Convergent Protocols for Production of Labeled Proteins in Auto-Induction Medium

Hassan K. Sreenath, Craig A. Bingman, Robert C. Tyler, Blake W. Buchan, Kory D. Seder, Brendan T. Burns, Holalkere V. Geetha, Ronnie O. Frederick, Frank C. Vojtik, Won Bae Jeon, David J. Aceti, Shanteri Singh, George N. Phillips, Jr., John L. Markley and Brian G. Fox

Center for Eukaryotic Structural Genomics and Department of Biochemistry, *University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin, USA 53706-1549,* http://www.uwstructuralgenomics.org

Abstract

Protocols were developed for high-throughput production of labeled proteins for X-ray and NMR studies. The protocols take advantage of the conditional auxotrophy of *Escherichia coli* B834 to provide rapid and reproducible growth of the scale-up culture. Protein expression was controlled by use of an auto-induction medium developed by Dr. F. W. Studier (Brookhaven National Lab) containing glucose, glycerol, and lactose. The entire cycle from inoculation of the culture bottle through the growth, induction, and expression was programmed to take ~24 h. Culture growth in the auto-induction medium gave an average final optical density at 600 nm of greater than 6 and an average wet cell mass yield of greater than 10 g/L. Correlations between the cell mass recovered, the level of protein expression, and the relative amounts of glucose, glycerol, and lactose in the auto-induction medium were noted. Analysis by mass spectrometry showed greater than 90% incorporation of SeMet, [U-^{15}N] , or [U-^{15}N, ^{13}C] in separate growths to express protein for the different structural techniques.

Introduction

The major goal of our structural genomic team was to screen a large number of unknown proteins to find novel folding patterns. Conventionally, labeled proteins have been produced by heterologous expression in *Escherichia coli* grown in minimal medium, which requires monitoring of liquid cell cultures as well as manual addition of an induction reagent. While such methodologies are reliable, they become overly labor-intensive in the setting of high-throughput protein production. Labeling proteins with selenomethionine (SeMet) is a well documented process which greatly facilitate analysis of structure by X-ray crystallography. One of the biggest obstacles to NMR-based structural genomics is the reliable and efficient production of stable-isotope labeled proteins.

The chemically defined, auto-induction medium developed by William Studier (Studier, F.W. Auto-induction for protein production in inducible T7 expression systems. Abstracts of the American Crystallographic Association, Series 2, Volume 30, page 43 (2004)), requires reduced handling time and, provides great simplicity and flexibility associated with recombinant protein production and expression within a bacterial system aimed at targeted range of cell density without laborious monitoring of cell concentrations. This medium, which uses lactose as an inducer facilitates large-scale production of SeMet labeled proteins as well as [U-^{15}N] and [U-^{15}N, ^{13}C]- labeled proteins with one cell line (B834-pLacIRARE) and one chemically defined medium that requires minimal adjustment for producing labeled proteins. We describe a high throughput method, using a Met auxotroph and codon-adapted strain *E. coli* B834 pLacIRare as the expression host to produce different labeled targets with an incorporation of near 100%. CESG large scale pipeline growths of a positive expressing target are carried out in a single C25-KC incubator shaker (Figure 1) in 24 h in 2L polyethylene terephthalate (PET) bottles, each containing 500 ml of culture.

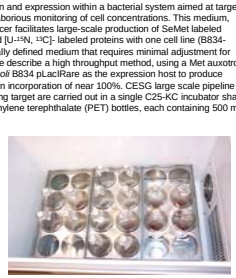


Figure 1 is a photo of modified C 25-KC shaker used to house the 2L PET bottles, each containing a 500 ml culture of *E. coli* transformed with the gene of interest.

Materials and methods

Labeled cell growth media

PA-0.5G medium: The defined medium for the starter culture growth contained (per 100 ml media) ~92.2 ml sterile water, 100 µl 1M $MgSO_4$, 10 µl trace metal mix, 1.25 ml 40% glucose, 5.0 ml 20x NPS, 0.4 ml methionine (25 mg/ml), 1 ml of 17 amino acid cocktail (each amino acid 10 mg/ml), 100 µl ampicillin (100 mg/ml), 100 µl chloramphenicol (34 mg/ml) and 100 µl vitamin cocktail including vitamin B12.
PASM-5052 medium: This medium for large-scale growth for SeMet labeling contained (per liter media) ~901 ml sterile water, 1 ml 1M Mg SO_4, 100 µl metal mix, 20 ml 5052, 50 ml 20x NPS, 0.4 ml methionine (25 mg/ml), 20 ml of 17 amino acid cocktail (each amino acid 10 mg/ml), 5 ml of SeMet (25 mg/ml), 1 ml ampicillin (100 mg/ml), 1 ml chloramphenicol (34 mg/ml), and 1 ml vitamin cocktail (with no vitamin B12).
[U-^{15}N]-P-5052 medium: This medium for large scale growth for ^{15}N labeling contained (per liter media) ~926 ml sterile water, 1 ml 1M $MgSO_4$, 100 µl metal mix, 20 ml 5052, 50 ml 20x [U-^{15}N]- NPS, 1 ml ampicillin (100 mg/ml), 1 ml chloramphenicol (34 mg/ml), and 1 ml vitamin cocktail (with vitamin B12).
[U-^{15}N, ^{13}C] -P-5052 medium: This medium was prepared for double labeling similar to [U-^{15}N] *P -5052 medium* except with a difference of adding 20 ml of [U-^{13}C]-P- 5052 [U-^{13}C glycerol], 0.5%, [U-^{13}C glucose],0.05% and α-lactose, 0.0125%) in a liter.

SeMet-labeled growth

For each target, colonies (1-3) from fresh PA-0.5 G agar media plates were picked to 3 ml PA 0.5 G medium in a test tube and grown 5 to 6 h. at 37°C, 300 rpm. Later, the test tube starter culture was poured into 100 ml PA 0.5G starter culture medium and grown for ~ 20 h at 25°C, 300 rpm. A 20 ml aliquot of this starter culture was used to inoculate 480 ml PASM-5052 medium in 2L PET bottles and was incubated in refrigerated shakers at 250 rpm, 25°C for 22 to 24 h. The cell growth OD was measured at 600 nm. The culture was harvested by centrifuging at 5000 xg for 20 min followed by washing with 40 ml of sonication buffer with 20% ethylene glycol. Protein expression was analyzed by SDS-PAGE (4-20% gradient gel).

[U-^{15}N] / [U-^{15}N, ^{13}C] -P-5052 labeled growth

For each target, colonies (1-3) were grown similar to SeMet growth in 3 ml of PA 0.5G medium in a test tube for 5 to 6 h at 37°C, 300 rpm. Later, the starter culture was poured to 100 ml of PA 0.5G medium and grown for ~20 h at 25°C, 300 rpm. A 20 ml aliquot of this starter culture was inoculated to 480 ml [U-^{15}N]-*P-5052* / [U-^{15}N, ^{13}C]-P-5052 medium in 2L PET bottles incubated at 250 rpm, 25°C for 24 ~ 26 h. The cell growth was analyzed prior to cell harvest similar to SeMet growth.

Table 1 below shows that the auto-induction method gave total yields of purified SeMet-labeled protein after proteolysis in the range of 2 to 70 mg (*73 successful purifications out of 85 attempts, 86%*).

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%
At1g78830.1	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	90%
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 PASM-5052 medium containing 125 mg/L SeMet as described in **Materials and methods**.
[b] nd, not determined.
[c] The average purified protein yield of 39.5 mg was obtained from 173 protein purification trials using cells grown with the auto-induction method and SeMet labeling. The range of protein yield was 0.2 mg to 184 mg.

Figure 2 below shows representative results of expression trials for total expression, solubility, and cleavage by TEV protease. Analysis of total protein expression, solubility, and proteolysis of fusion protein targets using denaturing gel electrophoresis: A) AK099872; 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). In this figure, A is an example of a scoring of high total protein expression, high solubility, and high percentage of proteolysis (H, H, H scoring), B is an example of high total protein expression, high solubility, and medium percentage of proteolysis (H, H, M scoring), and C is an example of medium total protein expression, medium solubility, and weak percentage of proteolysis (M, M, W scoring).

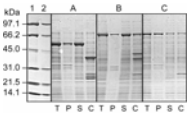


Figure 3 below shows that the induction of protein expression could be clearly detected after ~15 h. A time course of growth and expression in PASM-5052 medium at 25°C for 25 h. Panel A shows an increase in OD_{600} for At4g34215 (●) and At4g09670 (■); Panel B shows an SDS-PAGE analysis of protein expression of At4g34215 (●) and At4g09670 (■). Both gels show visible accumulation of the target protein by ~15 h and received an expression scoring of high at 25 h.

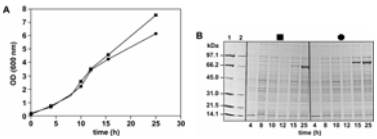


Table 2 below shows that Semet concentrations from 62 to 125 mg/L give faster cell growth without compromising the percentage incorporation.

Effect of SeMet Concentration on the Percentage Incorporation of SeMet

SeMet[a] (mg/ ml)	Cell yield (g L^{-1})	SeMet labeling (%)
62.5	9.9	92.8
125	9.6	95.6
250	6.6	97.1
500	8.2	97.7

[a] Cultures expressing At5g11950.1 in PASM-5052 medium with the indicated amount of SeMet were grown for 24 h at 25°C with shaking at 250 rpm.

Figure 4 below shows growth curves and SDS-scoring for protein expression at four SeMet concentrations. Effects of increasing SeMet concentration on the growth of *E. coli* B834 expressing pVP13-At5g11950. The cultures were grown at 25°C with shaking at 250 rpm for 24 h. Panel A shows growth curves. The SeMet concentrations were (mg L^{-1}): 62.5 (●); 125 (■); 250 (•); and 500 (□). Panel B shows the denaturing gel analysis: *lane* 1, 62.5; *lane* 2, 125; *lane* 3, 250; and *lane* 4, 500.

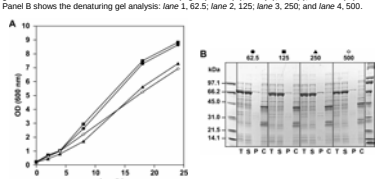


Figure 5 below shows that the highest OD_{600} was obtained after 24 h at 25°C. The cell growth at 20°C and 37°C gave noticeably lower OD_{600} values. Effect of temperature on the growth of *E. coli* B834 expressing At4g19003 from *p*VP13. The growth temperatures were 20°C (□), 25°C (■), 30°C (•) and 37°C (●).

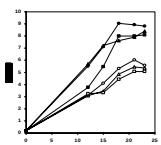


Figure 6 below shows that increased agitation gave a substantial increase in cell mass. Effect of agitation on the growth of *Arabidopsis* targets in PASM-5052 medium. At3g28950 (○ 250 rpm; ● 350 rpm), At4g19003 (△ 250 rpm; • 350 rpm), At1g22270 (□ 250 rpm; ■ 350 rpm) and were grown at 25 ºC for 23 h.

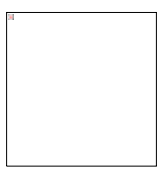


Table 3 below shows that [U-^{15}N]- labeled cell growth gave an average final OD of ~5, an average wet cell mass yield of ~9.5 g/L, and an average yield of ~20 mg of labeled protein

Summary of Expression Results using Auto-Induction Medium

Target Protein	Molecular Weight (Da)	[U-^{15}N] Label	[U-^{13}C] Label	OD$_{600}$[a]	Cell mass (g)	Purified protein (mg)
At4g35640	21,260	yes	no	4.5	17.0	29.6
At1g04240	21,520	yes	no	5.5	19.8	Cleavage –[b]
At5g06450	23,215	yes	no	6.4	22.8	37.5
At4g24380	23,514	yes	no	6.0	19.1	Cleavage –
At1g04250	25,288	yes	no	5.1	19.4	Cleavage –
At5g41910	20,726	yes	no	6.1	22.5	22.4
At1g01470	16,501	yes	yes	5.0	18.7	10.0
At1g77540	11,690	yes	yes	4.7	15.0	10.0
At5g01610	18,952	yes	yes	4.7	14.9	13.4
At3g03773	17,322	yes	yes	4.6	14.7	9.3
Average[c]				5.3 ± 0.7	18.8 ± 2.8	20.5 ± 11.4

[a] OD_{600} value obtained in 2-L culture upon completion of the growth protocol.
[b] Cleavage – indicates that little or no proteolysis of the target protein from maltose binding protein was observed.
[c] The average and standard deviation calculated from the Table 1 entries.

Figure 7 below shows representative expression gels from auto induction trials in minimal medium with ^{15}N-labeling.Denaturing polyacrylamide gel showing over-expression of fusion proteins from an auto-induction medium containing ^{13}C and ^{15}N. Six different *Arabidopsis* proteins were expressed as a fusion with *E. coli* maltose binding protein, with the molecular weight of each fusion being ~66 kDa. The letters at the top indicate the following *Arabidopsis* proteins: *A*, At4g15640; *B*, At1g04240; *C*, At5g06450; *D*, At4g24380; *E*, At1g04250; *F*, At5g41910. The three lanes shown for each protein represent total (T), soluble (S), and insoluble (I) fractions of the clarified cell lysates as described in Material and Methods. Molecular weight markers (kDa) are shown on the left.

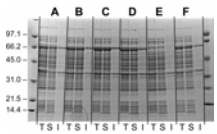


Figure 8 below shows that 0.0125% (w/v) α-lactose was sufficient in the double labeled auto-induction medium to permit expression of target proteins (a 16-fold reduction relative to the original P-5052 medium). Cell growth and expression results from an α-lactose titration experiment conducted with At5g11950. Panel A shows growth curves assembled from OD_{600} measurements at 10 h, 15 h, 22 h, 25 h, and 30 h. The auto-induction cultures were grown in medium containing 0.05% (w/v) glucose and 0.5% (w/v) glycerol with the following α-lactose levels (w/v): A, 0.2% (●); B, 0.1% (◆); C, 0.05% (■); D, 0.025% (•); E, 0.0125% (◆); F, 0.0% (○). Panel B shows SDS-PAGE analysis of fractions collected at 10 h (*left*) and 22 h (*right*) to evaluate protein expression at each α-lactose level. *Lanes* A-F have α-lactose levels as indicated for Panel A, *lane* G had no α-lactose and IPTG added to a final concentration of 1 mM after 12 h.

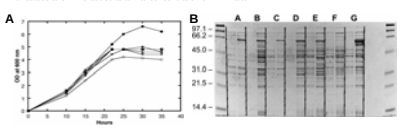


Figure 8 below shows that 0.0125% (w/v) α-lactose was sufficient in the double labeled auto-induction medium to permit expression of target proteins (a 16-fold reduction relative to the original P-5052 medium). Cell growth and expression results from an α-lactose titration experiment conducted with At5g11950. Panel A shows growth curves assembled from OD_{600} measurements at 10 h, 15 h, 22 h, 25 h and 30 h. The auto-induction cultures were grown in medium containing 0.05% (w/v) glucose and 0.5% (w/v) glycerol with the following α-lactose levels (w/v): Panel A 0.2% (●); B, 0.1% (◆); C, 0.05% (■); D, 0.025% (•); E, 0.0125% (◆); F, 0.0% (○). Panel B shows SDS-PAGE analysis of fractions collected at 10 h (*left*) and 22 h (*right*) to evaluate protein expression at each α-lactose level. *Lanes* A-F have α-lactose levels as indicated for panel A, *lane* G had no α-lactose and IPTG added to a final concentration of 1 mM after 12 h.

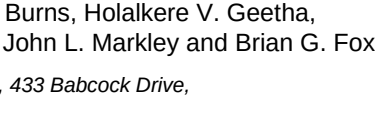


Figure 9 below shows that the OD600 of cell growths using 0.0125% lactose and 0.05% glucose was dependent on the glycerol level. Cell growth results from a glycerol titration experiment conducted with At5g11950. The cultures were grown in auto-induction medium containing 0.05% (w/v) glucose and 0.0125% (w/v) α-lactose with the following glycerol levels (w/v): 0.5%, (●); 0.4%, (■); 0.3%, (◆); and 0.2%, (•). The growth curves were assembled from OD_{600} measurements at 10 h, 15 h, 22 h, 25 h, and 30 h.

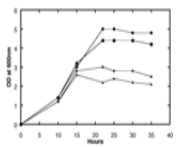


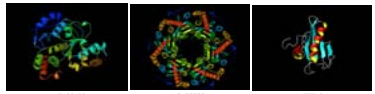
Table 4 below shows that [U-^{15}N]-labeled proteins expressed in the P-5052 auto-induction medium had isotopic incorporation greater than 95% (evaluated by electrospray ionization mass spectrometry) suitable for determination of NMR structure.

Labeling Efficiency Determined for [U-^{13}C, U-^{15}N]-Labeled *Arabidopsis* Proteins Expressed in 2-L Beverage Bottles and P-50501 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[a] (mg)	Labeling efficiency[b] (%)	BMRB Accession number	PDB Code
At1g01470	16,501	17,435	17482	10.0	95	6339	1XO8
At1g77540	11,690	12,336	12,360	10.0	96	6338	1XO4
At5g01610	18,952	19,998	20,027	13.4	97	In progress	In progress
At3g03773	17,322	18,284	18,286	9.3	96	6340	1XO9

[a] Recovered from purification of cell mass reported in Table 1.
[b] Isotopic labeling efficiency as determinate using Table 3 data.

Figure 10 below shows representative structures obtained from cells grown in auto-induction medium. Two structures developed by x-ray crystallographic analysis that have been solved as a result of SeMet labeling: 1) At5g06450.1 (left) is a hexamer with an unknown function; 2) At2g33760.1 (middle) is a putative steroid sulfotransferase; 3) At1g77540.1 (right) solved by NMR as a result of [U-^{13}C, U-^{15}N]-labeling.



At5g06450.1 At2g33760.1 At1g77540.1

Summary of results

Production of SeMet labeled proteins:

1. In addition to salts, amino acids and trace metals, 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.

2. 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 in greater than 150 expression trials.

3. Analysis by mass spectrometry showed greater than 90% incorporation of SeMet. So far 14 X-ray crystal structures were solved by multiwavelength anomalous diffraction phasing using cells grown by auto-induction.

Production of [U-^{15}N] and [U-^{15}N, ^{13}C] labeled proteins:

1. 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.

2. The cell growth in auto-induction medium at 25°C gave an average final optical density of ~5, an average wet cell mass yield of ~9.5 g/L, and an average yield of ~20 mg of labeled protein.

3. [U-^{13}C]-α-lactose was not used in the growth medium due to its cost (~$3200/g). A level of 0.0125% (w/v) α-lactose in the double labeled auto-induction medium was dependent on the glycerol level to permit expression and isotopic labeling efficiency of target proteins.

4. Mass spectral analysis showed that the purified proteins contained both [U-^{15}N] and [U-^{13}C] at levels greater than 95%.

Acknowledgments:

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

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