ABSTRACT

Optimization of Recombinant *E. coli* Protein Expression in Growth Medium

**Abstract:** With the onset of novel genomic sequence data, much effort is now being taken to discover the role of many genes and the corresponding proteins that remain unknown. One of the limiting steps in such efforts has been the production of recombinant proteins in *E. coli* using bacterial expression systems. In this current work, several studies were initiated to address the issues affecting recombinant protein expression. These studies specifically focus on the aspects of protein expression and solubility levels, bacterial growth medium design, and expression vector design. While earlier attempts were aimed at comparing commercial expression systems (T5 and T7 RNA polymerase-based systems), later work on vector design showed that vectors with lower lac repressor dosing were optimal for the expression of higher levels of protein in *E. coli*. Similar collaborative work focused on a factorial-designed growth medium experiment. Results showed optimal growth medium conditions for the expression of two model proteins, *Photinus* luciferase and enhanced green fluorescent protein (eGFP). Our current work is being directed at the comparison of expression levels and solubility for proteins expressed in vectors containing the Trigger Factor versus maltose binding protein fusion proteins.
COVER SHEET

TITLE: Optimization of Recombinant *E. coli* Protein Expression in Growth Medium

AUTHOR’S NAME: Petar Duvnjak

MAJOR: Biochemistry

DEPARTMENT: Biochemistry

MENTOR: Professor Brian G. Fox

DEPARTMENT: Biochemistry

MENTOR(2): N/A

DEPARTMENT(2): N/A

YEAR: 2007

(The following statement must be included if you want your paper included in the library’s electronic repository.)

The author hereby grants to University of Wisconsin-Madison the permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.
Optimization of Recombinant *E. coli* Protein

Expression in Growth Medium

*Petar Duvnjak*

College of Agricultural and Life Sciences

Honors in Research

Senior Honors Thesis

Laboratory of Dr. Brian G. Fox

Department of Biochemistry, University of Wisconsin
433 Babcock Drive, Madison, WI 53706

Spring, 2007
Abstract

With the onset of novel genomic sequence data, much effort is now being taken to discover the role of many genes and the corresponding proteins that remain unknown. One of the limiting steps in such efforts has been the production of recombinant proteins in \textit{E. coli} using bacterial expression systems. In this current work, several studies were initiated to address the issues affecting recombinant protein expression. These studies specifically focus on the aspects of protein expression and solubility levels, bacterial growth medium design, and expression vector design. While earlier attempts were aimed at comparing commercial expression systems (T5 and T7 RNA polymerase-based systems), later work on vector design showed that vectors with lower \textit{lac} repressor dosing were optimal for the expression of higher levels of protein in \textit{E. coli}.

Similar collaborative work focused on a factorial-designed growth medium experiment. Results showed optimal growth medium conditions for the expression of two model proteins, \textit{Photinus} luciferase and enhanced green fluorescent protein (eGFP). Our current work is being directed at the comparison of expression levels and solubility for proteins expressed in vectors containing the Trigger Factor versus maltose binding protein fusion proteins.
Introduction

Methods for the production of recombinant proteins in *E. coli* have become of great importance for many biological applications and fields of research. One such developing field is structural genomics, such as at the Center for Eukaryotic and Structural Genomics (CESG) where novel structural properties of proteins are solved through X-ray crystallography or nuclear magnetic resonance (1, 2). With growing knowledge of genomic DNA sequences from many model organisms, many efforts are currently being directed towards structural studies of novel gene products. An understanding for the role of such proteins can provide a path for the elucidation of information on human disease and their potential cures (3). To date, CESG has solved approximately 100 structures from model eukaryotes, and is currently working with over 100 potential targets involved with the differentiation of human embryonic stem cells (http://www.uwstructuralgenomics.org).

One of the greatest limitations on the success of structural genomics efforts is the efficiency of producing proteins to study. Conventionally, such efforts utilize either cell-free or *E. coli* based methods to express target genes of interest, followed by protein purification steps. This work focuses on the optimization of protein production using *E. coli* based methods. The general protocol for this process encompasses the following steps, each of which presents itself as a potential limiting step. First, target genes are cloned into a suitable bacterial expression plasmid, and then transformed into a proper *E. coli* expression host. Second, cells are grown in a medium that induces the expression of the target gene. Last, the cell cultures are harvested via centrifugation, the cells lysed, and the target proteins are purified from the cell lysates.

Previous work at CESG has been done to improve the efficiencies of this process at certain steps. A comparison of two cloning methods, Gateway and Flexi Vector, highlighted
some advantages of the Flexi Vector method over the more established Gateway method. Benefits included a faster and more cost-effective protocol, fewer PCR steps leading to fewer mutations introduced in target genes, and higher gene transfer efficiencies between vectors (4).

Two years ago, my efforts were aimed at the comparison of two commercially available bacterial expression systems: pQE (Qiagen) and pET (Novagen). The majority of expression plasmids utilized by CESG are pQE-based, which use a strong T5 phage promoter to drive protein expression. While the T5 promoter is recognized by an endogenous *E. coli* RNA polymerase, gene expression is regulated through two copies of the *lac* operator, which are binding sites for the repressor protein from the bacterial *lac* operon (5). In contrast, the T7 promoter is recognized by a highly specific viral T7 RNA polymerase, which is commonly introduced into the genome of an engineered *E. coli* expression strain. The engineered strains, called (DE3) lysogens, express the viral T7 RNA polymerase under the control of the *lac* operator, and then the recombinant polymerase recognizes the T7 promoter in the expression plasmid to initiate recombinant protein expression. In the T7 promoter system, gene expression can be regulated by a copy of the *lac* operator in the genome, which controls the expression of the T7 RNA polymerase (6).

The results from these earlier efforts indicated that the T7-based expression system was more suitable for expressing more target protein as indicated by luminescence activity of the *Photinus* luciferase reporter gene.

Recently, work by Blommel *et al* (7) addressed this question regarding plasmid type. Commercial T5-based vectors (pQE) contain a mutant *lacI* gene promoter that causes approximately ten times the amount of wild-type *lac* repressor found in *E. coli* to be expressed. In taking advantage of this, Blommel *et al* constructed a T5-based vector containing wild type, low dosage copy of the *lacI* gene, and conducted a comparison of *Photinus* luciferase expression in this vector with luciferase in the high-dosage *lacI*-T5 vector and in a T7 vector. As predicted,
the T7-based vector expressed higher levels of luciferase than the high-dosage lacI-T5 vector; however, the low-dosage lacI-T5 vector expressed ~4x higher levels of luciferase than the T7 vector (7). With these results, it has become clear that the low dosage lacI-T5-based expression vector is better suited for the high throughput expression of proteins in *E. coli*.

In addition to those issues affecting target gene expression already discussed, the requirement for a suitable bacterial growth medium is essential for the optimal expression of protein in *E. coli*. Recent data from the Fox lab has also shown that various growth media and conditions (such as pH, O₂ availability, and carbon sources) influence not only the growth of bacteria, but also the amount of target gene expression (7). With this work, it has been suggested that specific growth conditions can provide an optimum level target gene expression.

The use of various fusion proteins is a common practice in bacterial-based protein production. One frequent application for such tags is in aiding protein solubility in the *E. coli* cytoplasm by assisting in protein folding. Currently, CESG utilizes maltose binding protein (MBP) as an amino terminal fusion to serve this function (8, 9). Another solubility tag of interest is trigger factor, which is a bacterial ribosomal chaperone, and is known for its role as the first encountered molecular chaperone in the *E. coli* cytoplasm. Compared to MBP, trigger factor contains an amino terminal ribosomal binding domain, a middle peptidyl prolyl isomerase domain, and a C-terminal structural chaperone module (10). The crystal structure of trigger factor has shown that this chaperone forms a molecular cradle for nascent polypeptides on the ribosomal tunnel, thus shielding the polypeptide from cytoplasmic proteases (11).

While earlier work was targeted at the comparison of T5- and T7-containing expression vectors, this current work addresses the issues of bacterial growth medium composition and the properties of the trigger factor protein solubility tag. Collaborative work on a factorial design experiment showed optimal *E. coli* growth conditions for the high expression of enhanced green
fluorescent protein (eGFP). Results were interpreted to illustrate the dependence of eGFP expression on the medium composition (varying in glycerol, lactose, and glucose). Additionally, a novel protein expression vector (pVP70K) containing the trigger factor fusion protein was created. Current efforts are being taken to compare the expression and solubility of a subset of structural genomics targets from pVP70K and an MBP-containing expression vector, pVP65K. Preliminary results indicate that trigger factor may have a useful effect on protein expression distinct from maltose binding protein.
Material and Methods

Chemicals and Reagents:

General chemicals, bacterial growth reagents, antibiotics, and labware were from either Fisher (Pittsburgh, PA) or Sigma-Aldrich (St. Louis, MO). Gateway cloning reagents were from Invitrogen (Carlsbad, CA). The Qiagen Plasmid Miniprep Kit (Valencia, CA) was used for DNA plasmid extraction. The Bright-Glo luciferase assay system and all Flexi Vector cloning reagents were from Promega (Madison, WI). Primers for gene amplification were from Integrated DNA Technologies (Coraville, IA). The Tris-HCl precast gels used for denaturing gel electrophoresis were from Bio-Rad (Hercules, CA).

Bacterial Strains:

Escherichia coli Top10 competent cells used for DNA cloning procedures were from Invitrogen, and Escherichia coli B834 and B834 (DE3) strains used for the expression of plasmids containing either T5 or T7 promoters, respectively, were from EMD Biosciences/Novagen (Madison, WI). Top10 cells were plated onto LB-agarose plates with kanamycin, while B834 expression strains were plated onto MDAG non-inducing media with appropriate antibiotics.

Expression Vectors:

T5-lacI (aka pVP62K), T5 lacI9 (aka pVP61K), and T7-lacI (aka pET32) were used for the factorial-designed medium experiments in collaboration with Paul Blommel. LacI refers to the wild type dosing of lacI where lacI9 represents the mutated version of the lac repressor.
promoter with ~10x the LacI expression level. The vectors pVP61K and pVP62K were modified from pQE80 (Qiagen, Valencia, CA) as described by Blommel et al (7), and pET32 was available commercially from EMD Biosciences/Novagen. The vectors pVP65K and pVP70K were used in the current work on the trigger factor fusion protein: pVP65K contains a self-cleaving version of MBP from an incorporated gene for the tobacco vein mottling virus (TVMV) protease and a TVMV cleavage site (see supporting material, Figure S1-A); pVP70K was produced in this work by the insertion of a PCR-amplified BARCAT cassette (lethal barnase gene fused with a chloramphenicol acetyltransferase) from pVP65K into pVP66K using the restriction sites HindIII and PacI.

**Protein Targets:**

Enhanced green florescent protein (eGFP) was cloned into pVP61K, pVP62K, and pET32 as described by Blommel et al for work in the factorial experiment. Gene targets from CESG were introduced into pVP65K and pVP70K from destination clones (Workgroup 2123) in the Gateway vector pVP64A using a PCR with universal primers that incorporate SgfI and PmeI sites. The four target genes that were investigated were CESG cloned open reading frame (ORF) numbers 79976, 79978, 79979, and 79980, with their original source being the algae specie *Galderia sulphuraria*. The forward primer consisted of the sequence 5’GGG TCA TAC ACC TAC GTA CTT CCT CCA GTC C 3’ where the bold nucleotides represent the SgfI site. The reverse primer consisted of the sequence 5’ GTG TGA TTA AAC GTA CAA GAA AGC TGG GTG CTA where the bold nucleotides represent the PmeI site.
**Factorial Medium Design:**

A total of 53 *E. coli* expression growth medias that varied in the concentrations of three carbon sources: glucose, glycerol, and lactose were as described by Blommel et al (7).

**Protein Expression and Purification:**

*E. coli* inocula were grown overnight in 400 µL of non-inducing growth medium (MDAG) in 96-well growth blocks at ~25°C while shaking. Next, 400 µL of inducing small-scale expression medium were inoculated with 20 µL of the saturated overnight inoculum, and grown again overnight under the same conditions. To assess the expression, 400 µL of expression culture was mixed with 600 µL of 50mM Hepes, pH 7.5 and purified using a Maxwell automated purification system (Promega), which utilizes magnetic nickel particles to purify polyhistidine-containing proteins through affinity chromatography.

**Protein Analysis:**

Following an overnight growth, protein expression cultures were split into 100 µL aliquots and frozen at -80°C until further use. Frozen cell cultures were then thawed and combined with lysis buffer for sonication (20 mM Tris-HCl, pH 7.5, 20 mM NaCl, 3 kU/mL lysozyme, 0.7 U/mL benzonase, 0.3 mM triscarboxyethylphosphine, and 1 mM MgSO₄, final concentrations). Cell mixtures were incubated at 25 °C for 25 min, and then sonicated for 10-15 min on a Misonix plate sonicator (Farmington, NY). Lysed samples were then prepared for either SDS-PAGE analysis or a florescence assay for eGFP-containing plasmids. Cultures containing eGFP were diluted with a buffer (10 mM Tris-HCl, pH 7.5, 20 mM NaCl, 0.1 mg/mL acetylated bovine serum albumin) at a 1:16 dilution of lysate to buffer for optimal assay.
conditions. Florescence measurements were taken as described by Blommel et al (7). Cell cultures from the trigger factor work were lysed in the same manner, and total and soluble lysates were ran under denaturing gel electrophoresis conditions.
Results

Factorial Designed Growth Medium:

As described by Blommel et al., a five level, three factor factorial design approach was used to define 53 unique medium compositions with a total of 64 methionine containing medias and 32 selenomethionine containing medias. The concentrations (w/v) of glucose, glycerol, and lactose were varied from 0 to 0.1%, 0 to 1.2%, and 0 to 0.6%, respectively with all other medium constituents held constant. A representation of the growth medium arrangement within an 8 x 8 arrayed expression block is shown in the Supporting Material (Table S1).

Figure 1 shows a picture of a 96-well plated containing cellular eGFP lysates that were diluted as previously described. Each well corresponds to one of the factorial-designed growth mediums in which eGFP was expressed according to Table 1. Upon illumination with 340 nm light, the activity of eGFP is observed. It is important to note the variation of eGFP activity in between wells, as this represents a success of the factorial design method.

Figure 2 shows the response surfaces for the expression of eGFP in three different expression plasmids (pVP62K, pVP61K, and pET32) in growth medias containing either methionine (left side) or selenomethionine (right side). From this data, one can observe how medium composition alters expression levels. In the left-hand side (Figures 2A, 2C, and 2E), it can be observed that variation in the carbon-source composition results in a significant increase in the amount of eGFP expressed. In the poorest performing expression medium, eGFP was found to have a expression level near 100 µg/mL, whereas in the optimum growth medium, eGFP can be expressed at a level near 1500 µg/mL , which represents an ~15-fold increase. An SDS-PAGE analysis of the levels for eGFP from the optimal medium condition (0.025% glucose, 0.9% glycerol, 0.45% lactose) is shown in Figure 3. Expressed with the N-terminal
fusion MBP, eGFP in pVP62K shows high levels of total and soluble expression following self-cleavage from the TVMV protease present in pVP62K.

Figure 1: The expression of eGFP-containing lysates can be indicated by the varying florescence levels at 340 nm. The compositions of growth medium within each well correlates to those listed in Table S1.
Figure 2: Response surfaces showing the activity of eGFP in the various factorial-designed medium compositions. A and B, expression from the T5-lacI plasmid (pVP62K) in medium containing methionine (A) or selenomethionine (B). C and D, expression from the T5-lacIq plasmid (pVP61K) in medium containing methionine (C) or selenomethionine (D). E and F, expression from T7-lacI (pET32) plasmid in medium containing methionine (E) or selenomethionine (F). The response models were not extended to zero lactose concentration for A, E and F due to highly non-linear response with these medium compositions.
The trigger factor-containing expression plasmid (pVP70K) was constructed by introducing the BarCat cassette into the pVP66K vector. Using restriction digests with HindIII and PacI, the BarCat cassette and cut pVP65K fragments were generated as shown in Figure 4A.

Following a gel-purification of the pVP65K-recovered BarCat cassette and empty pVP66K plasmid, ligation of the BarCat cassette into pVP66K was accomplished using the HindIII and PacI restriction sites. The resulting product, pVP70K was produced and screened using the same restriction enzymes. Figure 4B shows the results from this digest screen.
Figure 4: Agarose gel electrophoresis (0.8%) for the digestion of pVP65K and pVP66K with HindIII and PacI (A), and analytical digest of pVP70K with HindIII and PacI (B). A, lanes 1 and 2 show the digestion products from pVP65K. The top band represents the empty vector while the bottom is the BarCat cassette. The bottom band was recovered from the gel. Lanes 3 and 4 show the digestion products from pVP66K. The top band represents the empty vector while the bottom represents the BarCat cassette containing the AIA-encoding sequence. The top band was recovered from the gel. B, digestion of pVP70K with HindIII and PacI yields the predicted fragments: empty vector (~6 kb) and the BarCat cassette (~1.4 kb) without AIA-encoding nucleotides.

Although pVP66K contains the trigger factor fusion tag, it also contains an undesirable artifact of Flexi cloning, namely a short AIA-encoding nucleotide sequence (SgfI restriction site) just upstream of the BarCat cassette and downstream of the TEV cleavage site. By introducing the BarCat cassette from pVP65K (which lacks a TEV cleavage site) into pVP66K, the AIA-encoding sequence of pVP66K can later be cleaved from the target protein by introduction of a
TEV cleavage site downstream of the SgfI site when cloning in target genes. Previous experiments have suggested that this AIA N-terminal fusion may affect protein behavior during crystallization trials, thus removal of this region provided the rationale for producing pVP70K. Figure 5 compares the nucleotide sequences for this region between pVP65K and pVP66K. The vector map and partial sequence for pVP70K can be seen in the supporting material section.

**Figure 5:** DNA sequences for the upstream regions of BarCat for pVP65K (A) and pVP66K (B). A, pVP65K contains a TVMV site and an His$_8$ tag sequence 5’ to the BarCat cassette. B, pVP66K contains a TVMV site, His$_8$ tag sequence, and a TEV protease site 5’ to the BarCat cassette. Note, the boxed region in B shows the AIA-encoding site 5’ to the SgfI site and BarCat cassette.
Following the construction of pVP70K, 53 target genes from CESG workgroup 2123 were inserted into the SgfI and PmeI sites, replacing the BarCat cassette. Target genes from workgroup 2123 in pVP64A were PCR-amplified using universal CESG primers, which generated SgfI and PmeI sites on the 5’ and 3’ ends of the target genes, respectively. Figure 6A shows these PCR products. Following the CESG-modified cloning protocol, both PCR products and vectors (pVP65K and pVP70K) were digested with SgfI and PmeI, and digested PCR products were then ligated into both pVP65K and pVP70K. From these ligations, the first 6 target genes in pVP65K and pVP70K were transformed into E. coli Top10 cells and plated onto kanamycin-containing LB-agar plates. From these transformants, 2 colonies from each plate were screened via colony PCR using primers specific to pVP65K and pVP70K. Figure 6B shows the results from this screen.
The above positive PCR screens were sequenced using the Big Dye sequencing reaction at the University of Wisconsin Biotechnology Center, and results yielded four targets from both pVP65K and pVP70K for possible expression and subsequent analysis. These four targets (each in pVP65K and pVP70K) were transformed into B834 pRARE2 E. coli cells, and expressed in Terrific Broth (TB) auto-induction medium for 24 hours. Growth cultures were subsequently harvested and purified using the Maxwell 16 automated purification system (Promega). Figure 7 shows SDS-PAGE results.

**Figure 6:** Agarose gel electrophoresis (0.8%) for target gene amplification (A) and a representative colony PCR for target-gene containing plasmids (B). A, lanes A2, A4, A5, and A6 represent orfs 79976, 79978, 79979, and 79980, respectively.
**Figure 7:** 10-20% SDS-PAGE results for the purified target proteins expressed from pVP65K and pVP70K without sonication (A) and with sonicated cell lysis (B). A, three of the four expressed target genes (79976, 79980, and 79979) show detectable protein levels following expression in TB auto-induction medium and Maxwell purification. Boxes represent expected protein band with expression from pVP65K on the left and pVP70K on the right. Target protein orf number is listed beneath each box. Two replicates of each target protein (expressed in different growth block wells) were ran. For all three cases, protein levels of the target gene is greater in pVP70K than in pVP65K. Two of the four target genes showed no detectable protein levels in pVP65K, but were detected in pVP70K. B, following cell lysis via sonication, Maxwell purification of the four expected proteins show similar results as without sonication.

<table>
<thead>
<tr>
<th>Expression Vector</th>
<th>ORF 79976</th>
<th>ORF 79978</th>
<th>ORF 79980</th>
<th>ORF 79979</th>
</tr>
</thead>
<tbody>
<tr>
<td>pVP16 (Gateway)</td>
<td>MMM</td>
<td>HHM</td>
<td>MMW</td>
<td>MMW</td>
</tr>
<tr>
<td>pVP33K (Flexi)</td>
<td>HHM</td>
<td>HHH</td>
<td>HMM</td>
<td>WMW</td>
</tr>
<tr>
<td>pVP65K</td>
<td>No purification</td>
<td>No purification</td>
<td>No purification</td>
<td>No purification</td>
</tr>
<tr>
<td>pVP70K</td>
<td>High purification</td>
<td>No purification</td>
<td>Medium purification</td>
<td>purification</td>
</tr>
</tbody>
</table>

**Table 1:** Small-scale expression screening results from pVP16 (a Gateway vector) and pVP33K (a Flexi vector) in TB plus glucose medium (Rows 1 and 2). Small-scale Maxwell purification yields following expression from pVP65K and pVP70K (Rows 3 and 4). Small-scale expression scores are listed for expression levels, solubility, and TEV cleavage as indicated by high (H), medium (M), or weak (W) scores. Small-scale purification yields are listed as no purification, low, medium, or high purification.
Discussion

In this work, efforts towards optimizing the expression and solubility of recombinant proteins in *E. coli* have been taken in a dual approach. From the factorial growth medium design experiments, the dependency of protein expression on carbon source availability can be seen. Figures 1 and 2 demonstrate optimal medium conditions for the expression of eGFP. From this work, several observations can be made. First, the dependence of protein expression on glucose, glycerol, and lactose concentrations in the growth medium suggests a strong correlation between growth medium constituents and protein expression. Statistical analysis on this data has shown that glycerol and lactose increase small-scale gene expression, while glucose negatively impacts small-scale expression (data not shown). This can be visualized by Figure 2, where wells containing no glucose but high glycerol and lactose (wells B6, C6, D6) show very high eGFP florescence, but decreases in glycerol and lactose results in decreased expression (wells E6, F6, G6). Secondly, eGFP expression was highest in T5-*lacI* (wildtype lacI dosing) plasmids, while T5-*lacI* plasmids (10x lacI dosing) showed the lowest levels of expression. This data confirms those in previous experiments at CESG. It is also important to note the significant difference in eGFP expression between mediums that contain methionine versus selenomethionine. In all cases (except T5-*lacI*), eGFP expression levels were higher in methionine based expression mediums. Since growth mediums must be supplemented with selenomethionine for the production of proteins in X-ray crystallographic studies, knowledge of this expression behavior in selenomethionine-supplemented mediums is important (12). This work shows that medium compositions containing selenomethione can also give high level expression of eGFP.

The basis for the observations in these experiments relies on an understanding of carbon-consumption patterns in auto-induction medium. Much work has been done in this field, and has revealed that the typical pattern of carbon consumption is as follows: glucose is consumed first
since it is the preferred carbon source by *E. coli*, then lactose, followed by glycerol (13-20). In auto-induction medium, expression induction corresponds with the consumption of lactose, since lactose hydrolysis occurs following the expression of genes in the *lac* operon. Additionally, high glucose levels have the effect of preventing target gene expression from the *lac* operon, since glucose flux in the glucose metabolic pathways result in low levels of cAMP being present, inhibiting cyclic-AMP receptor protein’s binding to the promoter (21). Thus, the negative effect of glucose on eGFP expression can be explained by its role in negatively regulating the expression of genes on the *lac* operon. Higher levels of lactose, on the other hand, are expected to result in a longer time of gene expression, due to its role in positively regulating the *lac* repressor. Upon hydrolysis of lactose into allolactose (D-galactopyranose (β,1 → 6)-D-glucose) in the *E. coli* growth medium, the *lac* repressor becomes bound and loses affinity for the operator site (22). There were also instances in which too much glycerol or lactose resulted in decreased eGFP expression; however, its mechanism is unknown.

In addition to the growth medium work, efforts on producing a vector which expresses the trigger factor chaperone were described. The major components of this expression plasmid (pVP70K) are highlighted in Figure S2. Comparison of the plasmid maps for pVP66K and pVP70K reveals a very slight difference in molecular weight, which is attributed to a region of nucleotides encoding the amino acids AIA just 5’ of the BarCat cassette. A comparison of figure 5B with figure S2-B reveals this slight difference in DNA sequence between pVP66K and pVP70K, respectively. This AIA tag is present as an artifact from the cloning of pVP66K. It is important to note that the removal of this N-terminal AIA tag was necessary in order to directly compare the trigger factor and MBP solubility tags. Since pVP65K does not contain this N-terminal AIA tag, but pVP66K does, a direct comparison could not be made since this AIA tag may affect protein solubility. From the preliminary comparison trials for the purified protein
levels of four CESG target genes expresses in pVP65K and pVP70K, results seem promising for use of trigger factor as a fusion tag for auto-cleavage expression. From Figure 7, it is shown that three of the four target genes studied showed dramatically enhanced protein levels when target genes were expressed from pVP70K rather than pVP65K (orfs 79976, 79980, and 79979). Interestingly, for two of the target genes, there were no protein levels present following expression from pVP65K and Maxwell purification, but these proteins were readily purified when expressed from pVP70K.

Previous small-scale expression screening results of these four target proteins from pVP16 (a Gateway vector) and pVP33K (a Flexi vector) in the same expression medium (TB plus glucose) showed higher levels of expression of ORFs 79976, 79978, and 79980 in the Flexi vector, and roughly the same solubility levels between the Gateway and Flexi vectors. Table 1 summarizes these results. Small-scale expression testing for these four target proteins was also found to yield the same expression patterns following expression from pVP64A, a Gateway vector (data not shown). These small-scale expression results from pVP64A are consistent with the observed patterns of expression from pVP65K, indicating the success of trigger factor in enhancing protein solubility. Since this work did not look at actual expression levels but rather levels of soluble Maxwell purified protein, it is possible that expression levels were higher in the MBP-containing vector (pVP65K) with a large fraction being expressed as an insoluble pellet. This is an important implication that needs further investigation. Although these experiments were limited to only four target genes, present results support the hypothesis that pVP70K, a trigger factor expression vector used in auto-cleavage mode, can enhance the production of soluble proteins. Combined with the factorial growth medium experiments, these data collectively provide substantial contributions towards the optimization of protein expression from E. coli.
Acknowledgments: The author thanks Paul G. Blommel and Dr. Russell L. Wrobel for their continuous support, mentorship, and directional leadership throughout this work; other members of CESG, including Dr. R.O. Fredrick for their continuous assistance; Dr. B. G. Fox for his mentorship and willingness to support my undergraduate research at the University of Wisconsin; and the University of Wisconsin College of Agricultural and Life Sciences Honors in Research Program.
References


Table S1: The concentration of glucose (top), lactose (middle), and glycerol (bottom) are shown in (w/v). Columns 1-8 represent 64 growth medium compositions containing methionine, while columns 9-12 contain 32 growth medias with selenomethionine.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
<td>0.10</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>B</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
<td>0.10</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>C</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
<td>0.10</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>D</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
<td>0.10</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>E</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
<td>0.10</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>F</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
<td>0.10</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>G</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
<td>0.10</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>H</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
<td>0.10</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Supporting Material
Figure S1: Annotated vector maps for pVP65K (A) and pVP66K (B) were used to construct pVP70K. The BarCat cassette from pVP65K was cut using HindIII/PacI and inserted into the HindIII/PacI sites of pVP66K to produce pVP70K. pVP66K serves as the precursor to pVP70K; however, contains an artifact nucleotide sequence encoding the amino acids AIA on the amino terminus of the target gene.
Figure S2: Annotated vector map (A) and partial DNA sequence (B) of pVP70K. A, the annotated vector map of pVP70K shows the key elements, including trigger factor and the BarCat cassette. B, the DNA sequence for the 3’ end of trigger factor, TVMV site, 8X His tag, and the 5’ end of the BarCat cassette is shown. Note, there is no AIA-encoding nucleotide sequence upstream to the BarCat cassette in pVP70K.
Enhanced Bacterial Protein Expression During Auto-Induction Obtained by Alteration of Lac Repressor Dosage and Medium Composition

Paul G. Blommel, Katie J. Becker, Petar Duvnjak, and Brian G. Fox*

Department of Biochemistry, University of Wisconsin, 433 Babcock Drive, Madison, Wisconsin 53706

The auto-induction method of protein expression in E. coli is based on diauxic growth resulting from dynamic function of lac operon regulatory elements (lacO and LacI) in mixtures of glucose, glycerol, and lactose. The results show that successful execution of auto-induction is strongly dependent on the plasmid promoter and repressor construction, on the oxygenation state of the culture, and on the composition of the auto-induction medium. Thus expression hosts expressing high levels of LacI during aerobic growth exhibit reduced ability to effectively complete the auto-induction process. Manipulation of the promoter to decrease the expression of LacI altered the preference for lactose consumption in a manner that led to increased protein expression and partially relieved the sensitivity of the auto-induction process to the oxygenation state of the culture. Factorial design methods were used to optimize the chemically defined growth medium used for expression of two model proteins, Photorus luciferase and enhanced green fluorescent protein, including variations for production of both unlabeled and selenomethionine-labeled samples. The optimization included studies of the expression from T7 and T7-lacI promoter plasmids and from T5 phage promoter plasmids expressing two levels of LacI. Upon the basis of the analysis of over 50 independent expression results, combinations of optimized expression media and expression plasmids that gave protein yields of greater than 1000 μg/mL of expression culture were identified.

Introduction

Reliable and reproducible methods for high-throughput production of proteins are required for structural genomics (1–3), functional proteomics (4), drug discovery (5, 6), and other current protein biochemistry and enzymology initiatives. As one approach to this problem, the auto-induction method has been used for production of recombinant proteins in E. coli (7–9). Auto-induction is based on the function of lac operon regulatory elements in mixtures of glucose, glycerol, and lactose under diauxic growth conditions (10). Control of this metabolic circuit in wild-type E. coli has been the subject of much research and is generally well understood (10–14). During the initial growth period, glucose is preferentially used as a carbon source and protein expression is low as a result of catabolite repression of alternative carbon utilization pathways (15–17) and binding interactions between lac repressors (LacI) and lac operators (lacO). As glucose is depleted, catabolite repression is relieved, which leads to a shift in cellular metabolism toward the import and consumption of lactose and glycerol. Lactose import results in the production of allolactose from lactose by a promiscuous reaction of β-galactosidase. Allolactose then acts as the physiological inducer of the lac operon.

An apparent advantage of the auto-induction method is that it places the transition from the uninduced to induced state under metabolic control of the expression host (7, 9). This minimizes the required handling of cultures from inoculation until cell harvest, which is a major advantage for high-throughput efforts. However, several factors complicate the use of auto-induction medium, their relative amounts and their patterns of usage are critically important contributors to the outcome of the auto-induction expression. Furthermore, for optimal utility, the auto-induction method should be easy to perform in both small-scale screening and large-scale production modes and should also provide correlation between the results obtained at the different scales of operation. This scaling requirement introduces variability arising from physical parameters such as the extent of aeration associated with different types of vessels used for cell culture. Indeed, it has been previously noted that the availability of O2 can affect the outcome of auto-induction experiments (7), but the origin of this effect was not clear.

For recombinant expression systems that operate under control of the lac operon, the appearance of allolactose during auto-induction also initiates the expression of heterologous proteins (7, 18). However, the construction of recombinant expression systems makes the circumstances of induction more complicated than in wild-type E. coli. For example, E. coli cells harboring a multicopy expression plasmid may produce LacI at levels 200-fold higher than that present in wild-type cells. Currently, there is only limited experimental information on the diauxic behavior of cells expressing high concentrations of LacI (19).

Given the importance of bacterial protein expression studies (20–22), it is important to more fully understand the underlying metabolic and physical constraints to reproducibility and productivity of auto-induction approaches. This work addresses some fundamental questions of E. coli growth and heterologous expression in auto-induction medium. The results show that the level of LacI, the oxygenation state of the culture, and the medium composition have profound effects on the carbon consumption patterns required for successful execution of the auto-induction process. Arising from these findings, a factorial

* To whom correspondence should be addressed. Ph: (608) 262-9708. Email: bgfox@biochem.wisc.edu.
design approach was undertaken to optimize protein expression for three expression plasmid variants and two model proteins. In the discussed cases, the combination of fundamental and empirical studies led to an improvement in the volumetric productivity of protein expression from less than 100 μg/mL to more than 1000 μg/mL.

Materials and Methods

Chemicals. Unless otherwise stated, bacterial growth reagents, antibiotics, routine laboratory chemicals, and disposable labware were from Sigma-Aldrich (St. Louis, MO), Fisher (Pittsburgh PA), or other major distributors. L-SeMet was from Acros (Morris Plains, NJ). Preparations of standard laboratory reagents were as previously described (23). The 2-L polyethyleneteraphthalate (PET) beverage bottles used for bacterial cell growth were from Ball Corporation (Chicago, IL).

Expression Strains. The methionine auxotroph E. coli B834 [genotype F′ompT hsdS(b−m−) gal dcm, (24, 25)] was used for expression studies with T7 promoter plasmids, and E. coli B834(DE3) [genotype F′ompT hsdS(b−m−) gal dcm, (DE3)] was used for studies with T7 promoter plasmids (EMD Biosciences/Novagen, Madison, WI). Both expression hosts were transformed with pRARE2 (EMD Biosciences/Novagen) for rare codon adaptation. The pRARE2 plasmid conferred chloramphenicol resistance.

Expression Vectors. Table 1 summarizes relevant properties of the expression vectors evaluated in this work. pFN6K (Promega, Madison WI) and pET32 (EMD Biosciences/Novagen) are commercially available. The vectors pVP38K, pVP58K, pVP61K, and pVP62K were created from pQE80, 5′-GTCGACCAAGAACCTGGCGGTATGGCATG-3′[the point mutation responsible for the lacI promoter was restored by PCR in pVP58K and pVP62K, 5′-GCGGAAACACCTTGGCGGTATGGCATG-3′], pVP61K, and pVP62K also incorporate the gene for tobacco vein mottling virus (TVMV) protease with low-level constitutive expression so that co-transformation with a separate plasmid encoding the protease is not needed to achieve in vivo proteolysis (27, 28).

<table>
<thead>
<tr>
<th>Table 1. Expression Vectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>expression vector*</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>pFN6K T7</td>
</tr>
<tr>
<td>pET32 T7-lacO</td>
</tr>
<tr>
<td>pVP38K T5-lacO-lacO2</td>
</tr>
<tr>
<td>pVP61K T5-lacO-lacO2</td>
</tr>
<tr>
<td>pVP62K T5-lacO-lacO2</td>
</tr>
</tbody>
</table>

a pFN6K is from Promega (Madison, WI). pET32 is from Novagen (Madison, WI). Other vectors were created as part of this work as described in Materials and Methods. b The promoter and operator construction used for expression of the target gene. In pET32, a single copy of lacO is located 3′ to the T7 promoter. In the T5 vectors, lacO is placed between the −35 and −10 regions of the promoter and lacO2 is located between the −10 region and the start codon of the expressed gene. lacO2 is truncated from the full length lacO. c Target gene in the expression plasmid. d Promoter used for expression of lacI from the expression plasmid. e Relative level of LacI expression as compared to E. coli BL21 containing pFN6K, which includes contributions from copy number of the plasmid and relative strength of the lacO or lacO2 promoters. f N-terminal fusion tag on the expressed target protein. g Abbreviation for the expression plasmid used in the text. h The fusion protein is cleaved in vivo by TVMV protease to release SerHisGluAsnLeuTyrPheGln-AlaIeAla-eGFP.

pFN6K expresses Photinus luciferase as an N-terminal fusion (HisGln1) under control of the T7 promoter. Photinus luciferase was also expressed in the T7-lacI plasmid (pET32) and T5 promoter plasmids conferring both high (T5-lacI, pVP38K) and medium (T5-lacI, pVP58K) levels of LacI. The luciferase gene was amplified by PCR from pFN6K and the appropriate restriction sites were incorporated into the 5′ and 3′ primers. Primers were from IDT (Corvala, IA), the Ndel and HindIII restriction sites were used for cloning into pET32, the Ncol and HindIII restriction sites were used for cloning into pVP38K and pVP58K. The luciferase expressed from each expression vector investigated had an identical primary sequence including an N-terminal (HisGln1) tag.

The enhanced green fluorescent protein (eGFP) gene was assembled by standard molecular biology protocols (23) to incorporate the F64L, S65T, R81Q, F99S, M153T, and V163A mutations (29–31). The eGFP gene was subsequently amplified to add the SgfI and Pmel restriction sites required for Flexi-vector cloning (32) and transferred into pVP61K and pVP62K. eGFP was initially expressed from these vectors with an N-terminal maltose binding protein fusion that underwent in vivo proteolysis by tobacco vein mottling virus (TVMV) protease to liberate SerHisAlaSerGluAsnLeuTyrPheGln-AlaIeAla-eGFP.

Media Formulations. The noninducing and auto-induction media are derived from earlier reports on the development and use of auto-induction (7–9). All media contained 34 μg/mL of chloramphenicol and either 100 μg/mL of ampicillin or 50 μg/mL of kanamycin, depending on the selectable marker of the expression plasmid.

A 50× amino acids solution (1 L) was prepared from 10 g each of sodium glutamate, lysine-HCl, arginine-HCl, histidine-HCl, free aspartic acid, and zwitterionic forms of alanine, proline, glycine, threonine, serine, glutamine, asparagine, valine, leucine, isoleucine, phenylalanine, and tryptophan.

A 5000× trace metals solution (100 mL) was prepared from 50 mL of 0.1 M FeCl3·6H2O dissolved in ~0.1 M HCl, 2 mL of 1 M CaCl2, 1 mL of 1 M MnCl2·4H2O, 1 mL of 1 M ZnSO4·7H2O, 1 mL of 0.2 M CoCl2·6H2O, 2 mL of 0.1 M CuCl2·2H2O, 1 mL of 0.2 M NiCl2·6H2O, 2 mL of 0.1 M Na2MoO4·5H2O, 2 mL of 0.1 M Na2SeO3·5H2O, and 2 mL of 0.1 M H3BO3 and 36 mL of deionized water.

A 1000× vitamins solution (100 mL) for the noninducing medium was prepared from 2 mL of 10 mM nicotinic acid, 2 mL of 10 mM pyridoxine-HCl, 2 mL of 10 mM thiamine-HCL, 2 mL of 10 mM p-aminobenzoic acid, 2 mL of 10 mM pantothenate, 5 mL of 100 μM folic acid, 5 mL of 100 μM riboflavin, 4 mL of 5 mM vitamin B12 solution and 76 mL of

Protein Targets. Structural genomics target At1g56502 is an unknown Arabidopsis protein, and BC058837 is an unknown human protein.
sterile water. A 1000× vitamins solution (100 mL) for the auto-induction medium was the same as above except that the volume of the vitamin B12 solution was replaced with sterile water.

A 20× source of nitrogen, sulfate, and phosphorus (1 L), was prepared using 68 g of KH$_2$PO$_4$, 71 g of Na$_2$HPO$_4$, 53.6 g of NH$_4$Cl, and 14.2 g of Na$_2$SO$_4$ dissolved in sterile water.

A noninducing medium for starting inocula (1 L) was prepared using 50 mL of 20× nitrogen, sulfate, and phosphorus mix, 0.5 g of MgSO$_4$ 20 mL of the 50× amino acids solution, 0.2 mL of the 5000× trace metals solution, 1 mL of the 100× vitamins solution for the noninducing medium, appropriate antibiotics, and 0.8% (w/v) glucose with the balance sterile water.

The auto-induction medium contained the ingredients listed above for the noninducing medium with the noted omission of B$_12$ from the 1000× vitamins solution (9) and changes in the amino acids and carbon sources as described next. For expression of unlabeled proteins, the medium contained 0.2 mg/mL of methionine. For expression of selenomethionine labeled proteins, the medium contained 0.01 mg/mL of methionine and 0.125 mg/mL of selenomethionine. The concentrations of the carbon sources (glucose, glycerol, lactose) in the auto-induction medium were varied as part of a five-level, three-parameter factorial design in the following range of carbohydrate concentrations (w/v): glucose, 0 to 1%, glycerol 0 to 1.2%, and lactose 0 to 0.6%. Sucinate was maintained at 0.375% for all media formulations. The design points were based on two full three-level cubic factorials, with one nested within the other (33). This gave a total of 53 independent medium compositions (the inner and outer factorial shared a common center point). In this design, the center points were replicated four times and the face-centered points along the lactose and glycerol axes were duplicated. These conditions were conveniently arranged into an 8 × 8 array within a 96-well growth block. The Supporting Information defines how the sugar concentrations were arranged in the growth block format.

Variations of the media containing either methionine alone or selenomethionine and methionine were tested separately. The composition of the media used for selenomethionine-labeling was tested in a factorial design space comprising the inner factorial (32 data points per experiment including replicates) except in the case of work with the T7-lacI (pET32) expression vector where the full nested factorial was tested. This combination gave a total of 512 expression experiments.

**Protein Expression.** Starting inocula were grown to saturation overnight in the noninducing medium using either 96-well growth blocks having a capacity of 2 mL per well (Qiagen) or in Erlenmeyer flasks. For the growth blocks, 400 μL of the medium was used per well. For the Erlenmeyer flasks, the volume of starting inoculum was less than 10% of the total flask volume in order to promote aerobic growth. All culture growth was done at 25 °C using either plate or platform shakers.

Small-scale expression trials were carried out in 96-well growth blocks. A 20-μL aliquot of the starting inoculum was transferred to 400 μL of the auto-induction medium and incubated for 24 h at 25 °C on a plate shaker. After the incubation period, an aliquot (100–200 μL) of each 400-μL culture was transferred into a 96-well PCR plate. These samples were directly frozen at −80 °C without a preliminary cell pelleting centrifugation step. The plates were stored at −80 °C until expression analysis. Large-scale expression was conducted in 2-L PET bottles containing 500 mL of culture medium (8, 9, 34). Samples for expression analysis were harvested and stored as for the small-scale expression trials.

**Stirred Vessel Fermentations.** A Sixfors parallel six fermenter system (Informs AG, Bottmingen, Switzerland) was used to investigate the influence of aeration on the auto-induction process. Two aeration states were developed to mimic the small- and large-scale cell culture environments. For the aerobic case, which best mimics the small-scale culture in the 96-well growth blocks, airflow and agitation rate were manually adjusted to maintain dissolved O$_2$ above 10% of saturation. For the O$_2$-limited condition, which best mimics the large-scale cell culture in shaken 2-L bottles, a fixed 12 vol of air/h was added with low agitation. Samples were taken periodically to determine cell density, protein expression, and concentration of carbon sources remaining in the growth medium. The temperature was maintained at 25 °C, and the pH was passively monitored during these experiments.

**Carbon Source Analysis.** An HPLC method was developed to measure the concentration of sugars and organic acids present in the expression medium. A 1-mL aliquot of the culture medium was centrifuged at 16,000 g for 3 min to pellet the cells. A 900-μL aliquot of the clarified medium was added to 100 μL of a saturated Al$_2$(SO$_4$)$_3$ solution to precipitate phosphate. This mixture was then heated to 90 °C for 5 min to inactivate any residual enzymatic activity. Samples were stable for at least 1 wk at 4 °C after this treatment. Prior to HPLC analysis, the samples were centrifuged briefly to remove aluminum phosphate precipitate. The clarified samples were analyzed using a Shimadzu 10A HPLC system (Shimadzu, Columbia, MD) with RID10A refractive index detector and Coregel 87H3 organic acid analysis column (Transgenomic, San Jose, CA). A 20-μL sample loop was used. An isocratic 0.08 N sulfuric acid mobile phase was used for elution. The elution times of the sugars, organic acids, and phosphate were determined using the known compounds as standards.

**Protein Expression Analysis.** For analysis of protein expression, the PCR plates of frozen cell cultures were thawed and mixed with lysis buffer to obtain a final sample composition of 20 mM Tris·HCl, pH 7.5, 20 mM NaCl, 3 kU/mL of lysozyme (EMD Biosciences/Novagen), 0.7 U/mL of benzonase (EMD Biosciences/Novagen), 0.3 mM triscarboxyethylphosphine, and 1 mM MgSO$_4$. The presence of culture media due to the lack of a centrifugation step prior to cell lysis did not interfere with the biological assays, SDS–PAGE, or capillary electrophoresis analysis. The samples were sonicated for 6–10 min on a plate sonicator (Misonix, Farmington, NY). Samples for total protein expression were prepared for analysis by LabChip90 capillary electrophoresis (Caliper Life Sciences, Hopkinton, MA) as recommended by the manufacturer and were prepared for SDS–PAGE analysis as previously reported (9). The soluble protein fraction used for the biological assays and LabChip90 analysis was obtained by centrifuging the sample plates for 30 min at 2200g. Expressed protein levels were determined by LabChip90 analysis (both eGFP and luciferase) and fluorescence (eGFP only).

**Protein and Enzyme Assays.** Assays for eGFP and luciferase were performed after dilution of the soluble lysate samples with buffer containing 10 mM Tris·HCl, pH 7.5, 20 mM NaCl, and 0.1 mg/mL of acetylated bovine serum albumin (Promega, Madison, WI). For eGFP, a 5-μL aliquot of the lysate sample was mixed with 75 μL of dilution buffer prior to measurement in the wells of a black Greiner 384-well plate (ISC Bioexpress, Kaysville, UT). Fluorescence measurements were conducted in duplicate using a Tecan Ultra 384 plate reader (Tecan Group LTD, Männedorf, Switzerland) with 485 nm (25 nm bandpass) excitation and 525 nM (20 nm bandpass) emission filters.
Luciferase luminescence assays were performed using the Bright Glo luciferase assay system (Promega) after appropriate dilution of samples to bring the luciferase concentration into the linear assay measurement range. A serial dilution of purified recombinant luciferase (Promega) was assayed as a standard on every plate. Measurements were performed in duplicate with 80 μL total volume in black Greiner 384 well plates using the Tecan plate reader in luminescence mode.

**Numerical Analysis.** Carbon source consumption patterns were analyzed using Microsoft Excel and the XLFit3 curve fitting add-in (ver. 3, ID Business Solutions Ltd., Guildford, U.K.). The changes in sugar and organic acid concentrations with respect to time and cell density were fitted to sigmoidal functions. The apparent carbon source consumption rate was determined by taking the first derivative of the sigmoidal curve fits. Results of factorial design experiments were analyzed with SAS version 9.1 (SAS Institute, Inc., Cary, NC). Where expression data was available for both eGFP and luciferase, the luciferase expression level was empirically found on average to be 1.58-fold higher than the eGFP expression level based on LabChip 90 quantitation of electropherograms. This factor was used to normalize expression of luciferase and eGFP and to simulate a response surface for eGFP from luciferase data as shown in Figure 3C. For model fitting purposes, the luciferase and eGFP expression data were merged into a single data set by normalizing the luciferase expression data to the eGFP expression data. This increased the number of observations available for model fitting. Expression levels were fit to either a first-order model with two factor interactions (eq 1) or a second-order model without factor interactions (eq 2), where sugar concentrations are expressed in % (w/v), EL is the expression level, RF are the fitted response factors for the different media constituents, and C is a fitting constant.

\[
EL = [\text{glycerol}]RF_{\text{Gly}} + [\text{lactose}]RF_{\text{Lac}} + [\text{glucose}]RF_{\text{Glu}} + [\text{glycerol}]^2[\text{lactose}]RF_{\text{GlyLac}} + [\text{lactose}][\text{glucose}]RF_{\text{LacGlu}} + [\text{glycerol}][\text{glucose}]RF_{\text{GlyGlu}} + C \quad (1)
\]

\[
EL = [\text{glycerol}]RF_{\text{Gly}} + [\text{lactose}]RF_{\text{Lac}} + [\text{glucose}]RF_{\text{Glu}} + [\text{glycerol}]^2[\text{lactose}]^2RF_{\text{Lac}}^2 + [\text{glucose}]^2RF_{\text{Glu}}^2 + C \quad (2)
\]

Both models contained seven fitted parameters, and the model with the higher \(R^2\) value was chosen for each data set. Data fits were significantly improved in some cases by excluding data at zero lactose concentration because of highly nonlinear expression responses observed at low lactose concentrations. In order to simplify the graphical representation of the response surfaces, the effect of glucose was removed before generation of response surface plots by subtracting the fitted model estimate of the glucose contribution from the response at each data point. Response surface plots were generated using MathCAD version 13.0 (Mathsoft Engineering and Education, Inc.).

**Results**

**Expression in Growth Blocks and 2-L Bottles.** Our initial experiments with published auto-induction media (7–9) and T5-lacI expression plasmids revealed substantial differences between small-scale expression trials run in 96-well blocks and large-scale expression trials run in 2-L bottles. Figure 1A shows three representative examples, which were typically characterized by low total expression in the small scale and more robust expression in the large scale. Surprisingly, higher cell densities were often obtained from the small-scale trials, which suggested more efficient use of the total carbon sources added. This poor correlation limited the predictive utility of the small-scale trials.

At first, we assumed that the large-scale trials had better aeration (34) than the small-scale and that \(O_2\) limitation led to lower protein expression in the smaller cultures. However, by comparing growth rates, pH profiles, and acetate production from the two growth methods, it became apparent that the opposite was true. In one representative experiment, the small-scale cultures reached saturation at OD_{600} of 22, did not produce acetate, and maintained a stable or increasing pH, whereas cultures grown in 2-L bottles attained an OD_{600} of 8, produced significant amounts of acetate, and showed a drop in pH from 6.7 to 5.0 after 24 h of incubation. By undertaking a limited investigation of the medium composition, we were able to identify other formulations of glucose, glycerol, and lactose that improved the correlation between small- and large-scale expression trials. Figure 1B shows this preliminary result for the three representative examples from Figure 1A. Although potentially useful, this finding did not yet clarify the origin of the differences in expression behavior dependent on culture scale.

**Properties of Expression Plasmids Studied.** Figure 2 shows maps of the four types of expression plasmids used in this study. Key elements of these plasmids related to the performance of auto-induction are the copy number of the plasmid, the promoter and regulator systems used to control inducible target expression, and the promoter used to control constitutive expression of LacI.

All of the expression plasmids used in this work contain the pBR322 origin of replication and have similar copy numbers of \(\sim 15\) to \(20\) per cell (35). Since only \(\sim 10\) molecules of LacI are present in wild-type *E. coli* (26), several strategies have been developed to control basal expression from lac operator-repressed expression systems. pFN6K provides a simple T7 promoter for control of expression and no contributions from lacO or recombinant LacI to control basal expression. In contrast, pET32 provides a T7 promoter with an associated lacO sequence and constitutive expression of LacI from the plasmid (designated T7-lacI in the following). In this case, the copy number of the plasmid and the wild-type lacI promoter serve...
to supplement the level of Lacl expression. Both pFN6K and pET32 plasmids require a lysogenic host containing T7 RNA polymerase under inducible control of the lacUV5 promoter such as E. coli B834(DE3) used here.

The pVP vectors used in this work have the T5 phage promoter (36–38) under control of two copies of the lac operator (lacO1 and lacO2 in Figure 2). The lacO1 sequence was truncated during the original construction of the T5 promoter vectors (37) to fit between the −35 and −10 regions of the promoter, so is distinct from lacO2, which retains the natural sequence from E. coli lac operon operator O1 (39). Biochemical data suggests that Lacl is able to bind the truncated operator with an affinity similar to the full length operator (40).

E. coli RNA polymerase recognizes the T5 promoter so many different E. coli expression strains can be used with this vector. pVP58K and pVP62K contain the strong lacI q promoter for overexpression of Lacl, whereas pVP58K and pVP62K contain the wild-type lacI promoter controlling expression of Lacl, whereas pVP58K and pVP62K contain the wild-type lacI promoter. Photinus luciferase was expressed from plasmids A–C. Enhanced green fluorescent protein was expressed from pVP61K and pVP62K, shown in D. pVP61K and pVP62K also contain the coding region for tobacco vein mottling virus protease (TVMV) under control of the tet promoter. The expression strains used in this study do not overexpress the tet repressor, leading to low level, constitutive expression of TVMV. Due to the presence of a TVMV recognition site between the MBP and eGFP, the fusion protein is cleaved in vivo to liberate His7-eGFP.

**Factorial Design of Medium Composition.** Since the preliminary results of Figure 1 showed that increasing the amounts of glycerol, lactose, and succinate and decreasing the amount of glucose could improve the correlation between small- and large-scale expression with the T5-lacI expression system, we applied a factorial design approach to individually optimize the media for small-scale expression using the T5-lacI, T5-lacI q, and T7-lacI plasmids. For this optimization, all media constituents were fixed except for glucose, glycerol, and lactose, and these were independently varied in a factorial design approach (22, 33). A 3-dimensional representation of this design and the individual medium compositions are presented in Supporting Information.

Figure 3 shows response surface models for expression results obtained in media containing methionine only (left side, including evaluation of 53 independent medium compositions) or selenomethionine (right side, including evaluation of 32 independent medium compositions for T5-lacI and T5-lacI q or 53 compositions for pET32). eGFP was used as an expression target for total soluble protein expression due to the ease of...
quantification through intrinsic fluorescence. Since eGFP requires O$_2$ for fluorophore formation (41), only small-scale expression experiments where O$_2$ was not limited were undertaken.

Figure 3A shows the response surfaces for expression using the T5-lacI-eGFP expression plasmid. The left response surface shows that variations of the carbon sources in a methionine medium can give a nearly 15-fold increase in soluble eGFP production based on the measured fluorescence, which corresponds to a range from $\sim$100 $\mu$g/mL of eGFP in the poorest performing composition to $\sim$1500 $\mu$g/mL of eGFP in the best performing composition. Figure 4A shows a SDS–PAGE analysis of the total eGFP expression obtained from one of the better performing media compositions (0.025% glucose, 0.9% glycerol, 0.45% lactose). In this case, stoichiometric proteolysis of the original fusion protein (70 kDa) to MBP (42 kDa) and the tagged-eGFP (29 kDa) was obtained from the constitutively expressed TVMV protease. The right side of Figure 3A shows the response surface for the same expression experiment in media containing selenomethionine. Overall, the response surfaces for T5-lacI-eGFP expression in the methionine and selenomethionine media tracked each other closely. Indeed, among the lesser number of compositions investigated for the selenomethionine medium, soluble eGFP expression was observed in excess of 1000 $\mu$g/mL (total recombinant protein expression exceeded 2000 $\mu$g/mL if MBP expression was also accounted for).

Table 2 shows the statistical factors for the model analysis of these two different media optimizations with the T5-lacI-eGFP expression vector. In both the methionine and selenomethionine media, a change in the lactose concentration was most strongly correlated to a positive expression response.
of the plasmid types. The unregulated T7-Luc plasmid (pFN6K) was maximized. Figure 3C shows an estimated response surface of basal expression, around 1% of that from the T7-Luc plasmid, and no difference in basal expression was observed in either noninducing or auto-induction media. The presence of two copies of lacO in the promoter region and overexpression of LacI from the plasmid contribute to this result. The T7- lacI plasmid (pET32-Luc) gave a 20-fold reduction in basal luciferase expression as compared to the T7 vector, but this level was still 5× higher than that observed with the T5- lacF plasmids. Results (not shown) from the T5- lacI-Luc plasmid (pVP38K-Luc) suggested an expression level in the noninducing medium similar to T7- lacI-Luc. Thus the higher basal expression observed for the T7- lacI and T5- lacI plasmids compared to T5- lacF is likely a result of a decrease in cellular LacI and corresponding lower occupancy of the promoter lacO sites. Overall, the presence of lactose in the medium did not significantly increase basal expression of luciferase, indicating that the effects of catabolite repression and inducer exclusion (13, 16, 42) are sufficiently strong to prevent premature induction of the lac operon.

Table 3. Basal Expression of Luciferase ($\mu g$/mL) from Different Expression Plasmids in Auto-Induction Media

<table>
<thead>
<tr>
<th>expression vector</th>
<th>$-\text{lactose}^a$</th>
<th>$+\text{lactose}^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7-Luc</td>
<td>2.7 ± 0.3</td>
<td>2.9 ± 0.4</td>
</tr>
<tr>
<td>T7- lacI-Luc (pET32-Luc)</td>
<td>0.19 ± 0.04</td>
<td>0.19 ± 0.04</td>
</tr>
<tr>
<td>T5- lacF-Luc</td>
<td>0.03 ± 0.008</td>
<td>0.03 ± 0.004</td>
</tr>
</tbody>
</table>

*a* Lucifer activity interpolated at a cell density of 2 (600 nm) based on measurements taken at lower and higher cell densities during exponential growth in a noninducing medium containing 0.8% (w/v) glucose. *b* Lucifer activity interpolated at a cell density of 2 (600 nm) based on measurements taken at lower and higher cell densities during exponential growth in an auto-inducing medium containing 0.8% (w/v) glucose and 0.1% (w/v) lactose.

accounting for an estimated 38% or 36% of the modeled effect, respectively. In the methionine medium, increasing lactose concentration was also correlated with the expression response, accounting for 21% of the modeled effect. In the selenomethionine media, increasing lactose had less influence on the expression response, accounting for 13% of the modeled effect, while other higher order terms had a larger influence. Possible implications of these results are considered in the Discussion.

Figure 3B shows the response surfaces for expression from T5- lacF-eGFP. This expression system gave lower total expression than T5- lacI-eGFP, with expression levels ranging from near zero at low lactose to ~600 $\mu g$/mL when glycerol and lactose were maximized. Figure 3C shows an estimated response surfaces for expression of eGFP from T7- lacI generated by normalization of scalar multiplication of the luciferase expression data with the factor 1.58 as described in Materials and Methods.

**Basal Expression Studies Using Luciferase.** Luciferase was used as an expression target due to the large linear range of the luminescence assay (5–6 orders of magnitude) and a low detection limit that was useful for quantifying basal expression. Table 3 compares the basal expression of luciferase from three of the plasmid types. The unregulated T7-Luc plasmid (pFN6K) gave the highest level of expression in noninducing medium and a small increase in basal expression in auto-induction medium. This result arose through expression of T7 RNA polymerase from the poorly repressed genomic lacUV5 promoter and subsequent transcription from the plasmid T7 promoter upstream of the luciferase gene. In contrast, the highly regulated T5- lacF-Luc plasmid (pVP38K) gave the lowest level of basal expression, around 1% of that from the T7-Luc plasmid, and no difference in basal expression was observed in either noninducing or auto-induction media. The presence of two copies of lacO in the promoter region and overexpression of LacI from the plasmid contribute to this result. The T7- lacI plasmid (pET32-Luc) gave a 20-fold reduction in basal luciferase expression as compared to the T7 vector, but this level was still 5× higher than that observed with the T5- lacF plasmids. Results (not shown) from the T5- lacI-Luc plasmid (pVP38K-Luc) suggested an expression level in the noninducing medium similar to T7- lacI-Luc. Thus the higher basal expression observed for the T7- lacI and T5- lacI plasmids compared to T5- lacF is likely a result of a decrease in cellular LacI and corresponding lower occupancy of the promoter lacO sites. Overall, the presence of lactose in the medium did not significantly increase basal expression of luciferase, indicating that the effects of catabolite repression and inducer exclusion (13, 16, 42) are sufficiently strong to prevent premature induction of the lac operon.

**Fermentation Approach.** An instrument-controlled fermenter was used to investigate the correlation between carbon source utilization, O2 saturation of the culture, and protein expression. In the fermenter, an aerobic growth condition was maintained during auto-induction by fixing O2 at greater than 10% of air saturation during the entire cell growth. The aerobic growth condition in the fermenter best represents growth of small-scale cultures in 96-well growth blocks. For comparison, a microaerobic growth condition was maintained by completing the growth phase under O2 limitation (43–45). The microaerobic growth condition best represents growth of large-scale cultures in 2-L bottles. Figure 6 shows dissolved O2 and pH profiles for growth and auto-induction under these two conditions.

**Carbon Source Consumption Patterns.** Figure 7A shows representative HPLC traces obtained from the culture medium during the course of a growth of *E. coli* B834(DE3) with the simple T7-Luc plasmid in auto-induction medium. At t = 0, lactose, glucose, succinate, and glycine are present. At t = 8 h (cell density of ~5), the glucose was entirely consumed and
luciferase expression levels were 1820 expression from the indicated luciferase expression plasmids. Reported later consumed. At accumulated early in the growth and auto-induction but was depleted from the culture medium. It is notable that acetate lactose had become the preferred carbon source, so it was being depleted from the culture medium. It is notable that acetate had become the preferred carbon source, so it was being depleted from the culture medium. It is notable that acetate lactose is consumed as a byproduct of lactose consumption, while acetate was a byproduct of anaerobic fermentation. (B) Sigmoidal curve fitting of the relationship between change in carbon source concentration and cell density. In all cases, glucose (C) was consumed first, followed successively by lactose (D), succinate (E), glycerol (F), and then acetate (G). Acetate was initially produced and later consumed as a carbon source. (C) First derivative of the sigmoidal curve fits, defined to be the specific consumption for each carbon source. These series have the same markers as in panel B. The filled circles (●) show luciferase expression from the T7-Luc expression plasmid as determined by luminescence assay.

Figure 7B shows the complete pattern of carbon source consumption during the aerobic growth of E. coli B834(DE3) transformed with T7-Luc (see Table 1) in the auto-induction medium. In these cells, LacI is only provided by low-level constitutive expression from the bacterial genome. The carbon source concentrations were fitted as sigmoidal functions (solid lines) for illustrative purposes, and Figure 7C shows the first derivative of these fits. In Figure 7B and 7C, the carbon consumption patterns are plotted relative to cell density (optical density at 600 nm), which provides a useful correlation between an easily measured experimental property and the status of the carbon sources during growth and auto-induction. This representation is independent of the time of the culture but represents the progress toward consumption of carbon required to enter and complete the auto-induction program. For example, the
transition from growth on glucose to growth on lactose occurs at a cell density of ~5, lactose consumption is complete at a cell density of ~7, and no consumable carbon sources are remaining when the cell density has reached ~13. The carbon consumption pattern of E. coli BL21 lacking an expression plasmid was indistinguishable (data not shown).

This pattern of carbon consumption is consistent with previous studies of E. coli diauxic growth (15). Thus glucose was preferentially consumed, followed by lactose, and finally glycerol. Furthermore, in these experiments, succinate was gradually consumed throughout the entire growth period and acetate was largely consumed by the end of the culture growth. In auto-induction, protein expression from the lac operon will be induced along with lactose consumption. For example, induction of T7 RNA polymerase expression under the control of a lacUV5 promoter in E. coli B834(DE3) would be expected to coincide with activation of the lac operon. Correspondingly, Figure 7C shows that luciferase activity was detected at a cell density of ~5 when lactose became the preferred carbon source, and continued to increase after lactose consumption was complete as glycerol and succinate were consumed.

**Effect of LacI Dosing on Carbon Consumption Patterns.**

Figure 8 shows the effect of different levels of LacI on the carbon consumption patterns during auto-induction. The consumption patterns for glycerol and lactose for E. coli B834 expressing T5-lacI-Luc (pVP58K) by aerobic auto-induction are shown in Figure 8A. This construct provides expression of plasmid-encoded LacI from the weak lacP promoter. Increasing LacI shifts the order of preference from glucose/lactose/glycerol to glucose/lactose/glycerol in aerobic culture. Thus there is a dramatic shift in the pattern of carbon consumption relative to the T7-Luc data shown in Figure 7C.

Figure 8B shows the consequences of LacI levels on lactose consumption for each of the luciferase expression plasmids from Table 1. T7-Luc, which provides no recombinant LacI, maximally consumed lactose at a cell density of ~10. In contrast, T7-lacI-Luc, with constitutive plasmid-encoded expression of LacI, shifted the maximal consumption of lactose to a cell density of ~15, while T5-lacI-Luc, also providing constitutive plasmid-encoded expression of LacI, behaved in a similar manner and shifted the maximal consumption of lactose to a cell density of ~18. Finally, T5-lacP-Luc, giving overexpression of plasmid-encoded LacI from the strong lacP promoter), the shift in carbon consumption pattern was so extreme that culture growth stopped in aerobic conditions before lactose could be substantially consumed (Figure 8B, x symbols).

**Consequences of O2 Availability During Auto-Induction.**

Figure 9A shows the consequences of aerobic or O2-limited growth on the lactose consumption pattern for T5-lacI-Luc expression in E. coli B834. During aerobic growth, the maximal lactose consumption occurred at a cell density of ~18, as shown in Figure 8.

The appearance of luciferase activity closely tracked this maximal consumption pattern, which is consistent with the relatively strong control of basal expression given by aerobic growth and the presence of recombinant LacI. For comparison, Figure 9A also shows that O2-limited growth during auto-induction shifted the maximal lactose consumption to a lower cell density. Thus changes in oxygenation state of the medium dramatically affected the preference for lactose consumption relative to other carbon sources. Furthermore, in the O2-limited growth, the appearance of luciferase activity no longer closely tracked the lactose consumption pattern, but was shifted to earlier in the overall growth period. These results are consistent with a weakening of catabolite repression and consequent increase in basal expression from both the genomic lac operon (generating allolactose) and from the recombinant expression system (generating luciferase).
the highest response for all cases. This simplified representation (v) glycerol and 0.6% (w/v) lactose, a trajectory that includes starting from zero glycerol and lactose and ending at 1.2% (w/v) lactose was higher than from T7-
lacI
expression and oxygenation state in the function of auto-induction systems for protein production in E. coli has been extensively studied (11). Figure 3A (T5-
lacI q
expression vector. Under aerobic auto-induction conditions, lactose utilization was only weakly initiated and ~70% of the initial lactose remained after ~40 h. After the time when glucose, glycero, and succinate were consumed (~15 h), little additional cell growth or protein expression were observed. For comparison, O2-limited auto-induction gave complete utilization of lactose between 10 and 30 h. During this time, continued cell growth and protein expression were both obtained.

**Discussion**

**Auto-Induction Method.** Auto-induction medium contains a mixture of carbon and energy sources. Glucose is the preferred source for E. coli and is utilized during the early stages of growth. Lactose and glycerol serve as carbon and energy sources during later stages of growth and recombinant protein production. Succinate (or other organic acids such as aspartate or glutamate) may be included to help maintain the culture pH and to act as additional sources of carbon and nitrogen. The consumption of these individual carbon sources by E. coli has been extensively studied (11–17, 46) and in some cases, in combination (as is the case for glucose-lactose diauxic growth). This work demonstrates the importance and possible advantages of considering the interactions between media composition, LacI expression and oxygenation state in the function of auto-induction systems for protein production in E. coli.

**Comparison of Response Surfaces.** Figure 10 shows a two-dimensional plane through the response surfaces of Figure 3 starting from zero glycerol and lactose and ending at 1.2% (w/v) glycerol and 0.6% (w/v) lactose, a trajectory that includes the highest response for all cases. This simplified representation offers a direct comparison of expression results achieved from the three expression systems. Expression from T5-
lacI (solid line) was higher than from T7-
lacI (pET32, dashed line) at all compositions except at the lowest lactose concentrations, where basal expression from T7-
lacI was higher (Table 3). T5-
lacI (diamonds, methionine medium; circles, selenomethionine medium) exhibited the lowest expression levels. Surprisingly, the combination of T5-
lacI with selenomethionine medium gave a higher level of expression than the same plasmid with methionine medium, and selenomethionine-labeled protein was obtained with yield of ~500 μg/mL. This enhanced performance occurred because the presence of selenomethionine shifted the maximal lactose consumption to lower cell density (data not shown), allowing more complete execution of the auto-induction program. How this response is mediated at the cellular level is not known.

Figure 11 shows a two-dimensional surface plot that reveals additional features about the composition of the optimal medium for the T5-
lacI plasmid. For this plot, the range of carbon source concentrations investigated was intentionally extended beyond that shown in Figure 3 and notably resulted in medium compositions that decreased the expression. These results indicate that with the present composition of non-carbon source components, maximum expression from the T5-
lacI plasmid was obtained near the limits of the lower factorial (dotted line), specifically 0.6% lactose and 1.2% glycerol and that slightly lower glycerol or lactose concentrations had little effect on expression while higher concentrations of glycerol adversely affected expression (region bounded by the dashed line). This behavior is also summarized in Table 2, where glycerol had a positive first-order scaled effect estimate but a negative second-order scaled estimate effect (qualitatively similar behavior was obtained from both methionine and selenomethionine media). The region where highest expression occurred is a broad plateau, indicating overall tolerance to minor variations in medium composition without altering the expression outcome. Figure 11 shows that there are choices for medium composition that give gradual change between lower and higher expression levels. Knowledge of this may be useful to maximize the soluble production of some proteins like luciferase that apparently have an intrinsic solubility limit within cells. Other choices for change in medium composition give precipitous changes in the expression level. Knowledge of these is important to avoid experimental conditions that are likely to give poor or irreproducible results.

Figure 11 also shows that additional increases in lactose and glycerol near the high end of the experimental range investigated did not increase expression but in some circumstances actually decreased expression. The reason(s) for the decrease in expression are not known. In the present media, the cell density appeared to be limited to OD600 ~25 and was not affected by further increases in lactose or glycerol, suggesting that some non-carbon source component may have become limiting at this cell density. Systematic evaluation of the contribution of other media components to expression results in a manner similar to
that used here for carbon sources may yield further increases in cell density and volumetric protein expression.

As indicated by Figures 7C and 8A, glucose was always the preferred carbon source. Thus changes in the level of glucose added to the medium control the cell density at which the auto-induction protocol will be initiated. This concept has been discussed previously in the context of lactose-induced high-density fed-batch fermentation (18) and use of this approach for isotopic labeling studies (47). However, increasing the level of glucose will increase the cell mass and biological demand for carbon sources, leading to more rapid consumption of lactose and glycerol during the auto-induction phase without compensating changes in the levels of lactose and glycerol. This would shorten the time of auto-induction. Depending on circumstances, this may or may not be beneficial.

**Influence of LacI on Auto-Induction.** LacI acts in two ways to delay the onset of lactose consumption required for auto-induction. First, high intracellular concentrations of LacI increase the occupancy of the lacO sites located upstream of the lac operon structural genes. This occupancy strongly decreases the basal expression of β-galactosidase and lac permease, which in turn decreases the rate of allolactose production. Second, a larger absolute amount of allolactose is required in order to dissociate intracellular LacI from lacO sites so that induction of the lac operon and heterologous protein expression can begin. These combined effects are sufficient to completely change the order of carbon source consumption from glucose/lactose/glycerol to glucose/glycerol/lactose for *E. coli* growths with each of the plasmids tested that supplement LacI expression.

Maximal lactose consumption occurred at a higher cell density for the growths with the T5-lacI plasmid (Figure 8) as compared with the T7-lacI (pET32) plasmid. Since both plasmids have the pBR322 origin, the copy number should not differ greatly (48). Moreover, since both use the lac promoter to express LacI from the plasmid, the level of LacI should be similar. Small differences in LacI expression due to positional effects in the plasmid may account for the difference in behavior. Thus a previous study has demonstrated that such positional effects can influence basal levels of heterologous protein expression (49, 50) and it is plausible that the effects could influence constitutive expression of LacI in a similar way.

It is not clear why expression levels from the T5-lacI plasmid were ~70% higher than those determined for the pET32 plasmid (T7-lacI, compare Figure 3A and 3C). The T5 promoter uses *E. coli* RNA polymerase, while pET32 requires that T7 RNA polymerase must also be made. It seems unlikely that this difference alone accounts for the lower expression from the pET32 plasmid. T7 RNA polymerase is highly active (51) and might be expected to make more mRNA than *E. coli* polymerase, especially upon considering that the T7 polymerase is dedicated to the production of target gene transcripts (52) while the T5 promoter must compete with other host promoters. Thus it is possible that high transcription levels may excessively direct energy fluxes toward mRNA production and away from protein expression. Furthermore, transcript instability due to a decoupling of transcription and translation caused by the high transcription rate of T7 RNA polymerase may play a role (53, 54).

**Influence of Oxygenation State on Auto-Induction.** The consequence of oxygenation state in the auto-induction culture is apparent from Figure 9. In all cases investigated, lactose consumption and protein expression were shifted to a lower cell density by O2 limitation. For T5-lacI, this effect was dramatic enough that the final expression levels were higher when O2 was limited.

*E. coli* is known to control glucose and lactose import as a response to O2 limitation through a variety of transcriptional and post-translational mechanisms. As one example, phosphorylated ArcA is a negative transcriptional regulator of the IICB\(^{Glc}\) component of the bacterial phosphoenolpyruvate:sugar phosphotransferase (PTS) system (46, 55). Decreased expression of IICB\(^{Glc}\) leads to an accumulation of the phosphorylated PTS enzyme component IIAGlc. Since dephosphorylated IIAGlc is a known inhibitor of lac permease (15–17), O2 limitation and accumulation of phosphorylated IIAGlc relieve the inhibition of lac permease, allowing higher lactose import rates.

The results elucidate the origin of the difference in expression behavior observed between small- and large-scale experiments. In the tests of Figure 1A, the elevated level of LacI postponed lactose utilization in the aerobic conditions of the small-scale, while the O2-limited conditions of the large-scale shifted lactose utilization to lower cell density and thus promoted protein expression. Reformulation of the carbon sources promoted growth to higher cell density by increases in glycerol and lactose, utilization of lactose at lower cell density through lowered glucose, and more complete utilization of the provided carbon sources, regardless of the culture oxygenation state (Figure 1B).

**Role in High-Throughput Protein Expression Studies.** For high-throughput studies, which provided the motivation for the development of auto-induction (7) and the experimental basis for this work (8, 9), it is often desirable to screen for suitable protein expression in small volumes using multiwell growth blocks. We found that this expression environment was aerobic but, surprisingly, led to significantly lower protein expression in the initially defined auto-induction medium. In contrast, O2 limitation was previously noted to increase the yield of recombinant protein (7), and we have found this limitation most closely correlated to the actual conditions in 2-L bottles used for large-scale protein expression (34). It is noted that fully anaerobic conditions are not conducive to either rapid cell growth or high-yield production of biomass.

This work was further prompted by the need for improved correlation between small-scale protein expression screening and large-scale protein production at the University of Wisconsin Center for Eukaryotic Structural Genomics (8, 9). Through an initial set of media optimization experiments, the discrepancy between small- and large-scale culture results was addressed by increasing the concentration of carbon sources available, which on average gave a 2- to 3-fold increase in target protein expression in the small-scale cultures. However, protein expression in large-scale cultures was only marginally improved with this initial change in medium composition. Therefore, we were interested in whether other manipulations of the biochemical apparatus used for auto-induction might also improve protein expression yields. By decreasing the LacI expression level provided by the expression plasmid, lactose consumption and heterologous protein expression were shifted to an earlier phase of growth. Although culture oxygenation still contributed to the timing of induction, both small- and large-scale growths were able to produce enough allolactose to derepress the lac operon, fully consume the available lactose, and achieve high levels of protein expression.

These studies also give insight into the use of auto-induction for production of 13C- and 15N-labeled proteins for NMR structure determination. First, efficient consumption of succinate during the auto-induction process limits the utility of these media
formulations for production of 13C-labeled samples for NMR structure determination, unless 13C-labeled succinate is used. The substitution of other amino acids (aspartate, glutamate) will not correct this problem and may potentially introduce problematic dilution of 15N-labeling unless the more expensive 13C- and 15N-labeled analogs are used for these components. Furthermore, our previously described changes in medium composition for NMR studies (8) are now recognized to fall into a low productivity region of the T5-lacP response surface shown in Figure 3B (0.5% glycerol, 0.2% lactose). In the previous study, unlabeled lactose (not cost-effective for use as a 13C-labeled compound) was intentionally minimized in order to avoid isotopic dilution of 13C-labeled glycerol. As an alternative, the response curve in Figure 3A suggests that this same mixture of glycerol and lactose may give considerably better expression results when coupled with a T5-lacI plasmid.

Other Possible Uses of Factorial Medium Design. Small-scale protein expression in 24- or 96-well blocks was originally intended to be a screening tool for numerous structural genomics targets, whose expression properties were not known. The array format for the various auto-induction media provides a simple way to test the performance of other plasmid vectors and host strains for conditions that maximize the expression of these unknown proteins. Furthermore, based on expression levels possibly exceeding 1000 µg/mL of culture fluid (actually exceeding 2000 µg/mL for the combination of in vivo proteolysis of MBP and eGFP from pVP62K; see Figure 4), it is reasonable to consider other applications for small-scale expression. For example, the amount of protein produced from a few milliliters of these cultures may be sufficient for automated protein purification (56), microfluidics-based crystallization screening (57, 58), initial nanoliter-scale crystallization trials (59, 60), 13N HSQC NMR measurements (61, 62), or functional and enzymatic characterizations (63, 64).

Other Experience with Use of Optimized Auto-Induction Conditions. The work presented here includes expression studies in 96-well growth blocks, 2-L shaken bottles, and automated stirred-vessel fermenters. In each case, the combination of a designed auto-induction medium and matched expression plasmid gave strong expression results, demonstrating utility in several different formats used to grow bacterial cells. The results presented here derive from study of only two target proteins, eGFP and luciferase, that were chosen because of the attractiveness of their assays. Nevertheless, our preliminary experience with other proteins suggests that these modifications to auto-induction media composition and LacI dosing may have general utility in improving the level of recombinant protein expression. Thus combination of a T5-lacI expression plasmid with a terrific broth medium supplemented with an auto-induction mixture of 0.015% glucose, 0.8% glycerol, 0.5% lactose, 0.375% aspartic acid and 2 mM MgSO4 contributed to a ~5-fold increase in expression of soluble TEV protease (64) when compared to previous reports (66, 67). Moreover, preliminary expression studies with other proteins such as toluene 4-monooxygenase hydroxylase (68), stearoyl-ACP A9 desaturase (18), cytochrome bs (Sobrado, P.; Fox, B. G., unpublished work) and various bacterial and plant FMN-containing oxidoreductases (Malone, T. E.; Fox, B. G., unpublished work) indicate the combination of a factorial designed auto-induction media with T5-lacI plasmids offers substantial promise for structural and functional work with known proteins.

Acknowledgment

We thank Dr. R. McClain (University of Wisconsin Chemistry Department) for the use of the HPLC refractive index detector and Dr. R. O. Frederick (University of Wisconsin Center for Eukaryotic Structural Genomics) for supplying the eGFP clone. The NIH Protein Structure Initiative (1 U54 GM 74901, J. L. Markley, Principal Investigator; G. N. Phillips, Jr. and B. G. Fox, Co-Investigators) and Promega Corporation (B. G. Fox, Principal Investigator) supported this work.

Supporting Information Available: Photograph of eGFP fluorescence in a 96-well plate format, composition of the auto-induction medium present in each well of the plate, and three-dimensional representation of factorial design experimental space. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(38) Gentz, R.; Bujard, H. Promoters Recognized by 

(36) Brunner, M.; Bujard, H. Promoter Recognition and Promoter 

(32) Blommel, P. G.; Martin, P. A.; Wrobel, R. L.; Steffen, E.; Fox, 

(31) Crameri, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. Improved 

(29) Patterson, G. H.; Knobel, S. M.; Sharif, W. D.; Kain, S. R.; Piston, 


(24) Wood, W. B. Host Specificity of DNA Produced by 

(22) Swalley, S. E.; Fulghum, J. R.; Chambers, S. P. Screening Factors 

(21) Terpe, K. Overview of Bacterial Expression Systems for Hetero-

in Vi

Transcription-Translation System for the Analysis of Proteins 

Strength in the 

qPCR Quantification of Plasmid Copy Number in 

2006 

Expression Purif. 

Chromophore Mutations Provide Enhanced Sensitivity with the 

90. 

Recombinant Proteins. 

Vector Tailored for Large-Scale, High-Throughput Purification of 

of Fusion Proteins by Tobacco Vein Mottling Virus Protease 

E.; Copeland, T. D.; Waugh, D. S. Efficient Site-Specific Processing 

274 

and 

in Vitro 

Nature 

122 

- 

173 

J. Mol. Biol. 

: Bacterial Mutations Affecting the Restriction and Modification 

Fundamentals to Commercial Systems. 

H.; Hard, T. Rapid Screening for Improved Solubility of Small 

Microbial 

Bacterial Mutations Affecting the Restriction and Modification 

Pyruvate Catabolism in 

679 

Systems: Comparison of Different Repressor Control Configurations 

1995 

- 

211 

EMBO J. 

11 

, 70. 

Using 2-Liter 

29 

,7 0 

2004 

- 

123 

, 273 

, 22. 

Nam, T. W.; Ryu, S.; 

Seok, Y. J. Expression of ptsG Encoding the Major Glucose 


(37) Studts, J. M.; Fox, B. G. Application of Fed-Batch Fermentation 

to the Preparation of Isotopically Labeled or Selenomethionyl-


(38) Polisky, B. ColE1 Replication Control Circuitry: Sense from 

Antisense. Cell 

1995, 82, 358–41. 

(39) Bahl, C. P.; Wu, R.; Stawinsky, J.; Narang, S. A. Minimal Length 


92, 12250–4. 

(40) Alexeeva, S.; Hellingwerf, K. J.; Teixeira de Mattos, M. J. 

Requirement of ArcA for Redox Regulation in Escherichia coli under 

Microaerobic but Not Anaerobic or Aerobic Conditions. J. Bacteriol. 

2003, 185, 204–9. 


(42) Saier, M. H., Jr. Protein Phosphorylation and AllostERIC 

Control of Inducer Exclusion and Catabolite Repression by the Bacterial 


(43) Alexeeva, S.; de Kort, B.; Sawers, G.; Hellingwerf, K. J.; de 

Mattos, M. J. Effects of Limited Aeration and of the ArcAB System 

on Intermediary Pyruvate Catabolism in Escherichia coli. J. Bacteri-


(44) Alexeeva, S.; Hellingwerf, K. J.; Teixeira de Mattos, M. J. 

Quantitative Assessment of Oxygen Availability: Perceived Aero-

biosis and Its Effect on Flux Distribution in the Respiratory Chain 


(45) Alexeeva, S.; Hellingwerf, K. J.; Teixeira de Mattos, M. J. 


(47) Studts, J. M.; Fox, B. G. Application of Fed-Batch Fermentation 

to the Preparation of Isotopically Labeled or Selenomethionyl-


(48) Polisky, B. ColE1 Replication Control Circuitry: Sense from 

Antisense. Cell 

1995, 82, 358–41. 

(49) Dubendorff, J. W.; Studier, F. W. Application of Fed-Batch Fermentation 

Methods Mol. Biol. 


(50) Bahl, C. P.; Wu, R.; Stawinsky, J.; Narang, S. A. Minimal Length 


92, 12250–4. 

(51) Makarova, O. V.; Makarov, E. M.; Sousa, R.; Dreyfus, M. 

Transcription of escherichia coli Genes with Mutant T7 RNA 

Polymerases: Stability of lacZ mRNA Inversely Correlates with 


(52) Studier, F. W.; Moffatt, B. A. Use of Bacteriophage T7 RNA 

Polymerase to Direct Selective High-Level Expression of Cloned 


(53) Iost, I.; Guillerez, J.; Dreyfus, M. Bacteriophage T7 RNA 


(54) Lopez, P. J.; Iost, I.; Dreyfus, M. The Use of a tRNA as a 

Transcriptional Reporter: The T7 Late Promoter Is Extremely 

Efficient in Escherichia coli but Its Transcripts Are Poorly Expressed. 


(55) Perrenoud, A.; Sauer, U. Impact of Global Transcriptional 

Regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on Glucose 


(56) Murphy, M. B.; Doyle, S. A. High-Throughput Purification of 

Hexahistidine-Tagged Proteins Expressed in E. coli. Methods Mol. 


(57) van der Woerd, M.; Ferree, D.; Pusey, M. The Promise of 

Macromolecular Crystallization in Microfluidic Chips. J. Struct. 


(58) Segelke, B. Macromolecular Crystallization with Microfluidic 


in Microfluidics to Facilitate and Understand Protein Crystallization. 


(60) Lion, N.; Reymond, F.; Girault, H. H.; Rossier, J. S. Why the 


Received January 10, 2007. Accepted April 24, 2007.

BP070011X