Published in final edited form as:

Curr Protoc Protein Sci. ; 90: 5.27.1–5.27.20. doi:10.1002/cpps.39.

Screening Fusion Tags for Improved Recombinant Protein Expression in *E. coli* with the Espresso® Solubility and Expression Screening System

Eric Steinmetz and

Lucigen Corporation, 2905 Parmenter St., Middleton, WI 53562. Phone 608-203-9512.

esteinmetz@lucigen.com

Michele Auldridge

Lucigen Corporation, 2905 Parmenter St., Middleton, WI 53562. Phone 608-203-9516.

mauldridge@lucigen.com

Abstract

The simplicity, speed, and low cost of bacterial culture make *E. coli* the system of choice for most initial trials of recombinant protein expression. However, many heterologous proteins are either poorly expressed in bacteria, or are produced as incorrectly folded, insoluble aggregates that lack the activity of the native protein. In many cases, fusion to a partner protein can allow for improved expression and/or solubility of a difficult target protein. Although several different fusion partners have gained favor, none is universally effective, and identifying the one that best improves soluble expression of a given target protein is an empirical process. This unit presents a strategy for parallel screening of fusion partners for enhanced expression or solubility. The Espresso Solubility and Expression Screening System includes a panel of 7 distinct fusion partners, and utilizes an extremely simple cloning strategy to enable rapid screening and identification of the most effective fusion partner.

Keywords

Protein expression; solubility; fusion tag; autoinduction; Espresso

INTRODUCTION

The exploitation of *E. coli* as a factory for the production of recombinant proteins is well within the capabilities of almost any moderately equipped research laboratory. The vast literature and variety of bacterial expression vectors available are testament to the power of *E. coli* as a protein production host. For all of its advantages, however, the expression of heterologous proteins in *E. coli* is not free from challenges. One of the most commonly encountered roadblocks is a failure to express recombinant protein in soluble form. In fact,

Correspondence to: Eric Steinmetz.

INTERNET RESOURCES

Additional background on the Espresso Solubility and Expression Screening System and related Espresso products, including vector sequence information, is available at <http://www.lucigen.com/Protein-Expression/>

estimates are that considerably more than half of heterologous proteins may fail to fold correctly when expressed in bacteria, or may even fail to express at all (Esposito and Chatterjee, 2006).

Fortunately, a number of strategies exist to rescue bacterial expression of recalcitrant proteins. The subject of this unit is a popular protein fusion strategy: the simple expedient of expressing a challenging target protein as a fusion to a specific protein partner can often improve its yield and/or solubility. Several such fusion partners have gained wide popularity, including maltose binding protein (MBP), glutathione S-transferase (GST), and small ubiquitin-like modifier (SUMO) among others. Each of these has a proven track record of improving expression and/or solubility of challenging recombinant proteins to which they are fused. Dozens of other fusion partners have also been shown to be highly effective at promoting improved expression or solubility.

Despite these successes, the fusion strategy is not universally effective. Each protein of interest may respond uniquely to the different fusion partner options available. Since there is no way to predict which fusion partner will perform best for a particular target protein, identifying the best partner from the large and growing number of available choices is an empirical process. Further compounding this challenge is the diversity of cloning strategies and induction systems that accompany different commercially available fusion vector systems. Evaluating each of these systems one-by-one can add significant time and cost to a project. The use of different vector backbones, promoter systems, and linker sequences also introduce variables into the screening process, making it difficult to directly compare fusion partner effectiveness.

The Espresso Solubility and Expression Screening System addresses these problems by providing a panel of fusion partners contained in a suite of expression vectors having a unified design. The System consists of 8 different vectors, 7 of which encode a unique solubility- or expression-enhancing partner joined to a His6 tag to enable metal-affinity purification. The eighth vector contains only the His6 tag, and serves as a control. The universal design of the vectors enables a single amplicon encoding the protein of interest to be cloned into each of the eight vectors in parallel for direct, side-by-side comparison of tag efficacy. Expression of each fusion protein is driven by the versatile rhamnose-inducible RhaP_{BAD} promoter (Egan et al., 1993, Haldimann et al., 1998, Giacalone et al., 1996; Wegerer et al., 2008).

This Unit presents a series of protocols to support the rapid screening for expression and solubility of challenging target proteins using the Espresso Solubility and Expression Screening System. The process begins with the extremely simple Espresso cloning strategy, which harnesses endogenous recombination machinery of *E. coli* to construct 8 different expression clones in parallel with minimal effort (Figure 1). Methods for rapidly screening the expression and solubility of the 8 fusion clones are presented. Analysis continues with the purification of soluble fusion proteins, followed by separation of the fusion partner from the protein of interest by cleavage with Tobacco Etch Virus protease.

BASIC PROTOCOL 1

CLONING A TARGET GENE INTO pSOL VECTORS

The Espresso Solubility and Expression Screening System consists of a panel of 8 vectors of unified design, each containing a different fusion partner (Figure 2). The universal design allows a single amplicon to be cloned in parallel into each of the 8 vectors, without extra residues encoded by restriction enzyme or recombinase sites that may negatively affect expression, solubility, or activity of the encoded protein. The method relies on endogenous recombination machinery of *E. coli*, which promotes homologous recombination between short tracts of identical sequence near the ends of linear DNA fragments (Bubeck et al., 1993). Each of the 8 pSol vectors is provided in a pre-processed linear format, with sequence encoding a unique solubility tag followed by a universal Tobacco Etch Virus (TEV) protease cleavage site at one end, and a stop codon followed by universal vector sequences at the other end. The gene of interest is amplified with primers that append 18 bp sequences matching these universal ends of the pre-processed vectors. Following confirmation of successful amplification, the PCR product is co-transformed with each of the vectors directly into chemically competent *E. coli* cells. Native homologous recombination machinery then joins the insert and vector, reconstituting a circular plasmid that encodes the target protein as a fusion to one of 7 unique solubility tags or to a control His6 tag. The cloning process is extremely simple, with no requirement for enzyme treatment. In most cases the PCR product need not be purified.

Materials

- *8 pre-processed pSOL vectors
- *E. cloni 10G chemically competent cells
- *Recovery Medium
 - *Included in Espresso Solubility and Expression Screening System kits.*
- DNA encoding protein of interest
- Custom primers for amplification of gene encoding protein of interest
- Proofreading PCR polymerase and reaction buffer
- dNTP solution
- Optional: Qiaquick Gel Extraction kit (Qiagen) or equivalent
- LB Lennox plates containing 30 µg/mL kanamycin
- Programmable thermal cycler
- Agarose gel electrophoresis equipment
- Ice
- 42°C water bath
- 37°C shaking incubator

37°C plate incubator

Protocol steps

1. Design custom primers to amplify the target gene for cloning into pSOL vectors. The primers must add flanking sequences specified below which match the sequences at the ends of the linear pSOL vectors:

The forward primer must have the general structure **5'-AAT CTG TAC TTC CAG GGT XXX XXX XXX XXX XXX XXX ...-3'**. The required sequence at the 5' end of the forward primer includes eighteen bases encoding 6 of the TEV recognition site amino acids within the pSol vectors. The 3' part of the forward primer (represented by **XXX**) matches the sequence of the target gene, usually beginning with the second codon. Because each pSol vector contains an ATG initiation codon immediately preceding the polyhistidine (6×His) codons and solubility tag, the ATG initiation codon from the target gene is not needed.

The reverse primer must have the general structure **5'-GTG GCG GCC GCT CTA TTA XXX XXX XXX XXX XXX XXX ...-3'**. The required sequence at the 5' end of the reverse primer matches 18 bases of the downstream end of the pSol vectors. The 3' part of the reverse primer (represented by **XXX**) is complementary to the sense strand of the target gene, usually including the last codons prior to the stop codon of the coding region. Because the pSol vectors include in-frame stop codons, it is not necessary to include the stop codon of the target gene.

For both the forward and reverse primers, the length of complementarity to the target gene will depend on the sequence. Factors affecting the length of the target-specific portions of the primers include: GC content, T_m, and potential for formation of hairpins or primer-dimers. We recommend that the target-specific portion of each primer be designed with a T_m of ~60°C. The annealing temperature used in amplification may be adjusted to accommodate primers with higher or lower T_m. Primer T_m can be determined using tools provided by most vendors of synthetic oligonucleotides.

2. Amplify the target gene using a proofreading PCR polymerase. The performance of the Expresso Solubility and Expression Screening System has been verified with PCR products from various proofreading polymerases. Follow the manufacturer's recommendations regarding extension times and annealing temperatures during amplification of the target gene.

The use of non-proofreading polymerases (e.g. Taq) is strongly discouraged. While such PCR products can be cloned efficiently into the pSOL vectors following the procedure below, the relatively low fidelity will increase the probability of introducing mutations into the product and may necessitate screening multiple clones by sequencing.

3. Analyze the amplified DNA by agarose gel electrophoresis. If the reaction yields a single product of the expected size at a concentration of 10 ng/μL or higher, the

product can be used directly without purification. If the reaction yields additional products or if the yield of the desired product is low, it may be necessary to gel purify the product before proceeding.

Note: If the gene to be amplified is carried on an intact plasmid encoding kanamycin resistance, high background is likely due to carryover of the template plasmid. Methods to reduce this background include:

- *Gel purify the PCR product prior to Expresso cloning.*
- *Treat the PCR product with DpnI restriction enzyme prior to Expresso cloning. This enzyme digests only methylated DNA, leaving the amplified product intact. For best results, the PCR product should be purified following DpnI treatment.*
- *Linearize the template by restriction digest and purify the linear template prior to amplification by PCR.*

4. Bring the Recovery Medium to room temperature. Remove chemically competent *E. cloni*[®] 10G cells from –80°C freezer and allow them to thaw on ice. The cells are provided in 40 µL aliquots, each sufficient for a single transformation. Pre-chill a 15 mL disposable polypropylene culture tube (17 × 100 mm) on ice for each transformation.

For optimal cloning results, we strongly recommend the use of Lucigen's E. cloni 10G Chemically Competent Cells, which are included with the Expresso Solubility and Expression Screening System. Electrocompetent cells are incompatible with the Expresso cloning procedure.

5. Combine the PCR product (25–100 ng in a volume of no more than 3 µL) with 25 ng (2 µL) of a pSol Vector and 40 µL competent *E. cloni*[®] 10G cells. This step may be performed in parallel for all 8 pSOL vectors. Transfer the cells and DNA to pre-chilled 15 mL polypropylene culture tubes. Incubate the tubes containing cells and DNA on ice for 30 minutes.

The DNA and cells may be combined in the tube in which the competent cells are provided. However, for maximal transformation efficiency the cells and DNA must be transferred to a disposable polypropylene culture tube before the heat shock step.

An amplicon encoding the human growth hormone GH1 is included in the kit and may be used as a control for the cloning procedure.

6. After the 30 minute incubation on ice, transfer the culture tubes to a 42°C water bath for 45 seconds. Return the tubes to ice for 2 minutes.

Note: For maximal transformation efficiency, the heat shock is performed in 15 mL disposable polypropylene culture tubes (17 × 100 mm). The use of other types of tubes may dramatically reduce the transformation efficiency.

7. Add 960 μL of room temperature Recovery Medium to the cells in the culture tubes. Place the culture tubes in a shaking incubator at 250 rpm for 1 hour at 37° C.
8. Plate 100 μL of transformed cells on LB agar plates containing 30 $\mu\text{g/mL}$ kanamycin.

Note: There are several different formulations of LB medium, differing primarily in NaCl concentration. The LB Lennox formulation (containing 5 g/L NaCl) is recommended for plating E. coli 10G transformants.

9. Incubate the plates overnight at 37° C. Colonies should appear and grow to a size sufficient for picking in 18 – 24 hours.

SUPPORT PROTOCOL 1

SCREENING CLONES BY COLONY PCR

The Expresso cloning method is highly efficient. In most cases, over 95% of the colonies obtained upon co-transformation of vector + insert DNA fragments will contain the gene of interest cloned into the pSOL vector. The use of different sequences at each end of the insert to direct homologous recombination ensures that the fragment is cloned in the correct orientation and in-frame with the solubility tag. Even so, it may be desirable to confirm presence of the insert before proceeding with sequencing and expression trials, especially in cases where the PCR product used for cloning may have contained additional nonspecific products or “primer dimers”. Colonies are selected at random and screened for the presence of inserts by colony PCR using primers flanking the insertion site in the vector. Products are evaluated by agarose gel electrophoresis to confirm the insertion of a fragment of the expected size into the vector. Colony PCR may also be performed using a vector-specific primer in combination with a screening primer that is specific to the gene of interest (e.g. pRham Forward and the gene-specific reverse primer from step 1.1).

Colony PCR provides only a preliminary confirmation that the desired fragment has been cloned. Sequencing of the cloned insert, using a combination of vector-specific primers included in the kit and user-provided gene-specific primers is recommended.

Materials

*pRham Forward primer

*pETite Reverse primer

**Included in Expresso Solubility and Expression Screening System kits.*

Plates containing candidate colonies from Protocol 1

CloneID 1× Colony PCR Master Mix (Lucigen)

Programmable thermal cycler

Agarose gel electrophoresis equipment

Protocol steps

1. Prepare a master mix of a volume sufficient for the number of colonies to be analyzed. For each colony, add 0.25 μL of a 50 μM stock of each primer to 25 μL of 1 $\times$ CloneID Colony PCR master mix. Distribute 20 μL of the master mix to a number of PCR strip tubes corresponding to the number of colonies to be analyzed.
2. Use a pipet tip to transfer a portion of a colony to each tube. Pipet up and down to disperse the cells into the PCR reaction mix.

After transferring the colony to the PCR tube, the same pipet tip may be used to inoculate an overnight culture for plasmid DNA prep for sequencing and/or to start a culture for induction of protein expression (BASIC PROTOCOL 2 or ALTERNATE PROTOCOL 2).

3. Run the PCR program below, using an extension time of 1 minute per kb of the longest expected product. The total amplicon size includes the size of target gene + a portion of the vector including the solubility tag. See Table 1 for the expected sizes of the vector-encoded portion of each amplicon.

Step	Temperature	Time
1. Initial Denaturation	98°C	2 minutes
2. Denaturation	94°C	15 seconds
3. Annealing	55°C	15 seconds
4. Extension	72°C	1 minute/kb
5. Repeat steps 2–4	25 total cycles	
6. Final extension	72°C	10 minutes
7. Hold	4°C	∞

4. Analyze a portion (3 to 5 μL) of the PCR reaction by agarose gel electrophoresis. Include a molecular weight marker to allow accurate evaluation of the PCR product sizes.

BASIC PROTOCOL 2

TESTING EXPRESSION AND SOLUBILITY OF FUSION PROTEINS

In order to evaluate the expression and solubility of candidate fusion clones, small-scale expression trials (10–20 mL) are recommended. Clones that have been confirmed to contain the desired insert by colony PCR, and ideally by sequencing, are inoculated into liquid culture medium to create starter cultures. Starter cultures are then used to inoculate expression cultures to a low initial density, and fusion protein expression is induced by the addition of rhamnose during early logarithmic growth. Samples of the culture are harvested before and after induction to evaluate expression, and samples of the induced cultures are lysed by sonication. After centrifugation to separate soluble and insoluble fractions of the lysate, the fractions are analyzed by denaturing polyacrylamide gel electrophoresis (SDS-

PAGE). Figure 3 presents an example of screening for soluble expression of growth hormone GH1.

Materials

*20% rhamnose solution

*15% glucose solution

**Included in Expresso Solubility and Expression Screening System kits.*

LB-Miller liquid medium containing 30 µg/mL kanamycin

Cell Lysis Buffer

10× SDS-PAGE gel running buffer

2× SDS-PAGE sample loading buffer

Coomassie brilliant blue R250 gel staining solution

Spectrophotometer and disposable spectrophotometer cuvettes

Centrifuge (e.g. Sorvall RC-5)

Microcentrifuge

Sonicator equipped with a 1/8 inch micro-tip, or with an 8-tip sonicator horn (e.g., Qsonica #4602)

SDS-PAGE electrophoresis equipment (e.g. BioRad MiniProtean)

SDS-PAGE gel (e.g. BioRad MiniProtean TGX, 4–20%)

Protocol steps

1. Inoculate starter cultures from colonies containing verified clones by transferring a portion of a colony into 3 mL of LB-Miller medium containing 30 µg/mL kanamycin. Incubate the starter cultures at 37°C overnight with shaking at 250 rpm.

A starter culture may be inoculated from the same pipet tip used to transfer a colony to a colony PCR reaction tube, as described above in SUPPORT PROTOCOL 1.

2. Use the starter cultures to inoculate fresh LB Miller + 30 µg/mL kanamycin medium for induction of recombinant fusion protein expression. Use a volume of saturated starter culture equivalent to 1/100th of the expression culture volume. The cultures for induction should NOT contain glucose unless you are following an autoinduction procedure (see ALTERNATE PROTOCOL 2).

An overnight culture grown in LB-Miller medium will typically saturate at an OD₆₀₀ of approximately 4 to 5. Inoculation of the induction culture with 1/100th volume of saturated overnight should result in an initial OD₆₀₀ of 0.04 – 0.05. A convenient starting volume for the induction culture is 10 to 20 mL, which will

include sufficient volume for monitoring the OD_{600} before induction, as well as for harvesting the uninduced and induced samples.

3. Incubate the cultures with shaking at 37°C and monitor the growth of the cultures by optical density at 600 nm (OD_{600}) until the OD_{600} reaches approximately 0.5. Prior to induction, collect a sample of each uninduced culture corresponding to 1 OD_{600} equivalent (1 OD_{600} equivalent = 1 mL of culture with OD_{600} of 1). Harvest the uninduced cells by centrifugation in a microcentrifuge tube. Resuspend the pellet in 0.1 mL of 2× SDS-PAGE sample loading buffer. These **Uninduced** samples will be used for SDS-PAGE analysis in comparison to induced samples to evaluate expression. The samples may be stored at –20°C until the induced samples are harvested.

Example: To harvest 1 OD_{600} equivalent of a culture having an OD_{600} of 0.5, centrifuge 2 mL of the undiluted culture. Resuspend the cell pellet in 0.1 mL of 2× SDS-PAGE buffer.

4. To the remainder of the culture, add 1/100 volume of 20% rhamnose solution to induce expression. Continue incubation with shaking at 37°C for at least 4 hours to allow expression and accumulation of the induced protein.

A final concentration of 0.2% rhamnose will induce maximal expression of the fusion protein. Lower amounts in the range of 0.001% to 0.1% can be used for lower levels of expression, which may improve the solubility of some proteins.

For some proteins, longer induction times of up to 8 hours may improve yield. Incubation at reduced temperature (20°C to 30°C) following induction may improve solubility for some proteins, but may also require longer induction times of up to 24 hours.

5. Measure the OD_{600} of the induced cultures. The cultures should first be diluted 10-fold in LB-Miller medium to ensure an accurate measurement.
6. Harvest two samples from each culture by centrifugation:
 - i. A smaller sample containing 1.0 OD_{600} equivalent of cells will serve as the **Induced** sample for analysis of expression by SDS-PAGE. Resuspend this cell pellet in 0.1 mL of 2× SDS-PAGE sample loading buffer.
 - ii. A larger sample of each induced culture, corresponding to 10 OD_{600} equivalents, will be used to prepare lysates for evaluation of solubility of the fusion protein. Resuspend these cell pellets in 1 mL of cell lysis buffer. These samples may be stored at –20°C at this point.

Example: For a culture that yields an OD_{600} of 0.4 when diluted 10-fold, the OD_{600} of the culture is 4.0. To harvest 1 OD_{600} equivalent for expression analysis, centrifuge 0.25 mL of the undiluted culture and resuspend the pellet in 0.1 mL of 2× SDS-PAGE sample loading buffer. To harvest 10 OD_{600} equivalents for lysate preparation, centrifuge 2.5 mL of the undiluted culture and resuspend in 1 mL of cell lysis buffer.

7. Evaluate the **Uninduced** and **Induced** whole-cell samples by SDS-PAGE. Heat the samples in 2× SDS-PAGE gel loading buffer to 95°C for 5 minutes. Centrifuge the samples briefly and load 5 µL of each sample (0.05 OD₆₀₀ equivalents) separately into the wells of a 15-well, 4–20% polyacrylamide gel. Include a molecular weight marker to facilitate identification of induced protein bands. Run the gel at 200 V for 35 min or as recommended by manufacturer until the dye front reaches the bottom of the gel.

*In this protocol, fusion protein expression is evaluated by SDS-PAGE to eliminate any non-expressing clones before proceeding with lysate preparation. Alternatively, steps 8 and 9 may be omitted and the **Uninduced** and **Induced** samples saved for later analysis alongside the **Total**, **Soluble** and **Insoluble** fractions of the lysate (steps 11 and 12).*

8. Stain the gel with Coomassie brilliant blue for at least 30 minutes. Destain the gel with water and evaluate expression of fusion proteins by comparing the uninduced and induced samples. Select candidates showing induced expression of the expected fusion protein for lysate preparation.

Estimate the expected molecular weight of the fusion protein by adding the molecular weight of the target protein to the molecular weight of the fusion partner encoded by the pSol vector. Table 2 presents the molecular weights of the vector-encoded portion of each fusion including 6×His, solubility tag and TEV site.

9. Prepare lysates from the larger (10 OD₆₀₀ eq.) samples harvested in step 8 by sonication. Keep samples on ice throughout sonication. Use repeated pulses of sonication separated by intervals of resting on ice (e.g. 10 cycles of sonication for 10 seconds, resting for 1 minute between cycles) to avoid buildup of heat or frothing of the sample. Refer to the sonicator manual for recommended settings.

Cell pellets from all 8 candidate fusion clones may be lysed in parallel using an 8 tip sonication horn equipped with 1/8" diameter mini tips (ex, Qsonica #4602). The tips are spaced to fit standard strip tubes or a 96 well plate. In this example the cell pellets containing 10 OD₆₀₀ equivalents of cells are resuspended in 0.2 mL of lysis buffer. A 0.1 mL portion of each resuspended pellet is transferred to a tube in an 8-tube strip and sonicated. Note that these samples are 5 times as concentrated as pellets resuspended in 1.0 mL of lysis buffer and should be diluted accordingly in Cell lysis buffer following sonication and before gel analysis.

In our experience using an 8-tube strip with an 8-tip sonication horn and a Sonics VibraCell instrument, a protocol of 12 cycles with 1 sec on/1 sec off pulses at 70% amplitude is sufficient for cell lysis.

10. Transfer a 50 µL sample of lysate to a fresh tube, and store on ice. This fraction will be used to assess the **Total** amount of protein in the cell lysate.

11. Transfer a second 50 μ L sample of lysate to a fresh tube and centrifuge at 15,000 $\times$ g for 15 minutes at 4° C. Transfer the supernatant to a fresh tube and store on ice. This fraction is the cleared cell lysate and contains the **Soluble** protein. Add 100 μ L of 2 $\times$ SDS Sample Buffer to the Pellet. This fraction will be used to assess the **Insoluble** protein in the cell lysate. Care should be taken to *fully* resuspend this pellet.

The remaining lysate may be centrifuged and the supernatant saved for small-scale purification trials (Basic Protocol 3). However, the amount of protein obtained may not be sufficient for downstream applications.

12. Prepare samples of the **Total** and **Soluble** fractions for SDS-PAGE analysis by combining equal volumes of each fraction and 2 $\times$ SDS-PAGE Sample Buffer (e.g. 20 μ L of sample + 20 μ L of 2 $\times$ SDS-PAGE Sample buffer). Heat these samples and the resuspended **Insoluble** fraction to 95°C for 5 minutes, and load 10 μ L each of the **Total**, **Soluble**, and **Insoluble** fractions into adjacent lanes of a 4–20% SDS-PAGE gel. Include molecular weight standards to estimate the molecular weight of the recombinant protein. Run the gel at 200 V for 35 min or as recommended by manufacturer until the dye front reaches the bottom of the gel.

*If sonication is performed with an 8-tip sonication horn using concentrated samples as described above in Step 10, dilute the lysate sample 5-fold by adding 0.4 mL of Cell Lysis buffer before centrifugation. Prepare samples of **Total**, **Soluble**, and **Insoluble** fractions by dilution with 2 $\times$ SDS-PAGE Loading buffer as described in steps 11 through 13. Each 10 μ L sample loaded in the gel will contain protein from an equivalent number of cells corresponding 0.05 OD₆₀₀ unit equivalents.*

13. Stain the gel with Coomassie brilliant blue and destain to reveal protein bands.
14. Choose one or more of the protein fusions that produce the greatest proportion of protein in the soluble fraction for scaled-up expression and affinity purification trials.

ALTERNATE PROTOCOL 2

AUTOINDUCTION

A convenient method for inducing protein expression, known as autoinduction, exploits the repression of the rhaP_{BAD} promoter by glucose (Wegerer et al. 2008). In a culture containing both glucose and rhamnose, cells will preferentially metabolize glucose during the early stages of growth, and the rhaP_{BAD} promoter will become active after glucose is depleted. This simplifies the induction procedure by eliminating the need to monitor cell density and to add rhamnose manually. The autoinduction procedure is especially convenient when many clones are being evaluated in parallel. The protocol below includes glucose at an initial concentration of 0.05%. The timing of induction by rhamnose can be controlled by varying the concentration of glucose between 0.05% and 0.15%. The higher concentration of glucose

leads to later onset of induction, which may be beneficial if the expressed protein is unstable, insoluble, or toxic to the host cells.

Materials

*20% rhamnose solution

*15% glucose solution

**Included in Expresso Solubility and Expression Screening System kits.*

LB-Miller liquid medium containing 30 µg/mL kanamycin

Protocol steps

1. Prepare autoinduction medium by adding 10 µL of 20% rhamnose solution and 3.3 µL of 15% glucose solution per mL of LB Miller medium containing 30 µg/mL of kanamycin.
2. Inoculate the autoinduction cultures with a starter culture or with a colony. Cultures may be inoculated directly using the same pipet tip used to transfer a colony to colony PCR reaction in Support Protocol 1.
3. Continue incubation with shaking at 37°C for 8–24 hours to allow expression and accumulation of the induced protein.
4. Measure the OD₆₀₀ of the autoinduced cultures as described in step 5 of Basic Protocol 2.

Cultures grown to saturation in LB-Miller autoinduction medium
5. Proceed with harvest of samples and analysis of expression and fusion protein solubility as described in Basic Protocol 2, steps 6 through 15.

BASIC PROTOCOL 3

PURIFICATION OF FUSION PROTEINS BY METAL-AFFINITY CHROMATOGRAPHY

Each of the pSol vectors encodes a 6×His tag at the amino terminus of the fusion protein. The resulting proteins may be purified by Immobilized Metal Affinity Chromatography (IMAC). Several different IMAC resins are commercially available, including Ni-NTA Affinity Resin (Qiagen), complete His-tag Purification Resin (Roche), TALON Metal Affinity Resin (Takara), and others. Since these products differ in binding affinity, capacity, and reagent compatibility, adherence to the manufacturers' recommendations is recommended.

The procedure below begins with cleared lysate from 100 mL of induced culture, and details a column purification procedure using Ni-NTA from Qiagen. The culture volume may be adjusted depending on the observed expression level and desired quantity of fusion protein. A general guideline of 5 mL of cell lysis buffer for every 1 g cell pellet weight may be followed.

Materials—Cleared cell lysate from 100 mL of induced culture

Chromatography column (e.g. Bio-Rad #731-1550)

Ni-NTA Resin (Qiagen)

Cell lysis buffer

Column wash buffer

Elution buffer

Dialysis buffer

Bradford Protein Assay reagents (ThermoFisher Scientific)

Spectrophotometer

Protocol steps

1. Prepare a cleared lysate by sonication of a sample of induced cell culture followed by centrifugation. Save samples of **Uninduced** and **Induced** cells, and **Total**, **Soluble** and **Insoluble** fractions of cell lysate for SDS-PAGE analysis.
2. Prepare resin. Determine column bed volume to be used. Qiagen Ni-NTA is supplied as a 50% slurry and generally has a 5–10 mg protein per mL resin binding capacity. Pipette 2 mL of a 50% slurry into a chromatography column, resulting in a 1 mL bed volume. Let the storage solution drain. Wash column with 5 column volumes (CVs) lysis buffer, careful to avoid disruption of the column bed.
3. Apply cell lysate to column. Collect flow-through and save for analysis by SDS-PAGE.
4. Wash column with 10 CVs wash buffer. Each CV can be collected as individual or pooled fractions. Save for analysis by SDS-PAGE.
5. Elute protein with 5 CVs elution buffer. It is beneficial to collect eluate as fractions because the protein of interest may elute at a different volume than contaminating proteins. Save fractions for analysis by SDS-PAGE.
6. Analyze samples by SDS-PAGE. Include samples of Soluble fraction of cell lysate, IMAC flow-through, IMAC wash fraction(s), and IMAC eluate(s). Mix 10 μ L of each sample with 10 μ L 2 $\times$ SDS-PAGE loading buffer. Loading 10 μ L of each sample per well should provide an amount of protein detectable by Coomassie stain. Heat all samples at 95°C for 5 minutes, centrifuge briefly and load on 4–20% polyacrylamide gel. Stain and destain gel to verify presence of protein of interest and to determine which eluate fractions to retain.
7. Pool eluate fractions that contain the most protein of interest and least contaminating protein. To remove imidazole and exchange to a buffer of choice, dialyze against a volume of dialysis buffer that is at least 200 times greater than the sample volume. Choose dialysis tubing with a molecular weight cutoff

(MWCO) that is at least 10 kDa lower than the molecular weight of the protein of interest. Replace the dialysis buffer after 2 to 4 hours, and continue dialysis overnight at 4°C for complete removal of imidazole.

8. Quantify protein by Bradford assay or by measuring absorbance at 280 nm. If using absorbance at 280 nm, determine extinction coefficient using an online tool such as ProtParam available through Expasy.org.

BASIC PROTOCOL 4

REMOVAL OF SOLUBILITY TAG BY CLEAVAGE WITH SelecTEV PROTEASE

A recognition site for the protease from Tobacco Etch Virus (TEV) is located between the fusion tag and the protein of interest. SelecTEV Protease recognizes the seven-amino-acid sequence Glu-Asn-Leu-Tyr-Phe-Gln-Gly and cleaves between Gln and Gly with high specificity. Cleavage by the SelecTEV Protease is optimal in SelecTEV Buffer at 30°C, but SelecTEV Protease is active between 4 – 30°C and pH 6.0 – 8.5. Optimization of the cleavage conditions may be necessary depending on the protein of interest. Following cleavage, the released target protein can be recovered by capturing the fusion partner, SelecTEV protease, and any residual uncleaved fusion protein on IMAC resin. Figure 4 presents an example of SelecTEV cleavage and recovery of released target protein by subtractive metal affinity chromatography.

Materials

*SelecTEV Protease, 10 U/μL

*20× SelecTEV Protease buffer

*100 mM DTT

**Included in Expresso Solubility and Expression Screening System kits, or available separately (Lucigen)*

Purified fusion protein from Basic Protocol 3

SDS-PAGE gel (e.g. BioRad MiniProtean TGX, 4–20%)

SDS-PAGE gel running buffer

SDS-PAGE sample loading buffer

Protocol steps

1. Set up a test cleavage reaction following the guidelines below. Be sure to include DTT at a final concentration of 1 mM.

Volume, μL	Component	Final Concentration/Amount
X	Fusion Protein	30 μg
5	SelecTEV 20× Buffer	1×

Volume, μL	Component	Final Concentration/Amount
1	DTT, 100mM	1 mM
1	SelecTEV Protease, 10 U/ μL	10 U
Y	Water	N/A
100 μL Total		

- Analyze a sample of the cleavage reaction by SDS-PAGE. Mix equal volumes of the cleavage reaction and 2 $\times$ SDS-PAGE gel loading buffer. Heat the sample to 95°C for 5 minutes and centrifuge briefly. Run 10 to 20 μL of the sample diluted in 2 $\times$ SDS-PAGE loading buffer (1.5 to 3 μg) on a 4–20% acrylamide gel. Include a sample of undigested protein for comparison. Stain and destain the gel to evaluate the efficiency of cleavage.

1 Unit of SelecTEV Protease is able to cleave > 85% of 3 μg of a control substrate in 1 hour at 30°C. The efficiency of cleavage may vary considerably depending on the nature of the substrate. If incomplete cleavage is observed, try increasing the incubation time or the amount of SelecTEV protease.

Once you have established conditions for efficient cleavage of the fusion protein, the cleavage reaction may be scaled up.

Digestion with SelecTEV protease may also be performed during dialysis following IMAC purification. It is important to bear in mind the potential impact of salt concentration and temperature on activity of the protease maximal TEV activity (Nallamsetty et al., 2004).

- Isolate the “de-tagged” target protein by subtractive metal affinity chromatography. Perform binding to the IMAC resin as described in the resin manufacturer’s protocol. The cleaved native protein will be in the flow-through fractions. The 6His-tagged solubility partner and the SelecTEV Protease will remain bound to the resin, along with any uncleaved fusion protein.

SelecTEV Protease contains a (6 $\times$ His) tag at its amino terminus, as does the cleaved solubility tag fragment and any uncleaved fusion protein. Imidazole remaining in the sample could prevent the 6 $\times$ His tags from binding to the IMAC resin. Remove or dilute imidazole prior to IMAC purification. Many IMAC resins do not tolerate 1 mM DTT. Dilute the sample in column binding buffer or remove DTT prior to purification.

REAGENTS AND SOLUTIONS

LB-Miller liquid culture medium—Per liter:

10 g Bacto-tryptone

5 g yeast extract

10 g NaCl

LB-Lennox Agar plates with 30 $\mu\text{g}/\text{mL}$ kanamycin—10 g Bacto-tryptone,

5 g yeast extract,

5 g NaCl,

15 g agar.

Mix components, autoclave and cool to 55°C.

Add kanamycin to a final concentration of 30 µg/mL. Pour into petri plates.

Store plates at 4°C.

10× SDS-PAGE gel running buffer—0.25M Tris Base

1.92M glycine

1% SDS

2× SDS-PAGE sample loading buffer—0.09M Tris·Cl, pH 6.8

20% glycerol

2% SDS

0.02% bromophenol blue

5% β-mercaptoethanol

Coomassie brilliant blue R250 staining solution—50% (v/v) methanol

10% (v/v) acetic acid

0.25% (w/v) coomassie brilliant blue R250

Cell lysis buffer—0.05M NaH₂PO₄

0.3M NaCl

Adjust pH to 8.0 using NaOH

(optional: 5 mM β-mercaptoethanol)

Column wash buffer—0.05M NaH₂PO₄

0.3M NaCl

0.02M imidazole (0.005–0.02M)

Adjust pH to 8.0 using NaOH

(optional: 5 mM β-mercaptoethanol)

Elution buffer—0.05M NaH₂PO₄

0.3M NaCl

0.25M imidazole (0.1–0.5M)

Adjust pH to 8.0 using NaOH

(optional: 5 mM β-mercaptoethanol)

Dialysis buffer—0.05M Tris-HCl

0.1M NaCl

0.001M DTT

Adjust pH to 8.0 using NaOH

COMMENTARY

Background Information

For decades, *E. coli* has been the go-to option when recombinant protein expression is required for structural and functional studies. However, a frequently encountered limitation of this system is the failure to produce the protein of interest in correctly folded, soluble form, or even to produce the protein in useful quantities at all. Estimates suggest that over half of heterologous proteins fail to be produced in soluble form upon initial trials in *E. coli* (Esposito and Chatterjee, 2006). Although numerous other expression host systems exist, the resources needed to implement these systems may be prohibitive to many researchers. The strong motivation to rescue bacterial expression of challenging proteins has resulted in a number of strategies, among the most effective of which is the use of fusion partners to promote expression and solubility.

Among the first fusion partners reported to enhance expression or solubility of a target protein in *E. coli* were GST, ubiquitin, thioredoxin, and MBP (Smith and Johnson, 1988; Butt et al., 1989; LaVallie et al., 1993; Pryor et al., 1997). Studies comparing the efficacy of different tags at enhancing target protein expression or solubility have often reached different conclusions, depending in part on the identity of the passenger protein(s) being evaluated (e.g., Kapust et al., 1999; Marblestone et al., 2006). The mechanisms underlying the improvements in solubility engendered by fusion partners are in general poorly understood, but their differential profiles of effectiveness against different targets suggest that more than one mechanism may be involved. MBP has been subjected to the most extensive mechanistic study, leading to the proposal that it may act at least in part as a passive chaperone. In this role, fully folded and highly soluble MBP may form transient physical interactions with incompletely folded segments of a fusion partner. Transient, iterative interactions with MBP may allow for spontaneous folding of the passenger protein in some instances, while other passengers may require action of endogenous chaperones to progress to a correctly folded state (Raran-Kurussi and Waugh, 2012).

Subsequently, rational approaches have led to the demonstration of solubility-enhancing properties of numerous other proteins tested as fusion partners. One set of studies focused on *E. coli* proteins that are induced in response to stress caused by heat shock or by protein denaturants, and identified SlyD and Tsf (Han et al., 2007a,b). SlyD has an intrinsic peptidyl-prolyl cis-trans isomerase activity, and Tsf is a translation initiation factor. Both have been proposed to contain passive chaperone functions, and were found to promote soluble expression of fused partners.

Proteins with a high content of acidic side chains have also been reported to promote enhanced solubility of fusion partners (Zhang et al. 2004; Zou et al. 2008). The high surface charge density of these proteins has been proposed to shield their fusion partners from aggregation by virtue of charge repulsion, providing an opportunity for their partners to fold correctly in isolation. Highly charged proteins from halophilic organisms, such as beta-lactamase (Bla) from *Chromohalobacter* sp. 560 (Tokunaga et al. 2010) may promote soluble expression through a similar mechanism. In our search for additional solubility-enhancing factors, we identified the hypothermostable and pH-resistant protein AFV1–99, a 13.5-kDa protein with a high content of charged residues from Acidianus filamentous virus 1 (Goulet et al. 2009).

The seven fusion partners included in the Espresso Solubility and Expression Screening System (MBP, GST, SUMO, SlyD, Tsf, Bla, and AFV1–99) were chosen to represent the variety of potential solubility-promoting mechanisms described above, and on the basis of demonstrated effectiveness at improving the expression and/or solubility of multiple targets from a panel of challenging proteins. Although there was some overlap in effectiveness, no two of the seven fusion partners showed identical profiles against this panel. While there can be no guarantee that the fusion strategy will prove successful for any particular challenging protein, the panel of fusion partners provided in the Espresso System is intended to maximize the success of this approach by representing a diversity of potential mechanisms for enhancement of expression and solubility.

Frequently, once an effective fusion partner has been identified which enables the improved expression or solubility of the target protein, the removal of the fusion partner is desirable for subsequent structural or functional analysis. Due to its relatively high recognition site specificity (Parks et al., 1994; Phan et al., 2002), the Tobacco Etch Virus (TEV) protease is commonly used for fusion tag removal. Most commercially available forms of TEV protease, including SelecTEV, have been engineered for improved stability and activity (Kapust et al., 2001). Each pSOL vector encodes a TEV protease recognition site (AA sequence ENLYFQ|G, where cleavage occurs between Q and G residues) at the C-terminus of solubility tag. Sequence encoding this site serves as one of the universal sequences for cloning the gene of interest into the pSOL vectors by homologous recombination.

In principle, the Espresso System cloning strategy is similar to the popular Gibson Assembly method, which relies on treatment with an enzyme cocktail to promote recombination *in vitro* between linear DNA fragments having short tracts of sequence homology at their ends (Gibson et al., 2009). However, the Espresso System relies instead on native recombination machinery contained within the *E. coli* host cells (Bubeck et al.,

1993). Since no prior enzymatic treatment of the DNA fragments to be joined is required, a target gene amplicon with vector-homologous ends appended during PCR is simply co-transformed with one of the 8 vectors (which is provided in a linearized, cloning-ready format) into the host bacterial cells. While the efficiency of this homologous recombination in *E. coli* is generally considered to be relatively low (in comparison to yeast, for example), it is more than sufficient for straightforward assembly of two linear DNA fragments. In our experience, the method may also be used to assemble at least three fragments containing appropriate homology, but at reduced efficiency. For reasons that are not clear, the method is dependent on transformation of chemically competent cells and is incompatible with electroporation (Bubeck et al., 1993).

In the Expresso system, fusion protein expression is driven by the *E. coli* rhaP_{BAD} promoter, which responds to the presence of rhamnose sugar in a highly concentration-dependent, “tunable” fashion (Egan et al., 1993; Haldimann et al. 1998; Giacalone et al. 2006). Low activity of this promoter in the absence of rhamnose allows for stable clone construction, without interference by potential toxicity of the target protein. The rheostat-like responsiveness of rhaP_{BAD} to a range of rhamnose concentrations provides a means to modulate expression of proteins that are prone to aggregation when expressed at high concentration. This tool may be used for fine-tuning expression levels to optimize the yield of soluble protein in some instances. Inhibition of rhaP_{BAD} by glucose can also be exploited to maintain repression of cloned genes, and is the basis for the convenient autoinduction strategy (Wegerer et al., 2008).

Critical Parameters and Troubleshooting

The Expresso Solubility and Expression Screening System is designed for the expression of proteins in the cytoplasm of *E. coli*. Integral membrane proteins or secreted proteins containing multiple disulfide bonds may not be appropriate candidates for screening with this system. Similarly, if specific maturation or modification pathways native to its original host organism are essential to the proper folding or activity of a protein, the fusion strategy may be unable to overcome these requirements.

Perhaps the most important limitation of the fusion strategy is the potential for “false” solubility. A proportion of target proteins that appear to be produced in soluble form during initial screening of fusions may actually be present as incorrectly folded aggregates, maintained in an artificially soluble form through shielding by a highly soluble fusion tag. Upon tag removal by protease cleavage in such cases, precipitates may form immediately or may accumulate over time. If this behavior is observed for a particular fusion construct it is advisable to test other fusion constructs that exhibit apparent solubility. If tag removal leads to precipitation of the target protein, it may be tempting to maintain solubility by leaving the tag intact. However, it should be noted that the formation of soluble aggregates may reflect incorrect folding of the target protein. Activity assays for the target protein, if available, may be useful to distinguish incorrectly folded aggregates from proteins that tend to aggregate in their correctly folded form.

In some instances, it may be possible to rescue solubility by varying parameters such as pH, salt concentration, or other buffer conditions during purification before or after TEV

cleavage. In particular, the isoelectric point (pI) of the protein should be considered. Using a buffer that differs from the pI by ± 0.5 pH units may improve solubility. However, a pH below 6 should be avoided as the 6×His tag will become protonated and not bind to the Ni-NTA resin. If simple adjustments to the pH or salt concentration do not prove effective, it may be more fruitful to investigate other expression systems such as yeast, insect or mammalian cell culture rather than to spend additional time troubleshooting bacterial expression.

If the results of expression and solubility screening show high levels of fusion protein expression but little or no soluble expression, it may be possible to rescue some solubility by reducing the rate of expression. A lower rate of accumulation may reduce aggregation and afford an opportunity for folding of the protein. Two such strategies are growth at reduced temperature (20°C to 30°C), and induction with a sub-saturating concentration of rhamnose (e.g. 0.001 to 0.1%).

Anticipated Results

The process of cloning a target gene amplicon into each of the 8 vectors of the Expresso Solubility and Expression Screening System is usually extremely simple and straightforward. In most cases >95% of the colonies resulting from co-transformation of the target gene amplicon with one of the pSol vectors will have the desired insert in the correct orientation. The outcome of screening the resulting fusion constructs for expression and solubility cannot be predicted, and will differ for each target protein of interest.

Time Considerations

Once primers designed to amplify the gene of interest for assembly with the pSol vectors have been obtained, the entire process of clone construction and initial screening of the expression and solubility of 8 different fusion constructs may be completed in as little as 3 to 4 days. Target gene amplification and co-transformation with the suite of pSol vectors are completed on the first day. Preliminary clone verification by colony PCR can be completed the second day, and candidates may be simultaneously inoculated directly into autoinduction medium. Autoinduced cultures may be harvested and samples prepared for evaluation of fusion protein expression and solubility on day 3. Following the standard induction procedure may require an additional day to complete the initial screening.

After selecting candidate fusion constructs for further analysis, scaled-up expression and affinity purification by IMAC can be completed in 2–3 days. Fusion partner removal by cleavage with SelectTEV protease and evaluation of solubility following cleavage can be completed in another 2 days.

Acknowledgments

Research by the authors was supported in part by SBIR grant 5R44GM089553 from the NIH.

LITERATURE CITED

Bubeck P, Winkler M, Bautsch W. Rapid cloning by homologous recombination in vivo. *Nucleic Acids Res.* 1993; 21(15):3601–3602. [PubMed: 8346047]

- Butt TR, Jonnalagadda S, Monia BP, Sternberg EJ, Marsh JA, Stadel JM, Crooke ST. Ubiquitin fusion augments the yield of cloned gene products in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 1989; 86(8):2540–2544. [PubMed: 2539593]
- Egan SM, Schleif RF. A regulatory cascade in the induction of rhaBAD. *J Mol Biol*. 1993; 234(1):87–98. DOI: 10.1006/jmbi.1993.1565 [PubMed: 8230210]
- Esposito D, Chatterjee DK. Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotechnol*. 2006; 17(4):353–358. DOI: 10.1016/j.copbio.2006.06.003 [PubMed: 16781139]
- Giacalone MJ, Gentile AM, Lovitt BT, Berkley NL, Gunderson CW, Surber MW. Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system. *Biotechniques*. 2006; 40(3):355–364. DOI: 10.2144/000112112 [PubMed: 16568824]
- Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009; 6(5):343–345. DOI: 10.1038/nmeth.1318 [PubMed: 19363495]
- Goulet A, Spinelli S, Blangy S, van Tilbeurgh H, Leulliot N, Basta T, Prangishvili D, Cambillau C, Campanacci V. The thermo- and acido-stable ORF-99 from the archaeal virus AFV1. *Protein Sci*. 2009; 18(6):1316–1320. DOI: 10.1002/pro.122 [PubMed: 19472363]
- Han KY, Song JA, Ahn KY, Park JS, Seo HS, Lee J. Enhanced solubility of heterologous proteins by fusion expression using stress-induced *Escherichia coli* protein, Tsf. *FEMS Microbiol Lett*. 2007a; 274(1):132–138. DOI: 10.1111/j.1574-6968.2007.00824.x [PubMed: 17608803]
- Han KY, Song JA, Ahn KY, Park JS, Seo HS, Lee J. Solubilization of aggregation-prone heterologous proteins by covalent fusion of stress-responsive *Escherichia coli* protein, SlyD. *Protein Eng Des Sel*. 2007b; 20(11):543–549. DOI: 10.1093/protein/gzm055 [PubMed: 17971396]
- Kapust RB, Waugh DS. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci*. 1999; 8(8):1668–1674. DOI: 10.1110/ps.8.8.1668 [PubMed: 10452611]
- Kapust RB, Tözsér J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS. Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Prot. Eng*. 2001; 14:993–1000.
- LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnology (N Y)*. 1993; 11(2):187–193. [PubMed: 7763371]
- Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci*. 2006; 15(1):182–189. DOI: 10.1110/ps.051812706 [PubMed: 16322573]
- Nallamsetty S, Kapust RB, Tozser J, Cherry S, Tropea JE, Copeland TD, Waugh DS. Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. *Protein Expr Purif*. 2004; 38(1):108–115. DOI: 10.1016/j.pep.2004.08.016 [PubMed: 15477088]
- Parks TD, Leuther KK, Howard ED, Johnston SA, Dougherty WG. Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal Biochem*. 1994; 216(2):413–417. DOI: 10.1006/abio.1994.1060 [PubMed: 8179197]
- Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HK 3rd, Kapust RB, Waugh DS. Structural basis for the substrate specificity of tobacco etch virus protease. *J Biol Chem*. 2002; 277(52):50564–50572. DOI: 10.1074/jbc.M207224200 [PubMed: 12377789]
- Pryor KD, Leiting B. High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein double-affinity fusion system. *Protein Expr Purif*. 1997; 10(3):309–319. DOI: 10.1006/prep.1997.0759 [PubMed: 9268677]
- Raran-Kurussi S, Waugh DS. The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. *PLoS One*. 2012; 7(11):e49589.doi: 10.1371/journal.pone.0049589 [PubMed: 23166722]
- Sambrook J., Russell, DW. *Molecular Cloning: A Laboratory Manual*. 3. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2001.

- Smith DB, Johnson KS. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene*. 1988; 67(1):31–40. DOI: 10.1016/0378-1119(88)90005-4 [PubMed: 3047011]
- Tokunaga H, Saito S, Sakai K, Yamaguchi R, Katsuyama I, Arakawa T, Tokunaga M. Halophilic beta-lactamase as a new solubility- and folding-enhancing tag protein: production of native human interleukin 1alpha and human neutrophil alpha-defensin. *Appl Microbiol Biotechnol*. 2010; 86(2): 649–658. DOI: 10.1007/s00253-009-2325-9 [PubMed: 19902204]
- Tozser J, Tropea JE, Cherry S, Bagossi P, Copeland TD, Wlodawer A, Waugh DS. Comparison of the substrate specificity of two potyvirus proteases. *FEBS J*. 2005; 272(2):514–523. DOI: 10.1111/j.1742-4658.2004.04493.x [PubMed: 15654889]
- Wegerer A, Sun T, Altenbuchner J. Optimization of an *E. coli* L-rhamnose-inducible expression vector: test of various genetic module combinations. *BMC Biotechnol*. 2008; 8:2.doi: 10.1186/1472-6750-8-2 [PubMed: 18194555]
- Zhang YB, Howitt J, McCorkle S, Lawrence P, Springer K, Freimuth P. Protein aggregation during overexpression limited by peptide extensions with large net negative charge. *Protein Expr Purif*. 2004; 36(2):207–216. DOI: 10.1016/j.pep.2004.04.020 [PubMed: 15249042]
- Zou Z, Cao L, Zhou P, Su Y, Sun Y, Li W. Hyper-acidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in *Escherichia coli*. *J Biotechnol*. 2008; 135(4):333–339. DOI: 10.1016/j.jbiotec.2008.05.007 [PubMed: 18599143]

Figure 1.

Solubility and Expression Screening Workflow. A single amplicon containing the target of interest is cloned into eight *E. coli* expression vectors in parallel. Seven of the vectors contain a unique fusion tag that may promote increased expression and solubility of the target as well as an amino-terminal 6×His tag for purification. The eighth “control” vector contains only the amino-terminal 6×His tag. The cloning reaction is simultaneously assembled and transformed into the *E. coli* expression strain. Resulting colonies are grown in liquid culture and protein expression is induced in parallel. Fusion tags promoting soluble protein expression are rapidly identified by comparing expression results to the control clone.

Figure 2. pSOL™ Expression Vectors

A suite of eight vectors, each containing a unique fusion partner, enables instant parallel cloning of a single amplicon containing the target of interest. The target gene amplicon is flanked by unique 18 bp sequences at each end that are homologous to the universal ends of the eight vectors. In the Expresso cloning method, the amplicon is simply combined with one of the eight vectors and transformed immediately into chemically competent *E. coli*. Endogenous recombination activity within the cells fuses the insert to the vector seamlessly. The amino-terminal cassette in each vector contains a translational start codon, 6×His and fusion tag sequence joined to the target sequence by a TEV Protease cleavage site, enabling IMAC purification and TEV Protease-mediated tag removal. A rhamnose-inducible promoter provides high-level, inducible protein expression of the target fusion. A strong

transcriptional terminator prevents unwanted transcription into or out of the cloned sequence.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3. Example of a solubility screen

A gene encoding the GH1 (human growth hormone) isoform from pituitary was cloned into the 8 pSOL vectors and evaluated for expression and solubility. Improved soluble expression was observed for each of the seven fusion tags compared to the 6×His control. Several fusion partners, especially Tsf, Bla, and MBP also promoted a dramatic increase in protein expression compared to the control. T: Total Lysate, S: Cleared Soluble Fraction, I: Insoluble Pellet.

Figure 4. Fusion-tag removal and purification of detagged GH1

A. Lanes 1 and 2 show purified Bla-GH1 fusion protein before and after cleavage with SelectTEV protease. Lanes 3, 4 and 5 show flow-through, wash and elution fractions from IMAC purification following the cleavage reaction. Free GH1 was present in the flow-through. The 6×His-tagged Bla fragment and SelectTEV protease are retained on the column and released by elution with buffer containing 0.25 M imidazole. M, marker proteins.

B. Purified, detagged GH1 remained soluble upon centrifugation after 72 h at 4 °C.

Table 1

Expected Sizes of Colony PCR Amplicons.

Vector	pRham Forward + pETite Reverse *	pRham Forward + GOI Reverse *
pSol AFV	490	439
pSol SlyD	781	730
pSol Tsf	1042	991
pSol SUMO	496	445
pSol Bla	1294	1243
pSol MBP	1297	1246
pSol GST	850	799
pSol His Control	193	142

* Size, in base-pairs, of the vector portion of an amplicon produced with the specified primers. To determine the expected size of the entire amplicon, the length of the gene of interest must be added to the value in the table.

Table 2

Amino Acid Length and Molecular Weight of Fusion Tags in pSOL vectors

Fusion Tag	AA Length	kDa*	pI	Description
6×His-AFV	113	13.5	5.0	AFV1–99 protein from Acidianus filamentous virus 1
6×His-SlyD	210	22.7	5.2	FKBP-type peptidyl-prolyl cis-trans isomerase
6×His-Tsf	297	32.2	5.7	<i>E. coli</i> elongation factor
6×His-SUMO	115	13.3	5.2	Small Ubiquitin-like Modifier
6×His-Bla	381	41.3	4.4	Beta-lactamase
6×His-MBP	382	42.1	5.5	Maltose-Binding Protein
6×His-GST	233	27.4	6.6	Glutathione S-transferase
6×His Control	14	1.8	7.0	Affinity tag

* Molecular weights listed are all based on amino acid composition. Some tags (e.g. SUMO and Bla) frequently display anomalous migrations on SDS-PAGE.