Pipeline production of labeled proteins in auto-induction medium

Hassan K. Sreenath, Brendan T. Burns, Teresa N. Tiradani, Kory D. Seder, Craig A. Bingman, Robert C. Tyler, David J. Aceti, Holalkere V. Geetha, Ronnie O. Frederick, John G. Primm, George N. Phillips, Jr., and Brian G. Fox

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

Abstract

The University of Wisconsin Center for Eukaryotic Structural Genomics (CESG) has developed a cell-based protein production pipeline that incorporates the auto-induction strategy introduced by Studier for first-pass production of SeMet-labeled proteins in *Escherichia coli* B834. The entire production cycle from receipt of freshly transformed expression hosts through the growth, induction, and expression was designed to take ~72 h, and capacity for up to 36 growths per week has been demonstrated with existing labor and equipment. The medium used for culture scale-up was adjusted to provide rapid and reproducible growth, while the production cell growth was optimized for timing, cell mass yield, total protein expression, and percentage of SeMet incorporation. At 250 rpm shaking in a standard refrigerated shaker, an average final optical density at 600 nm of greater than 6 and a yield of ~14 g of cell paste were obtained over ~500 growths. Cell growth with shaking at 350 rpm gave an ~1.6 fold increase in OD_{600} and a corresponding increase in the mass of cell paste with similar characteristics for the expressed fusion proteins (total expression, solubility, and proteolysis of the fusion protein to release the target). Increased agitation indicates a potential pathway to improvements in process yield. However, it was not possible to fully load the shaker at this higher agitation rate due to overheating before the 24 h growth cycle could be completed. Engineering efforts to improve the shaker and rack design are under way. Results of these and other features of our SeMet protein production pipeline will be discussed.

Large-scale cell growth team

The major goal of our pipeline team is to reliably produce sufficient mass of bacterial cells expressing recombinant proteins to facilitate subsequent protein purification efforts. By working closely with the Small-Scale Expression (see poster by Ronnie O. Frederick), Protein Purification (see poster by Won Bae Jeon), and X-Ray (see poster by Craig A. Bingman) teams, deliverable goals of 15 g of recombinant cell paste with favorable scoring for total expression, solubility of expressed protein, and percentage of proteolysis from the fusion construct using tobacco etch virus protease were established.

Our team produces SeMet-labeled proteins by heterologous expression in *Escherichia coli* grown using the auto-induction approach introduced by Studier [1]. As compared to other methods for growth and induction of recombinant cultures, we have found the auto-induction method to be reliable and scalable for all types of expression monitoring used at CESG [2]. Moreover, for large-scale cell growth, the auto-induction minimizes the amount of labor required for culture monitoring. As described below, our implementation of the large-scale cell growth pipeline is sufficiently robust that all proteins studied by CESG are labeled with selenomethionine (SeMet) as the first-pass production effort.

Growth vessel and modified floor shaker

CESG uses the 2-L polyethylene terephthalate (PET) beverage bottles, introduced by Donnelly and colleagues from the Midwest Center for Structural Genomics [3], as the growth vessels for bacterial cell culture. In practice, each bottle contains 500 mL of culture medium. CESG currently uses 4 culture bottles per protein target to meet project-defined production goals.

In order to most efficiently use the PET bottles for protein expression, a custom rack for the New Brunswick C25KC refrigerated shaker was developed (**Figure 1**). This rack, built from thick gauge aluminum with supporting plates, gives firm support to the PET bottles during shaking operations up to 400 rpm over at least a 24 h period. Schematics for this rack are available at http://www.uwstructuralgenomics.org and are published elsewhere [2].

The standard C25KC shaker has a 1/8 hp motor and delivers torque of 63 oz-inch. In practice, this configuration was unable to sustain maximum shaking rates at full bottle capacity. Efforts in collaboration with Big Sky Engineering (Middleton, WI) gave an improved shaker design with increased motor power, improved internal mechanics, and external variable speed control. This modified shaker became available in the 2nd quarter of grant year 4, and is now being tested in research mock-ups of the production environment. Results of these tests will be presented elsewhere.

Figure 1. A photograph of the inside of a New Brunswick C25KC refrigerated shaker equipped with a custom rack with capacity for up to 24 2-L PET beverage bottles used as growth vessels for bacterial cell culture.

Time course of auto-induction

Figure 2 shows the time course of protein expression obtained from the auto-induction method. It is important to note that no user intervention is required after inoculation of the final growth bottle until harvest of the culture. **Figure 2A** shows the change in OD_{600} over time for 2 protein targets, while **Figure 2B** shows the corresponding SDS PAGE analysis of the time-dependent changes in the level of protein expression. The strong repression of target gene expression early in the growth is provided by the presence of glucose, tandem *lac* repressor construction of the project pVP16 expression vector (derived from pQE80, Qiagen, Valencia, CA, see [4,5] and the poster by Paul G. Blommel for further details) and by catabolite repression. At ~12-15 h, when glucose has been consumed, the primary carbon metabolism of the culture shifts from glucose to the combination of glycerol and lactose, which causes induction of gene expression. The auto-induction culture can be predictably harvested after 24 h without further evaluation of the status of the cultures. This predictable behavior of the expression host is a major advantage in the structural genomics environment.

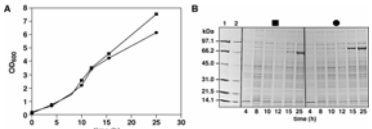

Figure 2. A time course of growth and expression in auto-induction medium at 25°C for 25 h. **A**, increase in OD_{600} for At4g34215.1 (●) and At4g09670.1 (■). **B**, SDS-PAGE analysis of protein expression of At4g34215.1 (●) and At4g09670.1 (■). The amount of total cell protein analyzed was normalized by OD_{600} measurements. Both gels show visible accumulation of the target protein by ~15 h and received an expression scoring of high at 25 h.

Scoring of protein expression

CESG uses 3-tier scoring to evaluate: (1) total protein expression, (2) the level of soluble expression, and (3) the percentage of total soluble fusion protein that can be proteolyzed by tobacco etch virus (TEV) protease.

Figure 3 shows three representative expression gels, and the following text provides an explanation of the scoring method. This scoring is based on visual comparison with the stained intensity of bovine serum albumin (66.2 kDa), which is present at ~8 µg in standard *lane* **1** and ~2 µg in standard *lane* 2. If the target protein band (~45-80 kDa) was more intense than bovine serum albumin band in *lane* 1, the total expression was scored high, "H". If the intensity of the target protein was between that of the marker bands in *lanes* 1 and 2, it was scored medium, "M," and if the intensity of the target protein was less than that of the marker band in *lane* 2, it was scored weak, "W". Similar assessments were made for the lanes containing the soluble fraction, the pellet fraction, and the TEV protease-treated soluble fraction.

MALDI-TOF mass spectrometry was used to identify endogenous *E. coli* proteins useful as control markers for cell fractionation. Chain A elongation factor complex Ef-TuEF-Ts (~45 kDa, GenBank gi:1942721, see **Figure 3**) was a highly soluble endogenous protein, while the insoluble outer membrane proteins 1a (~40 kDa, gi:/16128896) and 3a (~37 kDa, gi:/15800816) were reliably observed in pellet fractions. In process evaluation, complete cell breakage corresponds with the presence of the elongation factor in both the total expression, "T," and soluble, "S," lanes, but its absence from the pellet, "P," lane. In contrast, both of the 1a and 3a membrane proteins will be present in the "T" and "P" lanes and both will be absent from the "S" lane upon complete cell breakage. By including evaluation of these internal control proteins to assure consistency in cell lysis, the CESG scoring system has been successfully standardized across the expression of several hundreds of targets.

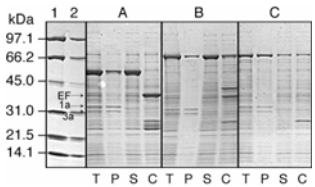

Figure 3. Analysis of total protein expression, solubility, and proteolysis of fusion protein targets using denaturing gel electrophoresis: **A**, AK098872; **B**, AK072216; and **C**, AK065519. The marked lanes were used to evaluate total expression (*T*), soluble expression (*S*), insoluble or pellet expression (*P*), and TEV proteolysis (*C*). Protein molecular weight markers were loaded in the first two lanes (8 µL in *lane* 1 and 2 µL in *lane* 2).

Implications of expression scoring

The 3-tier expression scoring system inherent in the method of **Figure 3** gives rise to 27 unique states that describe the expression, solubility, and proteolysis behavior of a given protein target. In our hands, this method is simple, reproducible, and can be used by a wide range of project personnel to achieve the same scoring with a minimal amount of training.

Our ongoing efforts to record these three scores for all proteins expressed have provided a large and growing database that is also linked to results from small-scale expression, protein purification, crystallomics, and structure determination. **Figure 4** shows how sorting based on the lowest score obtained in any of the three tiers is a reliable predictor of being able to obtain greater than 3 mg yield of purified protein.

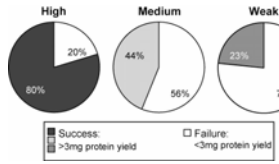

Figure 4. Success in protein purification sorted according to the lowest score given by SDS-PAGE analysis. Success defined to be greater than 3 mg of proteolyzed target protein obtained from standard pipeline purification.

Application of these scoring methods also provides an experimental basis for evaluation of results of computed predictions of solubility, such as those being developed by our collaborator Prof. A. Keith Dunker (Indiana University School of Medicine and Molecular Kinetics). Another emerging power of this database arises from the ability to make predictive responses to improve the outcome from expression trials that gave less than optimal results. These predictive responses are being executed by changes in vector design (see poster by Paul G. Blommel), expression hosts used for small-scale screening (see poster by Ronnie O, Frederick), or protein purification (see poster by Won Bae Jeon) and will become major contributing paths to pipeline production.

Results of SeMet labeling experiments

Table 1 shows results from expression of SeMet-labeled proteins as the first pass of the CESG protein production pipeline including the amount of cells used, the 3-tier score, the yield of purified protein obtained after TEV proteolysis, and the protein yield per gram of cells. **Table 1** also reports the percentage of SeMet incorporated into the protein as determined by mass spectrometry. On a routine basis, greater than 90% incorporation of SeMet is observed with proteins prepared by the CESG implementation of the auto-induction approach. This level of incorporation has proven suitable for structure determination at CESG.

It is also important to note that this percentage of labeling is achieved with a fixed, reproducible time schedule.

TABLE 1

Cell Mass, Protein Yield, and Incorporation of SeMet From Auto-Induction Medium

Protein designator[a]	Cells used (g)	Exp	Sol	Clv	Purified protein (mg)	Protein yield (mg/g)	SeMet incorporation
At1g01470	16.5	H	H	H	43	2.6	85%
At1g06000	18	H	H	M	1.5	0.08	95%
At1g27020	13.5	H	H	H	18.7	1.4	90%
At1g49820	21.0	H	H	H	68.6	3.3	92%
At1g70830.1	15.8	M	H	H	43.8	2.8	92%
At2g25570	16.2	M	M	nd[b]	3.1	0.19	80%
At2g34160	19.5	H	H	H	39.1	2.0	90%
At2g34160	21.3	H	H	H	22.8	1.1	85%
At2g35830	15.7	H	H	H	4.7	0.3	95%
At2g44060	20.8	H	M	M	18.6	0.9	85%
At2g45260	21.5	H	M	nd	15.4	0.71	95%
At2g45790	20.5	H	H	H	49.4	2.4	95%
At3g17820.1	16.6	M	H	M	19.3	1.2	92%
At3g22680	14.0	H	nd	H	36.0	2.6	90%
At5g01610	16.0	H	M	H	25.3	1.6	85%
At5g05060	18.5	H	H	H	32.2	1.7	90%
At5g16510	24.0	H	H	H	52.9	2.2	75%
At5g16510	18.5	H	M	M	49.2	2.7	90%
At5g16570	22.0	M	W	H	17.6	0.8	100%
At5g18200	18.2	H	W	H	6.4	0.4	96%
At5g44450.1	10.6	H	M	H	42.3	4.0	93%
Average[c]	18.0±3				29±19	1.7±1.1	90±6

[a] Large-scale cultures were grown in auto-induction medium containing 125 mg L^{-1} SeMet.

[b] nd, not determined.

Applicability to NMR labeling

CESG has also used auto-induction protocols for the production of [*U*-^{15}N]- or [*U*-^{13}C-, *U*-^{15}N]-labeled proteins [6]. In this case, the large-scale growth and expression uses a chemically defined auto-induction medium containing salts and trace metals, vitamins including vitamin B_{12}, and glucose, glycerol, and lactose. By the addition of vitamin B_{12}, this medium takes advantage of the conditional methionine auxotrophy of *E. coli* B834. Complementation of cobalamin-dependent methionine synthase provides growth rates indistinguishable from other expression hosts used for NMR labeling experiments such as *E. coli* BL21.

Table 2 shows the purified protein yields and labeling efficiency determined for four proteins produced by auto-induction in chemically defined and isotopically enriched medium. The NMR structure of each of these proteins has been solved.

TABLE 2

Efficiency of [*U*-^{13}C, *U*-^{15}N]-Labeling in Auto-Induction Medium

Target protein	Natural abundance molecular weight (Da)	Measured [*U*-^{13}C, *U*-^{15}N] molecular weight (Da)	Calculated [*U*-^{13}C, *U*-^{15}N] molecular weight (Da)	Purified protein (mg)[a]	Labeling efficiency (%)[b]
At1g01470	16,501	17,435	17482	10.0	95
At1g77540	11,690	12,336	12,360	10.0	96
At3g03773	17,322	18,249	18,286	9.3	96

[a] Recovered from standard purification.

[b] Isotopic labeling efficiency.

Costs of reagents versus labor

When the price of isotopic reagents used in the auto-induction medium is considered, it appears that auto-induction may have a higher cost, particularly for NMR studies. However, the savings in labor costs given by the auto-induction approach, and the use of disposable growth vessels at CESG offset the reagent costs. Growths using the auto-induction medium do not require monitoring of optical densities, or multiple centrifugation steps, both of which place significant time demands on project technicians. The use of disposable PET bottles as growth vessels, which do not require cleaning or sterilization, also lend themselves to lower labor costs. Less sample handling streamlines the production efforts and reduces sample attrition caused by human error. By using the auto-induction method described here, one technician with one incubator shaker can manage 12 separate 2-L growths using disposable growth vessels in a 40 h workweek. By comparison, two or more technicians would most likely be required to handle a similar workload using traditional labeling methods and shaker flasks.

Summary

- For SeMet labeling, the auto-induction medium contained a 9.4 fold molar excess of SeMet relative to Met and *no additional vitamin* B_{12} other than carry-over from the scale-up inoculum. This medium also contained glucose and glycerol as carbon sources and α-lactose as both a carbon source and an inducer.
- For NMR labeling, the auto-induction medium contains vitamin B_{12} at all stages of culture scale-up. This effectively complements the conditional auxotrophy of *E. coli* B834.
- Culture growth in the auto-induction medium at 25°C gave an average final optical density at 600 nm of ~6 and an average wet cell mass yield of ~14 g from 2 liters of culture.
- Analysis by mass spectrometry showed greater than 90% incorporation of SeMet and greater than 95% incorporation of ^{13}C and ^{15}N. Proteins prepared by these methods have been successfully used for structure determination.
- From our experience, we conclude that the auto-induction method is superior for high-throughput labeling efforts.

Acknowledgements

We thank Dr. William Studier (Department of Biology, Brookhaven National Laoratory) for generous personal communication on the composition of the auto-induction medium.

References

1. Studier, F. W. (2004). Auto-induction for protein production in inducible T7 expression systems. *2004 Annual Meeting of the American Crystallographic Association, Chicago, IL*. July 17 - 22, 2004.
2. Sreenath, H.K., Bingman, C.A., Buchan, B.W., Seder, K.D., Burns, B.T.,Geetha, H.V., Jeon, W.B., Vojtik, F.C., Aceti, D.J., Frederick, R.O., Phillips, G.N.J., and Fox, B.G. (2005) *Protein Expression Purif.*, in press.
3. Stols, L., Sanville Millard, C., Dementieva, I. & Donnelly, M. I. (2003) *J. Struct. Funct. Genom.* **5**, 95-102.
4. Thao, S., Zhao, Q., Kimball, T., Steffan, E., Newman, C.S., Fox, B.G., and Wrobel, R.L. (2005) *J. Struct. Funct. Genom.*, in press.
5. Blommel, P.G., and Fox, B.G. (2005) *Anal. Biochem.* **336**, 75-86.
6. Tyler, R.C., Sreenath, H.K., Aceti, D.J., Bingman, C.A., Singh, S., Markley, J.L., and Fox, B.G. (2005) *Protein Expression Purif.*, in press.

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative (P50 GM 64598).