Center for Eukaryotic Structural Genomics

Protein Structure Initiative

High-Throughput Methods for Producing ^{15}N- and ^{13}C;^{15}N-Labeled *Arabidopsis* Proteins from *E. coli*, for Screening for Optimal Conditions, and for Investigating their Suitability for NMR Structural Analysis

Robert C. Tyler, Blake W. Buchan, Hassan K. Sreenath, Craig A. Bingman, John L. Markley, and Brian G. Fox

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

Abstract

One of the biggest obstacles to NMR-based structural genomics is the reliable and efficient production of stable-isotope labeled proteins. Conventionally, ^{15}N- and ^{13}C;^{15}N-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. The chemically defined, auto-induction medium developed by William Studier, simplifies the production aspects associated with recombinant protein expression within a bacterial system. Some of the many advantages of this approach include: high-density cell growth, the use of α-lactose as the inducer, minimal handling of liquid culture, and virtual ease in maintaining several simultaneous bacterial growths. We present how this approach has been modified and applied to the production of ^{15}N- and ^{13}C;^{15}N-labeled proteins at CESG. Generally, two liters of culture (500 ml in each of four plastic soda bottles) is sufficient to produce ~10 mg of labeled protein. An initial run with [^{15}N]-ammonium chloride as the nitrogen source in the medium is used to produce protein labeled with ^{15}N. A small aliquot (0.5 to 1.0 mg) of this sample is used for micro-drop solubility screening to search for optimal solution conditions. The balance of the sample is used for ^{1}H-^{15}N HSQC NMR analysis to determine whether the protein is sufficiently soluble, folded, non-aggregating, and stable in solution over the time period needed to collect NMR data for a structure determination. If the target passes this screen, the same procedure is used to produce a sample labeled with ^{13}C;^{15}N. In this case the [*U*-^{13}C]-glucose and [*U*-^{13}C]-glycerol are used in the growth medium in addition to [^{15}N]-ammonium chloride. By optimizing the level of lactose, we have found it unnecessary to label it with ^{13}C.

Protein Expression Using an Auto-Induction System

The auto-induction medium developed by William Studier allows for high density cell growth as well as high expression levels of target protein. One of the challenges of CESG was to modify components of the growth media in order to permit efficient production of ^{15}N and ^{15}N;^{13}C labeled proteins using the auto-induction system. The auto-induction medium works with plasmids that exploit either the T7 or T5 promoter regions, which are both regulated by the lac repressor. Freshly transformed cells containing target protein are streaked onto the appropriate plate and allowed to grow overnight. A single colony is selected and placed in 3 ml of a chemically defined medium, which is supplemented with amino acids, and is allowed to grow for approximately 7 h. This entire culture is then scaled up to 100 ml and allowed to grow approximately 18 h to a relatively high cell density in a chemically defined medium that contains only glucose as a carbon source. A 20 ml aliquot of this culture is then used to seed 480 ml of an auto-induction medium, which contains a limiting amount of glucose, and excess amounts of α-lactose, and glycerol. **Table 1** displays a detailed list of all components used in the starter culture and the auto-induction medium. Because α-lactose is used as the inducer, the cell strain used should contain functional lac permease (lac Y) and β-galactosidase (lac Z) to ensure effective auto-induction. The expression cultures are grown for 22 h at which time they are harvested and processed using standard CESG protocols. Typically 2 liters of auto-induction media provides 15 grams of wet cell paste, with yields generally in the range of 10 to 30 mg of purified protein.

Table 1. Chemical Components of Starter and Auto-Induction Medium

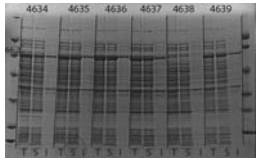
Small Scale Starter	Large Scale Starter	Auto-Induction Medium
1 mM $MgSO_4$	1 mM $MgSO_4$	1 mM $MgSO_4$
0.5% Glucose	0.5% Glucose	0.05% Glucose
100 mM PO_4^{3-}	100 mM PO_4^{3-}	0.5% Glycerol
25 mM SO_4^{2-}	25 mM SO_4^{2-}	0.2% α-Lactose
50 mM NH_4^+	50 mM NH_4^+	100 mM PO_4^{3-}
50 mM K^+	50 mM K^+	25 mM SO_4^{2-}
AA* (100 μg/ml)	Antibiotics (50 μg/ml)	50 mM NH_4^+
Antibiotics (50 μg/ml)		50 mM K^+
		Antibiotics (50 μg/ml)

All percentages are (w/v), all concentrations represented at final volume, AA* refers to amino acids

Uniform ^{15}N-Labeling Using an Auto-Induction Medium

Incorporation of ^{15}N isotopes into target proteins using the auto-induction medium is fairly straightforward. The initial small scale starter culture is brought to maturity using unlabeled chemical components. The large scale starter culture also uses components identical to those listed in **Table 1**, except that the 50 mM ammonium chloride is ^{15}N-labeled. The [^{15}N]-ammonium chloride is also substituted into the auto-induction medium at a final concentration of 50 mM and the cultures are allowed to grow for 22 hours. Because the cells are grown in a chemically defined medium, the only available nitrogen source is [^{15}N]-ammonium chloride; this ensures high level ^{15}N incorporation. **Figure 1** shows SDS-PAGE results from several target proteins grown in ^{15}N-labeled auto-induction medium by CESG.

Figure 1. ^{15}N-Labeled SDS-PAGE Results

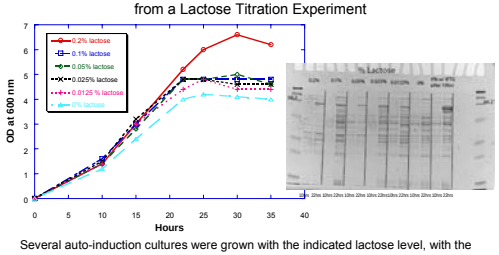


The results from ^{15}N-labeled cell growth using an auto-induction medium. Shown are six CESG targets expressed with a 46 kDa MBP fusion protein. Total weight of each target protein is approximately 66 kDa. The letters T,S, and I refer to the total, soluble, and insoluble protein fractions, respectively.

Uniform ^{13}C;^{15}N Labeling Using an Auto-Induction Medium

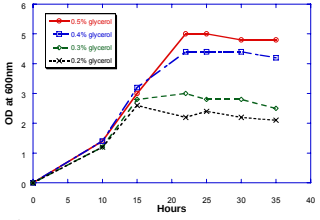
Due to the multiple carbon sources in the auto-induction medium (**Table 1**), ^{13}C-labeling can become problematic. One possibility would be to simply ^{13}C-label all carbon sources in the medium to ensure uniform ^{13}C incorporation. However given the expense of ^{13}C-labeled reagents, this approach is not cost effective in a high-throughput environment. Research was conducted to determine the minimal amount of α-lactose and glycerol required to maintain a high level of cell growth and protein expression. **Figure 2** displays the results from a titration experiment in which auto-induction cultures were allowed to grow in the presence of varying amounts of α-lactose. Results show that although a higher percentage of lactose produces a slightly higher cell density, expression levels are relatively uniform even at extremely low levels of lactose. This indicates that it is possible to supplement the auto-induction medium with [*U*-^{13}C]-labeled glucose and [*U*-^{13}C]-glycerol, while maintaining low levels of unlabeled lactose (e.g. 0.0125%) and still produce proteins that are greater than 95% ^{13}C-labeled based on the carbon mass present in the chemically defined medium. **Figure 3** displays the results from a glycerol titration experiment in which auto-induction cultures were allowed to grow in the presence of varying amounts of glycerol with a lactose level set at 0.0125%. Results indicate that cell density is greatly affected by glycerol concentration and that a higher level of glycerol is required to maintain a optimum level of cell growth. These experiments have revealed that it is possible to use the auto-induction system for double labeling of target proteins. The protocols adopted by CESG for ^{13}C;^{15}N labeling are similar to the procedures described above, expect that in the final auto-induction culture, unlabeled α-lactose is added to 0.0125%, with [*U*-^{13}C]- glucose and [*U*-^{13}C]-glycerol added to 0.05% and 0.5%, respectively. These conditions have been used for double labeling target proteins at CESG and have produced ^{13}C incorporation levels acceptable for standard triple resonance NMR experiments.

Figure 2. Cell Growth and Protein Expression Results from a Lactose Titration Experiment



Several auto-induction cultures were grown with the indicated lactose level, with the glycerol and glucose level set at 0.5% and 0.05%, respectively. 1 ml aliquots were taken at 10 h, 15 h, 22 h, 25 h and 30 h to determine growth curves. Fractions collected at 10 h and 22 h were analyzed by SDS-PAGE to determine protein expression.

Figure 3. Cell Growth Results from a Glycerol Titration Experiment



Cultures were grown with the indicated glycerol level, with the α-lactose and glucose level set at 0.0125% and 0.05%, respectively. 1 ml aliquots were taken at 10 h, 15 h, 22 h, 25 h, and 30 h to determine the growth curves.

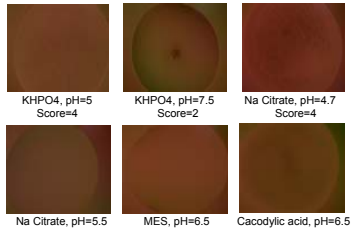
Screening for Optimal NMR Solution Conditions

Following production and purification of ^{15}N-labeled target proteins, samples are screened for their suitability for NMR structural analysis. A major limitation in the screening process is the difficulty in quickly identifying the optimum solution condition that will maximize solubility and chemical shift dispersion. CESG has adopted a procedure developed by C. Lepre and J. Moore of Vertex Pharmaceuticals (*Journal of Biomolecular NMR*, 12 493-499, 1998) called micro-drop analysis, which enables the screening of a large set of solution conditions using very little protein sample. This procedure relies on the use of a 24 well Linbro plate, in which each well is filled with 500 μl of test buffer of varying pH. Next, a siliconized glass cover slip is spotted with 1 μl of a protein solution, typically 10 mg/ml, and mixed with 2 μl of the test buffer. The glass cover slip containing the protein micro-drop is then inverted and sealed onto the corresponding well using silicone grease. The sealed micro-drop is allowed to equilibrate with the test buffer at room temperature. The 24 drops are monitored at regular intervals throughout the equilibration process with a microscope and scored on a scale of 0 to 4 for signs of precipitation. A score of 0 corresponds to no visible precipitation, and a score of 4 corresponds to a drop that is completely covered with precipitated protein. Once the optimum buffer is identified, the screening procedure can be repeated using known stabilizers such as salts, amino acids, and known cofactors. Using this methodology, testing against 24 different buffer conditions and stabilizers requires approximately 0.5 mg of protein. **Table 2** lists the 24 buffer conditions used in the screening procedure. **Figure 4** display examples of protein micro-drops in a variety of buffers with scoring result.

Table 2. Buffer Conditions

1. $KHPO_4$, pH=5.0	7. $NaHPO_4$, pH=7.5	13. Imidazole, pH=8.0	19. TRIS, pH=8.5
2. $KHPO_4$, pH=6.0	8. Na Acetate, pH=4.5	14. MES, pH=5.8	20. HEPES, pH=7
3. $KHPO_4$, pH=7.0	9. Na Acetate, pH=5.5	15. MES, pH=6.2	21. HEPES, pH=8.0
4. $KHPO_4$, pH=7.5	10. Na Citrate, pH=4.7	16. MES, pH=6.5	22. Bicine, pH=8.5
5. $NaHPO_4$, pH=5.5	11. Na Citrate, pH=5.5	17. TRIS, pH=7.5	23. Bicine, pH=9.0
6. $NaHPO_4$, pH=6.5	12. NH4 Acetate, pH=7.3	18. TRIS, pH=8.0	24. Cacodylic acid, pH=6.5

Figure 4. Micro-Drop Examples and Scoring Results



Results from solution optimization screen. Each test buffer listed in **Table 2** is at a concentration of 100 mM, with starting protein solution approximately 10 mg/ml in 10 mM phosphate buffer. Micro-drops and scoring results presented in **Figure 4** were obtained after 48 h at 25 C. The complete scoring results for a target protein are shown in the histogram plot recorded after 48 h. Well position numbers correspond to the buffering condition listed in **Table 2**.

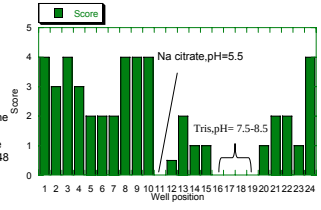
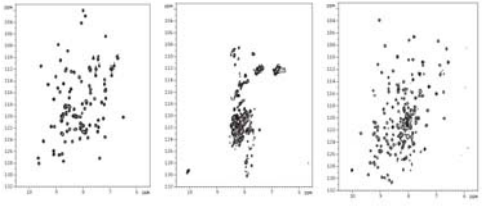


Figure 5. HSQC Spectra of *Arabidopsis* Proteins Used for Structural Analysis



^{15}N-HSQC Structural Analysis

All ^{15}N-labeled target proteins produced by CESG are initially screened for structural elements by NMR. Proteins that display HSQC spectra indicative of a folded state are monitored for 10 days by NMR for stability and solubility. Samples that pass these requirements are then double labeled for structural determination. Small portions of the protein sample are also used for solution optimization if initial solution conditions fail the 10 day quarantine. Most target protein samples are 1 mM in concentration, and require only 20 minutes for HSQC analysis at 400 MHz. **Figure 5** shows examples of HSQC spectra taken on three different *Arabidopsis* proteins all produced using the auto-induction system.

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