

# A generic protocol for the expression and purification of recombinant proteins in *Escherichia coli* using a combinatorial His<sub>6</sub>-maltose binding protein fusion tag

Sreedevi Nallamsetty & David S Waugh

Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, P.O. Box B, Frederick, Maryland 21702-1201, USA. Correspondence should be addressed to D.S.W. (waughd@ncifcrf.gov).

Published online 8 March 2007; doi:10.1038/nprot.2007.50

**We describe a generic protocol for the overproduction and purification of recombinant proteins in *Escherichia coli*. The strategy utilizes a dual His<sub>6</sub>-maltose binding protein (HisMBP) affinity tag that can be removed from the target protein by digestion of the fusion protein at a designed site by tobacco etch virus protease. The MBP moiety serves to enhance the solubility and promote the proper folding of its fusion partners, and the polyhistidine tag facilitates its purification to homogeneity. This protocol is divided into three stages, each of which takes approximately 1 week to complete: (i) construction of a HisMBP fusion vector; (ii) a pilot experiment to assess the yield and solubility of the target protein; and (iii) the large-scale production and purification of the target protein.**

## INTRODUCTION

Affinity tags and other types of genetically engineered fusion partners are widely employed as tools in molecular biology. Although these were originally developed to facilitate the detection and purification of recombinant proteins, in recent years it has become clear that certain tags can also improve the yield, enhance the solubility and even promote the proper folding of their fusion partners. The strengths and weaknesses of various tags have been reviewed recently<sup>1–3</sup>. As the insolubility of recombinant proteins, particularly in *Escherichia coli*, is a major bottleneck in the production of bioactive material for structural and functional studies<sup>4</sup>, the utilization of solubility-enhancing tags to avoid the formation of insoluble protein aggregates has been rapidly gaining in popularity.

Although many proteins that are highly soluble when overproduced in *E. coli* have been reported to possess solubility-enhancing qualities as fusion partners, in most cases the evidence to support these claims is scant<sup>1</sup>. Not all highly soluble proteins can function as solubility enhancers, so it is a mistake to assume that these two properties can be equated. Nevertheless, two proteins with a very well-established ability to function as solubility enhancers are *E. coli* maltose binding protein (MBP) and NusA<sup>5,6</sup>. The mechanisms by which these two proteins enhance the solubility of their fusion partners are not well understood but may be similar<sup>7</sup>.

Among the known solubility-enhancing fusion partners, MBP is unique in that it is also a natural affinity tag. MBP fusion proteins can be purified using amylose affinity chromatography<sup>8</sup>. However, we and others have observed that MBP fusion proteins often do not bind efficiently to amylose resin and that even when they do, this type of affinity chromatography does not routinely produce samples of satisfactory purity<sup>3,9,10</sup>.

We previously demonstrated that adding a polyhistidine tag (His<sub>6</sub>) to the N-terminus of MBP does not interfere with its ability to promote the solubility and proper folding of its fusion partners<sup>11</sup>. Immobilized metal affinity chromatography (IMAC) utiliz-

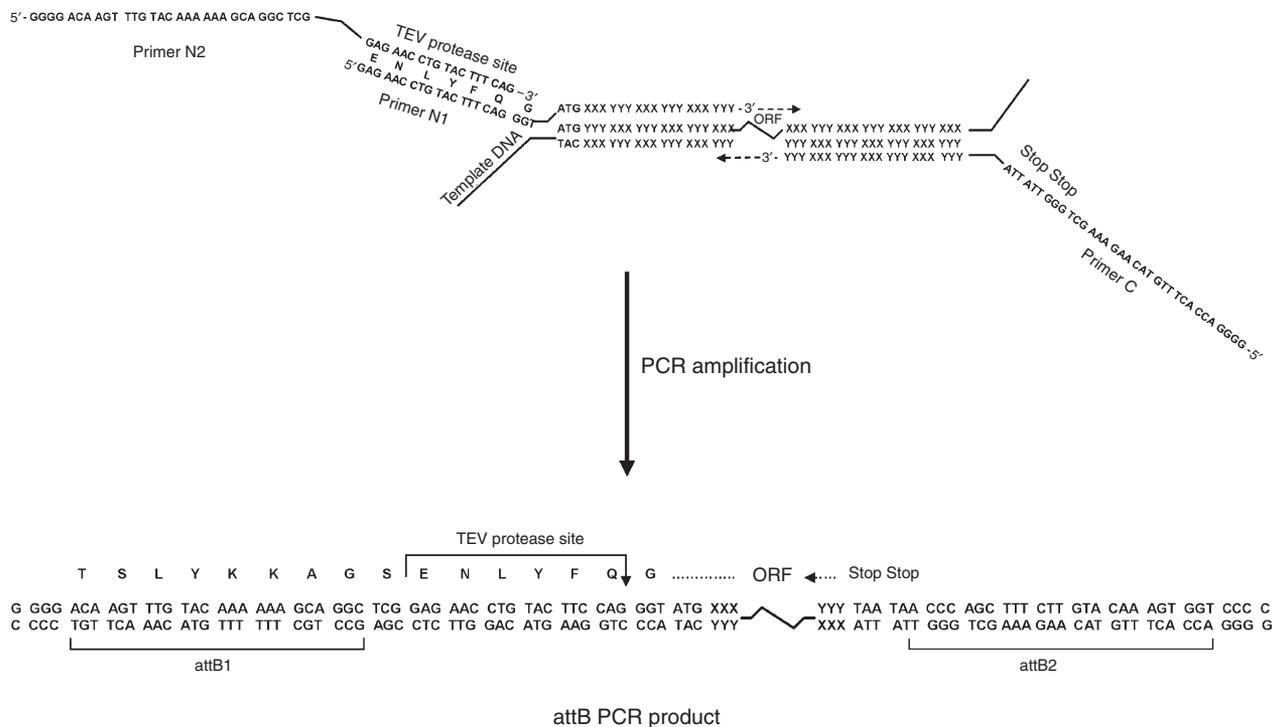
ing His<sub>6</sub>-tagged proteins has proven to be a very powerful method for protein purification and is routinely utilized by all of the large structural-genomics centers. Hence, by combining the His<sub>6</sub> and MBP tags, it is possible to exploit the unique benefits of each. The MBP moiety serves to enhance the solubility and promote the proper folding of its fusion partners, and the His tag facilitates their purification to homogeneity via an entirely generic protocol.

## Experimental design

This protocol describes how to (i) construct a dual His<sub>6</sub>-MBP (HisMBP)-tagged fusion protein expression vector by Gateway recombinational cloning; (ii) conduct a pilot experiment to assess the yield and solubility of the recombinant protein both before and after intracellular processing of the fusion protein by tobacco etch virus (TEV) protease; and (iii) purify the target protein on a large scale for structural or functional studies.

Gateway cloning technology utilizes the site-specific recombination reactions that mediate the integration and excision of bacteriophage lambda into and from the *E. coli* chromosome, respectively. For more detailed information on this system, the investigator is encouraged to consult the technical literature supplied by Invitrogen (<http://www.invitrogen.com/content/online%20seminars/gateway/gatewayhome.html>). Gateway recombinational cloning, which is featured in this protocol, greatly facilitates the construction of HisMBP fusion vectors by eliminating the need for restriction endonucleases and DNA ligase. However, alternative methods of cloning, such as ligation-independent cloning, may also be used provided that suitable vectors are available<sup>12</sup>.

To construct a HisMBP fusion vector by recombinational cloning, it is necessary to generate a PCR amplicon wherein the open reading frame (ORF) of interest is flanked by attB1 and attB2 recombination sites on its N- and C-termini, respectively. A recognition site for TEV protease is also added between the attB1 site and the N-terminus of the ORF. The PCR amplification is



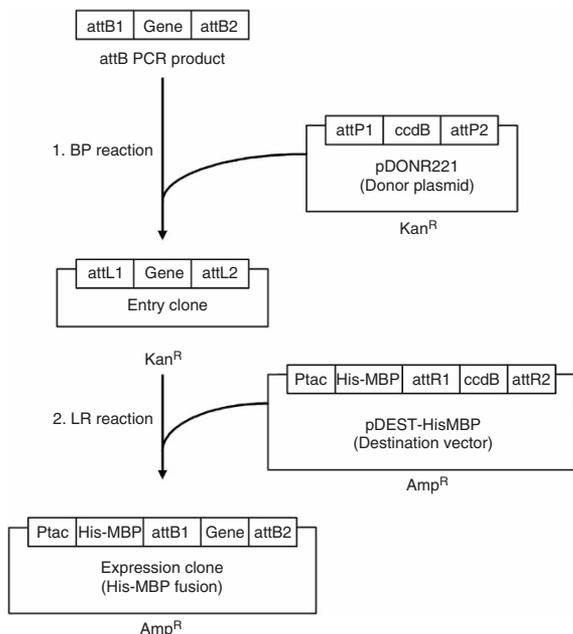
**Figure 1** | Schematic illustration of the strategy involved in generating a PCR product for recombinational cloning. Two gene-specific PCR primers (N1 and C), with constant regions indicated, are utilized in conjunction with one fixed primer (N2) that anneals to the tobacco etch virus (TEV) recognition site and adds the attB1 recombination site to the N-terminus of the ORF. The resulting PCR amplicon is suitable for use as a substrate in the BP Gateway reaction.

performed with two overlapping N-terminal primers and a single C-terminal primer, as outlined in **Figure 1**. Two gene-specific primers (N1 and C) are required for each ORF. The overlapping N-terminal primer (N2) can be used to add the attB1 recombination site to the N-terminus of any ORF to which the TEV protease recognition site has already been added.

To create the HisMBP fusion vector, the PCR product is first recombined by Gateway cloning into the donor vector pDONR221

to yield an entry clone intermediate (BP reaction), and then into pDEST-HisMBP (LR reaction) to generate the HisMBP fusion protein expression vector (**Fig. 2**). The BP and LR reactions are analogous to the recombination reactions that mediate integration and excision, respectively, of the bacteriophage lambda genome from the *E. coli* chromosome. Both positive (antibiotic resistance) and negative (deletion of the lethal *ccdB* gene) selections are utilized to ensure that the correct recombination product is obtained with high efficiency.

Before large-scale cell growth and purification of the target protein, it is advisable that a small-scale pilot experiment be performed to gather some important information. First, is the solubility of the HisMBP fusion protein adequate? Second, can the fusion protein be cleaved efficiently by TEV protease? Third, does the target protein remain soluble after it is separated from the HisMBP moiety? If the answers to any of these three questions is negative, then there is little point in scaling up the procedure without first attempting some troubleshooting (see **Table 1**). The latter two questions are answered using separate plasmid vectors to express TEV protease along with the fusion protein in *E. coli*. The



**Figure 2** | Gateway cloning steps involved in the construction of a His<sub>6</sub>-maltose binding protein (HisMBP) fusion protein expression vector. In the first step (BP reaction), the PCR amplicon (**Fig. 1**) is recombined with the donor vector pDONR221 to give rise to an entry clone intermediate. The entry clone can be recovered from kanamycin-resistant transformants of DH5α cells, if desired (see Step 8). Alternatively, the second recombination reaction (LR reaction), which involves the transfer of the ORF from the entry clone into the destination vector (pDEST-HisMBP), can be performed immediately after the BP reaction has been completed, thereby bypassing the isolation of an entry clone intermediate, as described in this protocol (Steps 4–15).



production of the fusion protein substrate and the production of TEV protease are induced by isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) and anhydrotetracycline, respectively. We have observed

that delaying the induction of TEV protease for 2 h after induction of the fusion protein often improves the solubility of the target protein after TEV digestion *in vivo*<sup>13</sup>.

## MATERIALS

### REAGENTS

- The Gateway destination vector pDEST-HisMBP<sup>8</sup> can be obtained from the non-profit distributor of biological reagents (AddGene, Inc., Cambridge, MA, USA) ▲ **CRITICAL** This product should be purchased from this company only.
- PCR cloning system with Gateway technology with pDONR-221 and BP clonase II enzyme cloning kit (Invitrogen, Valencia, CA, cat. no. 12535-029) ▲ **CRITICAL** This product should be purchased from this company only.
- Gateway LR clonase II enzyme mix (Invitrogen, cat. no. 11791-020) ▲ **CRITICAL** This product should be purchased from this company only.
- Anhydrotetracycline (ACROS Organics, Geel, Belgium; Fisher Scientific, Springfield, NJ) (see REAGENT SETUP)
- D(+)-Glucose monohydrate (Fluka Sigma-Aldrich, St. Louis, MO, USA)
- Cell lysis buffer (see REAGENT SETUP)
- Reagents and thermostable DNA polymerase for PCR amplification
- PCR primers (see Fig. 1; REAGENT SETUP)
- Tris-acetate-EDTA (TAE) agarose
- Ethidium bromide
- QIAprep spin miniprep kit (Qiagen, Valencia, CA) for small-scale plasmid DNA isolation
- QIAquick gel extraction kit (Qiagen) for the extraction of DNA from agarose gels
- *E. coli* BL21-Pro cells containing pRK603 (AddGene, Inc.) ▲ **CRITICAL** This product should be purchased from this company only.
- TE buffer (see REAGENT SETUP)
- Competent DH5 $\alpha$  and BL21(DE3) CodonPlus-RIL *E. coli* cells (Stratagene)
- Chemically competent One Shot CcdB Survival Competent cells for propagating pDEST-HisMBP and pDONR221 (Invitrogen, cat. no. C7510-03) ▲ **CRITICAL** This product should be purchased from this company only.
- LB medium (see REAGENT SETUP)
- LB agar plates (see REAGENT SETUP)
- Ni-NTA superflow resin (Qiagen)
- 20% (wt/vol) D(+)-glucose in H<sub>2</sub>O (see REAGENT SETUP)
- His<sub>6</sub>-TEV protease<sup>14</sup> (e.g., Ac-TEV protease; Invitrogen, cat. no. 12575023) ▲ **CRITICAL** This product should be purchased from this company only.
- Sodium phosphate buffer (see REAGENT SETUP)
- 150 mM sodium chloride buffer
- Imidazole
- Complete EDTA-free protease inhibitor cocktail tablets (Roche Diagnostics)
- SDS-polyacrylamide gel electrophoresis (SDS-PAGE) gel, 2 $\times$  SDS-PAGE sample buffer and running buffer (Invitrogen)
- $\beta$ -Mercaptoethanol
- Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)

### EQUIPMENT

- 0.22- and 0.45- $\mu$ m polyethersulfone filters
- Agarose gel electrophoresis system
- Gel photography system
- SDS-PAGE system
- Temperature-controlled shaking incubator
- Electroporator and electroporation cuvettes
- Mechanical device to disrupt *E. coli* cells (e.g., a sonicator, French press or cell homogenizer)
- ÄKTA explorer chromatography system (or equivalent)
- Amicon stirred ultrafiltration cell concentrator and ultrafiltration membranes
- Apparatus for submarine gel electrophoresis of DNA

### REAGENT SETUP

**TE buffer** 10 mM Tris-HCl (pH 8.0), 1 mM EDTA.

**Ampicillin** Prepare a stock solution of 100 mg ml<sup>-1</sup> in H<sub>2</sub>O and filter-sterilize. Store at 4 °C.

**Kanamycin** Prepare a stock solution of 30 mg ml<sup>-1</sup> in H<sub>2</sub>O and filter-sterilize. Store at 4 °C.

**IPTG** Prepare a 200 mM stock solution in H<sub>2</sub>O and filter-sterilize. Store at 4 °C.

**Anhydrotetracycline** Prepare a 1,000 $\times$  stock solution by dissolving in 50% ethanol at 100  $\mu$ g ml<sup>-1</sup>. Store in a foil-covered tube at -20 °C.

**Cell lysis buffer** 20 mM Tris-HCl (pH 8.0), 1 mM EDTA.

**20% (wt/vol) D(+)-glucose in H<sub>2</sub>O** Filter-sterilize and store at 4 °C.

**LB medium** Add 10 g bacto tryptone, 5 g yeast extract and 5 g sodium chloride to 1 l H<sub>2</sub>O and sterilize by autoclaving. If required, antibiotics can be added to a final concentration of 30  $\mu$ g ml<sup>-1</sup> for kanamycin, 30  $\mu$ g ml<sup>-1</sup> for chloramphenicol or 100  $\mu$ g ml<sup>-1</sup> for ampicillin.

**LB agar plates** Add 12 g bacto agar to 1 l LB medium before autoclaving. To prepare plates, allow medium to cool until flask or bottle can be held in hands without burning, then add 1 ml appropriate antibiotic stock(s), mix by gentle swirling and pour or pipette approximately 30 ml into each sterile Petri dish (100 mm diameter). The final concentration of kanamycin should be 30  $\mu$ g ml<sup>-1</sup>, and the final concentration of ampicillin should be 100  $\mu$ g ml<sup>-1</sup>.

**PCR primer design** The sequence of primer N2 is constant, as indicated in Figure 1. The 5' ends of primers N1 and C, encompassing the TEV protease recognition site and attB2 recombination site, respectively, are also constant. However, the 3' ends of the latter two primers will vary according to the ORF to be amplified. In general, the variable regions should be 20–25 nt in length. It is advisable to avoid dyad symmetry at the 3' ends.

**PCR optimization** If the PCR does not work when all three primers are added to the reaction at once, as described above, then it can be carried out in two steps instead, using primers N1 and C in the first step and primers N2 and C in the second step. The PCR amplicon from the first step is used as the template for the second PCR. Note, the PCR can be modified in numerous ways to optimize results, depending on the nature of the template and primers. See ref. 15 (vol. 2, chapter 8) for more information.

**Preparation of competent cells** Electrocompetent cells exhibit the highest transformation efficiency. However, transformation by electroporation requires expensive, specialized apparatus (the electroporator) that may not be available in every laboratory. Detailed protocols for the preparation of electrocompetent cells and electrotransformation can be obtained from the manufacturer (e.g., Bio-Rad, BTX, Eppendorf). In short, grow the cells in 1 l LB medium (with appropriate antibiotics) to mid-log phase (OD<sub>600 nm</sub> approximately 0.5) and then chill on ice. Pellet the cells at 4 °C, re-suspend in 1 l ice-cold water and pellet again. Repeat the wash with water (two to three times), re-suspend the cells in 3–4 ml 10% glycerol, divide into 50- $\mu$ l aliquots and then freeze immediately in a dry-ice/ethanol bath. Store the electrocompetent cells at -80 °C. Before electrotransformation, thaw the cells on ice and mix with 10–100 ng DNA (e.g., a plasmid vector or a Gateway reaction). Place the mixture in an ice-cold electroporation cuvette and electroporate according to manufacturer's recommendations (e.g., a 1.5-kV pulse in a cuvette with a 1-cm gap). Add 1 ml LB medium to the cells and shake (250 r.p.m.) at 37 °C for 1 h. Spread 5–200  $\mu$ l cells on an LB agar plate containing appropriate antibiotic(s). For the routine Gateway cloning procedures described in this protocol, high-efficiency (i.e., more than 1  $\times$  10<sup>7</sup> colony forming units per  $\mu$ g<sup>-1</sup>) chemically competent DH5 $\alpha$  cells (Invitrogen) can be used instead of electrocompetent cells, according to the manufacturer's instructions. Note that special cells with a CcdB-resistant *gyrA* allele (e.g., DB3 and DB5) or cells that have otherwise been rendered resistant to the DNA gyrase poison CcdB (e.g., One Shot CcdB Survival cells) must be used to propagate pDONR221 and pDEST-HisMBP. It will be necessary to prepare competent BL21-Pro cells containing pRK603 for the pilot experiment. Electrocompetent cells are recommended for this purpose owing to the inherently low transformation efficiency of *E. coli* BL21 cells.

## PROTOCOL

### PROCEDURE

#### Amplification of required ORF by PCR ● TIMING 1.5 h

1| Set up a 100- $\mu$ l PCR to amplify the required ORF as follows:

Component	Amount per reaction ( $\mu$ l)
Template (approximately 10 ng $\mu$ l <sup>-1</sup> )	1.0
Primer N1 (13 ng $\mu$ l <sup>-1</sup> )	2.5
Primer N2 (260 ng $\mu$ l <sup>-1</sup> )	2.5
Primer C (260 ng $\mu$ l <sup>-1</sup> )	2.5
dNTPs (1.25 mM each)	16.0
Thermostable DNA polymerase (2.5 U)	1.0
10 $\times$ reaction buffer	10.0
dH <sub>2</sub> O	64.5

2| Run the PCR using the program shown in **Table 2**.

▲ **CRITICAL STEP** The polymerization time at 72 °C and the annealing temperature may need to be adjusted, depending on the target and primer sequences, but the conditions given here should be applicable in most cases. See REAGENT SETUP for further optimization details.

3| Examine the product(s) of the PCR by agarose gel electrophoresis (1% TAE agarose) and ethidium bromide staining, and purify the PCR product to remove residual attB primers using a QIAquick gel extraction kit, according to the manufacturer's instructions.

■ **PAUSE POINT** The purified PCR product can be kept frozen at -20 °C for at least 1 month if desired.

#### ? TROUBLESHOOTING

#### Creating the ORF-entry vector by recombinational cloning ● TIMING 4–16 h

4| Set up the BP reaction (10  $\mu$ l) as follows:

Component	Amount per reaction
PCR product	300 ng
pDONR221 DNA	300 ng
TE buffer	To make 10 $\mu$ l total volume
BP Clonase II	2 $\mu$ l

5| Incubate the reaction on the bench top for at least 4 h.

■ **PAUSE POINT** The BP reaction can be incubated overnight (up to 16 h) on the bench top, if desired.

6| Remove a 5- $\mu$ l aliquot from the reaction mix and add it to 0.5  $\mu$ l proteinase K solution, included in the Gateway cloning kit. Incubate for 10 min at 37 °C.

7| Transform 2  $\mu$ l of the reaction mix from Step 6 into 50  $\mu$ l competent DH5 $\alpha$  cells using standard laboratory techniques<sup>15</sup>.

8| Pellet the cells by centrifugation (3,000*g* for 15 min), gently re-suspend the pellet in 100–200  $\mu$ l LB medium and spread on an LB agar plate containing 100  $\mu$ g ml<sup>-1</sup> kanamycin, the selective marker for pDONR221. Incubate the plate at 37 °C overnight. Entry clones can be recovered from these colonies in the event that no transformants are obtained after the subsequent LR reaction.

#### Creation of the HisMBP-ORF vector by recombinational cloning ● TIMING 3 d

9| To perform the LR reaction, add the following to the remainder of the BP reaction from Step 5:

Component	Amount per reaction ( $\mu$ l)
0.75 M NaCl	1
pDEST-HisMBP DNA (150 ng $\mu$ l <sup>-1</sup> )	3
LR Clonase II	2

▲ **CRITICAL STEP** Be sure to use the BP reaction that has not been treated with proteinase K.

10| Mix the contents by vortexing briefly and then incubate the reaction at room temperature for 3–4 h.

11| Add 1  $\mu$ l proteinase K solution to stop the reaction, and incubate for 10 min at 37 °C.

12| Transform 2  $\mu$ l of the reaction mix into 50  $\mu$ l competent DH5 $\alpha$  cells using standard techniques<sup>15</sup>.

**TABLE 2** | PCR program for amplification of required open reading frame

Cycle number(s)	Denaturation	Annealing	Elongation
1	5 min at 94 °C	30 s at 55 °C	1 min at 72 °C
2–29	30 s at 94 °C	30 s at 55 °C	1 min at 72 °C
30	30 s at 94 °C	30 s at 55 °C	7 min at 72 °C
31			Hold at 4 °C

**13|** Pellet the cells by centrifugation (3,000g for 15 min), gently re-suspend the pellet in 100–200 µl LB medium and spread on an LB agar plate containing 100 µg ml<sup>-1</sup> ampicillin, the selective marker for pDEST-HisMBP. Incubate the plate at 37 °C overnight.

**? TROUBLESHOOTING**

**14|** Select several ampicillin-resistant colonies and use them to inoculate 5 ml LB medium containing 100 µg ml<sup>-1</sup> ampicillin. Shake the cultures overnight at 250 r.p.m. and 37 °C.

**15|** Purify the plasmid DNA from each saturated overnight culture using a QIAquick spin miniprep kit, according to the manufacturer's instructions. Perform sequence analysis to confirm that the nucleotide sequence of the putative clones is correct.

■ **PAUSE POINT** Plasmid DNA can be stored at –20 °C for several months.

**Expression and purification of HisMBP fusion proteins**

**16|** To carry out a pilot expression experiment and TEV protease digest, proceed according to option A. The purpose of the pilot experiment is to assess the solubility of the HisMBP fusion protein in *E. coli*, to gauge how efficiently it is cleaved by TEV protease and to ascertain whether the target protein is likely to remain soluble after the HisMBP tag is removed by TEV protease. This information can be very valuable in deciding whether it is worthwhile to scale up the process. To perform large-scale cell growth and protein purification, proceed as described in option B.

**(A) Pilot expression experiment ● TIMING 3–4 d**

- (i) Transform 50 µl competent BL21-Pro cells that already contain the TEV protease expression vector pRK603 with 10–100 ng HisMBP fusion protein expression vector created in Step 15 and spread 5–200 µl on an agar plate containing 100 µg ml<sup>-1</sup> ampicillin and 30 µg ml<sup>-1</sup> kanamycin, as outlined in REAGENT SETUP. Incubate the plate overnight at 37 °C.
  - **PAUSE POINT** Plates can be stored at 4 °C for up to 1 month.
- (ii) Inoculate 2–5 ml LB medium containing 100 µg ml<sup>-1</sup> ampicillin and 30 µg ml<sup>-1</sup> kanamycin in a culture tube or shake flask with a single colony from the plate. Grow to saturation overnight at 37 °C with shaking at 250 r.p.m.
- (iii) The next morning, inoculate 50 ml of the same medium in a 250 ml baffle-bottomed flask with 0.5 ml saturated overnight culture.
- (iv) Grow the cells at 37 °C with shaking at 250 r.p.m. to mid-log phase (OD<sub>600 nm</sub> approximately 0.5).
- (v) Add IPTG (1 mM final concentration) and adjust the temperature to 30 °C.
  - ▲ **CRITICAL STEP** 30 °C is the optimum temperature for TEV protease activity. At 37 °C, the protease does not fold properly in *E. coli* and little processing will occur. Reducing the temperature also improves the solubility of some MBP fusion proteins.
- (vi) After 2 h, divide the culture into two separate flasks (approximately 20 ml in each). Label one flask '+' and the other '-'.
- (vii) Add anhydrotetracycline to the '+' flask (100 ng ml<sup>-1</sup> final concentration). Add nothing to the '-' flask. Grow for a further 2 h at 30 °C.
- (viii) Measure the OD<sub>600 nm</sub> of the cultures (dilute cells 1:10 in LB medium to obtain an accurate reading). An OD<sub>600 nm</sub> of approximately 3.5 is normal, although lower densities are possible. If the density of either culture is much lower than this, it may be necessary to adjust the volume of the samples that are analyzed by SDS-PAGE at Step (xvi).
- (ix) Transfer 10 ml of each culture to a 15-ml conical centrifuge tube and pellet the cells by centrifuging for 15 min at 4,000g and 4 °C.
- (x) Re-suspend the cell pellets in 1 ml lysis buffer and then transfer the suspensions to a 1.5 ml microcentrifuge tube.
  - **PAUSE POINT** Freeze and store the cell suspensions at –80 °C until further use (up to 1 month). When required, they can be thawed at room temperature and then placed on ice.
- (xi) Lyse the cells in a 1.5-ml microcentrifuge tube on ice with two or three 30-s pulses using a VCX600 sonicator with a microtip at 38% power. Cool the cells on ice for 1 min between pulses.
- (xii) Prepare samples of the total intracellular protein from the '+' and '-' cultures (T+ and T–, respectively) for SDS-PAGE by mixing 50 µl disrupted cell suspensions with an equal volume of 2× SDS-PAGE sample buffer containing 10% (vol/vol) β-mercaptoethanol.
- (xiii) Pellet the insoluble cell debris (and proteins) by centrifuging the sonicated cell suspension from each culture at maximum speed in a microcentrifuge for 10 min at 4 °C.



## PROTOCOL

- (xiv) Prepare samples of the soluble intracellular protein from the '+' and '-' cultures (S+ and S-, respectively) for SDS-PAGE by mixing 50  $\mu$ l of each supernatant from Step (xiii) with 50  $\mu$ l 2 $\times$  SDS-PAGE sample buffer containing 10% (vol/vol)  $\beta$ -mercaptoethanol.
- (xv) Heat the T-, T+, S- and S+ protein samples at 90 °C for approximately 5 min and then spin them at maximum speed in a microcentrifuge for 5 min.
- (xvi) Assemble the SDS gel in the electrophoresis apparatus, fill with SDS-PAGE running buffer, load the samples (10  $\mu$ l each) and carry out the electrophoretic separation according to standard lab practices<sup>14</sup>. Load T and S samples from each culture ('+' and '-') in adjacent lanes to allow easy assessment of solubility. Molecular weight standards can also be loaded on the gel if desired.
- (xvii) Stain the proteins in the gel with GelCode Blue reagent, PhastGel Blue R or a suitable alternative.

### ? TROUBLESHOOTING

#### (B) Large-scale cell growth and protein purification ● TIMING Cell growth, 2 d; protein purification, 3 d

- (i) Transform 50  $\mu$ l competent BL21(DE3) Codon Plus-RIL cells with 10–100 ng HisMBP fusion protein expression vector created in Step 15 and spread 5–200  $\mu$ l on an agar plate containing 100  $\mu$ g ml<sup>-1</sup> ampicillin and 30  $\mu$ g ml<sup>-1</sup> chloramphenicol, as outlined in REAGENT SETUP. Incubate the plate overnight at 37 °C.
- (ii) Inoculate 100 ml LB medium in a 500-ml baffled-bottomed shake flask with a single-drug-resistant colony from the transformation in Step 16B(i). Shake overnight at 250 r.p.m. and 37 °C.
- (iii) Add 25 ml of the saturated overnight culture to 1 l fresh LB medium containing 100  $\mu$ g ml<sup>-1</sup> ampicillin and 30  $\mu$ g ml<sup>-1</sup> chloramphenicol in a 4-l baffled-bottomed shake flask.
  - ▲ **CRITICAL STEP** To ensure an adequate yield of pure protein at the end of the process, grow at least four sets of 1-l cultures of cells at a time. Add sterile glucose to 0.2% to increase biomass production.
- (iv) Shake the flasks at 250 r.p.m. and 37 °C until the cells reach mid-log phase ( $OD_{600\text{ nm}}$  approximately 0.5), which should take approximately 90–120 min.
- (v) Shift the temperature to 30 °C and then add IPTG to a final concentration of 1 mM. Continue shaking for 4–6 h.
- (vi) Recover the cells by centrifuging for 15 min at 5,000g and 4 °C.
  - ▲ **CRITICAL STEP** Perform all of the following procedures at 4 °C.
  - **PAUSE POINT** Freeze the cell pellet at -80 °C until further use (this may be for as long as several months).
- (vii) Re-suspend the cell pellet in 50 mM sodium phosphate buffer, (pH 7.7 at room temperature), 150 mM sodium chloride and 25 mM imidazole containing complete EDTA-free protease inhibitor cocktail tablets (according to the manufacturer's instructions), using at least 10 ml g<sup>-1</sup> wet cell paste.
- (viii) Lyse the cell suspension using the APV-1000 homogenizer at 10,000–11,000 psi for two to three rounds and centrifuge the disrupted cell suspension for at least 30 min at 15,000g. Filtering through 0.45- $\mu$ m polyethersulfone membrane helps to remove residual particulates before chromatography.
- (ix) Apply the supernatant to a column of Ni-NTA resin equilibrated in 50 mM sodium phosphate buffer (pH 7.7), 150 mM sodium chloride and 25 mM imidazole. Wash the column with this buffer until a stable baseline is reached and then elute the bound fusion protein (HisMBP-passenger) with a linear gradient over 10 column volumes into 50 mM sodium phosphate buffer (pH 7.7), 150 mM sodium chloride and 250 mM imidazole. The fusion protein usually elutes between 100 and 150 mM imidazole. Identify fractions containing the fusion protein by SDS-PAGE<sup>15</sup>.
- (x) Pool the peak fractions containing fusion protein and concentrate the sample approximately tenfold using an Amicon stirred ultrafiltration cell fitted with a YM30 membrane according to the manufacturer's instructions. Remove any precipitate by centrifuging at 5,000g for 10 min.
- (xi) Add His<sub>6</sub>-TEV protease (1  $OD_{280}$  TEV protease per 100  $OD_{280}$  fusion protein) to the supernatant from Step 16B(x). Digest overnight at 4 °C. The progress of the cleavage reaction can be monitored by SDS-PAGE if desired<sup>15</sup>.
  - ? **TROUBLESHOOTING**
- (xii) Dilute the digest with 50 mM phosphate buffer (pH 7.7) and 150 mM sodium chloride buffer to reduce the imidazole concentration to 25 mM. Apply the TEV protease digestion products from Step 16B(xi) to a second Ni-NTA column under the same conditions used during application to the first column. This step serves to absorb the TEV protease, the HisMBP moiety and any residual uncleaved fusion protein to the column.
- (xiii) Collect the unbound (flow-through) fractions that pass through the Ni-NTA column, which should contain the target protein free of the HisMBP tag, and check the purity by SDS-PAGE<sup>15</sup>.
  - ? **TROUBLESHOOTING**
- (xiv) Concentrate the protein sample from Step 16B(xiii) approximately tenfold using the Amicon stirred cell fitted with a YM10 membrane according to the manufacturer's instructions and add the reducing agent TCEP to a final concentration of 2 mM. The TCEP will prevent the formation of intermolecular disulfide bonds.
- (xv) Apply 2 ml concentrated sample onto a HiPrep 26/60 Sephacryl S-100 HR column equilibrated with gel filtration buffer.
  - ▲ **CRITICAL STEP** The volume of the sample loaded should be no more than 4% of the column volume and contain no more than 50 mg protein.

- (xvi) Pool the peak fractions of pure protein from the gel filtration column(s) and concentrate to 1–5 mg ml<sup>-1</sup> using the Amicon stirred cell fitted with a YM10 membrane according to the manufacturer's instructions.
  - (xvii) Filter through a 0.2- $\mu$ m syringe filter, prepare 0.5–1-ml aliquots and flash-freeze with liquid nitrogen.
- **PAUSE POINT** Store at –80 °C indefinitely.

● **TIMING**

- Steps 1–3, amplification of ORF by PCR: 1.5 h
- Steps 4–8, creating ORF-entry vector by recombinational cloning: 4–16 h
- Steps 9–15, cloning HisMBP-ORF vector by recombinational cloning: 3 d
- Step 16A, pilot expression of HisMBP fusion protein: 3–4 d
- Step 16B, large-scale expression and purification of HisMBP fusion protein: 5 d

? **TROUBLESHOOTING**

Troubleshooting advice can be found in **Table 1**.

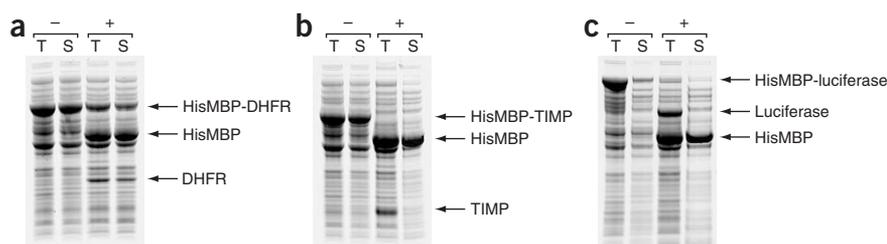
**TABLE 1** | Troubleshooting table.

Step	Problem	Possible solution(s)
3	No PCR product or very poor yield of PCR product	<ol style="list-style-type: none"> <li>1. Try a different thermostable DNA polymerase</li> <li>2. Vary the dimethyl sulfoxide (DMSO) and/or magnesium concentration</li> <li>3. Change the 3' ends of the PCR primers</li> </ol>
13	No ampicillin-resistant transformants are obtained	<ol style="list-style-type: none"> <li>1. Increase the amount of DNA in the Gateway reactions</li> <li>2. Extend the incubation times</li> <li>3. Perform the BP and LR reactions in separate steps</li> </ol>
16A(xvii)	The His <sub>6</sub> -maltose binding protein (HisMBP) fusion protein is poorly soluble	<ol style="list-style-type: none"> <li>1. Reduce the temperature to 20–25 °C after isopropyl-<math>\beta</math>-D-thiogalactopyranoside (IPTG) induction</li> <li>2. Reduce the IPTG concentration to approximately 20 <math>\mu</math>M and induce overnight</li> </ol>
16B(xi)	The fusion protein is not cleaved efficiently by tobacco etch virus (TEV) protease	<ol style="list-style-type: none"> <li>1. Increase ratio of protease to substrate</li> <li>2. Increase duration of digest</li> <li>3. Perform digest at room temperature</li> <li>4. Modify the expression vector by adding 3–4 'spacer' residues (e.g., Gly or Ala) between the TEV site and N-terminus of the target protein</li> </ol>
16B(xiii)	The target protein precipitates after it is cleaved from HisMBP	<ol style="list-style-type: none"> <li>1. Reduce the temperature to 20–25 °C after IPTG induction</li> <li>2. Explore alternative buffers and/or stabilizing additives such as glycerol or sorbitol</li> </ol>

**ANTICIPATED RESULTS**

The pilot expression experiment and TEV protease digest may yield a range of different results, depending on the nature of the passenger protein. The most common results are illustrated in **Figure 3**, using three different passenger proteins as examples. All three of these proteins—dihydrofolate reductase (DHFR), tissue inhibitor of metalloproteinases 2 (TIMP) and luciferase—

**Figure 3** | Intracellular processing of His<sub>6</sub>-maltose binding protein (HisMBP) fusion proteins by tobacco etch virus (TEV) protease. Dihydrofolate reductase (DHFR), tissue inhibitor of metalloproteinases 2 (TIMP) and luciferase (panels **a**, **b** and **c**, respectively) were expressed from derivatives of pDEST-HisMBP in BL21-Pro cells that also contained the TEV protease expression vector pRK603 as described (pilot expression experiment). 'T' and 'S' refer to the total and soluble fractions of the intracellular protein, respectively. All cultures were induced with isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) to initiate the production of the HisMBP fusion proteins. Samples marked '+' were induced with anhydrotetracycline to initiate the production of TEV protease 2 h after adding IPTG, whereas samples marked '-' were not induced with anhydrotetracycline. The results of this experiment are discussed in the text.



have previously been shown to be very insoluble when expressed in an unfused state and/or as glutathione-S-transferase fusion proteins in *E. coli*<sup>6,16</sup>.

The most desirable outcome is exemplified by the HisMBP-DHFR fusion protein. In this case, the DHFR is efficiently solubilized by fusing it to HisMBP. Moreover, a significant fraction of the DHFR remains soluble in the extract after intracellular processing of

the fusion protein by TEV protease, which is consistent with the existence of properly folded protein, and a similar result can be expected during scale-up and purification. Note, however, that unlike the other examples shown here, the HisMBP-DHFR fusion protein is not digested to completion. This indicates that more than the usual amount of enzyme and/or a longer digest may be required to cleave this substrate to completion *in vitro*.

The HisMBP-TIMP fusion protein is also mostly soluble, but none of the cleaved TIMP remains soluble after intracellular processing. This may indicate that the TIMP is improperly or incompletely folded. Alternatively, proteins that have a tendency to aggregate in their native states may also exhibit this type of behavior.

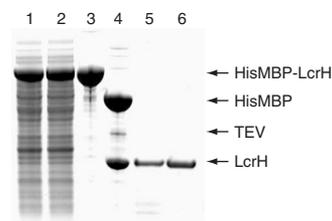
In the case of HisMBP-luciferase, only a very small fraction of the fusion protein is soluble. Not surprisingly, therefore, virtually none of the luciferase is soluble after intracellular processing. This example is included to demonstrate that not all passenger proteins can be rendered soluble by fusing them to HisMBP.

LcrH, a virulence factor encoded by the plague-causing bacterium *Yersinia pestis*, is used here as an example to illustrate the generic purification process based on two successive IMAC steps (Fig. 4). Before this, an intracellular processing experiment was conducted on the HisMBP-LcrH fusion protein that yielded results similar to those obtained with the HisMBP-DHFR fusion protein (data not shown). After the first IMAC column, the purity of the fusion protein is typically in the range of 70–80%. As any endogenous proteins that bind non-specifically to the resin during the first IMAC step also do so during the second round of IMAC, purity improves significantly at this stage.

We recommend using a gel filtration column as a final polishing step to exchange the protein into an appropriate buffer for the intended application and to detect and remove any high-molecular-weight aggregates that may be present. Occasionally some contaminants will persist after the second IMAC step. These may be chaperones (e.g., GroEL) or endogenous proteins with natural affinity for the target protein. Chaperones can sometimes be released from the target protein by incubating it with ATP and magnesium, but because they bind selectively to proteins that are incompletely or improperly folded, their presence as contaminants is never a good sign. In addition, in some cases, the HisMBP moiety may stick (bind non-covalently) to the target protein after digestion of the fusion protein by TEV protease. It is unclear exactly why this happens, but it may be related to the mechanism of solubility enhancement. When contaminants persist after the second IMAC step, additional forms of chromatography (e.g., ion exchange<sup>17</sup>) may be required to produce a homogeneous sample.

Not all proteins can be rendered soluble by fusing them to HisMBP, and of those that can, some will precipitate after digestion of the fusion protein by TEV protease. In addition, some fusion proteins are resistant to digestion by TEV protease. Nevertheless, the available evidence indicates that many aggregation-prone proteins can be recovered in a properly folded, biologically active form in this manner. For instance, in one recent study involving 632 HisMBP fusion proteins<sup>18</sup>, 80% were observed to be soluble. Approximately 70% of the soluble fusion proteins could be cleaved efficiently by TEV protease. Fifteen of the target proteins precipitated upon removal of the HisMBP moiety by TEV protease. The overall success rate was 42%.

**Figure 4** | Generic purification of *Yersinia pestis* LcrH, performed as described in Step 16B (protein purification). Lane 1, total intracellular protein; Lane 2, soluble cell extract, Step 16B(viii); Lane 3, peak fractions from the first round of immobilized metal affinity chromatography (IMAC), eluted with an imidazole gradient, Step 16B(ix); Lane 4, tobacco etch virus protease digest, Step 16B(xi); Lane 5, flow-through fractions after second round of IMAC, Step 16B(xiv); Lane 6, Pooled peak fractions after gel filtration and concentration, Step 16B(xvi). The results of this experiment are discussed in the text.



**ACKNOWLEDGMENTS** This research was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research.

**COMPETING INTERESTS STATEMENT** The authors declare that they have no competing financial interests.

Published online at <http://www.natureprotocols.com>  
Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

1. Waugh, D.S. Making the most of affinity tags. *Trends Biotechnol.* **23**, 316–320 (2005).
2. Esposito, D. & Chatterjee, D.K. Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **17**, 353–358 (2006).

3. Lichty, J.J., Malecki, J.L., Agnew, H.D., Michelson-Horowitz, D.J. & Tan, S. Comparison of affinity tags for protein purification. *Protein Expr. Purif.* **41**, 98–105 (2005).
4. Christendat, D. *et al.* Structural proteomics: prospects for high throughput sample preparation. *Progr. Biophys. Mol. Biol.* **73**, 339–345 (2000).
5. Davis, G.D., Elisei, C., Newham, D.M. & Harrison, R.G. New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnol. Bioeng.* **65**, 382–388 (1999).
6. Kapust, R.B. & Waugh, D.S. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668–1674 (1999).
7. Nallamsetty, S. & Waugh, D.S. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr. Purif.* **45**, 175–182 (2006).
8. Riggs, P. Expression and purification of recombinant proteins by fusion to maltose-binding protein. *Mol. Biotechnol.* **15**, 51–63 (2000).

9. Pryor, K.D. & Leiting, B. High-level expression of soluble protein in *Escherichia coli* using a His<sub>6</sub>-tag and maltose-binding protein double-affinity fusion system. *Protein Expr. Purif.* **10**, 309–319 (1997).
10. Routzahn, K.M. & Waugh, D.S. Differential effects of supplementary affinity tags on the solubility of MBP fusion proteins. *J. Struct. Funct. Genomics* **2**, 83–92 (2002).
11. Nallamsetty, S., Austin, B.P., Penrose, K.J. & Waugh, D.S. Gateway vectors for the production of combinatorially-tagged His<sub>6</sub>-MBP fusion proteins in the cytoplasm and periplasm of *Escherichia coli*. *Protein Sci.* **14**, 2964–2971 (2005).
12. Cabrita, L.D., Dai, W. & Bottomley, S.P. A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol.* **6**, 12 (2006).
13. Kapust, R.B. & Waugh, D.S. Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr. Purif.* **19**, 312–318 (2000).
14. Tropea, J.E., Cherry, S. & Waugh, D.S. Expression and purification of soluble His<sub>6</sub>-tagged TEV protease. *Methods. Mol. Biol.* (in press).
15. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
16. Fox, J.D., Routzahn, K.M., Bucher, M.H. & Waugh, D.S. Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. *FEBS Lett.* **537**, 53–57 (2003).
17. Scopes, R.K. *Protein Purification* 3rd edn. (Springer-Verlag, New York, 1994).
18. Jeon, W.B. *et al.* High-throughput purification and quality assurance of *Arabidopsis thaliana* proteins for eukaryotic structural genomics. *J. Struct. Funct. Genomics* **6**, 143–147 (2005).