Small-Scale Screening of Eukaryotic Protein Expression, Solubility, and TEV Protease Cleavage

Ronnie Frederick, Megan Riters, John Kunert, Andrew Olson, Frank Vojtik, 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

One of the main goals of the NIH Protein Structure Initiative (PSI-1 & 2) is to develop high-throughput technologies to assist protein structure determination. The University of Wisconsin-Madison Center for Eukaryotic Structural Genomics (CESG) has developed a rapid, small-scale, high-throughput screening method for identifying positively expressed cloned eukaryotic genes (from *Arabidopsis thaliana*, rice, human, yeast, zebrafish and mouse and others) suitable for downstream large-scale protein production and subsequent protein purification efforts. The genes are produced with maltose-binding protein (MBP) as an N-terminal fusion to enhance solubility and folding.

All protein fusion targets investigated by cell-based methods are now screened in a high-throughput format of 96-well growth blocks in 0.4 mL culture volume. Each target is scored as as a three-tiered scoring of total expression, solubility and protease cleavage and the results are recorded in the Sesame LIMS system. Targets are scored "suitable" if they are overexpressed, highly soluble and cleavable by TEV protease (MBP cut from protein target). The expression clones given suitable screening scores are then transformed into B834-pRARE2 (for selenomethionine, ^{15}N, or $^{13}C/^{15}N$ labeling) and grown at the 2-L scale to produce sufficient protein for purification and structural analysis.

For PSI-2, we continued to improve this small-scale screening method and are now able to successfully predict with ~86% success whether a protein fusion target will be "suitable" for large-scale protein production and protein purification.

Expression Plasmid Design

The cloned eukaryotic genes are expressed under control of either a T5 or T7 phage based promoter that is lactose inducible. The expressed proteins are fused to an N-terminal (His)n-tagged (n=6 or 8) maltose-binding protein (MBP), which enhances solubility and expression. The construct includes a TEV protease cleavage site (located between the MBP and target protein). The modular architecture of the expression plasmid allows for flexibility for future design modifications. For example, the MBP segment can now be readily replaced with many other solubility enhancers such as glutathione S-transferase (GST), thioredoxin or NusA. The *E.coli* host strain, B834-pRARE2 (supplemented with seven rare aminoacyl tRNAs) is used for both small scale expression screening and for large-scale protein production (2L cell growths). Figures 1 shows an overview of the CESG protein production pipeline.

CESG's *E. coli* Based Protein Production Pipeline

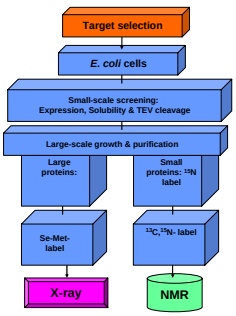


Figure 1. Overview of the CESG high-throughput protein production pipeline. Workgroups of 96 eukaryotic genes are processed in parallel from target selection through small-scale screening. Large-scale protein production (LSPP) produces isotopically labeled proteins for protein purification. Proteins are sent for structural analysis by both NMR spectroscopy and X-ray crystallography.

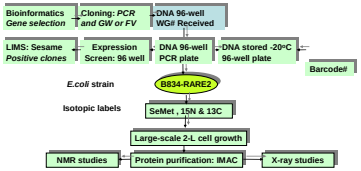


Figure 2. Details of the high-throughput protein production pipeline. Workgroups of 96 eukaryotic cloned genes are processed and screened in parallel, on the small-scale, for positive expression, solubility, and TEV protease cleavage using the *E. coli* host strain B834-pRARE2. Large-scale growths are used to produce isotopically enriched proteins for IMAC purification leading to samples for structural analysis by NMR spectroscopy and X-ray crystallography.

IPTG-Screen

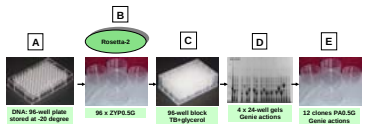


Figure 3. Original PSI-1 small-scale expression screening method: IPTG and auto-induction based screening using Rosetta-2 and B834-pRARE2 cells, respectively. Protein expression is performed in a 96-well format and analyzed by SDS-PAGE.

PSI-2 Small-Scale Auto-Induction Screen

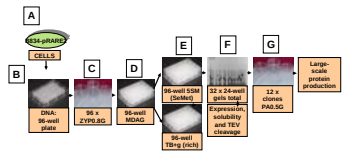


Figure 4. Our new PSI-2 small-scale auto-induction screening method using the methionine auxotroph, B834-pRARE2. Cells are processed in a 96-well format. Cell lysates are analyzed by SDS-PAGE (32 x 24-well 4-20% Biorad gels in total), and highly expressed, soluble, and TEV protease cleavable targets are passed on to large-scale growth. The advantages of 5SM medium are discussed in the poster by Blommel et al.

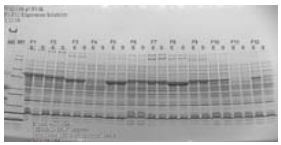


Figure 5. Small-scale expression and solubility screen method and results. The protein expression levels were analyzed by SDS-PAGE, and the total protein expression (E) and soluble fraction (S) were compared. Suitable targets were judged to be those proteins that had at least 30% of the total expression in the soluble fraction (for example target F1). The protein in lane F12 is an example of a highly expressed but insoluble and "unsuitable" target. For the protein targets to be "suitable" they also had to pass the TEV protease cleavage analysis (see **Figure 6**).

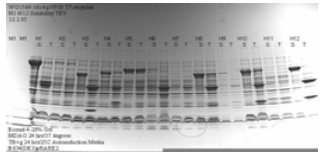


Figure 6. Small-scale solubility and TEV cleavage screen method. The expressed soluble proteins were treated with TEV protease to determine how well the target protein could be liberated from the N-terminal MBP fusion. Results were analyzed by SDS-PAGE. The target in lane H3 shows that the highly soluble fusion protein (H3-S) can be completely cleaved to give free MBP (lane H3-T) and target. The target H5 is an example of an unsuitable protein that fails TEV protease cleavage.

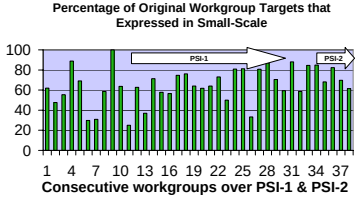


Figure 7. Expression rates of protein targets as a percentage of 96 proteins present in workgroups (for PSI-1 and PSI-2). In PSI-1, 64.5% of workgroup target proteins expressed on the small-scale (0.4 mL) from combination of both IPTG and auto-induction method results. From PSI-2, this fraction has increased to 67.7% using a single host, B834-pRARE2, and auto-induction methods.

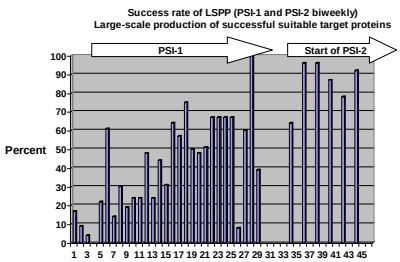


Figure 8. This figure shows the percentage of successive workgroups that are scored suitable after the 2 liters cell growths for large-scale protein production (LSPP). For the last year of PSI-1 (CESG Year-4), the average LSPP success rate was 44.5%. In PSI-2, the average success rate for producing suitable targets on the LSPP is 85.5% (about a two-fold improvement). This change is attributed to better selection of suitable targets by the small-scale screening, i.e., the screening is more effective in identifying targets that will be successfully scaled up to the production level.

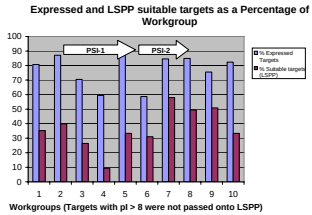


Figure 9. Comparison of the expressed targets from a workgroup (wg) versus the large-scale suitable targets. In the PSI-2, workgroups are providing an ~2-fold greater number of targets (~60% of a wg) suitable for large-scale protein production as compared to PSI-1.

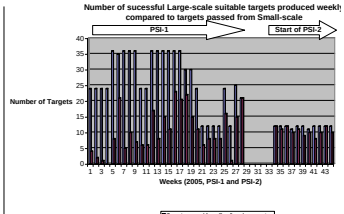
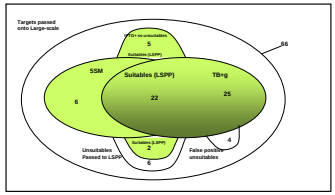


Figure 10. For PSI-2, the productivity of large-scale protein production has increased, in part, due to improvements in the small-scale screening methods used to decide which proteins are suitable for large-scale production. In PSI-2, we have decreased the number of targets passed onto LSPP by about one third. However, the number of these targets that are successfully produced in large-scale has increased to greater than 80%.

A detailed analysis of individual workgroups is ongoing. We have determined false positives and negatives for some PSI-2 workgroups. For example, for a workgroup of biomedically important eukaryotic protein, the small-scale prediction success rate was 83.3%, i.e., 55 out of 66 targets gave the same behavior in the large-scale as predicted from the small-scale work. We also determined that the small-scale screening gave a false positive frequency of 6.1% (targets judged suitable by small-scale that failed in large-scale), and a false negative frequency of 10.6% (targets that were judged unsuitable on the small-scale, but were subsequently suitable in large-scale).



Expression Analysis WG1323

Total LSPP growths = **66**
Total LSPP failures = **10**
Unsuitable in LSPP = **6**
Unsuitable (small-scale error) = **4**

Small-scale predicted **49** *(suitable)* **+ 6** *(unsuitable)* correct

(True positive + true negative), therefore success rate

= **49+6/66 = 55/66 = 83.33%**

False negative rate = 7/66 =10.6%

False Positive rate = 4/66=6.1%

Conclusion

For PSI-2, small-scale screening at CESG has achieved an 85% success rate for predicting whether a protein fusion target will also succeed in large-scale protein production. Going forward, we count only those protein fusion targets that are highly expressed, soluble and TEV protease cleavable at both small-scale and large-scale. Successful large-scale targets are passed onto protein purification, where the project has an ~55% success rate for successful purification of proteins from the suitable category.

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative (1 U54 GM074901)