Published in final edited form as:

Biochem Eng J. 2013 October 15; 79: 129–135. doi:10.1016/j.bej.2013.07.011.

The effects of protein solubility on the RNA Integrity Number (RIN) for recombinant *Escherichia coli*

Mary Alice Salazar^{#1}, Lawrence P. Fernando^{#1,#}, Faraz Baig², and Sarah W. Harcum^{2,*}

¹Department of Chemistry, Clemson University

²Department of Bioengineering, Clemson University

[#] These authors contributed equally to this work.

Abstract

High quality, intact messenger RNA (mRNA) is required for DNA microarray and reverse transcriptase polymerase chain reaction analysis and is generally obtained from total RNA isolations. The most widely recognized measure of RNA integrity is the RNA Integrity Number (RIN) obtained from the Agilent Bioanalyzer, as it provides sizing, quantification, and quality control measures. This work describes comparisons of the RIN values obtained for recombinant *E. coli*. Uninduced recombinant *E. coli* cultures were examined, as well as induced cultures that produced either a soluble or insoluble recombinant protein. The uninduced cultures and the induced cultures producing soluble protein had higher RIN values than the induced cultures producing insoluble protein. These lower RIN values for *E. coli* producing the insoluble protein indicate that cellular degradation of the ribosomal RNA species is the likely cause of the lower RIN values. As the use of DNA microarrays and other gene expression tools increase in usage in the industrial recombinant protein production community, these results suggest the need for further studies to determine acceptable RIN ranges for gene expression analysis and effects of various culture conditions on RIN values for recombinant *E. coli*.

Keywords

Aggregation; Protease; Purification; Recombinant DNA; RNA; Inclusion bodies

1. Introduction

Escherichia coli is used to produce a wide range of recombinant proteins, many of which are used as therapeutic agents [1, 2]. *E. coli* has many advantages over the other host organisms for the large-scale production of recombinant proteins including genome simplicity, well understood genetics and metabolism, and fast growth rates on inexpensive growth medium [3, 4]. One major disadvantage of recombinant protein production in *E. coli* is its tendency to produce insoluble inclusion bodies of the desired target recombinant protein [4–14]. Many efforts to control inclusion body formation include overexpression of chaperones [15–

© 2013 Elsevier B.V. All rights reserved

*Corresponding author information S. W. Harcum Department of Bioengineering 301 Rhodes Clemson University Clemson, SC 29634-0905.

[#]Present address: Department of Biomedical Engineering, University of Florida

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

20], codon optimization [21, 22], and decreased culture temperatures [23–25]. Despite these advances to control inclusion bodies, it is still not possible to *a priori* predict the solubility state of a new recombinant protein [16, 26–30].

To gain a better understanding of inclusion body formation, DNA microarrays have been used [27, 31]. In order to conduct DNA microarray or any gene expression analysis, high quality messenger RNA (mRNA) is required [32–34]. Most RNA purification techniques for prokaryotes isolate and amplify the total RNA, which includes mRNA, ribosomal RNA (rRNA), and transfer RNA (tRNA) species, since prokaryotic mRNA lacks a stable poly(A) tail [35]. There are several methods that are routinely used to evaluate RNA quantity and quality, the most common being RNA absorbance [36]. Absorbance methods indicate purity and concentration, but cannot distinguish the RNA species or intactness [36]. Electrophoresis methods allow visualization of the ribosomal RNA species, but are limited in quantification precision [36–38]. Due to the limitation of the absorbance and traditional electrophoresis methods, DNA microarray manufacturers (*i.e.*, Affymetrix, Illumina, Agilent, and Roche Nimblegen) highly recommend analysis of RNA quality using the Agilent Bioanalyzer.

The Agilent Bioanalyzer is a microfluidics-based platform that generates an electropherogram and simulated gel image that provides sizing, RNA quantification, and quality control, which is reported as the RNA Integrity Number (RIN) [37, 39]. The RIN value is calculated from the proportion of expected RNA fragment sizes and is independent of variance in sample concentrations [37, 39]. Low RIN values are usually attributed to ribosome degradation during the isolation and purification steps; however, these detailed characterization studies were largely focused on eukaryotic RNA [37, 39–44]. Two recent investigations examined RIN values in prokaryotes; wild-type *E. coli* in meat samples [45] and several bacterial species found in human stool samples [46]. To illustrate the lack of knowledge for RIN values for recombinant *E. coli*, the Agilent RNA Integrity Database (RINdb) contains only one *E. coli* electropherogram out of the 646 entries (http://www.chem.agilent.com/rin/_rinDetail.aspx?rID=4156, accessed May 30, 2013); and this electropherogram is for a “normal” wild-type *E. coli* culture, and has a RIN value of only 1.0.

In this study, RIN values were evaluated for recombinant *E. coli* cultures in preparation for DNA microarray analysis. Both induced and uninduced cultures were evaluated. Additionally, the effects of soluble and insoluble protein productions were compared. All cultures were synchronized with respect to growth phase and cell densities at induction. Samples were harvested in parallel from the soluble and insoluble protein producing cultures. The total RNA was isolated and purified in parallel using standard RNA isolation and purification techniques. The isolated total RNA was evaluated by standard absorbance techniques. And, the Agilent Bioanalyzer (2100) with the Prokaryotic Total RNA Nano software was used to obtain RIN values for all samples. A statistical comparison of the RIN values obtained for the total RNA samples was conducted.

2. Materials and Methods

2.1 Bacterial strains and plasmids

E. coli MG1655 obtained from American Type Culture Collection (ATCC) were transformed with either pTVP1GFP or pGFPCAT plasmids. Both plasmids are isopropyl β -D thiogalactopyranoside (IPTG) inducible via a *trc* promoter and encode ampicillin resistance. The pTVP1GFP plasmid (donated by A. Villaverde) encodes for a fusion protein (VP1GFP) which contains the VP1 capsid protein from the Foot and Mouth disease virus [47] fused to a green fluorescent protein (mGFP) [48]. The pGFPCAT plasmid encodes a

chloramphenicol acetyltransferase (CAT) and mGFP fusion and was constructed by replacing the GFP_{uv} in the pTrcHis-GFP_{UV}/CAT plasmid (donated by W.E. Bentley [49]) with the mGFP from pTVP1GFP. The primers used for the mGFP substitution were: Forward: 5' G ATC CAT ATG AGC AAA GGA GAA GAA CTT TTC 3' and Reverse: 5' G ATC CAT ATG TGT AGA GCT CAT CCA TGC CAT GTG TAA TCC 3'.

2.2 Culture conditions

Cells were cultured in a minimal medium as described previously [50, 51]. One mL of *E. coli* from frozen stocks was added to minimal medium in the presence of ampicillin (40 µg/mL, Hyclone) [36] and grown overnight at 37°C and 250 rpm (New Brunswick Scientific, C24 incubator shaker) to an optical density (OD₆₀₀) of 2.5 OD, where 1 OD is equivalent to 0.50 g dry cell weight per L. From the overnight cultures, *E. coli* pGFPCAT and pTVP1GFP were added separately to 500 mL shake flasks (120 mL working volume) at 37°C in a water bath shaker at 200 rpm (New Brunswick Scientific, C76 incubator shaker) for an initial cell density of 0.05 OD. Cell growth was monitored by optical density using a spectrophotometer (Spectronic 20 Genesys), where samples were taken without stopping agitation or removing the flasks from the water bath. Samples for cell densities were diluted with deionized water to obtain absorbance readings in the linear range (0 – 0.25 OD units). Cultures were induced (1 mM IPTG) in the mid-exponential phase (OD₆₀₀ = 0.5). For the RNA isolations, samples were collected prior to induction (time 0-minutes) and 5-, 20-, 40-, and 60-minutes post-induction for the induced cultures and at 60-minutes for parallel uninduced cultures. The uninduced samples were obtained only for times 0- and 60-minutes. The entire culture broth sample was immediately stabilized in RNAProtect™ Bacteria Reagent (Qiagen) and processed as per manual instructions. These protected samples were centrifuged, and growth media and RNAProtect™ Bacteria Reagent (14,500 × g, 10 minutes, Hermle Labnet Z383K centrifuge) were removed. Cell pellets were stored at –80°C until used for RNA isolation. Additionally, parallel samples were harvested and processed for protein characterization [52]. All culture conditions were conducted in biological triplicates.

2.3 Protein Expression Analysis

The CAT activity of the fusion GFPCAT protein was measured using the kinetic assay described by Rodriguez and Tait (1983) and adapted to a 96-well plate format [52, 53]. Additionally, CAT biological activity was confirmed by the ability of induced *E. coli* pGFPCAT to grow on high-levels of chloramphenicol (0.61 mM) containing LB plates. VP1GFP protein production was confirmed by obtaining fluorescence emission spectra for whole cells fixed with 2% formaldehyde and immediately assayed [54–56]. Fluorescence measurements were taken using a BD Influx Cell Sorter (formerly Cytopeia) with a 488-nm Argon excitation laser and a 530/540 emission filter. The mean signal acquisition from 50,000 cells was used to characterize sample intensity. VP1GFP inclusion bodies were confirmed by fluorescence microscopy (Nikon Ti, 60X TIRF oil) [57].

2.4 RNA isolation and Characterization

Total RNA was isolated using the RNeasy® Bacteria Kits (Qiagen, #74524 and #75144) were used depending on the cell numbers to be processed). RNA was quantified by a Nanodrop spectrophotometer (ND 1000, Thermo Scientific). The Agilent 2100 Bioanalyzer and Agilent RNA 6000 Nano chip kits were used to assess the total RNA quality as per manual instruction (Agilent). The Agilent 2100 Expert software (Version B.02.07.SI532) was used with the Prokaryote total RNA series II assay settings. No significant differences in RIN values were observed between different dilutions of the same sample.

2.5 Gene Expression Analysis

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 the second strand synthesis. Custom *Escherichia coli* DNA microarrays (12 arrays per slide $\times$ 135K probes per array) were prepared by Roche NimbleGen with 4,281 *E. coli* genes, mGFP, TVP1, ampicillin resistance gene (Amp^r), and CAT probes (45–60mer, 10 probes per target, 3 copies of each probe on array). 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 [58–60]. 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™ from the RMA.calls files. Technical replicate expression levels were scaled using the “global averaging” data transformation. An ANOVA test ($p = 0.10$) was conducted on the gene expression values for the biological triplicates to screen for differentially expressed genes (ArrayStar™).

Statistical Analysis—Statistical analysis was conducted using the generalized linear model (GLM) method ($p = 0.05$) with the JMP 10 software (SAS Institute Inc.). For the RIN values and the growth rates, the effectors were protein carried on the plasmid (VP1GFP and GFPCAT) and the induction state (uninduced and induced) of the cultures. Additionally, for the RIN values, interaction effects were examined for the protein and induction conditions, as well as any time dependency effects. For the RIN values, post-hoc testing was conducted using the LSMeans Tukey's HSD test.

3. Results and Discussion

3.1 Growth Profiles and Protein Expression

The overall objective of these studies was to characterize the dynamics of gene expression variability in *E. coli* due to insoluble and soluble protein production and specifically to examine the RIN values. The pTVP1GFP and pGFPCAT plasmids and the VP1GFP and GFPCAT proteins were selected to minimize the differences between the culture conditions, except for protein solubility. Specifically, the reasons for the selection of these two systems are: 1) VP1GFP is a well-characterized inclusion body-prone protein with fluorescence in both insoluble and soluble conformations [15, 48]; 2) CAT and GFP are both well-characterized soluble proteins [53, 55, 61–64]; 3) the two fusion proteins (GFPCAT and VP1GFP) have similar molecular weights; and 4) both plasmids are very similar with a common lineage (*e.g.*, ampicillin resistance, pBR322 origin, *lacI* expression, and *trc* promoter) [15, 65]. Thus, any differences observed in the RNA electropherogram profiles and RIN values can be attributed to the protein solubility.

In order to evaluate RIN values for the different culture conditions, total RNA was isolated from synchronized cultures expressing either the soluble protein GFPCAT or the inclusion body-prone protein VP1GFP. The growth profiles for these cultures are shown in Figure 1 for triplicate cultures. Each culture condition was conducted in triplicate and the culture times have been aligned to the induction time, which corresponds to a cell density of approximately 0.5 OD and mid-exponential growth. The samples used for the total RNA isolation were taken just prior to induction (time 0-minutes) and 5-, 20-, 40-, and 60-minutes

post-induction. The time 0-minute represents uninduced cultures only. Additionally, the uninduced cultures were sampled at 60-minutes past the synchronization cell density. The pre-induction and post-induction growth rates were analyzed using the GLM method. It was determined that the growth rates were not different for the cultures due to the protein produced (VP1GFP and GFPCAT) or the induction state (uninduced and induced) ($p > 0.05$). The exponential growth rate model fit for all cultures prior to and up to 1-hour post-induction were $0.568 \pm 0.013 \text{ h}^{-1}$ (standard error), and is highlighted in Figure 1 by the overlapping exponential model line fits, one for each of the four culture conditions examined (uninduced GFPCAT, induced GFPCAT, uninduced VP1GFP, and induced VP1GFP). Additionally, induction with 1 mM IPTG did not significantly alter the growth rate of the cultures ($p > 0.05$). In minimal medium, it is commonly observed that induction does not alter the culture growth rate [50, 53]. In contrast, in richer medium decreased growth rates have been observed at the time of induction [66]. In this study, the growth rates and cell densities were not different for the cultures producing either the GFPCAT (soluble) or VP1GFP (insoluble) protein for any of the analyzed time points.

The CAT activity of the GFPCAT protein was quantified (Figure 1) and shows an approximately 8-fold increase in protein enzyme activity by 60-minutes post-induction and an approximately 20-fold increase at 4-hours post-induction. Additionally, induced GFPCAT cells grew on chloramphenicol (0.61 mM) LB plates, whereas no growth was observed for uninduced cells (data not shown). The VP1GFP protein production was quantified by fluorescence intensity per cell using flow cytometry (Figure 1). Flow cytometry analysis of the induced VP1GFP cultures showed the fluorescent intensity increased approximately 8-fold at 60-minutes post-induction and approximately 30-fold at 3.5-hours post-induction. Fluorescence microscopy showed the VP1GFP protein was localized in the flagella end of the cells, which was also reported and characterized by Dr. Villaverde's research group as inclusion bodies [57]. The induction levels for both GFPCAT and VP1GFP were similar under the experimental conditions, which was expected for these very similar plasmid constructs.

3.2 Total RNA Isolation and Characterization

The cell pellet samples were harvested from synchronized cultures and were processed in parallel. To obtain total RNA from the cell pellets, a set of samples representing all the triplicate conditions was also processed in parallel. Thus, one set of the uninduced and induced, and GFPCAT and VP1GFP conditions were processed in parallel from the synchronized culture for total RNA analysis. Once the total RNA was obtained, the isolated RNA was quantified using a Nanodrop spectrophotometer. The 260/280 nm and the 260/230 nm ratios demonstrated that cellular contaminants had been sufficiently removed (> 2.0) and that the purity was sufficient (2.0 to 2.2), respectively, for all samples with no statistical differences [67, 68]. The isolated total RNA was then examined using the Agilent Bioanalyzer for the biological triplicates, again grouped to include at least one complete set of the 12 conditions. Representative electropherograms obtained using the Agilent Bioanalyzer are shown for VP1GFP cultures in Figure 2; uninduced 0-minute and induced 5-minute post-induction. The uninduced VP1GFP culture had a RIN value of 7.7, which is considered acceptable [45], while the induced culture had a much lower RIN value of 4.8 [37, 39].

The RIN algorithms were entirely developed using 1208 eukaryotic data set, where the 5S, 18S and 28S peaks were used [37]. For prokaryotes, the RIN values are calculated from the entire electrophoretic trace with weighting given to the 1) total RNA ratio, 2) 23S peak height, 3) 23S area ratio, 4) comparison of the 16S and 23S area to the fast region area, 5) a linear regression of the fast region end point, 6) detected fragment amounts in the fast region, 7) the presence or absence of the 16S peak, and 8) a comparison of the overall mean

value to the median value [37]. The fast region is defined as the area between the 5S and 18S peaks; however, for *E. coli* a 5S peak is not detected, and thus not used in the calculation; and the 16S peak replaces the 18S peak [45]. Therefore, for *E. coli*, the fast region is between the dye front and the 16S peaks and includes the elution times between approximately 25 and 40 seconds. Due to the parallel processing for all samples, including sample harvesting, sample protection, sample storage, and RNA isolation procedures, this anomaly in the RIN values could not be attributed to ribosome degradation alone during processes. A statistically analysis was conducted to determine if the lower RIN values could be attributed to protein solubility.

To evaluate the RIN values statistically, the RIN values for the uninduced time 0- and 60-minutes samples as well as the induced 5-, 20-, 40-, and 60-minute samples were used. The average RIN values with standard error are shown for the VP1GFP and GFPCAT cultures in Figure 3. A statistical analysis was conducted to determine the significance of the RIN values using the generalized linear model (GLM) method ($p \leq 0.05$) with the JMP 10 software (SAS Institute Inc.). Both the protein and the induction state were determined to be statistically significant for the RIN values. Post-hoc analysis using the least square mean differences with Tukey HSD determined that the uninduced culture RIN values, regardless of the protein encoded by the plasmid, were not significant ($p > 0.05$). The post-hoc analysis also determined that the induced GFPCAT culture RIN values were not significantly different from the uninduced cultures ($p > 0.05$). However, the VP1GFP induced culture RIN values were determined to be statistically different ($p \leq 0.05$) from the uninduced cultures and the induced GFPCAT cultures. Additionally, the observed decrease in RIN values was at 5-minute post-induction for the VP1GFP cultures and had no time dependence up to 60-minutes post-induction ($p > 0.05$). This rapid RIN value decrease followed by no change in RIN values would indicate a counterbalancing cellular response allowing for stabilized overall RNA integrity.

Ribosomes are composed of both protein and RNA, and ribosome abundance has been observed to decrease due to recombinant protein production in *E. coli* [66, 69–73]. Gene expression analyses of *E. coli* producing insoluble versus soluble proteins have demonstrated higher levels of the heat shock response genes [27, 31]. And, it is well known that the heat-shock response activates protease activity in response to insoluble recombinant protein production [27, 64, 74], where recombinant protein production has been observed to have elevated protease activity as well [75]. Increased protease activity toward the ribosomes could cause the lower RIN values for the cultures producing the insoluble protein. A counteracting cellular response that could increase ribosome abundance would be increased ribosomal subunit gene expression.

In this study, 25 of the 55 ribosomal subunits genes (*rplACDEFILMNQRX*, *rpmDII*, and *rpsABFGHLMNPU*) had statistically significant gene expression levels ($p \leq 0.10$). The dynamic behavior for all 25 ribosomal subunit genes was coordinated across the 12 samples. The average ribosomal subunit gene expression profiles are shown in Figure 3 for the uninduced and induced GFPCAT and VP1GFP cultures (Supplemental Data has the profiles for all 25 ribosomal subunit genes by culture condition). The ribosomal subunit gene expressions levels increased linearly for the GFPCAT cultures due to the protein induction ($p \leq 0.05$), and remained unchanged for both the uninduced VP1GFP and GFPCAT cultures ($p > 0.05$). For the induced VP1GFP cultures, the 25 ribosomal subunit gene expression levels at 60-minutes post-induction were higher than any other condition ($p \leq 0.05$). As expected, the housekeeping genes (*mdoG* and *crp*) [76] were not significantly affected ($p > 0.10$) by the conditions. The RIN and ribosomal subunit gene expression profiles support the hypothesis that VP1GFP production causes increased ribosome degradation due to heat

shock response proteases, and these negative effects are counterbalanced by increased ribosomal subunit gene expression levels, resulting in a plateau in the RIN values.

4. Conclusions

Recombinant *E. coli* prior to induction of a target protein has RIN values that indicate acceptable RNA integrity for gene expression analysis. Induction of an insoluble protein resulted in lower RIN values compared to soluble protein induction, as well as to the uninduced cultures. The observed RIN values for the induced insoluble protein cultures would normally indicate ribosome degradation during the isolation and purification steps; however, since parallel processing was used for the induced soluble protein and the uninduced cultures, the apparent ribosome degradation most likely occurred in the cells in response to the insoluble protein production. Specifically, increased protease activity due to a heat shock response caused a rapid decrease in the RIN value, indicative of ribosome degradation. Interestingly, the RIN values stabilized, suggesting a counterbalancing cellular response, which is likely due to increased ribosomal subunit gene expression levels. Since the DNA microarray community uses the RIN values to assess RNA integrity as a critical first step in gene expression analysis, and because there is limited data for both wild-type and recombinant *E. coli* under various culture conditions, these widely different RIN values observed between the induced insoluble protein cultures and both the induced soluble protein and the uninduced cultures indicate that further analysis is needed to determine the effects of various culture conditions on RIN values for recombinant *E. coli*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The pTVP1GFP plasmid was generously provided by E. Garcia-Fruitos and A. Villaverde, Universitat Autònoma de Barcelona. The pGFPCAT plasmid was constructed by M. T. Morris, Clemson University. The pTrcHis-GFP_{UV}/CAT plasmid was donated by W. E. Bentley, University of Maryland. The authors also would wish to thank N. Vyavahare, Clemson University, for the use of the Agilent Bioanalyzer 2100; Dr. Terri Bruce of the Clemson Light Imaging Facility for microscopy assistance; and Arthur Nathan Brodsky for reviewing the manuscript. Funding was provided by the National Science Foundation under NSF Award CBET 0738162 and by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103444.

References

- [1]. Demain AL, Vaishnav P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* 2009; 27:297–306. [PubMed: 19500547]
- [2]. Walsh G. Biopharmaceutical benchmarks 2010. *Nat. Biotechnol.* 2010; 28:917–924. [PubMed: 20829826]
- [3]. Swartz JR. Advances in *Escherichia coli* production of therapeutic proteins. *Curr. Opin. Biotechnol.* 2001; 12:195–201. [PubMed: 11287237]
- [4]. Anselment B, Schoemig V, Kesten C, Weuster-Botz D. Statistical vs. Stochastic experimental design: An experimental comparison on the example of protein refolding. *Biotechnol. Prog.* 2012; 28:1499–1506. [PubMed: 22961887]
- [5]. Hoffmann F, Rinas U. Plasmid amplification in *Escherichia coli* after temperature upshift is impaired by induction of recombinant protein synthesis. *Biotechnol. Lett.* 2001; 23:1819–1825.
- [6]. Sahdev S, Khattar SK, Saini KS. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry.* 2008; 307:249–264. [PubMed: 17874175]

- [7]. Lim WK, Smith-Somerville HE, Hardman JK. Solubilization and renaturation of overexpressed aggregates of mutant tryptophan synthase α -subunits. *Appl. Environ. Microbiol.* 1989; 55:1106–1111. [PubMed: 2547334]
- [8]. Ventura S, Villaverde A. Protein quality in bacterial inclusion bodies. *TRENDS in Biotechnology.* 2006; 24:179–185. [PubMed: 16503059]
- [9]. Carrio M, Villaverde A. Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. *J. Bacteriol.* 2005; 187:3599–3601. [PubMed: 15866952]
- [10]. Idicula-Thomas S, Balaji PV. Protein aggregation: a perspective from amyloid and inclusion-body formation. *Curr. Sci.* 2007; 92:758–767.
- [11]. Middelberg APJ. Preparative protein folding. *TRENDS in Biotechnology.* 2002; 20:437–443. [PubMed: 12220907]
- [12]. Medwid RD, Krebs L, Welch S. Evaluation of *Escherichia coli* cell disruption and inclusion body release using nucleic acid binding fluorochromes and flow cytometry. *BioTechniques.* 2007; 43:777–781. [PubMed: 18251254]
- [13]. Bade PD, Kotu SP, Rathore AS. Optimization of a refolding step for a therapeutic fusion protein in the quality by design (QbD) paradigm. *J. Sep. Sci.* 2012; 35:3160–3169. [PubMed: 23086665]
- [14]. 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
- [15]. 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*. *J. Mol. Biol.* 2007; 374:195–205. [PubMed: 17920630]
- [16]. Fahner, B. Using Folding Promoting Agents in Recombinant Protein Production: A Review. In: Lorence, A., editor. *Recombinant Gene Expression: Reviews and Protocols.* Third Edition. Humana Press Inc; Totowa: 2012. p. 3-36.
- [17]. Gonzalez-Montalban N, Carrio MM, Cuatrecasas S, Aris A, Villaverde A. Bacterial inclusion bodies are cytotoxic *in vivo* in absence of functional chaperones DnaK or GroEL. *J. Biotechnol.* 2005; 118:406–412. [PubMed: 16024126]
- [18]. Jhamb K, Sahoo DK. Production of soluble recombinant proteins in *Escherichia coli*: Effects of process conditions and chaperone co-expression on cell growth and production of xylanase. *Bioresour. Technol.* 2012; 123:135–143. [PubMed: 22940310]
- [19]. Guenther I, Zolkiewski M, Kedzierska-Mieszkowska S. Cooperation between two ClpB isoforms enhances the recovery of the recombinant beta-galactosidase from inclusion bodies. *Biochem. Biophys. Res. Commun.* 2012; 426:596–600. [PubMed: 22982305]
- [20]. Kyrtatous, CA.; Panagiotidis, CA. Heat-Shock Protein Fusion Vectors for Improved Expression of Soluble Recombinant Proteins in *Escherichia coli*. In: Lorence, A., editor. *Recombinant Gene Expression: Reviews and Protocols.* Third Edition. 2012. p. 109-129.
- [21]. Angov E. Codon usage: Nature's roadmap to expression and folding of proteins. *Biotechnol J.* 2011; 6:650–659. [PubMed: 21567958]
- [22]. Sorensen HP, Mortensen KK. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* 2005; 115:113–128. [PubMed: 15607230]
- [23]. Tyedmers J, Mogk A, Bukau B. Cellular strategies for controlling protein aggregation. *Nature Reviews Molecular Cell Biology.* 2010; 11:777–788.
- [24]. Song JM, An YJ, Kang MH, Lee YH, Cha SS. Cultivation at 6–10 degrees C is an effective strategy to overcome the insolubility of recombinant proteins in *Escherichia coli*. *Protein Expression and Purification.* 2012; 82:297–301. [PubMed: 22333528]
- [25]. Weickert MJ, Pagratis M, Curry SR, Blackmore R. Stabilization of apoglobin by low temperature increases yield of soluble recombinant hemoglobin in *Escherichia coli*. *Appl. Environ. Microbiol.* 1997; 63:4313–4320. [PubMed: 9361418]
- [26]. Weiss WF, Young TM, Roberts CJ. Principles, Approaches, and Challenges for Predicting Protein Aggregation Rates and Shelf Life. *Journal of Pharmaceutical Sciences.* 2009; 98:1246–1277. [PubMed: 18683878]

- [27]. 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; 15:153–160. [PubMed: 11917152]
- [28]. Smialowski P, Martin-Galiano AJ, Mikolajka A, Girschick T, Holak TA, Frishman D. Protein solubility: sequence based prediction and experimental verification. *Bioinformatics*. 2007; 23:2536–2542. [PubMed: 17150993]
- [29]. Smialowski P, Doose G, Torkler P, Kaufmann S, Frishman D. PROSO II - a new method for protein solubility prediction. *FEBS J*. 2012; 279:2192–2200. [PubMed: 22536855]
- [30]. Agostini F, Vendruscolo M, Tartaglia GG. Sequence-Based Prediction of Protein Solubility. *J. Mol. Biol.* 2012; 421:237–241. [PubMed: 22172487]
- [31]. Smith H. The transcriptional response of *Escherichia coli* to recombinant protein insolubility. *J Struct Funct Genomics*. 2007; 8:27–35. [PubMed: 17992580]
- [32]. Jurgen B, Lin HY, Riemschneider S, Scharf C, Neubauer P, Schmid R, Hecker M, Schweder T. Monitoring of genes that respond to overproduction of an insoluble recombinant protein in *Escherichia coli* glucose-limited fed-batch fermentations. *Biotechnol. Bioeng*. 2000; 70:217–224. [PubMed: 10972933]
- [33]. Bachoon DS, Chen F, Hodson RE. RNA recovery and detection of mRNA by RT-PCR from preserved prokaryotic samples. *FEMS Microbiol. Lett*. 2001; 201:127–132. [PubMed: 11470350]
- [34]. Ling D, Salvaterra PM. Robust RT-qPCR Data Normalization: Validation and Selection of Internal Reference Genes during Post-Experimental Data Analysis. *Plos One*. 2011; 6
- [35]. Dreyfus M, Regnier P, Superieure EN. The poly(A) tail of mRNAs: Body guard in eukaryotes, scavenger in bacteria. *Cell*. 2002; 111:611–613. [PubMed: 12464173]
- [36]. Sambrook, J.; Fritsch, EF.; Maniatis, T. *Molecular Cloning A Laboratory Manual*. 2nd ed.. Cold Spring Harbor Laboratory Press; 1989.
- [37]. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassman M, Lightfoot S, Menzel W, Granzow M, Ragg T. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol*. 2006; 7
- [38]. Lehrach H, Diamond D, Wozney JM, Boedker H. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry*. 1977; 16:4743–4751. [PubMed: 911786]
- [39]. Imbeaud S, Graudens E, Boulanger V, Barlet X, Zaborski P, Eveno E, Mueller O, Shroeder A, Auffrey C. Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res*. 2005; 33:e56, 51–12. [PubMed: 15800207]
- [40]. Jahn CE, Charkowski AO, Willis DK. Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *J. Microbiol. Methods*. 2008; 75:318–324. [PubMed: 18674572]
- [41]. Kiewe P, Gueller S, Komor M, Stroux A, Thiel E, Hofmann WK. Prediction of qualitative outcome of oligonucleotide microarray hybridization by measurement of RNA integrity using the 2100 Bioanalyzer (TM) capillary electrophoresis system. *Annals of Hematology*. 2009; 88:1177–1183. [PubMed: 19424697]
- [42]. Auer H, Lyianarachchi S, Newsom D, Klisovec MI, Marcucci G, Kornacker K. Chipping away at the chip bias: RNA degradation in microarray analysis. *Nat. Genet*. 2003; 35:292–293. [PubMed: 14647279]
- [43]. Haller AC, Kanakapelli D, Walter R, Alhasan S, Eliason JF, Everson RB. Transcriptional profiling of degraded RNA in cryopreserved and fixed tissue samples obtained at autopsy. *BMC Clinical Pathology*. 2006; 6
- [44]. Copois V, Bibeau F, Bascoul-Mollevis C, Salvatet N, Chalbos P, Barell C, Candeil L, Fraslon C, Conseiller E, Granci V, Maziere P, Kramar A, Ychou M, Pau B, Martineau P, Molina F, Del Rio M. Impact of RNA degradation on gene expression profiles: Assessment of different methods to reliably determine RNA quality. *J. Biotechnol*. 2007; 127:549–559. [PubMed: 16945445]
- [45]. Kjeldgaard J, Henriksen S, Cohn MT, Aabo S, Ingmer H. Method Enabling Gene Expression Studies of Pathogens in a Complex Food Matrix. *Appl. Environ. Microbiol*. 2011; 77:8456–8458. [PubMed: 21984236]

- [46]. Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ, Livny J, Earl AM, Gevers D, Ward DV, Nusbaum C, Birren BW, Gnirke A. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol.* 2012; 13
- [47]. 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 Immunol.* 2006; 19:202–209. [PubMed: 16817763]
- [48]. Garcia-Fruitos E, Gonzalez-Montalban N, Morell M, Vera A, Ferraz RM, Aris A, Ventura S, Villaverde A. Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microbial Cell Factories.* 2005; 4
- [49]. Cha HJ, Wu CF, Valdes JJ, Rao G, 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; 67:565–574. [PubMed: 10649231]
- [50]. Sharma SS, Blattner FR, Harcum SW. Recombinant protein production in an *Escherichia coli* reduced genome strain. *Metab. Eng.* 2007; 9:133–141. [PubMed: 17126054]
- [51]. Korz DJ, Rinas U, Hellmuth K, Sanders EA, Decker WD. Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. *J. Biotechnol.* 1995; 39:59–65. [PubMed: 7766011]
- [52]. Rodriguez, RL.; Tait, RC. Recombinant DNA Techniques: An Introduction. Addison-Wesley Publishing Company; Reading, Massachusetts: 1983.
- [53]. Sharma SS, Campbell JW, Frisch D, Blattner FR, Harcum SW. Expression of two recombinant chloramphenicol acetyltransferase variants in highly reduced genome *Escherichia coli* strains. *Biotechnol. Bioeng.* 2007; 98:1056–1070. [PubMed: 17497738]
- [54]. García-Fruitós E, Martínez-Alonso M, González-Montalbán N, Valli M, Mattanovich D, Villaverde A. Divergent Genetic Control of Protein Solubility and Conformational Quality in *Escherichia coli*. *J. Mol. Biol.* 2007; 374:195–205. [PubMed: 17920630]
- [55]. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. *Science.* 1994; 263:802–805. [PubMed: 8303295]
- [56]. Ward, WW. Green Fluorescent Protein (2nd Edition). 2 ed.. John Wiley & Sons, Inc.; 2005.
- [57]. Garcia-Fruitos E, Aris A, Villaverde A. Localization of functional polypeptides in bacterial inclusion bodies. *Appl. Environ. Microbiol.* 2007; 73:289–294. [PubMed: 17085715]
- [58]. Workman C, Jensen L, Jarmer H, Berká R, Gautier L, Nielser H, Saxild H-H, Nielsen C, Brunak S, Knudsen S. A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol.* 2002; 3:research0048.0041–research0048.0016. [PubMed: 12225587]
- [59]. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 2003; 4:249–264. [PubMed: 12925520]
- [60]. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics.* 2003; 19:185–193. [PubMed: 12538238]
- [61]. Shaw WV. Chloramphenicol Acetyltransferase: Enzymology and Molecular Biology. *Critica Reviews in Biochemistry.* 1983; 14:1–35.
- [62]. Cohen JD, Eccleshall TR, Needleman RB, Federoff H, Buchferer BA, Marmur J. Functional expression in yeast of the *Escherichia coli* plasmid gene coding for chloramphenicol acetyltransferase. *Proceedings of the National Academy of Sciences of the United States of America.* 1980; 77:1078–1082. [PubMed: 6987664]
- [63]. Kim SJ, Jeon HY, Kim HB. Chloramphenicol acetyltransferase expression of *Escherichia coli* is increased at 42 degrees C. *Biotechnol. Tech.* 1997; 11:435–438.
- [64]. 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*. *J. Bacteriol.* 1992; 174:6938–6947. [PubMed: 1356969]
- [65]. Chae HJ, DeLisa MP, Cha HJ, Weigand WA, Rao G, Bentley WE. Framework for online optimization of recombinant protein expression in high-cell-density *Escherichia coli* cultures using GFP-fusion monitoring. *Biotechnol. Bioeng.* 2000; 69:275–285. [PubMed: 10861407]

- [66]. Haddadin FT, Harcum SW. Transcriptome profiles for high-cell-density recombinant and wild-type *Escherichia coli*. *Biotechnol. Bioeng.* 2005; 90:127–153. [PubMed: 15742388]
- [67]. Short protocols in protein science: a compendium of methods from Current protocols in protein science. John Wiley & Sons, inc.; 2003.
- [68]. Baelde HJ, Cleton-Jansen A-M, van Beerendock H, Namba M, Bovee JVMG, Hogendoorn PCW. High quality RNA isolation from tumours with low cellularity and high extracellular matrix component for cDNA microarrays: application to chondrosarcoma. *J. Clin. Pathol.* 2001; 54:778–782. [PubMed: 11577126]
- [69]. Borth N, Mitterbauer R, Mattanovich D, Kramer W, Batyer K, Katinger H. Flow cytometric analysis of bacterial physiology during induction of foreign protein synthesis in recombinant *Escherichia coli* cells. *Cytometry.* 1998; 31:125–129. [PubMed: 9482281]
- [70]. Wood TK, Peretti SW. Depression of protein synthetic capacity during cloned-gene expression in *E. coli*. *Biotechnol. Bioeng.* 1990; 36:865–878. [PubMed: 18597287]
- [71]. Wood TK, Peretti SW. Effect of chemically-induced, cloned-gene expression on protein synthesis in *Escherichia coli*. *Biotechnol. Bioeng.* 1991; 38:397–412. [PubMed: 18600776]
- [72]. Richins R, Chen W. Effects of FIS overexpression on cell growth, rRNA synthesis, and ribosome content in *Escherichia coli*. *Biotechnol. Prog.* 2001; 17:252–257. [PubMed: 11312701]
- [73]. Li X, Taylor KB. Effect of glucose on the expression parameters of recombinant protein in *Escherichia coli* during batch growth in complex medium. *Biotechnology Prog.* 1994; 10:160–164.
- [74]. 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, M.; Umbarger, HE., editors. *Escherichia coli and Salmonella*. ASM Press; Washington D. C.: 1996. p. 1382-1399.
- [75]. Harcum SW, Bentley WE. Protease detection, characterization, and quantification in recombinant *Escherichia coli*. *Biotechnol. Tech.* 1993; 7:441–447.
- [76]. Ma Q, Yin Y, Schell MA, Zhang H, Li G, Xu Y. Computational analyses of transcriptomic data reveal the dynamic organization of the *Escherichia coli* chromosome under different conditions. *Nucleic Acids Res.* 2013

Highlights

- Recombinant *E. coli* gene expression analysis requires high quality RNA
- RNA Integrity Number (RIN) evaluates RNA quality for gene expression analysis
- Soluble protein production does not affect RIN values relative to uninduced cells
- Insoluble protein production results in low RIN values
- Further analysis of effects on RIN values are needed for recombinant *E. coli*

Figure 1.

Growth profiles and protein levels for synchronized *E. coli* cultures. A) *E. coli* pTVP1GFP cultures uninduced (○) and induced (□); B) *E. coli* pGFPCAT cultures uninduced (○) and induced (□). The solid and dashed lines represent the exponential growth rate model fits for the induced and uninduced cultures, respectively, where all cultures had growth rates equal to 0.55 h^{-1} . C) CAT activity for GFPCAT (lighter bars) and fluorescence for VP1GFP (darker bars). Error bars represent 95% confidence intervals.

Figure 2. Representative electropherograms for total RNA samples using the Agilent Bioanalyzer. A) *E. coli* pTVP1GFP uninduced at time 0-minute; and B) *E. coli* pTVP1GFP 5-minutes post-induction. The electropherogram peaks include the dye front (dye), 16S rRNA (16S), and 23S rRNA (23S). The RIN values and RNA sample concentrations are indicated for each sample. Peak intensities are shown as fluorescence units (FU).

Figure 3.

RNA Integrity Number (RIN) values, and ribosomal subunit and housekeeping gene expression levels. A) RIN values; B) Normalized gene expression for the average of the 25 ribosomal subunit genes (*rplACDEFJLMNQRX*, *rpmDII*, and *rpsABFGHLMNPU*); and C) Normalized gene expression for two housekeeping genes, *mdoG* and *crp*. *E. coli* pTVP1GFP cultures uninduced (○) and induced (●); and *E. coli* pGFPCAT cultures uninduced (○) and induced (●). Error bars represent the standard error.