

Small-scale platform for high-throughput identification of proteins suitable for X-ray structure determination

Ronnie O. Frederick, Won Bae Jeon, Frank C. Vojtik, Megan A. Ritters, John Kunert, Paul G. Blommel, Fred R. Blattner, Russell L. Wrobel, Craig A. Bingman, Craig S. Newman, John G. Primm, George N. Phillips Jr., John L. Markley, and Brian G. Fox

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

Introduction: Small-scale cell based expression screening and preliminary identification of eukaryotic proteins suitable for structural determination

The Center for Eukaryotic Structural Genomics (CESG) has developed and currently uses a rapid small-scale high-throughput screen for identifying positively expressed cloned genes isolated from various eukaryotic sources such as *Arabidopsis thaliana*, rice, human, yeast, and mouse. So far, of over 2,500 genes that have been screened on the small-scale using *E. coli* host strains Rosetta-2 and B834-pRARE2, we have found that ~64% are successfully expressed at a small-scale (0.4 mL Terrific Broth or TB). Of ~1,600 cloned genes that were identified to be expressed on a small-scale, ~71% have been attempted on a large-scale (2 L) using methionine auxotroph *E. coli* host, B834-pRARE (that is required for selenomethionine, ¹⁵N, or ¹³C/18O enrichment). Of these, ~55% have yielded a soluble fusion protein that was finally suitable for X-ray crystallization screening trials (see Table 1). Our present screening protocol has been designed for speed and enables positive expressing targets to be identified within 3 days (see Figure 1 and 2). We are currently developing and implementing new protocols that increase the total number of successful soluble targets going through to X-ray crystallization screening and reduce the numbers of failed targets that occur downstream (see Figure 3). Table 1 shows that there is a significant attrition of target proteins from the current CESG pipeline scheme. We are currently implementing new protocols and methodologies that enable earlier screening and removal of unsuitable target proteins at the small-scale expression screening stage. The changes in the CESG pipeline design required to achieve these improvements are: cell culture production (using various *E. coli* host strains) and growth at a small-scale, micro-scale protein production and purification, TEV cleavage (of fusion enhancers), protein detection using novel fluorescent stains, and protein solubility and mass analysis. We have successfully used these new approaches to identify and purify large amounts of several different target proteins using small-scale or micro-scale cell based protein production and purification techniques. These proteins were then used successfully in X-ray crystallization screens using a Fluidigm microcircuit system.

Background: Design of expression plasmid and strategies for producing soluble target proteins

The eukaryotic genes selected by CESG are fused to an N-terminal (His)₆-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). The transcription of the cloned genes is controlled by a T5 phage based promoter that is IPTG or lactose inducible. The eukaryotic genes were amplified using PCR and inserted at the C-terminal end of this (His)₆-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)₆-tag segment has been increased from 6 to 8 histidine repeats (this has improved downstream protein purification steps) and the MBP segment can be readily replaced with thioredoxin, glutathione S-transferase (GST), NusA, or other solubility enhancers. In addition, with all of the cloned genes stored in inert "entry vectors", it will be easy to re-insert the genes into any of the future iterations of the expression plasmid.

For small-scale expression screening, we use Rosetta-2, an *E. coli* host strain supplemented with seven rare tRNA codons (codons provided by a plasmid, B834-pRARE2). B834-pRARE2 is used for large-scale 2-liter cell growths. Figures 1 and 2 show the overview CESG pipeline workflow and small-scale expression screening methods. The small-scale expression trials are performed in 96-well growth blocks, at 25°C for 24 hours using either 1 mM IPTG for induction of an autoinduction method developed by Professor Studier (Novagen Overnight Express Autoinduction System™). Presently, all genes that expressed successfully on a small-scale were scored "expression +" and recorded in a laboratory information management system (LIMS) called Sesame. These positively expressed genes were then used to screen B834-pRARE2 and grown on a 2-liter large-scale to produce protein for purification and structural studies.

Table 1. Summary of expression screening results.

Stage in CESG pipeline	Total	Percent %
Genes isolated by PCR	3,238	100
Genes cloned in expression plasmid	2,632	81
Small-scale expression positive	1,691	52
2-L large-scale protein production	1,361	42
Soluble proteins	1,181	37
Protein purification and structural analysis	350	11

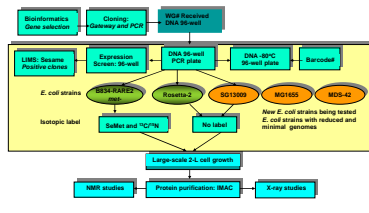


Figure 1. Summary of the high-throughput protein production CESG pipeline. Workflows (WGS) of 96 individual eukaryotic cloned genes are processed in this system at one time. Three types of *E. coli* host strains (Rosetta-2, SG13009 and B834-pRARE) are used by the large-scale protein production section (LSP) to produce proteins. Protein purification processes the cell pastes generated by LSP using IMAC methodologies and provides the pure proteins for structural analysis by NMR spectroscopy and X-ray crystallography.

Comparison of the current small-scale expression screen and the new small-scale protein expression and purification platform being implemented at CESG

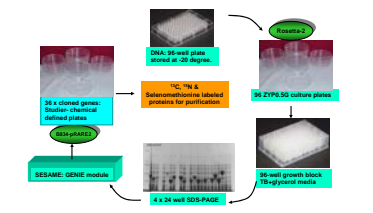


Figure 2. Rapid small-scale high-throughput protein expression screening method. On day 1, Rosetta-2 cells are transformed and grown on 96 culture plates. On day 2, colonies are picked and transferred into a 96-well growth block (0.4 mL of TB media), and induced for 24 hours (with 1 mM IPTG) at 25°C. To screen for positive expressing clones, the *E. coli* lysates are analyzed by SDS-PAGE gels on day 3. The positive expressing clones are recorded in Sesame on day 4. Large-scale growths are started with freshly transformed B834-pRARE2 cells in the following week.

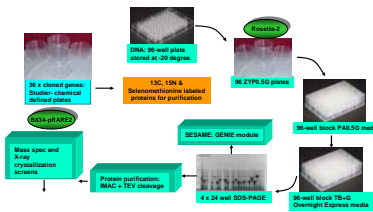


Figure 3. New scheme for small-scale high-throughput protein expression screening and purification. On day 1, Rosetta-2 cells are transformed and grown on 96 culture plates. On day 2, single colonies are picked and transferred into a 96-well growth block (0.4 mL of TB media), and grown for 24 hours at 27°C. To screen for positive expressing clones, 20 mL of cell culture are transferred into a second growth block of autoinduction TB-glycerol Novagen Overnight express media. After growth at 25°C for 24 hours, the cells are harvested and *E. coli* lysates were analyzed by SDS-PAGE gels on day 4. IMAC chemistry was used to isolate the positively expressed proteins. Soluble target proteins are cleaved from their fusion partners (MBP) and re-purified, and those that were still soluble and well behaved are checked by mass spec and then sent for X-ray crystallization trials. Large-scale growths (using freshly transformed B834-pRARE2 cells) are initiated only of those targets that showed favorable results on the small-scale.

Small-scale purification

The cells used for expressing target proteins were Rosetta-2 cells (using Novagen Overnight Express Autoinduction System™). One 96-well growth block was processed (16 times 0.5 mL of culture per target protein). For X-ray crystallization trials ~50-200 micrograms of purified target protein is needed (10 mg/mL) in a volume range of 5-50 microliters.

For protein purification, we use an automated AKTA HPLC. Fusion proteins are initially purified from ~8 ml cell volume by 1st IMAC capture using 1 mM Ni affinity column (Amersham Biosciences). After desalting, fusion proteins are subjected to TEV proteolysis. After cleavage, (His)₆-MBP fusion tag and (His)₆-TEV protease are separated from target proteins by a 2nd IMAC capture. Target proteins are desalted against the buffers specified according the pI values of targets. Target are concentrated to approximately 10 mg/ml by using Amicon Microcon™ YM-3 or YM-10. Final protein concentrations are determined by BCA protein assay using BSA as a standard protein.

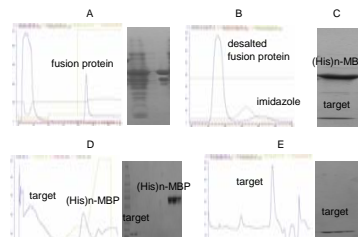
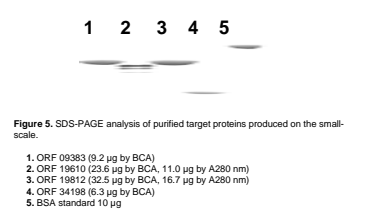


Figure 4. Chromatogram and SDS-polyacrylamide gel for the small-scale purification of mouse protein, BC026994. (A) IMAC capture of His-tagged MBP fusion protein, (B) desalting of fusion protein, (C) TEV proteolysis of fusion tag isolation of target by subtractive 2nd IMAC, and (E) final desalting of target protein.

Summary of small-scale purification yield

ORF	1 st IMAC yield after desalting ^a (mg)	Purity of fusion protein (%)	Purity of target protein (%)	Final volume (μl)	Final conc. (μg/ml)	Expected yield ^b μg	Final yield ^c μg (%)
00605 ^d	2.5	90	95	-	-	0.27	-
09383	3.1	90	95	21	9.2	0.57	193 (33.9) ^e
14940 ^f	4.8	90	>40	-	-	0.79	-
19610	4.3	80	70 ^g	18	23.6	0.39	198 (50.8)
19812	6.4	90	95	21	32.5	1.86	683 (36.7)
34198	5.7	70 ^g	95	10	13.3	0.65	133 (20.5)
34255 ^f	ca. 2.0 ^h	90	95	-	-	-	-

^aFrom 2 ml (8 ml cell culture) of soluble fraction.
^bEstimated yield was based upon the yield achieved in large-scale purification.
^cDetermined by BCA protein assay.
^dPurity of target was greater than 95%, however, protein disappeared during gel filtration (1st batch) and concentration (2nd batch).
^eFinal yield/Expected protein yield x 100.
^fTarget was found in flow-through fraction after 2nd IMAC and aggregated with uncleaved fusion protein. Purification failed.
^gGreen fluorescent protein. Purified target was not homogenous. Truncation occurred? Need to analyze MS data carefully.
^hMajor contaminant was truncation product that migrates at the size of MBP on SDS-Polyacrylamide gel.
ⁱPurity 95% after gel filtration but too little amount to continue. Purification stopped after gel filtration.
 Yield was estimated from the SDS-Polyacrylamide gel.



TEV protease cleavage analysis of targets at the small-scale using whole cell lysates

We are also using small-scale protein expression screening to evaluate the likely success of TEV cleavage. This protocol uses cell lysates without initiating IMAC based protein purification, and offers a rapid testing of the success rate of TEV cleavage of target proteins. We were able to match ~90% of the successful TEV cleavages achieved at the large pipeline-scale using previously assessed target proteins.

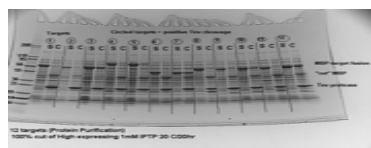


Figure 6. SDS-PAGE gel (4-20%) of 36 targets (soluble vs. cut with TEV). The cleaved target protein can be seen in lanes labeled 8c, 9c, 10c and 11c.

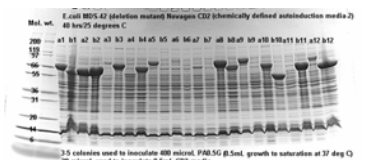


Figure 7. Alternative *E. coli* host strains have been used to screen for positive target protein expression. For example, we have been using an *E. coli* deletion mutant strain, MDS42-pRARE2 to screen for expression using Novagen Overnight Express™ Autoinduction System media (supplemented with selenomethionine). Growth was at 25°C for 40 hours and the final OD at 600 nm reached ~15. Higher levels of expression have been seen for selected target proteins in the MDS42-pRARE2 *E. coli* strain (deduced genome with extra insertion elements and several genes deleted).

Summary

We are improving our protein production system to allow earlier screening of potentially suitable protein targets by expressing and purifying proteins on the small-scale. We are also developing new methodologies to increase the total numbers of protein targets that make their way to the final structural determination stages. These new protocols are being designed and implemented around a semi-automated and robotically based system, to increase the total numbers of iterations and permutations of parameters that can varied and screened.

Acknowledgements

Cris Cowan at Promega, Madison, Wisconsin, USA.
 Professor William Studier at the Brookhaven National Laboratory for assistance and advice on autoinduction media and auxotrophic *E. coli* strain B834.
 Dr. David A. Frisch and Scarab Genomics, LLC, Madison, Wisconsin, USA for assistance with *E. coli* MG1655 and MDS-42 strains.