Center for Eukaryotic Structural Genomics
Protein Structure Initiative

Factorial Design Improvement of Auto-Induction Media

Paul G. Blommel, Katie Becker, and Brian G. Fox

University of Wisconsin-Madison, Department of Biochemistry, 433 Babcock Drive, Madison, Wisconsin, USA 53706-1549, http://www.uwstructuralgenomics.org

Abstract

Auto-induction, introduced to structural genomics by Bill Studier of Brookhaven National Lab [1], employs different carbon sources to support cell growth and protein expression without the requirement to monitor the culture growth state. Although conceptually simple, auto-induction arises from a complex set of changes in growth conditions and host regulatory responses. During our initial pipeline efforts with auto-induction medium [2], we observed that protein expression in small-scale screening was often drastically lower than obtained from large-scale culture. This issue prompted our detailed examination of the interplay between medium composition, host strain, and genotype of expression plasmids that might contribute to this discontinuity. A factorial design approach was first applied to optimize the carbon source concentrations for small-scale expression culture based on the CESG target scoring method. The altered medium obtained by the factorial design was then evaluated for comparability of protein expression results in large-scale culture. After several cycles of adjusting the carbon source concentrations, we obtained a greater than 80% agreement for target gene expression in small- and large-scale culture. To more fully understand the consequences of the changes in medium composition, parallel, instrumented fermenter studies were undertaken to correlate carbon source consumption patterns with cell growth and protein expression. For systems with higher than wild type expression of *lac* repressor (T5/*lac* and T7/*lac* systems), culture oxygenation was a dominant factor in determining protein expression levels from the auto-induction method. Surprisingly, fully aerobic culture conditions resulted in lower protein expression levels than obtained with a partial oxygen limitation. By monitoring carbon source consumption patterns, we found that the preferred order for use of carbon sources shifted from glucose/lactose/glycerol to glucose/glycerol/lactose for expression in the presence of high levels of *lac* repressor. Furthermore, oxygen limitation decreased the stringency of the glycerol to lactose shift, resulting in earlier onset of expression and higher overall protein expression levels. Auto-induction media have also been investigated for T7 promoter based vectors (with no *lac* operator) with expression controlled by rhamnose and arabinose. Since the strains used for these latter studies do not consume the inducing sugar, they may provide advantages for ^{13}C-labeling proteins with relatively inexpensive ^{13}C-glycerol. In addition, the "plain" T7 promoter system appears to be more tolerant of different oxygenation states than the T7/*lac* and T5/*lac* promoters.

Introduction

Auto-induction protocols for are attractive for both small and large-scale growths due to the reduction in process monitoring required and the higher achievable cell density compared to traditional IPTG induction. Although originally optimized for T7 promoter based expression, CESG has successfully implemented the auto-induction method for large-scale culture with the T5/lac2 promoter [2,3].

CESG uses *Escherichia coli* B834-pRARE2 for first pass production of selenomethionine-labeled proteins. Figure 1 shows the T5/*lac2* promoter expression vector used for these studies. Overexpression of *LacI* provides strong attenuation of basal expression, thus contributing Small-scale expression screening can provide crucial information on the suitability of target genes for large-scale production. As the small-scale results must be reliably predictive of the large-scale to have utility, CESG has been interested in developing culture methods that incorporate the same host strain, expression plasmid and growth medium for both scales of culture growth.

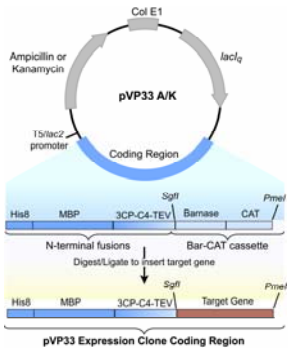

Figure 1: Current expression vector used at CESG for *E. coli* cell based protein production. The coding region contains N-terminal fusions to promote solubility (MBP), allow standardized IMAC purification (His8), allow enhanced detection of fusion proteins (C4 motif provided by the amino acids CCPGCC), and liberation of the target protein using either TEV or 3C proteases. For more information on this and other expression vectors used at CESG, see the poster presented by Russell Wrobel.

Initial Expression Results

CESG uses a three-tier scoring system based on total expression, degree of solubility, and TEV cleavage of the fusion protein in the soluble lysate fraction to assess target suitability. We found that a complex auto-induction medium based on terrific broth (TB) gave comparable expression levels at either scale. However, the use of this medium for selenomethionine labeling was problematic due to the presence of methionine in the medium components. **Figure 2** shows typical expression results for small and large-scale growths using the carbon source concentrations of the original chemically defined auto-induction medium [1]. Our attempts to correlate the results of small-scale screening and the large-scale production growths were unsatisfactory. Often, small-scale expression levels were not high enough to judge the suitability for large-scale expression.

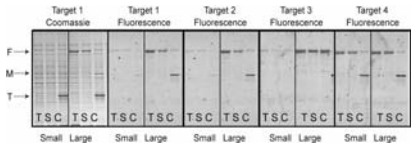

Figure 2: Initial auto-induction expression results shown by SDS-PAGE gels of small-scale screening and large-scale production conducted in a defined auto-induction medium containing 0.05% glucose, 0.2% lactose, 0.5% glycerol, and 0.25% aspartic acid [1]. For target 1, a Coomassie-stained image is shown for comparison. The other gels were imaged after reaction of the fluorophore FLAsH with the tetra cysteine (C4) motif incorporated into the fusion protein. Arrows on the left indicate the location of the fusion protein (F), MBP after TEV cleavage (M) and TEV protease (T, visible only on the Coomassie-stained gel). Expression levels for targets 1-3 were considerably lower for small-scale than for large-scale, while only target 4 exhibited similar expression.

Factorial Design Experiments

Upon the basis of the **Figure 2** results, we began a program to improve the small-scale performance of the auto-induction medium through empirical experiments. The goal was to define conditions that would give consistent screening results, regardless of the expression scale, so as to increase the predictive reliability of small-scale screening.

As a starting point, we considered the large difference observed in the cell density achieved at saturation in small-scale auto-induction experiments (OD_{600} ~ 20-25) and small-scale defined medium experiments (OD_{600} ~10). To determine if the problem was due to some limitation in carbon sources, we investigated the relationship between protein expression levels and the concentration of glucose, glycerol, lactose and aspartic acid or succinic acid using a factorial design approach. **Figure 3** illustrates this experimental space, where concentrations of the different carbon sources were systematically varied as increased, no change, and decreased from the initial state. The changes in protein expression was determined by assay of two proteins expressed from the T5/*lac2* and T7/*lac* expression systems: green fluorescent protein (GFP) and human rhinovirus 14 3C protease (3CP).

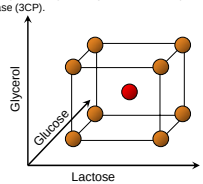

Figure 3: A schematic representation of the experimental space for single step factorial change (*increase, no change, decrease*) of three carbon sources, with the starting point shown in *red*. After each round of experiments, a new center point can be chosen based on the best previous result and the factorial process can be continued.

A linear response model may be used to describe the consequences of the changes in the variables being studied, according to the equation

$$E = C_0 + C_1X_1 + C_2X_2 + C_3X_3 + C_4X_4,$$

where the E is the measured total response, X_i is the variable being changed and C_i represents the partial response coefficient for that variable. Values of the C_i obtained from a factorial experiment for glucose and lactose are shown in **Table 1**. These representative values associate the percent change of GFP fluorescence derived from a 1% change in (wt/vol) of the two carbon sources. For example, in the T5/*lac2* system, a 1% increase in glucose concentration is correlated with a 660% decrease in GFP expression. Likewise, a 1% increase in lactose concentration is associated with a 288% increase in GFP expression. These values can only be considered valid within the range of independent variables evaluated. The response coefficients for glycerol and aspartic acid were evaluated in other, separate experiments.

Values of C_i	T5/*lac2*	T7/*lac*
Glucose	-660 ± 130	-900 ± 350
Lactose	288 ± 37	104 ± 100

Table 1: Response of GFP expression to changes in glucose and lactose concentrations around a midpoint of 0.05% glucose and 0.2% lactose.

Expression Results with Refined Medium

Figure 4 shows representative expression results for small- and large-scale expression using the medium formulation refined for the T5/*lac2* expression system through two rounds of factorial design experiments and several other experiments on single variables.

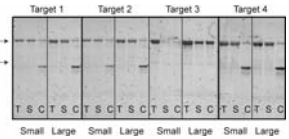

Figure 4: Expression results obtained from modified medium on either small or large-scale for the same four targets shown in **Figure 2**. The carbon source concentrations were: glucose, 0.015%, lactose, 0.5%, glycerol, 0.8%, aspartic acid, 0.375%. Three of the four targets shown had identical small and large-scale scoring results. The typical correlation for production is ~80%, as compared to ~50% correlation with the media used before factorial improvement.

Carbon Source Consumption Patterns

With the modified medium, our PSI-2 small-scale screening and large-scale production results have matched ~80% of the time. This compares to an ~50% match observed from PSI-1 using a TB auto-induction medium for small-scale screening and a defined selenomethionine medium for large-scale production. In order to further define the origin of this improvement, we undertook a more detailed examination of the interactions between the host cell, the expression plasmid and growth conditions on the consumption patterns of different carbon sources in *E. coli*.

For this work, a parallel, six fermenter system (Infors, Bottingen, Switzerland) was used to monitor and maintain the environmental state of cultures under different specified conditions. Oxygenation, plasmid type, and *E. coli* strains were varied during studies of the consumption of each carbon source. The investigation of O_2 is relevant as the differences in growth vessels used for the small- and large-scale experiments suggests significant differences in oxygenation.

Figure 5A shows that the individual carbon source concentrations can be readily determined by HPLC. **Figure 5B** shows that the correlation of carbon source concentration with cell density can be fitted with a sigmoidal curve. **Figure 5C** shows the first derivative of the sigmoidal carbon source concentration plot, corresponding to a consumption pattern for each carbon source.

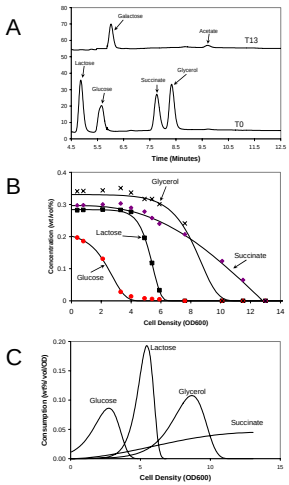

Figure 5: Analysis of carbon source consumption patterns in controlled culture fermentation of *E. coli* BL834(DE3) pRARE2 containing a T7 expression plasmid with no expression of exogenous *LacI* and with O_2 limitation. Sugars and organic acids in the clarified culture medium were analyzed by HPLC as shown in **Panel 5A**. The T0 sample was taken at the beginning of the fermentation and the T13 sample at saturation of growth. **Panel 5B** shows sigmoidal fits to the experimental time course.

The specific consumption pattern, shown in **Panel C**, is obtained by taking the first derivatives of the each of the fits shown in **Panel B**. The pattern of carbon consumption shown; glucose, then lactose, then glycerol is also observed with wild type *E. coli* B834 during growth with O_2 limitation.

Influence of LacI and O_2

Figure 6 shows changes in the specific consumption patterns for lactose (*red*) and glycerol (*blue*) in expression systems providing various levels of *LacI* and O_2. These results demonstrate the interplay between *LacI* and O_2 during carbon consumption leading to auto-induction, and suggest the origin of differential behavior during small-scale screening and large-scale production efforts. The carbon source concentrations were chosen to illustrate the consumption patterns but are different from the concentrations employed for protein production.

Panel 6A shows the same results as in **Figure 5C**. Again, with low *LacI*, lactose is preferentially consumed before glycerol. **Panel 6B** shows that this preference is weakened with high O_2, and lactose and glycerol are consumed simultaneously, providing induction and two carbon sources for protein expression.

With high concentrations of *LacI* (T5/lac^2, **Panel 6C** and **Panel 6D**), the opposite pattern of carbon consumption is observed, as glycerol is preferentially utilized before lactose. Moreover, high O_2 exacerbates this diauxic utilization pattern. *These shifts have important implications, particularly if the oxygenation states in small-scale and large-scale are not well matched.* Furthermore, for NMR studies, glycerol is the main carbon source for ^{13}C-labeling. The carbon consumption profiles such as in **Panel 6C** and **Panel 6D** suggest low efficiency for conversion of ^{13}C-glycerol into expressed protein as *LacI* and O_2 are increased.

Panel 6E and **Panel 6F** show consumption patterns for rhamnose and glycerol. Rhamnose and glycerol consumption coincide for both oxygen states, suggesting that glycerol would be the major souce of carbon during induction for a rhamnose auxotroph where T7 polymerase is under control of the rhamnose operon.

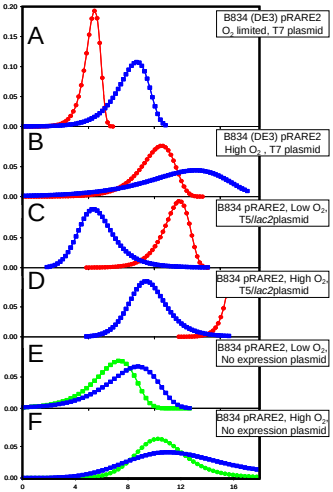

Figure 6: Effect of *lac* repressor dosing and oxygenation on the consumption of lactose. Conditions for each panel are shown. High oxygen indicates the dissolved oxygen was maintained above 20% for the duration of the experiment. Low O_2 refers to a condition where oxygen is added but is the limiting factor in the growth rate. The expression strain for rhamnose autoduction cannot consume rhamnose. Instead, B834 strain *E. coli* (rhamnose autotroph), was used to examine the timing of rhamnose consumption.

Conclusions

A single auto-induction medium composition may not be optimal for all expression systems or growth conditions. This work shows that expression from a T7 promoter system is tolerant of different culture conditions but suffers from previously identified high basal expression. By comparison, we have found that T7/*lac* and T5/*lac*^{2} offer better control of basal expression but suffer from lowered expression with auto-induction under aerobic conditions (typical of small-scale screening). In contrast, rhamnose gives a similar pattern for carbon consumption in either low or high O_2 conditions, offering potential for further unification of small-scale screening and large-scale production efforts. Basal and fully induced expression for strains of *E. coli* with T7 polymerase under control of the rhamnose operon (Promega, Madison, WI), are currently under evaluation.

References

1. Studier, F. W. (2005), Protein Expression Purif. 41, pp. 207-234.
2. Sreenath, H. K. et al, (2005),Protein Expression Purif. 40, pp.256-267.
3. Bujard, H.B., et al, (1987), Methods Enzymology, 155, pp. 416-433.

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative (1 U54 GM 74901).
This work was supported in part through a sponsored research agreement from Promega Corporation to B.G. Fox