturation of proteins from inclusion bodies. Postepy Higieny I Medycyny Doswiadczalnej. 2012; 66:322–329. [PubMed: 22706118]

Prescott M, Nowakowski S, Nagley P, Devenish RJ. The length of polypeptide linker affects the stability of green fluorescent protein fusion proteins. Analytical Biochemistry. 1999; 273(2):305–307. [PubMed: 10469502]

Quan S, Zhang N, French S, Squires CL. Transcriptional polarity in rRNA Operons of Escherichia coli nusA and nusB mutant strains. Journal of Bacteriology. 2005; 187(5):1632–1638. [PubMed: 15716433]

Richmond CS, Glasner JD, Mau R, Jin H, Blattner FJ. Genome-wide expression profiling in *Escherichia coli* K-12. Nucleic Acids Research. 1999; 27(19):3821–3835. [PubMed: 10481021]

Riley M, Abe T, Arnaud MB, Berlyn MKB, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, others. *Escherichia coli* K-12: a cooperatively developed annotation snapshot - 2005. Nucleic Acids Research. 2006; 34(1):1–9. [PubMed: 16397293]

Rinas U, Hoffmann F, Betiku E, Estape D, Maren S. Inclusion body anatomy and functioning of chaperone-mediated *in vivo* inclusion body disassembly during high-level recombinant proetin production in *Escherichia coli*. Journal of Biotechnology. 2007; 127:244–257. [PubMed: 16945443]

Rodriguez, RL.; Tait, RC. Recombinant DNA Techniques: An Introduction. Reading. Massachusetts: Addison-Wesley Publishing Company; 1983.

Rodriguez-Carmona E, Villaverde A. Nanostructured bacterial materials for innovative medicines. Trends in Microbiology. 2010; 18(9):423–430. [PubMed: 20674365]

Rohlin L, Oh MK, Liao JC. DNA Microarray for microbial biotechnology: Gene expression profiles in *Escherichia coli* during protein overexpression. Journal of the Chinese Institute of Chemical Engineers. 2002; 33(1):103–112.

Rokney A, Shagan M, Kessel M, Smith Y, Rosenshine I, Oppenheim AB. E. coli Transports Aggregated Proteins to the Poles by a Specific and Energy-Dependent Process. Journal of Molecular Biology. 2009; 392(3):589–601. [PubMed: 19596340]

Rozkov A, Schweder T, Veide A, Enfors SO. Dynamics of proteolysis and its influence on the accumulation of intracellular recombinant proteins. Enzyme and Microbial Technology. 2000; 27(10):743–748. [PubMed: 11118580]

Sabate R, de Groot NS, Ventura S. Protein folding and aggregation in bacteria. Cellular and Molecular Life Sciences. 2010; 67(16):2695–2715. [PubMed: 20358253]

Salazar MA, Fernando LP, Baig F, Harcum SW. The effects of protein solubility on the RNA Integrity Number (RIN) for recombinant *Escherichia coli*. Biochemical Engineering Journal. 2013; 79:129–135. [PubMed: 24151430]

Schlieker C, Bukau B, Mogk A. Prevention and reversion of protein aggregation by molecular chaperones in the *E. coli* cytosol: implications for their applicability in biotechnology. Journal of Biotechnology. 2002; 96(1):13–21. [PubMed: 12142139]

Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. Genome Research. 2003; 13(2):216–223. [PubMed: 12566399]

Sharma SS, Blattner FR, Harcum SW. Recombinant protein production in an *Escherichia coli* reduced genome strain. Metabolic Engineering. 2007a; 9(2):133–141. [PubMed: 17126054]

Sharma SS, Campbell JW, Frisch D, Blattner FR, Harcum SW. Expression of two recombinant chloramphenicol acetyltransferase variants in highly reduced genome *Escherichia coli* strains. Biotechnology and Bioengineering. 2007b; 98(5):1056–1070. [PubMed: 17497738]

Singh SM, Sharma A, Upadhyay AK, Singh A, Garg LC, Panda AK. Solubilization of inclusion body proteins using n-propanol and its refolding into bioactive form. Protein Expression and Purification. 2012; 81(1):75–82. [PubMed: 21964443]

Smialowski P, Doose G, Torkler P, Kaufmann S, Frishman D. PROSO II - a new method for protein solubility prediction. Febs Journal. 2012; 279(12):2192–2200. [PubMed: 22536855]

Smialowski P, Martin-Galiano AJ, Mikolajka A, Girschick T, Holak TA, Frishman D. Protein solubility: sequence based prediction and experimental verification. Bioinformatics. 2007; 23(19): 2536–2542. [PubMed: 17150993]

Smith H. The transcriptional response of *Escherichia coli* to recombinant protein insolubility. Journal of Structural and Functional Genomics. 2007; 8(1):27–35. [PubMed: 17992580]

Strandberg L, Enfors SO. Factors Influencing Inclusion Body Formation in the Production of a Fused Protein in *Escherichia coli*. Applied and Environmental Microbiology. 1991; 57(6):1669–1674. [PubMed: 1908208]

Striedner G, Cserjan-Puschmann M, Potschacher F, Bayer K. Tuning the transcription rate of recombinant protein in strong *Escherichia coli* expression systems through repressor titration. Biotechnology Progress. 2003; 19(5):1427–1432. [PubMed: 14524702]

Swartz JR. Advances in *Escherichia coli* production of therapeutic proteins. Current Opinion in Biotechnology. 2001; 12(2):195–201. [PubMed: 11287237]

Thomas JG, Baneyx F. Divergent effects of chaperone overexpression and ethanol supplementation on inclusion body formation in recombinant Escherichia coli. Protein Expression and Purification. 1997; 11(3):289–296. [PubMed: 9425634]

Tseng CP. Regulation of fumarase (*fumB*) gene expression in *Escherichia coli* in response to oxygen, iron, and heme availability: role of the *arcA*, *fur*, and *hemA* gene products. FEMS Microbiology Letters. 1997; 157:67–72. [PubMed: 9418241]

Tseng CP, Yu CC, Lin HH, Chang CY, Kuo JT. Oxygen- and Growth Rate-Dependent Regulation of *Escherichia coli* Fumarase (FumA, FumB, and FumC) Activity. Journal of Bacteriology. 2001; 183(2):461–467. [PubMed: 11133938]

Upadhyay AK, Murmu A, Singh A, Panda AK. Kinetics of Inclusion Body Formation and Its Correlation with the Characteristics of Protein Aggregates in Escherichia coli. Plos One. 2012; 7(3)

Van der, Rest ME.; Frank, C.; Molenaar, D. Functions of the Membrane-Associated and Cytoplasmic Malate Dehydrogenases in the Citric Acid Cycle of *Escherichia coli*. Journal of Bacteriology. 2000; 182(24):6892–6899. [PubMed: 11092847]

Ventura S, Villaverde A. Protein quality in bacterial inclusion bodies. TRENDS in Biotechnology. 2006; 24(4):179–185. [PubMed: 16503059]

Villaverde A, Carrio MM. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. Biotechnology Letters. 2003; 25(17):1385–1395. [PubMed: 14514038]

Walker, GC. The SOS Response of *Escherichia coli*. In: Neidhardt, FC.; Curtiss, R.; Lin, ECC.; Low, KB.; Magasanik, B.; Reanikoff, WS.; Riley, M.; Schaechter, M.; Umbarger, HE., editors. Escherichia coli and Salmonella. Washington D. C: ASM Press; 1996. p. 1400-1416.

Weber J, Senior AE. ATP synthesis driven by proton transport in F1F0-ATP synthase. FEBS Letters. 2003; 545(1):61–70. [PubMed: 12788493]

Welter BH, Laughlin RC, Temesvari LA. Characterization of a Rab7-like GTPase, EhRab7: a marker for the early stages of endocytosis in *Entamoeba histolytica*. Molecular and Biochemical Parasitology. 2002; 121(2):254–264. [PubMed: 12034459]

Wendisch VF, Zimmer DP, Khodursky A, Peter B, Cozzarelli N, Kustu S. Isolation of *Escherichia coli* mRNA and comparison of expression using mRNA and total RNA on DNA microarrays. Analytical Biochemistry. 2001; 290(2):205–213. [PubMed: 11237321]

Yoon SH, Han MJ, Lee SY, Jeong KJ, Yoo JS. Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. Biotechnology and Bioengineering. 2003; 81(7):753–767. [PubMed: 12557308]

Zimmer M. Green fluorescent protein (GFP): Applications, structure, and related photophysical behavior. Chemical Reviews. 2002; 102(3):759–781. [PubMed: 11890756]

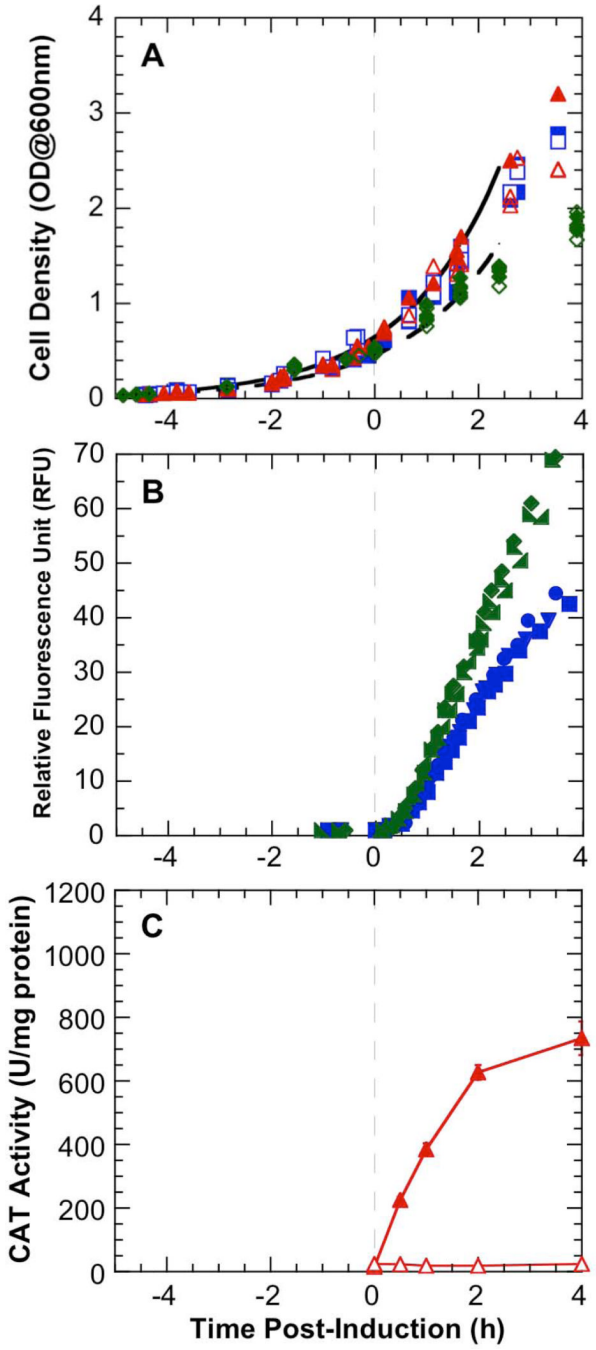

Figure 1. Growth and protein production profiles for *E. coli* pTVP1GFP and pGFPCAT

A) Cells were cultured in minimal medium with and without induction and with and without ethanol treatment. VP1GFP (■, □), GFPCAT (▲, △), and ethanol-treated VP1GFP (◆, ◇). Uninduced (□, △, ◇) and Induced (■, ▲, ◆). B) Fluorescent profiles for induced *E. coli* VP1GFP (■, ●, ▼) and ethanol-treated induced *E. coli* VP1GFP (◆, ◣, ◢) cultures. Triplicate data are shown. C) Specific CAT activity profiles for *E. coli* GFPCAT Uninduced (△) and Induced (▲) cultures. Error bars represent 95% confidence intervals. Cultures were synchronized to Time 0 at the time of induction.

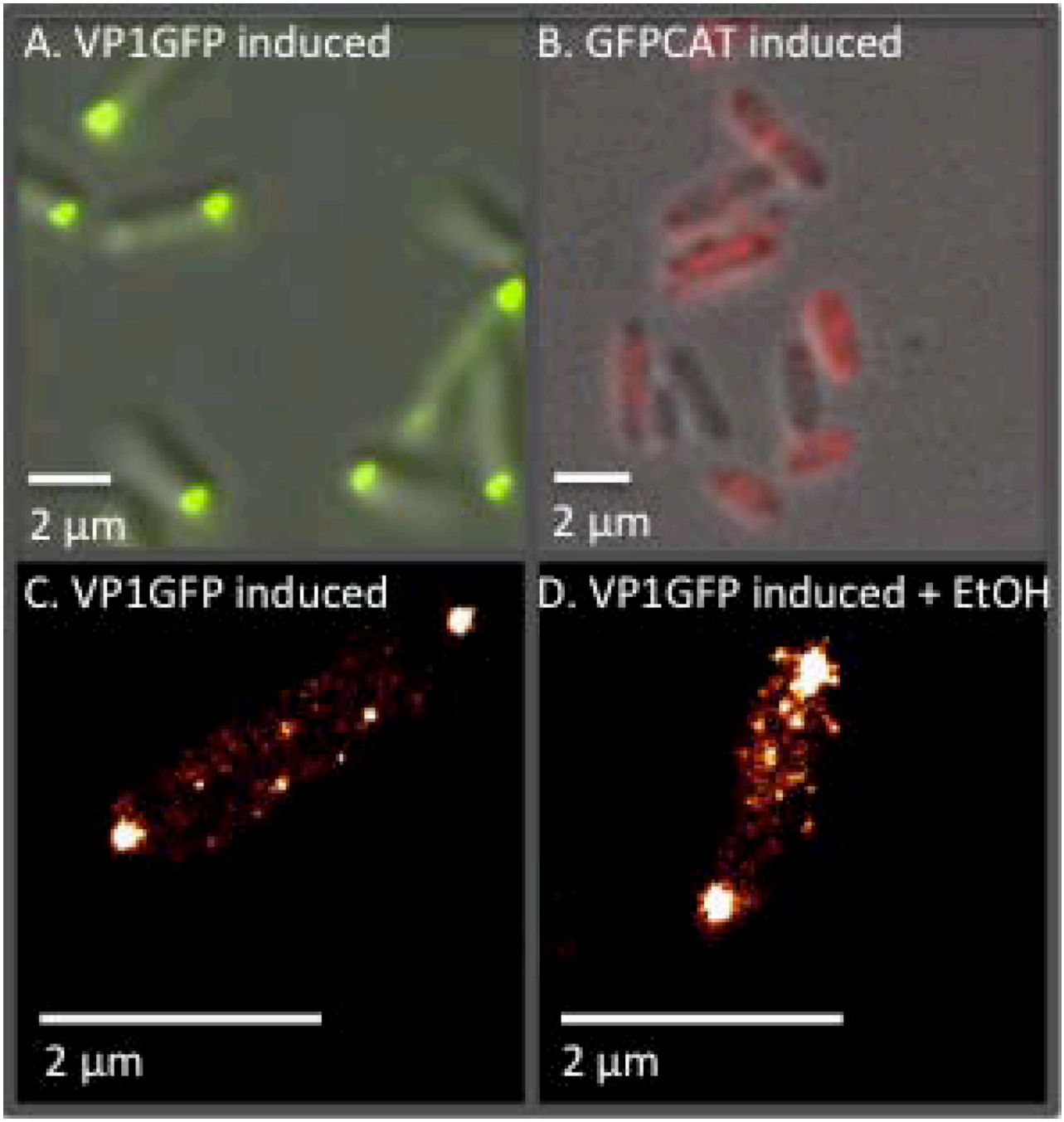

Figure 2. Protein localization within*E. coli*pTVP1GFP and pGFPCAT was observed using traditional and super-resolution widefield fluorescence microscopy

A) Induced *E. coli* pTVP1GFP (unstained); Merged DIC and fluorescence (GFP, green) images. B) Induced *E. coli* pGFPCAT (immunolabeled with anti-CAT antibody, shown in red). Merged DIC and fluorescence images. C) Induced *E. coli* pTVP1GFP (super-resolution and unstained); and D) Ethanol-treated induced *E. coli* pTVP1GFP (super-resolution and unstained). Areas of multiple fluorescence events in a particular location are identified in white, while fewer fluorescent events are represented in red.

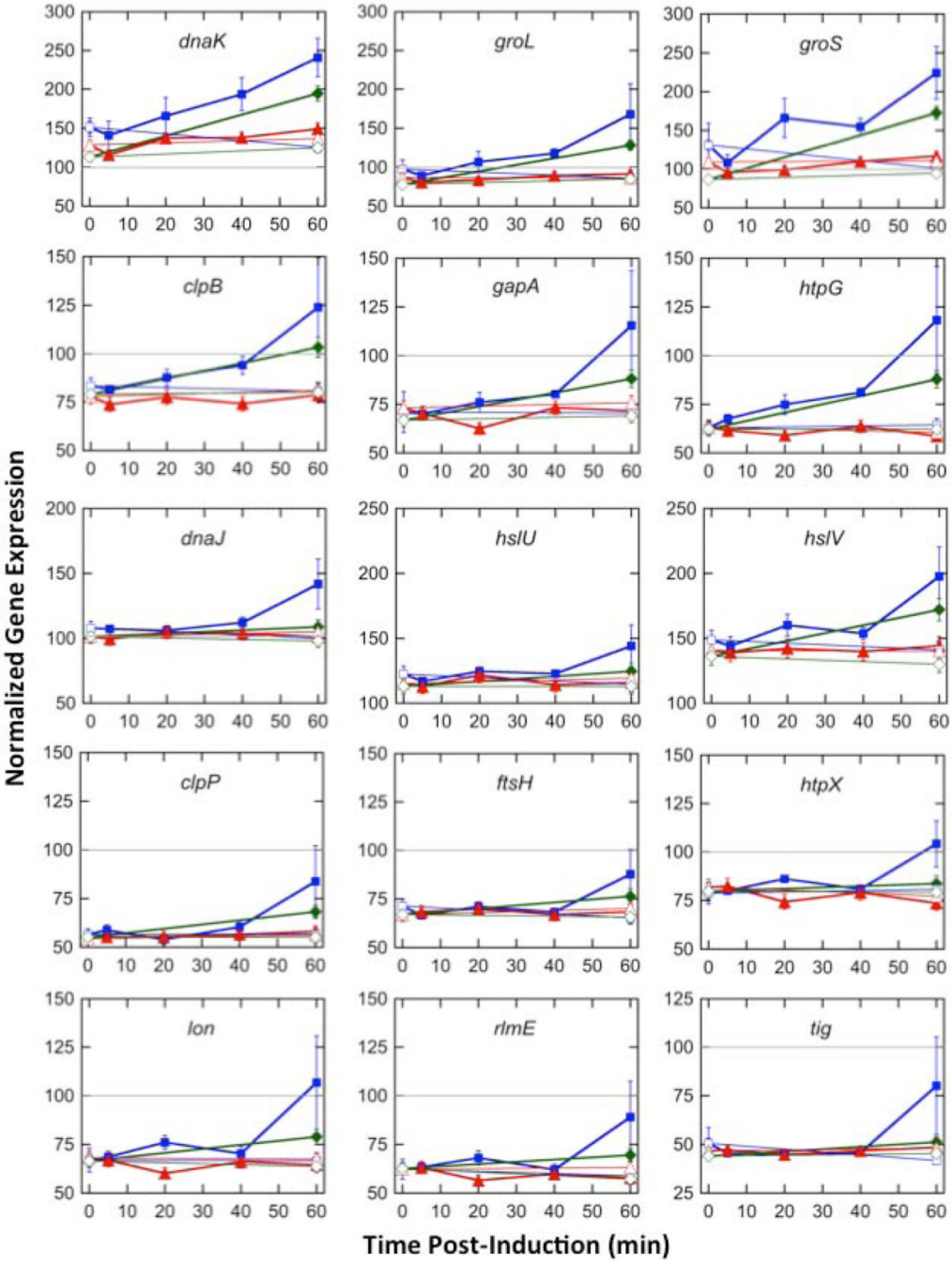

Figure 3. Classical heat shock and*tig*gene expression profiles for*E. coli*pTVP1GFP and pGFPCAT

The dynamic gene expression profiles for heat-shock and *tig* genes. VP1GFP (■, □), GFPCAT (▲, △), and ethanol-treated VP1GFP (◆, ◇); Uninduced (□, △, ◇) and Induced (■, ▲, ◆). Gene expression levels were normalized to 100, which represents the "average" gene expression intensity on the DNA microarray. Standard error bars are shown.

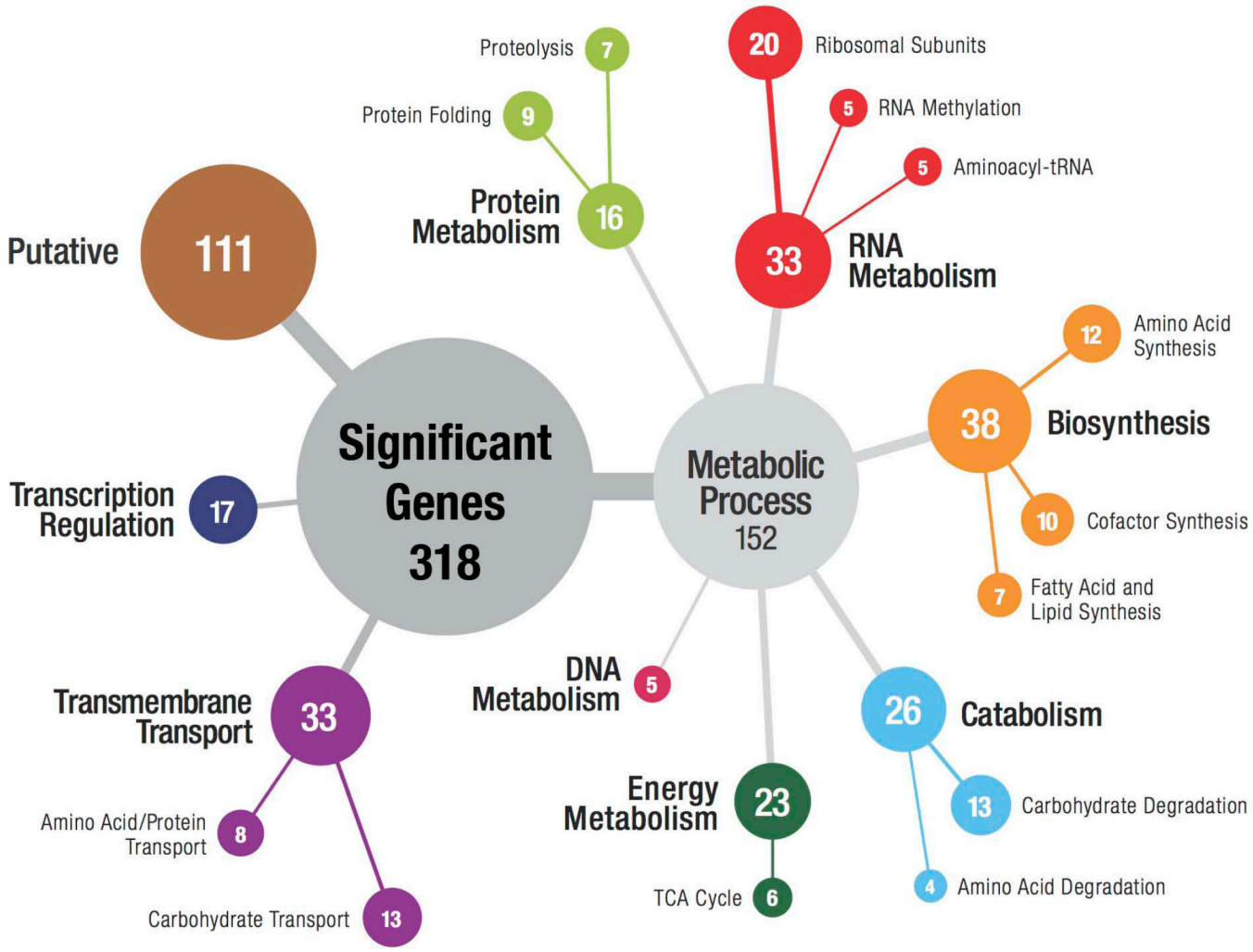

Figure 4. Classification of genes found to be affected by protein production

These genes were identified to be significantly effected by VP1GFP and/or GFPCAT protein production by ANOVA analysis ($p \leq 0.10$) followed by regression analysis ($p \leq 0.05$) and/or Tukey's W analysis ($p \leq 0.05$). These genes have been grouped by function using gene ontology (GO) terms from EcoCyc and annotations from the ASAP database.

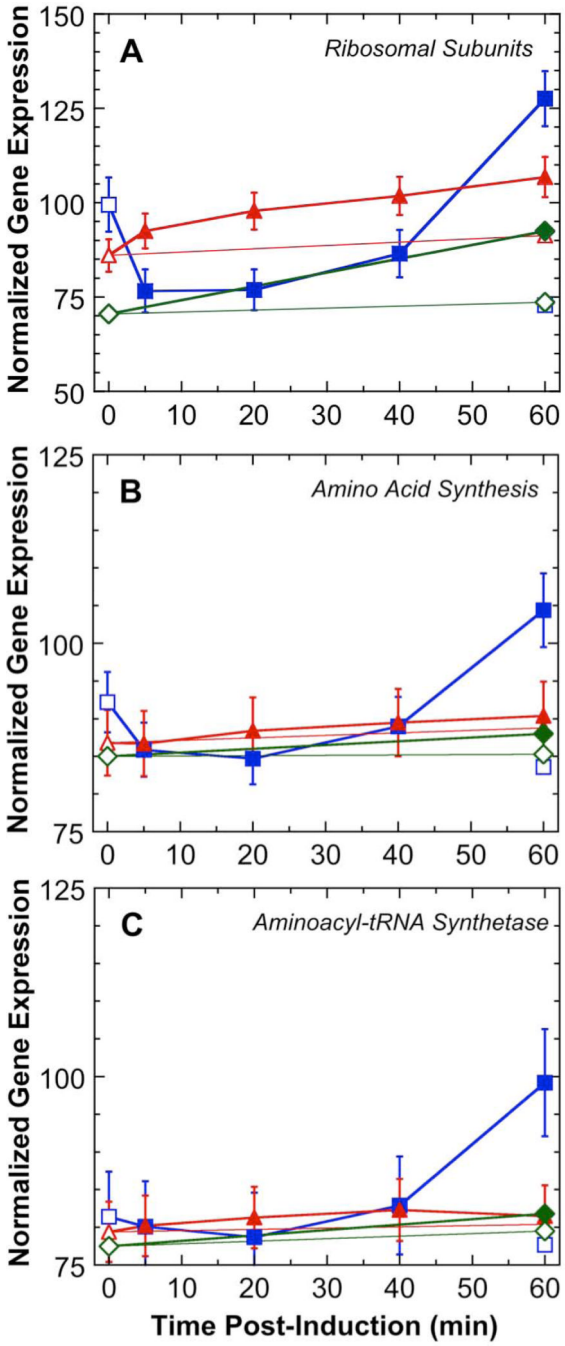

Figure 5. Gene expression profiles for protein synthesis-related genes with differential expression in*E. coli*pTVP1GFP and pGFPCAT

A) Ribosomal subunit genes (average of 20 genes). B) Amino Acid Synthesis genes (average of 12 genes). C) Aminoacyl-tRNA Synthetase genes (average of 5 genes). The scales for panels B and C are half of the scale length used for panel A to improve visualization. VP1GFP (■, □), GFPCAT (▲, △), and ethanol-treated VP1GFP (◆, ◇); Uninduced (□, △, ◇) and Induced (■, ▲, ◆). Gene expression levels were normalized to 100, which represents the "average" gene expression intensity on the DNA microarray. Standard error bars are shown.

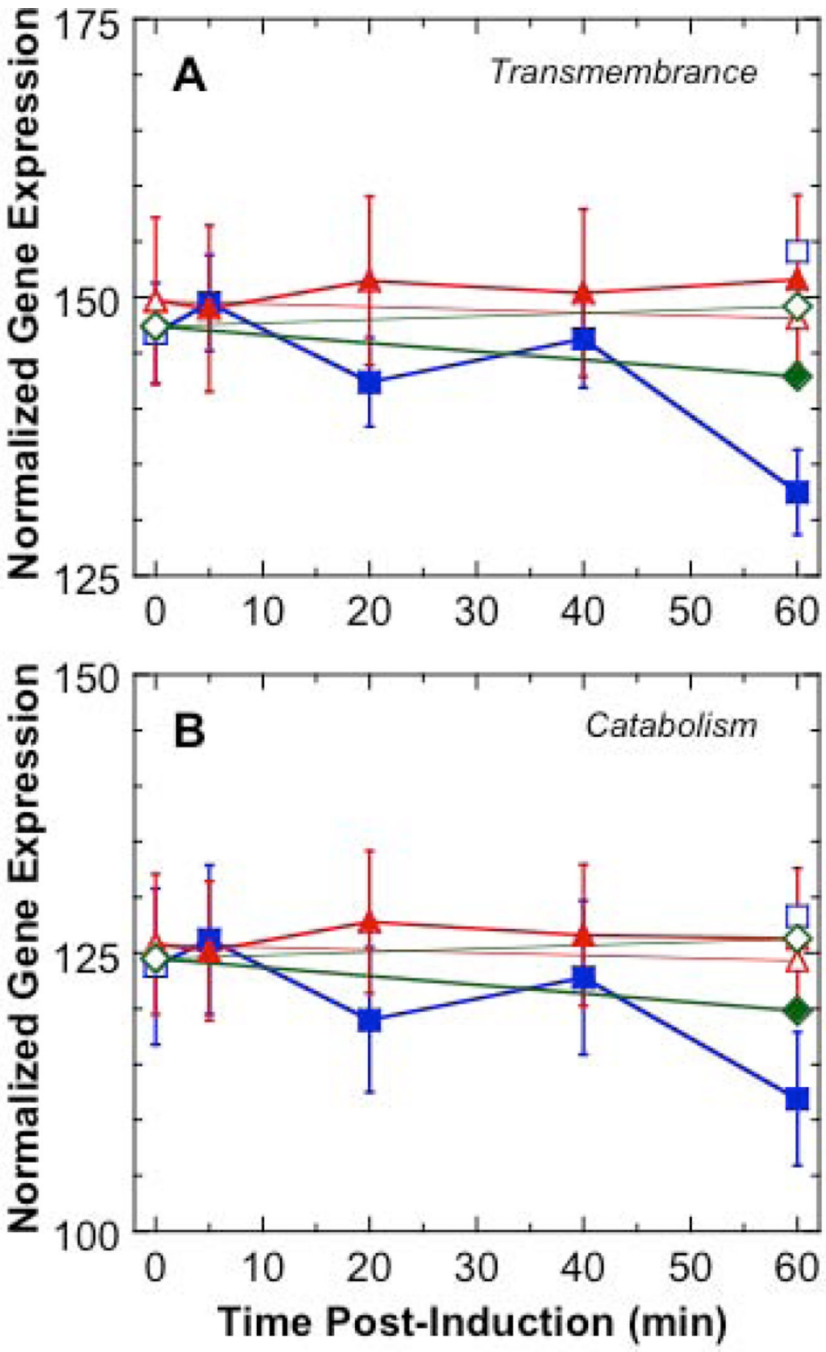

Figure 6. Gene expression profiles for transmembrane transport and catabolism genes with differential expression in*E. coli*pTVP1GFP and pGFPCAT

A) Transmembrane transport genes (*aaeB*, *acrF*, *actP*, *bglF*, *blc*, *dgoT*, *dtpB*, *eamB*, *fucP*, *glpF*, *gltS*, *gntT*, *gspDL*, *lamB*, *malF*, *modB*, *purP*, *tdcC*, *tsgA*, *ugpB*, *uhpT*, and *ulaA*) with decreased gene expression due to VP1GFP (average of 23 genes). B) Catabolism genes (*bglB*, *caiB*, *chiA*, *cpdB*, *dgoDK*, *dtd*, *glcD*, *gudD*, *maoC*, *mtlD*, *nudE*, *rhaBM*, *tdh*, *tnaA*, *treF*, *uxuB*, and *yiaS*) with decreased gene expression due to VP1GFP (average of 19 genes). VP1GFP (■, □), GFPCAT (▲, △), and ethanol-treated VP1GFP (◆, ◇); Uninduced (□, △, ◇) and Induced (■, ▲, ◆). Gene expression levels were normalized to 100, which

represents the “average” gene expression intensity on the DNA microarray. Standard error bars are shown.

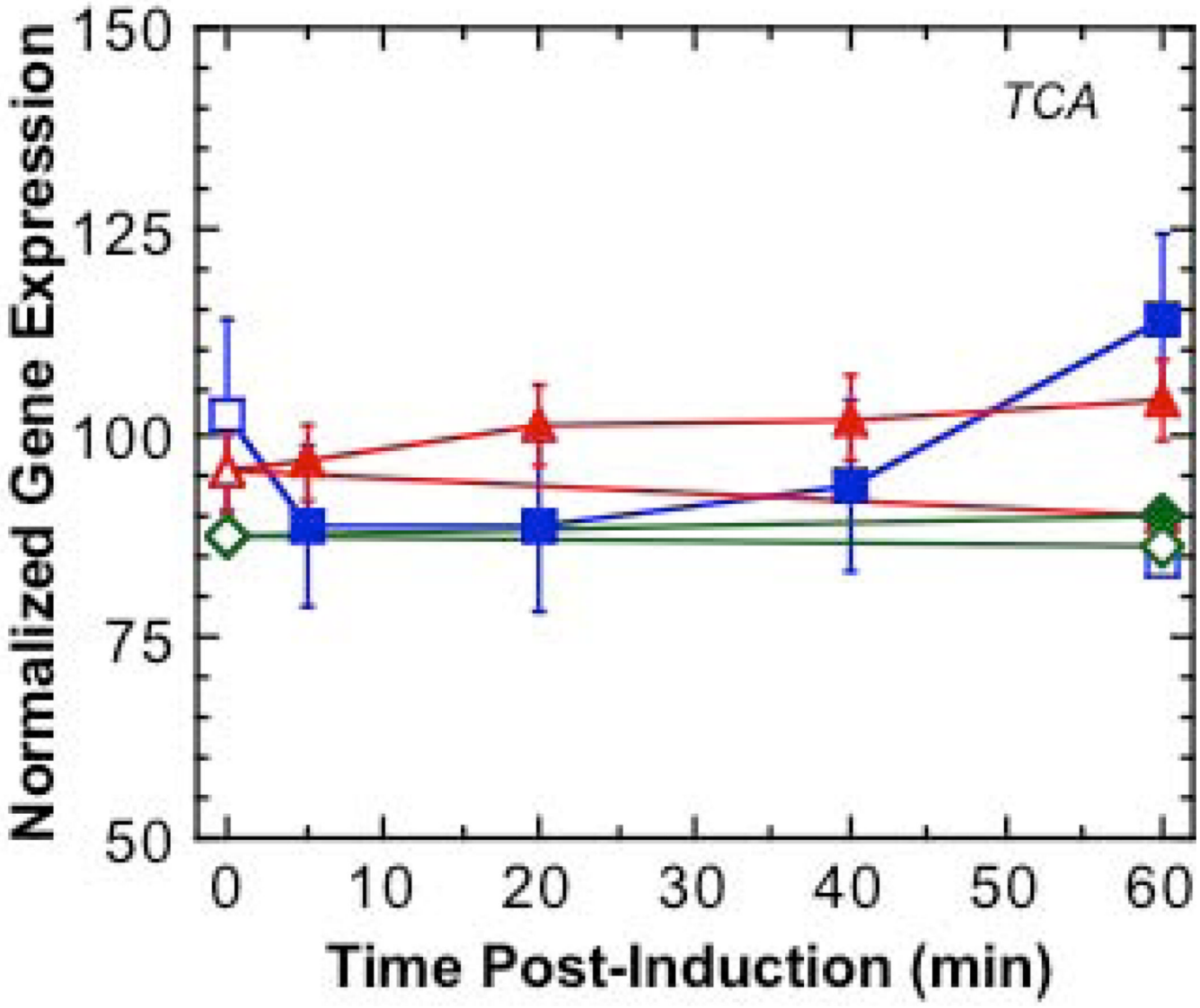

Figure 7. Gene expression profiles for TCA cycle genes with differential expression in*E. coli*pTVP1GFP and pGFPCAT

A) TCA cycle genes (*acnB*, *gltA*, *sdhAB*, and *sucBC*) (average of six genes). VP1GFP (■, □), GFPCAT (▲, △), and ethanol-treated VP1GFP (◆, ◇); Uninduced (□, △, ◇) and Induced (■, ▲, ◆). Gene expression levels were normalized to 100, which represents the "average" gene expression intensity on the DNA microarray. Standard error bars are shown.

Biotechnol Bioeng. Author manuscript; available in PMC 2015 May 01.

Table 1

Tukey's W pairwise comparisons

Condition	VP1GFP Time 0 Uninduced	VP1GFP Time 60 Uninduced	VP1GFP Time 60 Induced	GFPCAT Time 0 Uninduced	GFPCAT Time 60 Uninduced	GFPCAT Time 60 Induced	VP1GFP Time 0 Ethanol Uninduced	VP1GFP Time 60 Ethanol Uninduced	VP1GFP Time 60 Ethanol Induced
VP1GFP Time 0 Uninduced		2	**22 (5 HS)**	2	0 (C)	1 (X)	1	1 (C)	1 (C)
VP1GFP Time 60 Uninduced			**162 (11 HS)**	0 (X)	0 (C)	1 (C)	1 (X)	2	5(C)
VP1GFP Time 60 Induced				65 (X)	45 (C)	**126 (10 HS)**	164 (C)	76 (C)	**9 (0 HS)**
GFPCAT Time 0 Uninduced					0	2	0	4 (X)	5 (X)
GFPCAT Time 60 Uninduced						1	5 (X)	2 (C)	5 (C)
GFPCAT Time 60 Induced							23 (X)	6 (C)	11 (C)
VP1GFP Time 0 Ethanol Uninduced								1	2
VP1GFP Time 60 Ethanol Uninduced									0
VP1GFP Time 60 Ethanol Induced									

Tukey's W pairwise comparisons ($p \leq 0.05$) were used to identify genes with differential gene expression between culture conditions; the 961 significant genes previously identify by the ANOVA analysis ($p \leq 0.10$) were analyzed. The solubility-sensitive pairwise comparisons are indicated in **bold**. The number of heat shock (HS) genes that were identified by a comparison is indicated. Pairwise comparisons with limited biological meaning (X) or with multiple condition differences (C), *i.e*. confounding effects, are indicated. Self-comparisons form the diagonal, which is blank.

Table 2

Classical heat-shock gene behavior due to soluble and insoluble protein production.

				Regression Slope		Fold changes* Current Study			Fold changes: Past Studies	
bname	***Gene***	***Gene Description***	***GO Term***	***VP1GFP***	***GFPCAT***	***A***	***B***	***C***	**Lesley*et al*, 2002**	**Smith, 2007**
b2592	*clpB*	protein disaggregation chaperone	Protein Folding	(+)			1.5	1.5	9.7	8.3
b0437	*clpP*	proteolytic subunit of ClpA-ClpP and ClpX-ClpP ATP-dependent serine proteases	Proteolysis		(+)				11.8	
b0015	*dnaJ*	chaperone Hsp40, co-chaperone with DnaK	Protein Folding			1.3	1.4	1.4	5.6	3.6
b0014	*dnaK*	chaperone Hsp70, co-chaperone with DnaJ	Protein Folding	(+)			1.9		16.6	7.4
b3178	*ftsH*	subunit of integral membrane ATP-dependent zinc metallopeptidase	Proteolysis				1.3			
b1779	*gapA*	glyceraldehyde-3-phosphate dehydrogenase A	Energy Metabolism	(+)		1.6				
b4143	*groL*	Cpn60 chaperonin GroEL, large subunit of GroESL	Protein Folding	(+)			1.9	1.8		10.3
b4142	*groS*	Cpn10 chaperonin GroES, small subunit of GroESL	Protein Folding	(+)			2.2	1.9		9.8
b3931	*hslU*	molecular chaperone and ATPase component of HslUV protease	Protein Folding				1.2	1.2	4.7	6.1
b3932	*hslV*	peptidase component of the HslUV protease	Protein Folding				1.4	1.4	7.4	7.8
b0473	*htpG*	molecular chaperone HSP90 family	Protein Folding	(+)		1.8	1.7	1.9	13.2	14.3
b1829	*htpX*	heat shock protein, integral membrane protein	Proteolysis			1.3		1.4	4.9	
b0439	*lon*	DNA-binding ATP-dependent protease La	Proteolysis			1.5	1.5	1.6	6	
b3179	*rlmE*	23S rRNA U2552 ribose 2'-O-methyltransferase	RNA Methylation				1.5	1.5	3.2	2.5
b3400	*hslR*	ribosome-associated heat shock protein Hsp15	RNA Metabolism						28.3	8.2
b3686	*ibpB*	heat shock chaperone	Protein Folding						40	29.1
b4129	*lysU*	lysine tRNA synthetase, inducible	Aminoacyl-tRNA Synthetase							
b2573	*rpoE*	RNA polymerase, sigma 24 (sigma E) factor	Transcription Regulation							

These classical heat-shock genes were identified by an ANOVA analysis ($p \leq 0.10$) across the fifteen sample conditions. The regression analysis slopes ($p \leq 0.05$) are indicated (+) for positive slopes due to VP1GFP or GFPCAT protein production due to induction (no negative slopes were identified among these genes). Fold changes for Tukey's W pairwise comparisons ($p \leq 0.05$) are indicated. The fold changes observed by Lesley *et al*. (2002) and Smith (2007) for insoluble protein culture conditions relative to control culture conditions are indicated.

*****For fold changes, A = VP1GFP 60-minute induced vs. VP1GFP 0-minute uninduced. B = VP1GFP 60-minute induced vs. VP1GFP 60-minute uninduced. C = VP1GFP 60-minute induced vs. GFPCAT 60-minute induced.

Table 3

Transmembrane transporters and catabolic genes with common substrates.

Substrate	Genes with Decreased Expression Due to VP1GFP Production	
	Transmembrane Transport Gene	**Catabolic Pathway Gene**
D-galactonate	*dgoT*	*dgoDK*
Glycolate	*actP*	*glcD*
Arbutin-6-phosphate	*bglF*	*bglB*
Methyl-β-D-glucoside-6-phosphate	*bglF*	*bglB*
Threonine	*tdcC*	*tdh*
Cysteine	*eamB*	*tnaA*

These linked transmembrane transporters and catabolic genes had decreased gene expression due to VP1GFP production. The common substrates were determined using the PathwayTools software Version 13.0 (Karp et al. 2010).