High-throughput screening and semi-automated purification of eukaryotic proteins

Ronnie Frederick, Won Bae Jeon, Paul Blommel, John Kunert, Megan Riters, Frank Vojtik, Andrew Olson, Janet McCombs, Jason Ellefson, Craig A. Bingman, John L. Markley, George N. Phillips Jr., and Brian G. Fox

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

Introduction: cell based expression screening of eukaryotic proteins

The Center for Eukaryotic Structural Genomics (CESG) has developed a rapid small-scale high-throughput screen for identifying positively expressed cloned genes isolated from *Arabidopsis thaliana,* rice, human, yeast, and mouse. We have screened expression of over 2,500 genes using *E. coli* host strains, Rosetta-2, and B834-pRARE2, and found that ~64% are successfully expressed at a small-scale 0.4 mL in Terrific Broth (TB). The genes are produced with a N-terminal fusion protein, MBP, and the expression trials were performed in 96-well growth blocks, at 25ºC for 24 hours and using 1 mM IPTG for induction. The genes that expressed were scored "expression +" and recorded in a laboratory information management system (LIMS) called Sesame. These genes were then used to transform a methionine auxotroph *E. coli* host, B834-pRARE (for selenomethionine, ^{15}N, or $^{13}C/^{15}N$ enrichment), and grown on a 2 liter large-scale to produce protein for purification and structural studies. Of ~1,600 cloned genes that were identified to be expressed on a small-scale, ~71% have been attempted in large-scale (2 L), and ~55% have yielded a soluble fusion protein.

Design of expression plasmid

The cloned genes are under the control of a T5 phage based promoter that is IPTG or lactose inducible. The cloned genes are fused to an N-terminal (His)n-tagged (n=6 or 8) maltose binding protein (MBP which enhances solubility and expression levels), and a TEV protease cleavage site is located between the MBP and target protein (just in front of the cloned gene segment). Eukaryotic genes were cloned using PCR and then inserted at the C-terminal end of this (His)n-tag-MBP-TEV cassette fusion using a novel site specific recombination methodology. The architecture of this expression plasmid allows for flexibility in manipulating any future design modifications of the plasmid. For example, the (His)n-tag segment has been increased from 6 to 8 histidine repeats (to improve downstream purification requirements) and the MBP segment can be readily replaced with thioredoxin, glutathione S-transferase (GST), NusA, or other solubility enhancers. Also, with all of the cloned genes being stored in inert "entry vectors" it will be easy to re-insert the genes into any of the future iterations of the expression plasmid. Rosetta-2 is an *E. coli* host strain supplemented with seven rare tRNA codons (codons provided by a plasmid pRARE2) that is used in the small-scale expression screening. A second *E. coli* host, B834-pRARE2, is used for large-scale 2 L cell growths. Figures 1, 2, and 3 show the small-scale expression screening method used at the CESG.

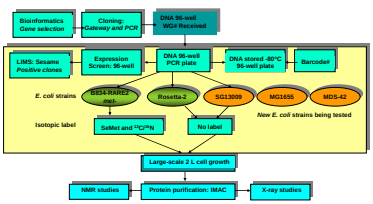

Figure 1. Overview of high-throughput protein production at the CESG. Workgroups (WG) of 96 eukaryotic cloned genes are processed in this pipeline system. Three *E. coli* strains (Rosetta-2, SG13009 and B834-pRARE) are used by large-scale protein production section (LSPP) to produce labeled and wild-type proteins. Protein purification processes the cell pastes generated by LSPP using IMAC methodologies and provides the proteins for structural analysis by NMR spectroscopy and X-ray crystallography.

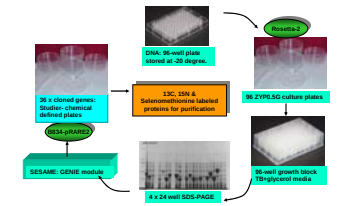

Figure 2. Small-scale high-throughput protein expression screening method. Rosetta-2 cells were transformed and grown on 96 culture plates. The next day, colonies were picked and transferred into a 96-well growth block (0.4 mL of TB media), and induced for 24 hours (with 1 mM IPTG) at 25ºC to screen for positive protein expression. The *E. coli* lysates were analyzed by SDS-PAGE gels. The positive expressing clones were recorded in Sesame.

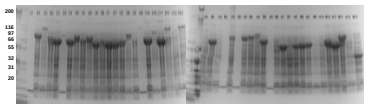

Figure 3. SDS-PAGE gels of high throughput small-scale protein expression screens. The *E. coli* Rosetta-2 cells were harvested from the 96-well growth block, then sonicated, and the cell lysates were analyzed by SDS-PAGE. The expressing proteins were fusions with a N-terminal maltose binding protein (MBP=~46 kDa). These fusion proteins are seen as distinct bands (larger than 45 kDa), and are positioned above the 55 kDa molecular weight band.

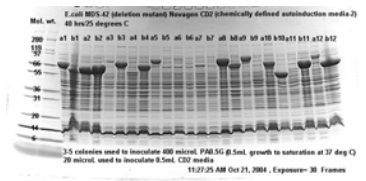

Figure 4. Expression screening using an *E. coli* deletion mutant strain, MDS42-pRARE2, grown in Novagen Overnight Express with selenomethionine. Growth was at at 25ºC for 40 hours (OD-600nm = ~15).

Summary of expression screening results

Stage in CESG	Total	Percent %
Genes isolated by PCR	2,655	100
Genes cloned in expression plasmid	2,252	85
Small-scale expression positive	1,507	57
2 L large-scale protein production	1,070	40
Protein purification and structural analysis	330	12.4

Table 1. Summary of expression screening results for CESG (November, 2004).

Protein purification pipeline

The protein purification pipeline at CESG provides highly purified proteins from *Arabidopsis*, rice, mouse, and human genomes. Eukaryotic proteins are fused with (His)n-MBP double tags (n = 6 or 8) at its N-terminus and expressed in *E. coli*. Using an automated AKTA Purifier system, fusion proteins are initially purified by 1st IMAC capture. After cleavage of (His)n-MBP tags by TEV protease, (His)n-MBP tags are separated from target proteins by 2nd IMAC capture.

Purification of His-tagged MBP fusion proteins

His-tagged MBP fusion proteins were loaded onto a Ni-column directly through a sample pump. Ethylene glycol (20%, w/v) was chosen as a chaotropic agent which decreases nonspecific protein-protein interaction during sample loading. The current protocol applies a linear imidazole gradient to elute fusion proteins. Fusion proteins were eluted from the column between 100 and 300 mM imidazole. The purities of fusion were at least 90% as judged by SDS-PAGE (Fig. 5).

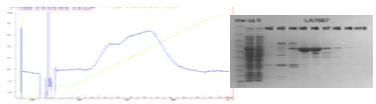

Figure 5. Chromatogram and SDS-polyacrylamide gel of the 1st IMAC capture of His-tagged MBP fusion protein from mouse genome, BC026994.

Cleavage of fusion proteins by TEV protease

TEV protease is a cysteine protease. Thus, it is crucial to keep the TEV protease under reducing conditions and to chelate metal ions which may catalyze the oxidation of the active site Cys residue. After the 1st IMAC capture, EDTA (1 mM final concentration) was added to the combined fractions of fusion proteins to strip any Ni^{2+} ions leached from the Ni-IDA column. TEV protease reaction buffer also contains 0.3 mM TCEP and 100 mM NaCl which minimize the precipitation of fusion or target proteins during TEV proteolysis. In most cases, cleavage of His-MBP tag by TEV protease was complete after 3-5 hrs (Fig. 6). Poor cleavage of fusion proteins was occasionally observed (Fig. 6, LA7419). Uncleaved fusion proteins were analyzed by analytical gel filtration chromatography (Fig. 7). Some target proteins contained misfolded domains or formed soluble protein aggregates, physically blocking the TEV protease cleavage site.

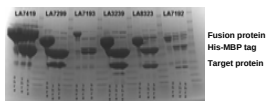

Figure 6. Kinetics of His-tagged fusion protein cleavage by TEV protease. Reaction was performed at 17ºC with a TEV protease to fusion protein ratio of 1 to 100. All proteins were from the mouse genome.

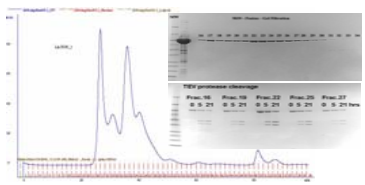

Figure 7. Analytical gel filtration chromatography was used to separate soluble aggregate (fraction 16 and 17), multimeric (fraction 18-24), and monodispersed forms of fusion protein (fractions 25-28). When TEV proteolysis was performed with the fractions of 16, 19, 22, 25, and 27, only 30% of fusion protein was cleaved after 21 hrs. The protein was from the *Arabidopsis thaliana* genome. At1g66200.1 (SeMet-labeled).

Purification of target proteins

During the subtractive IMAC removal of His-MBP tags and (His)6-TEV protease, some target proteins eluted in flow-through fractions (Fig. 8A), and some proteins bound to the Ni-IDA column (Fig. 8B). Target proteins bound to the column might contain intrinsic, surface exposed His residues which form a metal complex with Ni^{2+} ions. Hydrophobic or ionic interaction between target proteins and Ni-IDA column matrix may also slow target elution. Column-bound target proteins were eluted at a concentration of 70 mM imidazole. Generally, the purity of most of the *A. thaliana* proteins was greater than 90%.

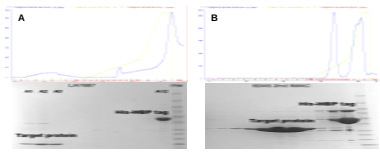

Figure 8. Chromatograms and SDS-polyacrylamide gels of the 2nd IMAC removal of His-MBP tags from target proteins. (A) Protein from mouse genome, BC026994, 7687. (B) Protein from *Arabidopsis thaliana* genome, At2g22530.1, 6345.

Polishing strategies to save target proteins

When the purity of target protein was less than 90%, a polishing step using ion exchange (Fig. 9) and size exclusion column was employed to improve the purity to greater than 90%. Forty-nine proteins were rescued from 123 attempts.

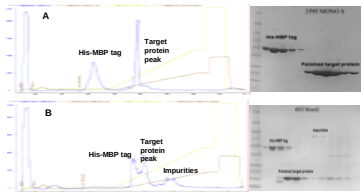

Figure 9. Chromatograms and SDS-polyacrylamide gels of target proteins polished by MonoQ (A) or MonoS (B) column chromatography. (A) Protein from the *Arabidopsis thaliana* genome, At2g14110, 3595. (B) Protein from the *Arabidopsis thaliana* genome At2g14045.1, 6032.

Statistics of protein production

A total of 450 proteins have been produced from 973 purification trials (success rate 46%) with an average yield of 40.1 mg from 2L culture volume (Fig. 10).

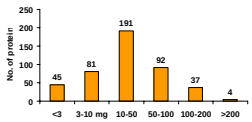

Figure 10. Distribution of target protein production level.

Acknowledgements

Professor William Studier at the Brookhaven National Laboratory for assistance and advice on autoinduction media and auxotrophic *E. coli* strain B834.

Professor Frederick R. Blattner, Dr. David A. Frisch, and Scarab Genomics LLC Madison WI for assistance with *E. coli* strains MDS-42 and MG1655.

PSI

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