

Published in final edited form as:

Methods Mol Biol. 2014 ; 1140: 89–105. doi:10.1007/978-1-4939-0354-2\_7.

## Protein Production for Structural Genomics Using *E. coli* Expression

Magdalena Makowska-Grzyska, Youngchang Kim, Natalia Maltseva, Hui Li, Min Zhou, Grazyna Joachimiak, Gyorgy Babnigg, and Andrzej Joachimiak

### Abstract

The goal of structural biology is to reveal details of the molecular structure of proteins in order to understand their function and mechanism. X-ray crystallography and NMR are the two best methods for atomic level structure determination. However, these methods require milligram quantities of proteins. In this chapter a reproducible methodology for large-scale protein production applicable to a diverse set of proteins is described. The approach is based on protein expression in *E. coli* as a fusion with a cleavable affinity tag that was tested on over 20,000 proteins. Specifically, a protocol for fermentation of large quantities of native proteins in disposable culture vessels is presented. A modified protocol that allows for the production of selenium-labeled proteins in defined media is also offered. Finally, a method for the purification of His<sub>6</sub>-tagged proteins on immobilized metal affinity chromatography columns that generates high-purity material is described in detail.

### Keywords

Protein expression; Protein purification; Disposable vessel fermentation; Selenomethionine-labeling; IMAC; His-tag; High-throughput

## 1 Introduction

Structural genomics uses X-ray crystallography and NMR spectroscopy for structure determination and requires milligram quantities of high-quality proteins. Protein samples must be soluble, highly homogeneous, and free of critical contaminants (especially proteases, metals, lipids, and other impurities). In structural genomics it is also important that numerous proteins or their constructs are produced in parallel and in a highly reproducible manner. To achieve these objectives, every step in the production of a protein must be carefully evaluated, optimized, and accelerated, including gene cloning, protein expression, and purification. Over the years, much effort has been put into optimizing *Escherichia coli* as an expression host for proteins from other bacteria and higher organisms [1–8]. A variety of classes of proteins, from prokaryotic to human and from single domain to protein complexes, can be produced in *E. coli* [1, 2, 9–12]. Because these proteins vary greatly in their properties, to facilitate purification each protein is typically fused with an N- or a C-terminal affinity tag. A hexahistidine (His<sub>6</sub>)-tag is most commonly used as it allows

the use of immobilized metal affinity chromatography (IMAC) for all soluble proteins [4, 7, 13–15].

Rapid technological advances have been made during the last decade in structural genomics pipelines that use robotic hardware, software, process parallelization, improved fermentation, and semiautomated purification methods [16, 17]. Developments have also been made to reduce the time and labor of protein production methods [13, 18–21]. An example of an improvement in the fermentation step of protein production is the growth of high cell density bacterial cultures in polyethylene terephthalate (PET) bottles [19]. Expression of proteins using optimized minimal media in non-sterilized disposable PET bottles reduces the time required for media preparation and the fermentation volume by a factor of 2–3 and eliminates much of the cleanup that follows cell harvesting. Disposal of the autoclaved bottles after use reduces the risk of cross-contamination of subsequent cultures.

Another key development includes the availability of specifically labeled proteins for X-ray crystallography or NMR spectroscopy. This protocol also uses disposable fermentation vessels for production of selenomethionine-labeled proteins [1, 22]. It involves a production of in vivo-labeled proteins in optimized minimal media. The method relies on forcing selenomethionine incorporation via inhibition of endogenous methionine biosynthesis by the addition of inhibitory amino acids [23]. Selenium-labeled protein crystals are then used for X-ray diffraction data collection and de novo phasing for structure determination.

The choice of purification and handling procedures plays a critical role in obtaining high-quality protein samples. This standard purification protocol has been designed for His<sub>6</sub>-tagged proteins and implemented on a semiautomated chromatography platform with minimal human involvement [1, 9]. Purification starts with the preparation of crude extract (or lysate) followed by two chromatography steps. The first affinity chromatography step (IMAC-I) is done on a Ni<sup>+2</sup>-charged column and immediately followed by buffer exchange on a desalting column. The IMAC-I step produces proteins with purity typically above 80–95 % as judged by SDS-PAGE. After IMAC-I, the His<sub>6</sub>-tag from the target protein is cleaved with a highly specific recombinant tobacco etch virus (TEV) protease (developed by Dr. B. G. Fox, University of Wisconsin; this protease is fused to a non-cleavable His<sub>7</sub>-tag and is resistant to auto-inactivation) at a designed cleavage site (ENLYFQ↓S). This cleavage produces a protein with three amino acids (SNA) added to its N-terminus [24, 25]. The second chromatography step, a subtractive IMAC (IMAC-II), is also performed on the Ni<sup>+2</sup>-charged column. During the IMAC-II process, the cleaved protein elutes in the flow through and is separated from the released His<sub>6</sub>-tag, undigested protein, persistent endogenous *E. coli* proteins having affinity for IMAC column, and His<sub>7</sub>-tagged TEV protease. This step is also followed by buffer exchange. The buffer exchange process is important as it removes low-molecular-weight contaminants and transfers the protein into buffer conditions suitable for downstream steps, such as protease treatment, protein concentration, crystallization, and storage. The protein is then concentrated using centrifugal concentrators. If higher purity protein sample is needed, the buffer exchange step after IMAC-I and IMAC-II or after both processes can be replaced with size-exclusion chromatography (SEC). Batches of different proteins can be purified in parallel using IMAC and buffer exchange or SEC columns. The

purification process described above typically produces high-quality protein samples with 95–98 % purity.

One of the key elements of large-scale protein production is the proper long-term storage of protein samples. We recommend that the concentrated protein solutions (20–200 mg/mL) are pipetted dropwise (30–50  $\mu$ L) directly into liquid nitrogen and the resulting small pellets are collected into plastic tubes and stored in liquid nitrogen or at  $-80$  °C. These small aliquots can be rapidly thawed at room temperature and used for subsequent crystallizations and biophysical characterizations or in biochemical experiments, while the bulk of the material remains safely stored.

It is important to keep in mind that the high-throughput approach described here involves processing a large number of proteins (or their variants) using standardized protocols that are not optimized for any given protein but work reasonably well for many proteins. Consequently, these protocols may need to be altered for a specific protein that failed the standard high-throughput purification procedure. This also includes efficient protocols for the purification of proteins without affinity tags and stable protein–protein and protein–nucleic acid complexes.

In this chapter, a method for large-scale production of heterologous fusion proteins in *E. coli* is described. The protocol for the production of native proteins uses 2-L PET bottles as disposable culture vessels. In addition, a modified expression protocol that allows for the production of selenium-labeled proteins in defined minimum medium is included. Finally, a method for the purification of numerous proteins that generates high-purity samples for structural biology and biochemical applications and a technique utilized to prepare protein samples for long-term storage are presented.

## 2 Materials

Prepare all solutions using ultrapure water (resistance of 18 M $\Omega$  in a 1 cm cell; conductivity of less than 5.6  $\mu$ S/m) and analytical grade reagents. Follow all waste disposal guidelines when disposing waste materials.

### 2.1 Materials for Media Preparation for the Production of Native Proteins

1. Ampicillin, sodium salt: Prepare a 100 mg/mL stock by dissolving 1 g of ampicillin in water to a final volume of 10 mL. Store aliquots at  $-20$  °C. 1 mL of 100 mg/mL solution is required per 1 L of culture to achieve a final concentration of 100  $\mu$ g/mL.
2. Kanamycin sulfate: Prepare a 30 mg/mL stock by dissolving 300 mg kanamycin sulfate in water to a final volume of 10 mL. Store aliquots at  $-20$  °C. 1 mL of 300 mg/mL solution is required per 1 L of culture to achieve a final concentration of 30  $\mu$ g/mL.
3. Luria Bertani (LB) broth: Dissolve 25 g of preformulated LB powder in water to a final volume of 1 L. Add ampicillin and kanamycin sulfate to 100 and 30  $\mu$ g/mL, respectively (*see* Note 1).

## 2.2 Materials for Media Preparation for the Production of Selenomethionine-Labeled Proteins

1. Ampicillin, kanamycin sulfate, B1, and B12 vitamins, premixed solids in a foil packet (Orion Enterprises, Inc): Prepare a stock solution by dissolving contents of one foil pouch (ampicillin, 2.5 g; kanamycin sulfate, 1.5 g; thiamine (B1), 50 mg; and vitamin B12, 135 mg) in 50 mL of water. Store aliquots at  $-20^{\circ}\text{C}$ , and avoid light during storage. 1 mL of this solution is required per 1 L of culture to achieve a final concentration of 50  $\mu\text{g}/\text{mL}$  for ampicillin, 30  $\mu\text{g}/\text{mL}$  for kanamycin sulfate, 1  $\mu\text{g}/\text{mL}$  for thiamine, and 2.7  $\mu\text{g}/\text{mL}$  for vitamin B12.
2. Trace Metal Supplement, as concentrated solution (Orion Enterprises, Inc): Store at  $4^{\circ}\text{C}$ , and avoid light. Use 10 mL for 1 L culture to achieve final concentrations of 5 mg EDTA, 430 mg  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ , 5 mg  $\text{MgSO}_4$ , 10 mg NaCl, 1 mg  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ , 1 mg  $\text{Co}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ , 11 mg  $\text{CaCl}_2$ , 1 mg  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.1 mg  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , 0.1 mg  $\text{AlK}(\text{SO}_4)_2$ , 0.1 mg  $\text{H}_3\text{BO}_3$ , 0.1 mg  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ , 0.1 mg  $\text{Na}_2\text{SeO}_3$ , 0.1 mg  $\text{Na}_2\text{WO}_4 \cdot 2\text{H}_2\text{O}$ , 0.2 mg  $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ .
3. Vacuum filtration system with a 0.22  $\mu\text{m}$  polyethersulfone membrane such as sterile Stericup® vacuum filtration unit (Millipore Corporation).
4. 50 % Glycerol solution (Sigma Aldrich): Use ultrapure water and a ReagentPlus grade of glycerol to prepare the 50 % solution. Filter the solution using sterile vacuum filtration unit. Use 10 mL of this solution per 1 L of culture to achieve a final glycerol concentration of 0.5 %.
5. M9 “Pink” medium and non-inhibitory amino acids (NIAAC), premixed solids in a foil packet (Orion Enterprises, Inc): Dissolve one foil pouch ( $\text{Na}_2\text{HPO}_4$ , 6.8 g;  $\text{KH}_2\text{PO}_4$ , 3.0 g; NaCl, 0.75 g; glucose monohydrate 4.4 g;  $\text{NH}_4\text{Cl}$ , 1.0 g;  $\text{Na}_2\text{SO}_4$ , 0.28 g; 11 amino acids (L-glutamate, L-aspartate, L-arginine, L-histidine, L-alanine, L-proline, L-glycine, L-serine, L-glutamine, L-asparagine, L-tryptophan, 200 mg each)) in water to a final volume of 1 L (*see* Note 2).
6. Selenomethionine and inhibitory amino acids (IAAC), premixed solids in a foil packet (Orion Enterprises, Inc): Dissolve one foil pouch in 250 mL of water. Use 30 mL for 1 L of cell culture resulting in 150 mg each/L of L-valine, L-isoleucine, L-leucine, L-lysine, L-threonine, L-phenylalanine, and 90 mg/L of selenomethionine.

## 2.3 Other General Materials for Protein Production

1. Freshly transformed *E. coli* BL21(DE3) cells (Stratagene) carrying plasmid pMAGIC and containing a gene for the target protein in vector pMCSG7 [25] or one of its derivatives (*see* Note 3). Alternatively, a glycerol stock can be used. For details on cloning protocols *see* Chapter 5 in this book.

<sup>1</sup>The use of antibiotics and freshly prepared media alleviates the need for sterilization of media and significantly reduces the labor involved. We have observed that un-inoculated controls exhibited no growth during the time required for protein expression in experimental cultures [19].

<sup>2</sup>Keep the culture volume at exactly or below 1,000 mL for convenient centrifugation in 1-L centrifugation bottles.

2. Isopropyl- $\beta$ -D-thio-galactopyranoside (IPTG), premixed solid in a foil packet (Orion Enterprises, Inc). Prepare a 1 M stock by mixing 11.9 g of IPTG (one pouch) with water to a final volume of 50 mL and store the aliquots at  $-20^{\circ}\text{C}$ . 0.5 mL of this solution is required per 1 L of culture to achieve a final concentration of 0.5 mM.
3. Lysozyme from chicken egg white. Prepare a 50 mg/mL stock by mixing 500 mg of lysozyme with water to a final volume of 10 mL. Store at  $-20^{\circ}\text{C}$ . 1 mL of 50 mg/mL solution is required per 50 mL of a re-suspended culture pellet to achieve a final concentration of 1 mg/mL.
4. Protease inhibitor cocktail such as Complete Protease Inhibitor Cocktail tablets (Roche). Use one tablet per 8 g of wet cell paste (or 50 mL of resuspended cells).
5. Sterile tubes for starter cultures such as 5-mL round bottom polypropylene culture tubes.
6. 500-mL polyethylene terephthalate beverage bottles (PET, Continental Glass and Plastics, Inc). Use one bottle per target protein for a small overnight culture. One bottle contains enough culture to inoculate two 1-L cultures [19, 22].
7. 2-L polyethylene terephthalate beverage bottles (PET, Continental Glass and Plastics, Inc.). Use one bottle per liter of culture [19, 22].
8. Foam tube plugs such as Identi-Plugs<sup>®</sup> (Jaece Industries).
9. 1-L centrifuge bottles (such as Nalgene<sup>®</sup> high speed polycarbonate centrifuge bottles with polypropylene-ether screw/blue silicone gaskets).
10. Incubator/shaker.
11. Centrifuge.
12. A spectrophotometer and cuvette that can measure absorbance at 600 nm and 10 mm disposable semi-micro cuvettes.

## 2.4 Materials for Protein Purification

1. Purification buffers. All buffers must be filtered through a 0.22  $\mu\text{m}$  polyethersulfone filter:
  - a. Lysis buffer (50 mM HEPES pH 8.0, 500 mM NaCl, 5 % glycerol, 20 mM imidazole, 10 mM  $\beta$ -mercaptoethanol).
  - b. Desalting buffer (50 mM HEPES pH 8.0, 500 mM NaCl, 5 % glycerol, 10 mM  $\beta$ -mercaptoethanol).
  - c. Elution buffer (50 mM HEPES pH 8.0, 500 mM NaCl, 5 % glycerol, 250 mM imidazole, 10 mM  $\beta$ -mercaptoethanol).

---

<sup>3</sup>Plasmid pMAGIC encodes one rare *E. coli* tRNA specific for two Arg codons (AGG/AGA) [1]. The Midwest Center for Structural Genomics has developed over 80 expression vectors for large-scale protein production that are optimized for structural genomics applications [14, 25]. Many of these vectors are available to the scientific community from the PSI Biology Materials Repository (<http://psimr.asu.edu>).

- d. Crystallization buffer (20 mM HEPES pH 8.0, 250 mM NaCl, 2 mM 1,4-dithio-DL-threitol).
2. Recombinant tobacco etch virus (TEV) protease carrying a non-cleavable His<sub>7</sub>-tag [24] and mutated to reduce self-inactivation [26]. Use purified TEV protease as a 5–10 mg/mL solution in 20 mM HEPES pH 8.0, 300 mM NaCl, 10 % glycerol, 2 mM β-mercaptoethanol. For His<sub>6</sub>-tag cleavage, incubate at a protein: protease mass ratio of 1:50 (*see* Note 4).
3. Sonicator.
4. Centrifuge.
5. Polyvinylidene fluoride (PVDF) syringe filter (0.45 μM).
6. ÄKTAexpress chromatography system (GE Healthcare), or the equivalent.
7. 5-mL HisTrap HP column (GE Healthcare) or the equivalent.
8. 53-mL HiPrep 26/10 desalting column (GE Healthcare) or the equivalent.
9. Superdex 200 HiLoad 26/60 size exclusion chromatography (SEC) column (GE Healthcare) or the equivalent.
10. 96-well deep well blocks such as sterile, 96-well 2-mL square top, round bottom polypropylene plates (BioExpress).
11. Slide-A-Lyzer™ dialysis cassette (Thermo Scientific) of proper liquid capacity or the equivalent dialysis device.
12. Centrifugal filter devices.
13. A spectrophotometer absorbance at 280 nm (such as NanoDrop 1000, for which no cuvette is required)
14. Cryogenic dewar (such as foam dewar FD-800, Spearlab) with small (10–20 mL) glass beakers.
15. Plastic cryogenic screw-cap tubes (such as 2.0-mL Nalgene® cryogenic vials).
16. Conical centrifuge tubes.

### 3 Methods

It is important to mention that large-scale protein fermentation is one of the many components of the protein production pipeline at a structural genomics center. Before proteins are produced on milligram-scale, target proteins are processed through several stages of the pipeline. First, protein targets are carefully selected, cloned into appropriate expression vectors, and protein expression and solubility is assessed in a small-scale (5 mL).

---

<sup>4</sup>The TEV protease typically used in our procedures was expressed from vector pRK508, which carries a non-cleavable His<sub>7</sub>-tag (a gift from Dr. D. Waugh, NCI). More recently, we started using a construct expressed from vector pMHT238 (developed by Dr. B. G. Fox, University of Wisconsin). This plasmid, in addition to non-cleavable His<sub>7</sub>-tag, also carries several mutations and a C-terminal truncation that improve TEV protease solubility and produce high levels of the protein (100–130 mg/L of LB medium). This plasmid can be obtained from the PSI Biology Materials Repository [28].

Only the targets with confirmed high-level expression and good solubility are moved into the large-scale fermentation process described here. Failed constructs are salvaged [1].

The incubators, shakers, and sonicators should be disinfected biweekly using 10 % bleach. All disposables should be autoclaved, while non-disposables should be disinfected in 10 % bleach. When disinfecting with 10 % bleach, ensure at least 20 min of contact time. Wear nitrile gloves, safety glasses and a lab coat during these procedures.

### 3.1 Production of Native Proteins

1. To prepare a starter culture, inoculate 1 mL of LB medium containing 100 µg/mL ampicillin and 30 µg/mL kanamycin sulfate in a 5-mL plastic tube with a fresh transformant or a glycerol stock of pMCSG7/pMAGIC transformed *E. coli* BL21(DE3) cells. Place in an incubator and shake at 220 rpm and 37 °C for 4–6 h. Use one tube per protein.
2. To prepare small overnight culture, add 50 µL of the turbid starter culture to 50 mL of fresh LB medium containing 100 µg/mL ampicillin and 30 µg/mL kanamycin sulfate in a 500-mL PET beverage bottle. Place in an incubator and shake at 150 rpm and 37 °C overnight.
3. Prepare LB medium in 2-L PET bottles using 990 mL of ultrapure water. Add ampicillin and kanamycin sulfate to a final concentration of 100 and 30 µg/mL, respectively.
4. Add 10 mL of the saturated overnight culture to each 1 L of fresh LB medium containing 100 µg/mL ampicillin and 30 µg/mL kanamycin sulfate. Shake the bottles in an incubator at 37 °C at 180 rpm until cells grow to OD<sub>600nm</sub> 0.8–1.0. Start measuring OD after about 2.0–2.5 h.
5. Transfer bottles to a pre-cooled 4 °C incubator and shake for 1 h (*see* Note 5).
6. Increase incubator temp to 18 °C. Induce the cells by adding 0.5 mL of 1 M IPTG solution to a final concentration of 0.5 mM.
7. Grow cells at 18 °C and shake at 180 rpm overnight.
8. Pour cell culture in 1-L centrifuge bottles. Centrifuge for 10 min at 7,880 × g (in a rotor such as Sorvall F8S-6x1000y). Remove the supernatant.
9. Perform all procedures at 4 °C from here on.
10. Re-suspend the cells by adding 25 mL of ice-cold lysis buffer and gentle shaking (150 rpm).
11. Add one tablet of protease inhibitor cocktail per 8 g of wet cell paste.
12. Transfer the re-suspended cells into a 50-mL tube and bring volume up to 40 mL with lysis buffer (*see* Note 6).

---

<sup>5</sup>We routinely grow 16 L of cultures at once. Transferring all 16 cultures to an incubator that is kept at 4 °C allows them to cool quickly within 45–60 min to achieve a final culture temperature of 16–18 °C. When a smaller number of cultures are used (such as four to eight 2-L bottles) the incubator can be pre-cooled to 10 °C to achieve the same final effect.

13. Add 50 mg/mL lysozyme solution to a final concentration of 1 mg/mL (for example, 0.8 mL of the solution for 40 mL of cell suspension).
14. Cryo-cool the cell suspension with liquid nitrogen and store at  $-80^{\circ}\text{C}$ .

### 3.2 Production of the Selenomethionine Labeled Proteins

1. Prepare 1-mL starter cultures in LB medium with antibiotics as described in **step 1** of Subheading 3.1.
2. Dissolve 1 pouch of M9 “Pink” salts and NIAAC in 980 mL of ultrapure water in 2-L PET beverage bottles. Add 10 mL of 50 % glycerol, 10 mL of Trace Metal Supplement, 1 mL of ampicillin, kanamycin sulfate, B1, and B12 vitamins stock solution.
3. To make small overnight cultures, aliquot 50 mL of the prepared M9 “Pink” Medium into 500-mL PET beverage bottles.
4. Add 50  $\mu\text{L}$  of the turbid starter LB culture to each 50 mL of M9 “Pink” medium. Place the bottles in an incubator and shake at 150 rpm and  $37^{\circ}\text{C}$  overnight.
5. Prepare M9 “Pink” medium in 2-L PET bottles described in **step 2** of this Subheading 3.1. using 940 mL of ultrapure water (*see* Note 2).
6. Add 25 mL of small culture from 500-mL PET bottle to each 2-L bottle of the prepared M9 “Pink” medium.
7. Shake in an incubator at  $37^{\circ}\text{C}$  at 180 rpm until cells grow to  $\text{OD}_{600\text{nm}}$  1.0–1.4. Start measuring OD after about 2.5–3 h.
8. Transfer bottles to a precooled  $4^{\circ}\text{C}$  incubator and shake for 1 h (*see* Note 5).
9. Add 30 mL of selenomethionine and IAAC solution per 1 L of culture (*see* Note 7).
10. Increase incubator temp to  $18^{\circ}\text{C}$ . Induce the cells by adding 0.5 mL of 1 M IPTG solution at a final concentration of 0.5 mM.
11. Grow cells at  $18^{\circ}\text{C}$  and shake at 180 rpm overnight.
12. Recover the cells by centrifugation, and proceed as described in **step 8** of Subheading 3.1.

### 3.3 Protein Purification

#### 3.3.1 IMAC-I Purification

1. Lyse the cell suspension. Thaw the cells on ice with gentle mixing and sonicate on ice for 5 min by means of a program or with manual control using a 4-s on and 20-s

---

<sup>6</sup>Keep the volume of the cell resuspension below 40 mL to account for an increase of solution volume upon immersing the sonication probe during the lysis step.

<sup>7</sup>We routinely use 90 mg of selenomethionine per 1 L of culture (30 mL of selenomethionine solution per 1 L) with the assumption that methionine residues comprise 2 % of the protein amino acid composition. If the methionine content is higher (5 % or more) the amount of selenomethionine added to the culture needs to be increased to 120 mg/L [29].



off mode and power/voltage settings according to the manufacturer's instructions (see Note 8).

2. Transfer the cell lysate to 45-mL centrifuge tubes. Centrifuge for 1 h at  $29,500 \times g$  in an SS-34 rotor. Alternatively, centrifugation can be performed in the 50-mL tubes for 1 h and 20 min in a conical fixed-angle rotor (such as Sorvall F13–14x50cy). Filter the supernatant into a clean 50-mL tube using 0.45  $\mu\text{m}$  PVDF or polyethersulfone syringe filter to remove residual particulates.
3. Apply the filtered supernatant with a maximum flow rate of 1 mL/min to a 5-mL HisTrap HP column (see Note 9) that was equilibrated with lysis buffer. Wash the column at 4 mL/min with 30 column volumes (CV) of lysis buffer to remove contaminating endogenous *E. coli* proteins. Elute the bound His<sub>6</sub>-target proteins with elution buffer over 5 CV at 4 mL/min. Collect fractions into a 10 mL loop or a 96-well deep-well block (Fig. 1, top). If high-purity protein sample is needed do not use the desalting column in the next step, but proceed to Subheading 3.3.3. Otherwise, see **step 4** of this section.
4. Apply the eluted protein to a 53-mL HiPrep 26/10 desalting column (GE Healthcare) pre-equilibrated with desalting buffer. Wash with 2.5 CV of desalting buffer at 8 mL/min. Collect 1.7-mL fractions into a 96-well deep-well block (see Note 10).
5. Pool the peak fractions, and measure the volume. Take a 50  $\mu\text{L}$  aliquot of each protein sample. These samples will be later used as “uncut controls” for SDS-PAGE. Estimate the concentration of the partially pure protein spectrophotometrically at 280 nm using an appropriate molar extinction coefficient.
6. Calculate the amount of TEV protease needed (protein:protease mass ratio 1:50). Add fresh  $\beta$ -mercaptoethanol to the protein pool to replenish the reducing agent. Add the required amount of protease.
7. Incubate proteins with TEV protease for 1–2 days at 4 °C. Run SDS-PAGE to determine if the cleavage was successful using the “uncut” samples as a reference. If the His<sub>6</sub>-tag was cleaved, run IMAC-II (see Note 11).

---

<sup>8</sup>We routinely break cells using sonicators from Qsonica equipped with a converter and a dual horn that allows us to process two samples simultaneously. Other lysis techniques such as French press, detergent lysis, or manual shearing should yield comparable results.

<sup>9</sup>For routine purifications, we use an ÄKTExpress chromatography system (GE Healthcare) and 5-mL HisTrap columns, HiPrep 26/10 desalting columns, and/or Superdex 200 26/60 SEC columns (all columns from GE Healthcare). The ÄKTExpress system has the capability of purifying four different proteins per purification unit. Each ÄKTExpress unit has five column positions. We use four of these positions for Ni-affinity columns and one for a desalting or an SEC column. There are five loops available to store peak fractions from the IMAC step before the protein is loaded onto the desalting or the SEC column. If a chromatography system is not available, the IMAC can be performed using manually poured Ni-NTA columns with a peristaltic pump or by gravitational flow.

<sup>10</sup>The buffer exchange step is introduced to prepare the protein sample for a cleavage with TEV protease and a subsequent IMAC-II step. We find that using a desalting column instead of dialysis speeds up the purification process, as the buffer exchange step is included in our ÄKTExpress program following the IMAC-I step. Specifically, the protein is eluted from a Ni-affinity column using elution buffer into a 10 mL loop and then applied onto a desalting column. To prevent Ni<sup>2+</sup> ion leakage from the metal affinity column, 2 mL of 5 mM EDTA solution in the desalting buffer is injected onto the desalting column prior to the injection of the protein. This creates an EDTA zone that sequesters released Ni<sup>2+</sup> ions.

### 3.3.2 IMAC-II Purification

1. Apply the IMAC-I protein pool and TEV protease mixture onto a 5-mL HisTrap HP column (GE Healthcare) pre-equilibrated with lysis buffer with a maximum flow rate of 1 mL/min. Wash the column at 3 mL/min with 8 CV of lysis buffer to elute the cleaved target protein. Next, increase the imidazole concentration to 32.5 mM (5 % elution buffer), and wash the column with 5 CV of that buffer at 4 mL/min to elute any residual target protein that was not washed out in the lysis buffer wash. Finally, elute the uncut His<sub>6</sub>-tag protein (if any) and the His<sub>7</sub>-TEV protease with 100 % of elution buffer over 3 CV at 4 mL/min. Collect all peaks from the flow-through, wash, and elution in 1.7-mL fractions into a 96-well deep-well block (Fig. 1, bottom).
2. Pool the peak fractions from the flow-through, and wash and measure the volume. Estimate the concentration of the protein spectrophotometrically at 280 nm using appropriate molar extinction coefficient. Run SDS-PAGE to determine sample purity (Fig. 2). If a protein sample of high purity is needed, proceed to the next section (Subheading 3.3.3). Otherwise, *see step 3* of this section.
3. Dialyze against crystallization buffer overnight at 4 °C in dialysis cassettes. Choose an appropriate dialysis cassette membrane capacity for the nominal molecular weight limit of the protein.
4. Concentrate to a final volume of ~1 mL using a centrifugal concentrator depending on protein solubility (*see Note 12*). Measure protein concentration at 280 nm.
5. Set up crystallizations (*see Chapter 12* in this book).
6. Freeze the remaining protein via drop freezing by dispensing the protein sample as a series of single drops. Use a P200 pipet and aliquot the protein into small beakers filled with liquid nitrogen and placed in a foam dewar. Dispense from a height of approximately 4 in. [27]. Collect the frozen droplets into a pre-chilled plastic cryogenic screw-cap tube and store at –80 °C (*see Note 13*).

**3.3.3 Size-Exclusion Chromatography Purification**—If purification of a single oligomeric state is critical, the purity of the sample is not satisfactory, or protein aggregation is suspected, the buffer exchange step after IMAC-I, IMAC-II, or both can be replaced with

---

<sup>11</sup>We frequently find that His<sub>6</sub>-tag cleavage is complete after overnight incubation with TEV protease. Sometimes we encounter fusion proteins that are only partially cleaved. This may be due to steric occlusion of the cleavage site. In some cases that problem can be mitigated by using larger amount of the protease and performing the reaction at higher temperature. On occasion, we deal with proteins that do not undergo cleavage of N-terminal His<sub>6</sub>-tag with TEV protease despite being designed to do so. In such cases, we do not proceed with the IMAC-II step but purify the protein using SEC and set up crystallizations with the uncleaved form. We also typically attempt to re-clone such protein into a vector that has a longer linker between the TEV cleavage site and the N-terminus of the protein or a vector with a C-terminal His<sub>6</sub>-tag.

<sup>12</sup>Because high protein concentrations are needed for crystallization, we typically concentrate proteins to a concentration of ~1 mM or higher, depending on protein solubility. On average we obtain 50–100 mg of protein from 1 L of culture (thus it is not uncommon to have, for example, a final volume and concentration of 1 mL of 60 mg/mL solution for a 30-kDa bacterial protein).

<sup>13</sup>Long-term protein storage is an important issue. The common practice of freezing proteins in a glycerol solution is problematic because glycerol can possibly interfere with crystallization and may need to be removed by dialysis after thawing. To overcome this problem we have implemented a drop-cryoprotection procedure, in which small aliquots (30–50 µL) of protein solution are dropped directly into small cylindrical containers (10–20 mL beakers) filled with liquid nitrogen. Dropped aliquots form spherical pellets, which sink to the bottom of the container. These pellets are retrieved from the bottom of the container with a pair of forceps, placed into pre-chilled Nalgene<sup>®</sup> cryogenic vials, and kept at liquid nitrogen temperature or at –80 °C for long-term storage.

SEC step. The choice of the running buffer for this step will depend on the condition and destination of the protein sample. In the case where the SEC procedure is the final purification step, it is recommended to use crystallization buffer. If the protein sample needs to undergo His<sub>6</sub>-tag cleavage with TEV protease, using the desalting buffer is recommended.

1. Equilibrate the SEC column with at least 2 CV of desalting or crystallization buffer at a flow rate recommended by the manufacturer (for Superdex 200 HiLoad 26/60 SEC column a flow rate of 2.5–3.0 mL/min is used).
2. Apply the protein sample onto an SEC column (for Superdex 200 HiLoad 26/60 use a flow rate of 1.5–2.0 min/mL). The volume of the samples should be no more than 2 % of the column volume and be free of precipitation (if needed, concentrate to a smaller volume and remove the precipitation by centrifugation or filtration through a 0.22 µm PVDF or polyethersulfone syringe filter). Collect 1.8 mL fractions into a 96-well deep-well block (*see* Note 14).
3. Pool the peak fractions from the size-exclusion column. If SEC purification was performed after IMAC-I, follow Subheading 3.3.1, **step 6**. If SEC purification was performed after IMAC-II, follow Subheading 3.3.2, **step 4**.

### 3.3.4 Evaluation of Bacterial Protein Expression in *E. coli* System—*E. coli*

strains are commonly used as a host for expression of proteins of microbial origin (and many eukaryotic proteins as well). The method is cost effective and remains the first choice for protein expression in structural genomics centers [1, 2, 4]. Over the years, over 20,000 proteins from over 150 different bacterial strains have been produced using the standardized protocols described here. Large-scale expression and solubility levels for a sequence diverse set of proteins from over 60 bacterial strains that were produced in *E. coli* as an expression host have been analyzed. Results of the analysis are presented in Fig. 3. It was observed that, while good expression levels (40 % or higher) were obtained for proteins from more than 40 of the examined bacterial species, solubility levels for these proteins were generally lower (20–30 %). This method was very successful in expressing over 80 % of target proteins from Gram-negative *Burkholderia thailandensis* E264, of which almost 50 % were soluble. Interestingly, proteins from Gram-positive *Bacillus cereus* ATCC 10987 were also expressed at high levels (over 70 %) with 40 % of them soluble, whereas *Bacillus anthracis* str. Sterne was more problematic, with 40 % of tested proteins expressed and only 5 % soluble. A very limited success was achieved producing proteins from *Rhodopseudomonas palustris* BisB18, for which only about 5 % of targets were overexpressed and none were soluble. These results suggest that many bacterial species may produce proteins that are less compatible with standard protein production protocols and may require a more customized approach for expression and purification. Alternatively, these difficult proteins may need to be tested in other expression systems like cell-free expression and insect or mammalian cells.

---

<sup>14</sup>The preparative SEC can be performed on any commercially available purification workstation or with a simple peristaltic pump system. We have developed protocols for ÄKTApexpress and ÄKTAexplorer chromatography systems (GE Healthcare), in which the buffer exchange step after IMAC-I or IMAC-II is replaced with preparative SEC.

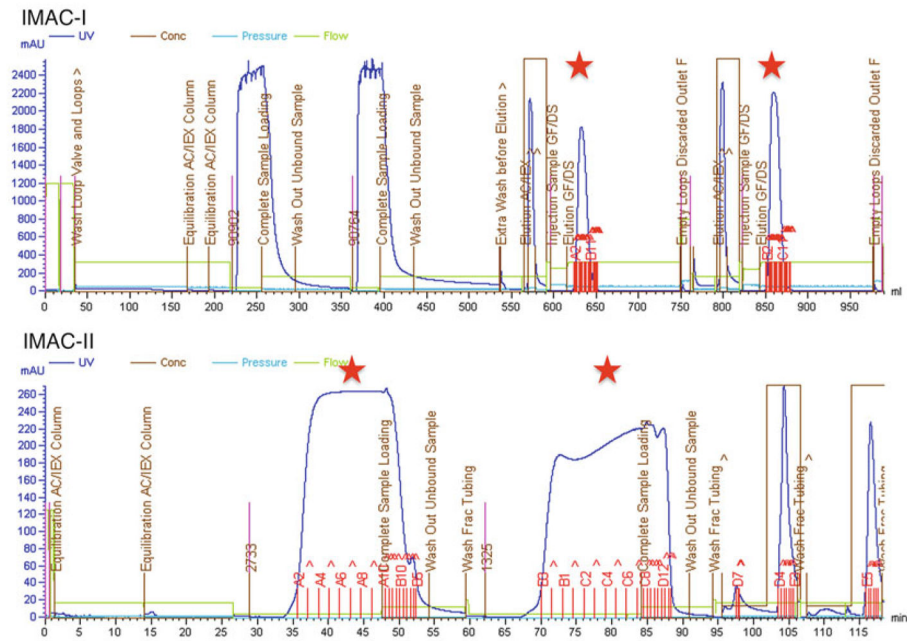
## Acknowledgments

This work was supported by the National Institute of Allergy and Infectious Diseases Contract (HHSN272200700058C and HHSN272201200026C), the National Institute of Health Grant GM094585, and the US Department of Energy office of Biological and Environmental Research under Contract No. DE-AC02-06CH11357. This work has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”). Argonne, a US Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The US Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or behalf of the government.

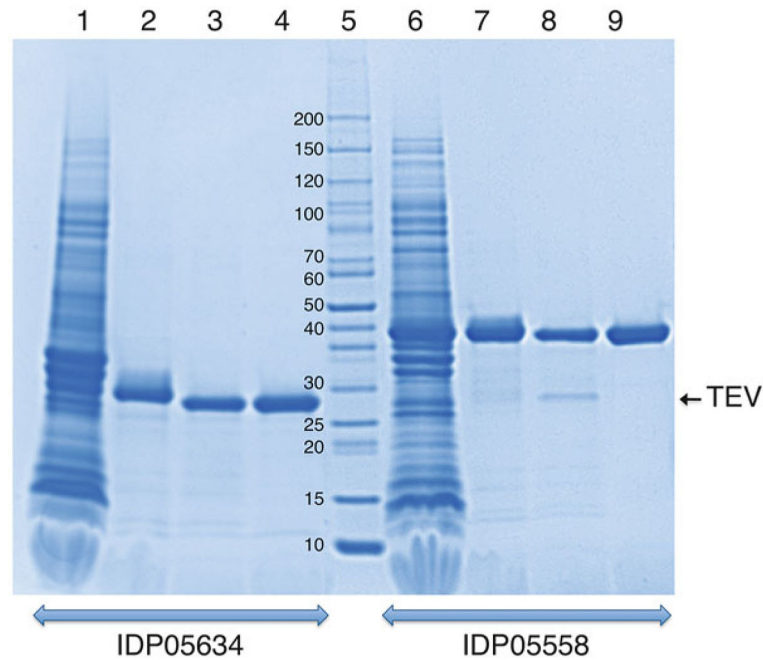
## References

- Kim Y, Babnigg G, Jedrzejczak R, et al. High-throughput protein purification and quality assessment for crystallization. *Methods*. 2011; 55:12–28. [PubMed: 21907284]
- Gräslund S, Nordlund P, Weigelt J, et al. Protein production and purification. *Nat Methods*. 2008; 5:135–146. [PubMed: 18235434]
- Vincentelli R, Bignon C, Gruez A, et al. Medium-scale structural genomics: strategies for protein expression and crystallization. *Acc Chem Res*. 2003; 36:165–172. [PubMed: 12641473]
- Elsiger M-A, Deacon AM, Godzik A, et al. The JCSG high-throughput structural biology pipeline. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2010; 66:1137–1142.
- Peti W, Page R, Moy K, et al. Towards miniaturization of a structural genomics pipeline using micro-expression and microcoil NMR. *J Struct Funct Genomics*. 2005; 6:259–267. [PubMed: 16283429]
- Price WN II, Handelman SK, Everett JK, et al. Large-scale experimental studies show unexpected amino acid effects on protein expression and solubility in vivo in *E. coli*. *Microb Inform Exp*. 2011; 1:1–20.
- Xiao R, Anderson S, Aramini J, et al. The high-throughput protein sample production platform of the Northeast structural genomics consortium. *J Struct Biol*. 2010; 172:21–33. [PubMed: 20688167]
- Studier FW. Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif*. 2005; 41:207–234. [PubMed: 15915565]
- Kim Y, Dementieva I, Zhou M, et al. Automation of protein purification for structural genomics. *J Struct Funct Genomics*. 2004; 5:111–118. [PubMed: 15263850]
- Lesley SA, Wilson IA. Protein production and crystallization at the joint center for structural genomics. *J Struct Funct Genomics*. 2005; 6:71–79. [PubMed: 16211502]
- Lesley SA, Kuhn P, Godzik A, et al. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc Natl Acad Sci*. 2002; 99:11664–11669. [PubMed: 12193646]
- Page R, Peti W, Wilson IA, et al. NMR screening and crystal quality of bacterially expressed prokaryotic and eukaryotic proteins in a structural genomics pipeline. *Proc Natl Acad Sci U S A*. 2005; 102:1901–1905. [PubMed: 15677718]
- Donnelly MI, Zhou M, Millard CS, et al. An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. *Protein Expr Purif*. 2006; 47:446–454. [PubMed: 16497515]
- Eschenfeldt W, Maltseva N, Stols L, et al. Cleavable C-terminal His-tag vectors for structure determination. *J Struct Funct Genomics*. 2010; 11:31–39. [PubMed: 20213425]
- Porath J. Immobilized metal ion affinity chromatography. *Protein Expr Purif*. 1992; 3:263–281. [PubMed: 1422221]
- Burley SK, Joachimiak A, Montelione GT, et al. Contributions to the NIH-NIGMS protein structure initiative from the PSI production centers. *Structure*. 2008; 16:5–11. [PubMed: 18184575]
- Joachimiak A. High-throughput crystallography for structural genomics. *Curr Opin Struct Biol*. 2009; 19:573–584. [PubMed: 19765976]
- Kim Y, Bigelow L, Borovilos M, et al. High-throughput protein purification for X-ray crystallography and NMR. *Adv Protein Chem Struct Biol*. 2008; 75:85–105. [PubMed: 20731990]

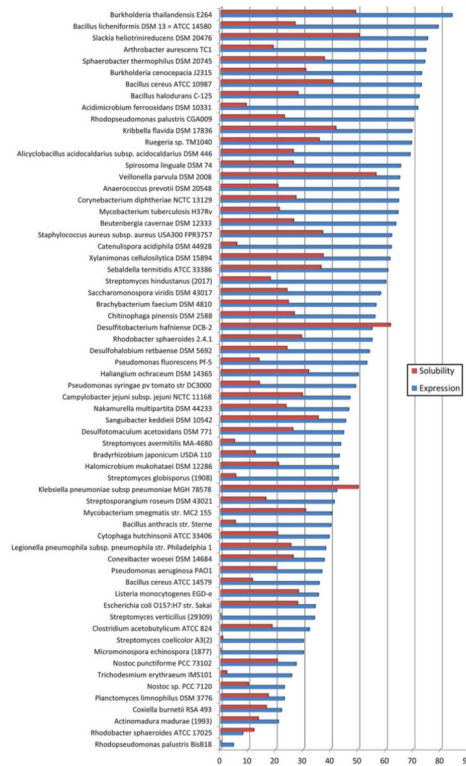
19. Millard CS, Stols L, Quartey P, et al. A less laborious approach to the high-throughput production of recombinant proteins in *Escherichia coli* using 2-liter plastic bottles. *Protein Expr Purif.* 2003; 29:311–320. [PubMed: 12767824]
20. Abdullah JM, Joachimiak A, Collart FR. “System 48” high-throughput cloning and protein expression analysis. *Methods Mol Biol.* 2009; 498:117–127. [PubMed: 18988022]
21. Acton TB, Xiao R, Anderson S, et al. Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods Enzymol.* 2011; 493:21–60. [PubMed: 21371586]
22. Stols L, Millard CS, Dementieva I, et al. Production of selenomethionine-labeled proteins in two-liter plastic bottles for structure determination. *J Struct Funct Genomics.* 2004; 5:95–102. [PubMed: 15263848]
23. Van Duyne GD, Standaert RF, Karplus PA, et al. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol.* 1993; 229:105–124. [PubMed: 7678431]
24. Kapust RB, Waugh DS. Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr Purif.* 2000; 19:312–318. [PubMed: 10873547]
25. Stols L, Gu M, Dieckman L, et al. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr Purif.* 2002; 25:8–15. [PubMed: 12071693]
26. Blommel PG, Fox BG. A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expr Purif.* 2007; 55:53–68. [PubMed: 17543538]
27. Deng J, Davies DR, Wisedchaisri G, et al. An improved protocol for rapid freezing of protein samples for long-term storage. *Acta Crystallogr D.* 2004; 60:203–204. [PubMed: 14684931]
28. Cormier C, Park J, Fiacco M, et al. PSI:Biological-materials repository: a biologist’s resource for protein expression plasmids. *J Struct Funct Genomics.* 2011; 12:55–62. [PubMed: 21360289]
29. Sauder MJ, Rutter ME, Bain K, et al. High throughput protein production and crystallization at NYSGXRC. *Methods Mol Biol.* 2008; 426:561–575. [PubMed: 18542890]



**Fig. 1.** Purification chromatograms of two His<sub>6</sub>-labeled proteins on IMAC-I (*top*) and IMAC-II after cleavage with TEV protease (*bottom*). Red stars indicate collected protein peaks



**Fig. 2.** Purification of two His<sub>6</sub>-labeled proteins monitored by SDS-PAGE (Lonza PAGEr™ Gold 4–20 % gradient precast gel). *Lanes 1 and 6:* Soluble cell extracts. *Lanes 2 and 7:* Pooled peak fractions after IMAC-I. *Lanes 3 and 8:* Cleavage of His<sub>6</sub>- tag with TEV protease. *Lanes 4 and 9:* Pooled peak fractions after IMAC-II. *Lane 5:* Molecular weight standards (kDa). IDP0564 (gi|126699790; amino acid ABC transporter, substrate-binding protein [*Clostridium difficile* 630]) and IDP05558 (gi|16803797; hypothetical protein lmo1757 [*Listeria monocytogenes* EGD-e]) are targets from the Center of Structural Genomics of Infectious Diseases (CSGID)



**Fig. 3.** Bacterial protein expression and solubility in the *E. coli* system. The expression and solubility values of more than 1,400 proteins from 25 species are shown on a normalized 0–100 scale. Expression (*blue bars*) and solubility (*red bars*) levels are calculated as a percentage of the relative abundance of the target protein in the cell lysate or the soluble fraction, respectively. The highest possible level of expression, when the relative abundance is above 80 %, is set as 100 %