



Fusing an insoluble protein to GroEL apical domain enhances soluble expression in *Escherichia coli*

Prasad T. Reddy^{a,b,*} and William B. O'Dell^{a,b} 

^aBiomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, MD, United States

^bBiomolecular Labeling Laboratory, Institute for Bioscience and Biotechnology Research, Rockville, MD, United States

*Corresponding author: e-mail address: tirumuruprasad@yahoo.com

Contents

1. Introduction	172
2. Before you begin	173
2.1 Considerations for using the pRE vectors for protein expression	173
2.2 Protein(s) of interest	173
2.3 Considerations for source of coding DNA sequences (CDSs)	174
3. Key resources table	175
4. Materials and equipment	177
5. Step-by-step method details	178
5.1 Design synthetic genes for DNA synthesis	178
5.2 Construct the pRE:GroEL fusion vector	180
5.3 Clone GroEL(191–345) protein fusions	181
5.4 Test GroEL(191–345)–adenylyl cyclase fusion protein expression	182
5.5 Express and purify GroEL(191–345)–CAMLp fusion protein	183
6. Expected outcomes	184
7. Advantages	186
8. Limitations	186
Disclaimer	187
References	187

Abstract

A protocol for increasing soluble protein expression by fusing the chaperone GroEL apical domain with a gene of interest is described herein. GroEL apical domain, the minichaperone that functions independently of GroES and ATP in protein folding, is cloned downstream of the lambda CII ribosome binding site in the parent pRE vector. The pRE vector has tightly controlled transcription suitable for expressing toxic proteins. The GroEL minichaperone is fused to a glycine–serine rich linker followed by the

enterokinase protease recognition sequence. A number of genes that are recalcitrant to protein production in the parent pRE vector 5were cloned into the pRE:GroEL fusion vector and successfully expressed as fusion proteins in *Escherichia coli*.



1. Introduction

In many protein overproduction endeavors initiated with the intention of easy protein expression and purification for downstream biochemical and biophysical investigations, we often observe four possible end scenarios: (1) successful protein expression with natural protein folding, (2) expression of protein into insoluble aggregates called “inclusion bodies” (Jaenicke, 1998; Schein, 1989), (3) no accumulation of overexpressed protein due to proteolytic degradation, or (4) protein toxicity to the host cell prior to induction due to “leaky” transcription (Reddy, Peterkofsky, & Mckenney, 1989) or to the presence of small but non-trivial concentrations of an inducer or inducer-like compound in the growth medium (Studier, 2005). To address these problems associated with protein production, a number of plasmid vectors have been constructed to alleviate some of these problems. However, there is no “one” protein overproduction plasmid that solves all the above mentioned difficulties in protein production.

There are two major protein folding systems in *Escherichia coli*, namely the HSP 60 GroEL/GroES system (Ellis & Hartl, 1996) and the Hsp70 DnaK/DnaJ/GrpE system (Langer et al., 1992). These refolding systems prevent aggregation of their client proteins and yield folded proteins during normal cell growth. However, in plasmid-driven overexpression of some genes, the protein production is so rapid and abundant that chaperones from both systems, also in high abundance intracellularly, fail to facilitate proper folding of the overexpressed protein product. Strategies such as slowing down protein synthesis by decreasing the incubation temperature during induction, induction in minimal medium and reducing inducer concentrations can improve the yield of properly folded protein in some cases. However, these strategies are not universal for every protein of interest.

Functional GroEL is a tetradecameric protein composed of two stacked, seven-subunit rings that works in conjunction with the heptameric GroES and ATP cofactor. Interestingly, a 155 amino acid (AA) segment of *E. coli* GroEL, GroEL(191–345), also called the minichaperone, has been observed to function as a monomer in the folding of unfolded rhodanase independently of either GroES or ATP (Zahn et al., 1996). We considered this minichaperone as a potential target to fuse to any protein that is

recalcitrant to expression in *E. coli* as such fusion protein might be expected to be expressed in a properly folded form amenable to downstream manipulation. Here we provide the methods for constructing a minichaperone fusion *E. coli* expression vector (pRE:GroEL) and for expressing GroEL(191–345)–protein-of-interest fusions to test potential improvements to the frequently encountered problems in protein overexpression described above.



2. Before you begin

2.1 Considerations for using the pRE vectors for protein expression

The pRE expression vector (Fig. 1) drives protein expression from the bacteriophage λ P_L promoter (Reddy et al., 1989). Expression is controlled using a host expressing a temperature sensitive mutant of lambda repressor (λ cI) protein known as λ cI857 (Isaacs, Echols, & Sly, 1965; Lieb, 1966). Expression of λ cI857 repressor from either a chromosomal (i.e., a λ 857 lysogen) or a plasmid-borne cassette is sufficient for P_L regulation (Remaut, Stanssens, & Fiers, 1981). If a toxic protein is encoded by the pRE plasmid, as is demonstrated here, the *E. coli* strain used for subcloning and/or propagating the pRE plasmid should express wild-type λ cI to repress expression.

1. Expression strain—The method described here uses *E. coli* MZ1 as expression strain. Other λ cI857 expressing strains, such as K12 Δ H1 Δ trp or M5219 (accessions and LMBP 69 and LMBP 130 in the Belgian Co-ordinated Collections of Micro-Organisms (BCCM), respectively) may also be suitable. Expressing λ cI857 from a co-transformed plasmid would require that plasmid to be compatible with the pMB1 origin of replication on the pRE vectors.
2. Subcloning strain(s)—The method described here uses *E. coli* C600 λ lysogen strain for subcloning/propagation (Reddy et al., 1989). However, strains DH5 α F(λ) or MC1061(λ) (BCCM accessions LMBP 2846 and LMBP 1061, respectively) may be suitable alternatives

2.2 Protein(s) of interest

The method for expressing GroEL(191–345) fusion proteins is here demonstrated using the cytoplasmic domains of adenylyl cyclases from *Mycobacterium tuberculosis* (Mtb1625) and *Mycobacterium leprae* (Mlp0201 and Mlp1399) which are toxic to *E. coli* growth (Cole et al., 1998, 2001) as well as the calmodulin-like protein (CAMLP) from *Mycobacterium smegmatis*.

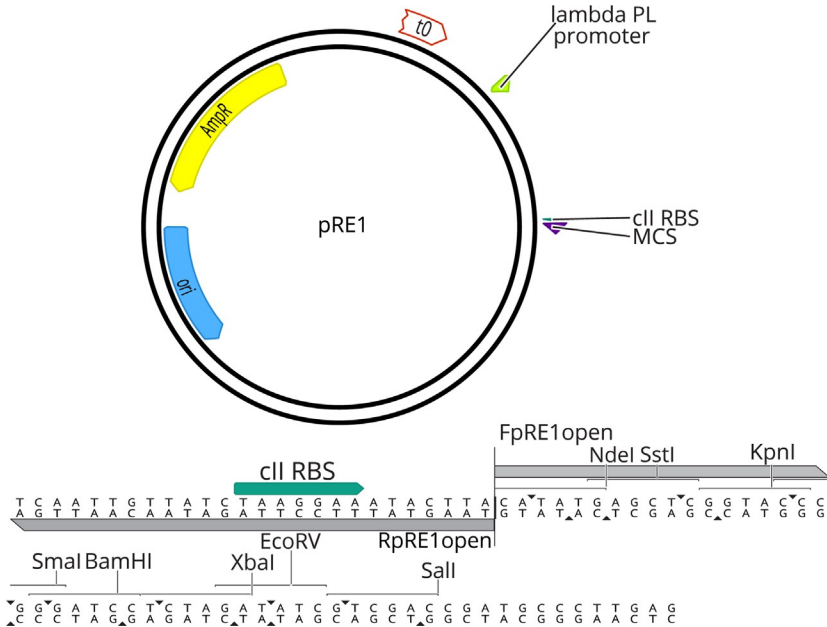


Fig. 1 pRE1 vector plasmid map and multiple cloning site sequence. The plasmid map schematically depicts the relative positions of important pRE1 plasmid elements including the *bla* ampicillin resistance gene (Amp^R, yellow) and the pMB1 origin (blue), both derived from pBR322, and the λ *t0* transcription terminator, λ P_L promoter, λ cII ribosome binding site (RBS) and multiple cloning site (MCS) specific to the pRE vector. PCR linearization of the pRE vector using primers FpRE1open and RpRE1open immediately 5' of the *NdeI* restriction site allows scarless homology cloning of the CDS for N-terminal GroEL(191–345) fusions to create the resulting pRE:GroEL vector.

2.3 Considerations for source of coding DNA sequences (CDSs)

The vector construction and cloning methods described below were originally performed using CDSs obtained by polymerase chain reaction (PCR) amplification using as template genomic DNA from the organism of interest (e.g., *E. coli*, *M. tuberculosis*, *M. leprae* or *M. smegmatis*). However, given the current speed and low cost of *de novo* gene synthesis, the methods below have been updated to demonstrate the techniques beginning with amino acid sequences of interest that are back translated and codon optimized prior to DNA synthesis to obtain the CDSs. It should be noted that changes in codon usage between native CDS and back translated and/or optimized CDSs can potentially impact the amount and/or rate of protein over-expression with downstream consequences including changes to protein folding, solubility, etc.



3. Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
<i>Escherichia coli</i> MZ1 (or alternative)	Coli Genetic Stock Center	7798
<i>E. coli</i> C600(λ) (or alternative)		Reddy et al. (1989)
Chemicals, Peptides, and Recombinant Proteins		
Enterokinase	New England Biolabs	P8070S (or similar)
NdeI	New England Biolabs	R0111S (or similar)
BamHI-HF [®]	New England Biolabs	R3136S (or similar)
T4 DNA ligase	New England Biolabs	M0202S (or similar)
Critical Commercial Assays		
NEBuilder [®] HiFi DNA Assembly Cloning Kit	New England Biolabs	E5520S
NEB Q5 [®] Hot Start High-Fidelity 2 × Master Mix	New England Biolabs	M0494S
Deposited Data		
<i>Escherichia coli</i> GroEL amino acid sequence	National Center for Biotechnology Information (NCBI)	NP_418567.1

Continued

—cont'd

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Mycobacterium tuberculosis</i> Rv1625c adenylyl cyclase amino acid sequence (Mtb1625)	NCBI	AAB70274.1
<i>Mycobacterium leprae</i> ML0201 adenylyl cyclase amino acid sequence (Mlp0201)	NCBI	CAA18817.1
<i>M. leprae</i> ML1399 adenylyl cyclase amino acid sequence	NCBI	CAC31780.1
<i>Mycobacterium smegmatis</i> calmodulin-like protein amino acid sequence	NCBI	AAP88233.1
Oligonucleotides		
5'-CATATGAGCTCGGTACCC-3'		FpRE1open
5'-TAAGTATTTTCCTTAGATAACAATTGA-3'		RpRE1open
Recombinant DNA		
pRE expression vector		Reddy et al. (1989)



4. Materials and equipment

- LB Broth (10 g/L tryptone, 5 g/L yeast extract and 10 g/L NaCl)
- LB-agar plates (15 g/L agar in LB Broth)
- 50 mg/mL ampicillin stock solution (sterile filtered)
- 1 × TAE buffer (40 mmol/L Tris-acetate, 1 mmol/L EDTA)
- 4 × SDS-PAGE loading buffer
- 10 mmol/L Tris-HCl, pH 7.5
- Lysis buffer (1 mmol/L CaCl₂, 1 mmol/L dithiothreitol (DTT), 0.1 mmol/L phenylmethylsulfonyl fluoride (PMSF) and 50 mmol/L Tris-HCl, pH 7.5)
- Buffer A (1 mmol/L CaCl₂, 50 mmol/L Tris-HCl, pH 7.5)
- Buffer B (1 mmol/L CaCl₂, 400 mmol/L NaCl, 50 mmol/L Tris-HCl, pH 7.5)
- Spin column-based PCR cleanup kit
- Spin column-based plasmid miniprep kit
- Spin column-based gel extraction kit
- 1% agarose-TAE gels containing nucleic acid stain (e.g., SYBR safe, etc.)
- Bacterial cell lysis reagent, if using (e.g., BugBuster[®] from MilliporeSigma, B-PER[™] from ThermoFisher, etc.)
- 12% Tris-glycine SDS-PAGE Gels
- 35 mL DEAE-cellulose column
- 500 mL Sephadex G-75 (superfine) column (Cytiva)
- 12.2 × 600 mm BIOSEP-SEC-S300 HPLC column (Phenomenex)
- 5 kDa and 1 kDa molecular weight cut-off (MWCO) centrifugal concentrators
- PCR thermal cycler
- 250 mL baffled culture flasks
- 2.8 L baffled Fernbach culture flasks
- Electroporator and cuvettes, if using
- Incubators for 32 °C, 37 °C and 65 °C
- Shaking incubators for 32 °C, 37 °C and 42 °C
- Heat block or water bath for 70 °C and 95 °C
- Small probe sonicator, if using
- French press with large volume pressure cell
- Microcentrifuge and benchtop centrifuge
- High-speed centrifuge and rotor (100,000 × g)
- High-performance liquid chromatography (HPLC) system



5. Step-by-step method details

5.1 Design synthetic genes for DNA synthesis

1. Design GroEL(191–345)–linker–enterokinase site fusion CDS as a double strand DNA (dsDNA) fragment for homology cloning into pRE1 vector.

- a) Residues 191–345 of the *E. coli* GroEL amino acid sequence (NP_418567.1, underlined below) should be appended with an initiation/methionine codon at the N-terminus and a glycine–serine linker sequence followed by the enterokinase protease recognition site (DDDDK) at the C-terminus resulting in a target amino acid sequence such as the following:

<u>MEGMQFDRGY</u>	<u>LSPYFINKPE</u>	<u>TGAVELESPF</u>	<u>ILLADKKISN</u>	<u>IREMLPVLEA</u>	50
<u>VAKAGKPLLI</u>	<u>IAEDVEGEAL</u>	<u>ATLVVNTMRG</u>	<u>IVKVAAVKAP</u>	<u>GFGDRRKAML</u>	100
<u>QDIATLTGGT</u>	<u>VISEEIGMEL</u>	<u>EKATLEDLGO</u>	<u>AKRVVINKDT</u>	<u>TTIIDGVGEE</u>	150
<u>AAIQGRGSGS</u>	<u>GSGSGSDDDD</u>	K			171

- b) The resulting amino acid sequence must then be back translated to the CDS and, at minimum, codon-optimized to prevent the introduction of *NdeI* and *BamHI* restriction sites. (These transformations of the amino acid and DNA sequences are routinely supported by molecular biology software and websites of DNA synthesis vendors.) Codon-optimization for *E. coli* codon usage and/or ease of synthesis may also be performed, if desired.
- c) Append ≥ 18 nt of homologous sequence from the pRE1 vector upstream of the PCR opening site (see Fig. 1) to the 5' end of the resulting DNA sequence. Also append ≥ 18 nt of homologous sequence from the pRE1 vector downstream of the PCR opening site to the 3' end of the resulting DNA sequence. An example of a possible resulting DNA sequence follows with the homology overhangs shown in bold and GroEL(191–345) again underlined:

TATCTAAGGA	AATACTTAAT	<u>GGAGGGCATG</u>	<u>CAGTTCGATC</u>	<u>GTGGTTATCT</u>	50
<u>CTCTCCGTAC</u>	<u>TTCATTAACA</u>	<u>AGCCAGAGAC</u>	<u>CGGCGCAGTT</u>	<u>GAGCTGGAAT</u>	100
<u>CCCCGTTTAT</u>	<u>CCTGCTCGCG</u>	<u>GACAAGAAAA</u>	<u>TCAGCAACAT</u>	<u>CCGCAGATG</u>	150
<u>CTGCCTGTCC</u>	<u>TGGAAGCGGT</u>	<u>AGCGAAGGCC</u>	<u>GGTAAACCGC</u>	<u>TCCTGATTAT</u>	200
<u>CGCAGAAGAC</u>	<u>TGCAAGGTG</u>	<u>AAGCCCTGGC</u>	<u>TACTCTGGTA</u>	<u>GTCAACACCA</u>	250
<u>TGCGTGGCAT</u>	<u>TGTAAAAGTT</u>	<u>GCAGCCGTGA</u>	<u>AAGCTCCGGG</u>	<u>CTTCGGCGAT</u>	300
<u>CGCCGTAAAG</u>	<u>CGATGCTGCA</u>	<u>GGACATCGCT</u>	<u>ACCCTGACCG</u>	<u>GCGGTACCGT</u>	350

<u>GATTTCCGAG</u>	<u>GAAATCGGCA</u>	<u>TGGAACTGGA</u>	<u>AAAAGCGACT</u>	<u>CTGGAAGACC</u>	400
<u>TGGGTCAGGC</u>	<u>CAAGCGTGTA</u>	<u>GTTATCAACA</u>	<u>AAGATACCAC</u>	<u>TACCATCATT</u>	450
<u>GACGGCGTAG</u>	<u>GCGAGGAAGC</u>	<u>TGCAATCCAG</u>	<u>GGTCGTGGTA</u>	<u>GCGGTTCTGG</u>	500
<u>CTCTGGTTCC</u>	<u>GGTTCCGACG</u>	<u>ATGACGACAA</u>	G <u>CATATGAGC</u>	T <u>CGGTACCC</u>	549

- d) Submit construct for DNA synthesis as a dsDNA fragment.

Note: If the vendor of choice provides dsDNA fragments with appended sequences at either end (e.g., synthesis adapters, etc.) also design and obtain PCR primers that will precisely amplify the desired sequence from the synthesized DNA fragment. The designed CDS may also be subcloned into a plasmid by the vendor, if desired. Again, precise PCR primers would be needed to amplify the desired sequence.

Note: An affinity tag, such as hexahistidine, could be introduced at the N-terminus of the GroEL(191–345) construct, after the start codon, at this point if desired for downstream use in affinity purification of the fusion product. Affinity purification efficiency would likely benefit from including an additional glycine–serine linker between the affinity tag and the first residue of GroEL(191–345).

2. Design Mtb1625, Mlp0201, Mlp 1399 and *Ms*CAMLp CDSs for restriction cloning using *Nde*I and BamHI sites into pRE1:GroEL vector.
 - a) Obtain amino acid sequences for the cytoplasmic domains of Mtb1625 (AAB70274.1 residues 201 to C-terminus), Mlp0201 (CAA18817.1 residues 262 to C-terminus) and Mlp1399 (CAC31780.1 residues 100 to C-terminus) and the full amino acid sequence of *Ms*CAMLp (AAP88233.1).
 - b) Back translate each desired amino acid sequence to the corresponding CDS, and perform codon optimization, at minimum, to remove any *Nde*I or BamHI restriction sites from the CDS. Codon-optimization for *E. coli* codon usage and/or ease of synthesis may also be performed, if desired.
 - c) To the designed CDSs append a 5' *Nde*I restriction site and a 3' BamHI restriction site. These sites can be used to subclone the CDS by the synthesis vendor and will be used for restriction enzyme cloning into pRE:GroEL.
 - d) Submit construct for DNA synthesis and subcloning into a propagation vector.

Note: Obtaining small CDSs (such as those described here) as dsDNA fragments may be within the capabilities of the chosen DNA synthesis vendor, if desired. However, if ordering sequences

for restriction cloning as dsDNA fragments, be certain to append ≈ 5 nt flanks before the 5' restriction site and after the 3' restriction site if the chosen restriction enzymes require additional bases for optimal activity. (In this example, both *NdeI* and *BamHI* benefit from ≥ 3 additional bases.)

5.2 Construct the pRE:GroEL fusion vector

4. Linearize pRE1 vector 5' of the multiple cloning site by PCR using primers FpRE1open and RpRE1 open (Fig. 1) and a high fidelity DNA polymerase. The following reaction is optimized for NEB Q5[®] Hot Start polymerase supplied as a 2 \times master mix in high-fidelity buffer.
 - a) Prepare the PCR reaction by combining 10 ng pRE1 vector, primers to final concentration of 0.5 μ mol/L each and nuclease-free water to one half of the intended reaction volume (e.g., to 25 μ L for a 50 μ L reaction). Add an equal volume of 2 \times polymerase master mix.
 - b) Perform the following PCR thermal cycle with steps 2 to 4 repeated for 30 cycles:
 - a) Initial denature: 98 $^{\circ}$ C, 30 s
 - b) Denature: 98 $^{\circ}$ C, 10 s
 - c) Anneal: 57 $^{\circ}$ C, 30 s
 - d) Extend: 72 $^{\circ}$ C, 150 s (≈ 30 s per kilobase of template)
 - e) Final extension: 72 $^{\circ}$ C, 120 s
5. Purify the linearized pRE1 vector using a spin column-based PCR cleanup kit according to the manufacturer's directions.
6. Perform the homology cloning reaction.
 - a) In PCR tube, combine to a final volume of 10 μ L ≈ 50 ng of linearized pRE1 vector (≈ 4.9 kb) with 3-fold molar excess of the GroEL(191–345)-linker-enterokinase site CDS DNA. For the sequence shown above, 16.5 ng insert would achieve the 3:1 insert: vector ratio.
 - b) Add 10 μ L of NEBuilder[®] HiFi DNA assembly master mix.
 - c) Mix the reaction, and incubate in a PCR thermal cycler at 50 $^{\circ}$ C for 15 min to 60 min. Then, chill the reaction on ice for transformation.
7. Perform a high-efficiency transformation into a subcloning strain of *E. coli* with 2 μ L homology cloning reaction. The NEBuilder[®] HiFi

DNA Assembly cloning kit includes High Efficiency NEB 5-alpha chemically competent cells, and the manufacturer's protocol should be followed.

8. Plate the recovered, transformed cells onto LB-agar plates containing 50 µg/mL ampicillin, and incubate overnight at 37 °C.

Critical: The homology cloning reaction must be transformed using high efficiency methods ($\geq 1 \times 10^9$ colony forming units (CFU) per µg control plasmid) such as “ultracompetent” chemically competent cells or electroporation. Transformation of “standard efficiency” ($\approx 1 \times 10^6$ CFU/µg control plasmid) chemically competent cells is unlikely to yield any transformed colonies.

9. Pick transformed colonies. Inoculate 50 mL LB containing 50 µg/mL ampicillin in 250 mL baffled flasks, and incubate overnight with shaking at 37 °C. Prepare pRE1:GroEL plasmid DNA by miniprep according to kit manufacturer's instructions.
10. Sanger sequencing of purified plasmid from 5' of the homology insertion site may be performed to verify that the correct sequence has been obtained.

5.3 Clone GroEL(191–345) protein fusions

11. Perform NdeI–BamHI “double digest” of pRE1:GroEL vector using DNA and enzyme concentrations, buffer and incubation time recommended by restriction enzyme manufacturer. Purify linearized vector using a spin column-based PCR cleanup kit according to the manufacturer's instructions.
12. Perform NdeI–BamHI double digests of plasmids containing CDSs for adenyl cyclase cytoplasmic domains and MsCAMLP.
13. Separate NdeI–BamHI CDS fragments from subcloning vector backbone by electrophoresis on a 1% (w/v) agarose–TAE gel containing nucleic acid stain. Excise bands containing CDS fragments, and purify DNA using a spin column-based gel cleanup kit according to the manufacturer's instructions.
14. Perform pRE:GroEL–CDS ligation reactions with T4 DNA ligase according to manufacturer's directions with insert and vector mixed in 3:1 mole ratio.

Note: The GroEL(191–345)-linker-enterokinase site CDS adds an additional 549 bp for a total vector length of ≈ 5.5 kb.

15. Transform ligation products bearing toxic genes, such as the adenylyl cyclase cytoplasmic domains, into an *E. coli* subcloning strain expressing functional λ cI repressor (e.g., C600(λ), DH5 α F(λ), MC1061(λ), etc.).
16. Plate transformation mixtures onto LB agar plates containing 50 μ g/mL ampicillin. Incubate overnight at 37 °C.
17. Pick transformed colonies. Inoculate 50 mL LB containing 50 μ g/mL ampicillin in 250 mL baffled flasks, and incubate overnight with shaking at 37 °C. Prepare pRE1:GroEL-CDS plasmid DNA by miniprep according to kit manufacturer's instructions.
18. Sanger sequencing of purified plasmids should be performed to verify that the correct sequence has been obtained.

Note: Colony PCR with insert-specific primers may be performed to verify successful cloning after step 14 using transformed colonies as template, if desired.

5.4 Test GroEL(191–345)–adenylyl cyclase fusion protein expression

19. Transform the pRE:GroEL–adenylyl cyclase plasmids into an *E. coli* strain expressing the temperature sensitive λ cI857 repressor such as MZ1 as described below:
 - a) Transformation may be accomplished by electroporation or by preparing chemically competent cells using a procedure such as that of Hanahan (1983). If using chemically competent cells, perform the heat shock at 32 °C.
 - b) Plate recovered transformed cells on LB agar plates containing 50 μ g/mL ampicillin, and incubate overnight at 32 °C.

Note: The reduced 32 °C heat shock and growth temperatures are used to maintain repression of toxic gene expression by the λ cI857 repressor.
20. Pick transformed colonies, and inoculate into 50 mL of LB containing 50 μ g/mL ampicillin in a 250 mL baffled flask. Incubate cultures at 32 °C with shaking until the optical density at 600 nm (OD_{600}) \approx 0.5. During this incubation phase, prepare 30 mL aliquots of LB for each culture, and preheat to 65 °C.
21. Induce protein expression by shifting culture temperature:
 - a) Withdraw 20 mL from the culture at 32 °C.
 - b) Add a 30 mL aliquot of LB preheated to 65 °C to each culture, and supplement with 30 μ L of 50 mg/mL ampicillin.

- c) Increase culture incubation temperature to 42 °C to maintain induction.
22. Determine the optimal induction time for, and estimate solubility of GroEL(191–345)–adenylyl cyclase fusion proteins.
- a) Withdraw 10 mL aliquots from each culture every hour post-induction. Equally divide each aliquot into total protein and soluble fraction protein samples. Pellet cells by centrifugation at $\geq 4000 \times g$ for 10 min, decant supernatant and store pellets at $-20\text{ }^{\circ}\text{C}$ until further processing. When ready to proceed, thaw cell pellets on ice.
- b) Total protein samples can be prepared for sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) analysis by resuspending in $1 \times$ SDS–PAGE loading buffer, supplemented with reducing agent and incubating at 95 °C to 100 °C for 10 min.
- c) Soluble fraction protein samples should be prepared by a non-denaturing cell lysis method such as probe sonication on ice or detergent and/or enzymatic lysis (e.g.: BugBuster[®], B-PER[™], etc.). Once cells are disrupted, clarify the supernatant by centrifugation at $\geq 12,000 \times g$ at 4 °C for 20 min. Withdraw aliquots of clarified supernatant, and mix with appropriate volume of $4 \times$ SDS–PAGE loading buffer supplemented with reducing agent. Incubate samples at 95 °C to 100 °C for 10 min.
- d) Load total protein and soluble protein fraction samples onto a 12% SDS–PAGE gel, and separate by electrophoresis. Stain and destain gel. Visualize to estimate peak expression time point and relative protein solubility by comparing band intensities from soluble protein fraction sample to corresponding total protein sample.

5.5 Express and purify GroEL(191–345)–CAMLP fusion protein

23. Transform *E. coli* MZ1 with the pRE:GroEL–CAMLP plasmid as described above.
24. Inoculate a seed culture of 50 mL LB supplemented with 50 $\mu\text{g}/\text{mL}$ ampicillin in 250 mL baffled flasks and incubate at 32 °C overnight with shaking.
25. Using seed culture, inoculate cultures of 1 L LB supplemented with 50 $\mu\text{g}/\text{mL}$ ampicillin in 2.8 L baffled Fernbach flasks. Incubate at 32 °C with shaking until culture $\text{OD}_{600} \approx 0.5$. For each culture flask, preheat to 65 °C a corresponding 600 mL volume of LB in a 2.8 L Fernbach flask.

26. Once the appropriate OD_{600} is reached, transfer 400 mL from each culture at 32 °C into a 600 mL LB aliquot at 65 °C. Supplement the cultures with 600 μ L of 50 mg/mL ampicillin, and incubate at 42 °C with shaking to maintain induction.
27. Harvest induced cells at the appropriate time point by centrifugation at $\geq 10,000 \times g$ for 10 min at 4 °C. Decant supernatant, wash the cell pellet with 10 mmol/L Tris-HCl pH 7.5 and decant.
28. Lyse cells to recover soluble GroEL(191–345)–CAMLP fusion protein.
 - a) Resuspend cell pellet to a ratio of 1 g wet cell weight per 10 mL of lysis buffer.
 - b) Lyse cells by passing suspension twice through a French pressure cell operated at 70 MPa.
 - c) Incubate the lysate at 70 °C for 2 min with mild stirring, and then immediately cool on ice.
 - d) Clarify lysate by centrifugation at $100,000 \times g$ for 15 min at 4 °C.
29. Purify GroEL(191–345)–CAMLP fusion protein
 - a) Equilibrate a DEAE-cellulose column with buffer A, and load lysate. Wash with ≥ 3 column volumes (CVs) of buffer A. Perform a linear gradient from 0% to 100% buffer B over 10 column volumes. GroEL(191–345)–CAMLP should elute in a sharp peak. Pool fractions.
 - b) Equilibrate a Sephadex G-75 superfine column (500 mL bed volume) with buffer A, and load pooled fractions onto column. Elute with ≈ 1.5 CVs of buffer A. Pool fractions, and concentrate using a 5 kDa MWCO centrifugal concentrator.
30. Digest GroEL–CAMLP fusion with enterokinase using buffer and recommended conditions from enzyme manufacturer.
31. Purify soluble CAMLP by gel filtration HPLC using a BIOSEP-SEC-S3000 column. Pool fractions, and concentrate using a 1 kDa MWCO cutoff centrifugal concentrator.
32. Analyze GroEL–CAMLP and CAMLP samples by SDS-PAGE as described above.



6. Expected outcomes

The ≈ 30 kDa cytoplasmic domains of adenylyl cyclases (ACCD) Mtb1625, Mlp0201 and Mlp1399 demonstrate the solubility benefit of N-terminal fusion to GroEL(191–345) for the overexpression of these proteins as demonstrated in Fig. 2. Comparing total protein from

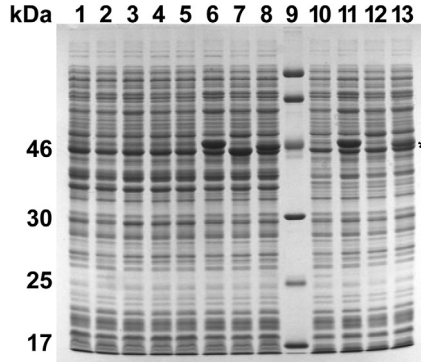


Fig. 2 Mycobacterial adenyl cyclase cytoplasmic domains overexpressed as GroEL(191–345) fusions. Proteins were separated by electrophoresis using a 12% SDS-PAGE gel with all lanes containing 30 μ g total protein except for lane 9 which is the molecular weight marker. Lanes 1 and 2 show total protein from induced *E. coli* transformed with pRE and pRE-GroEL(191–345) empty vectors, respectively. Lanes 3 through 5 show total protein from induced *E. coli* transformed with pRE-ACCDs Mtb1625, Mlp0201 and Mlp1399, respectively. No overexpressed band is apparent at \approx 30 kDa relative to the induced empty vector control (lane 1). Lanes 6 through 8 show total protein from induced *E. coli* transformed with pRE-GroEL(191–345)-ACCDs. Overexpressed bands are apparent at \approx 46 kDa (starred) relative to the induced empty vector control (lane 2). Lanes 11 through 13 show the soluble protein fraction from the same induced pRE-GroEL(191–345)-ACCDs in *E. coli* as lanes 6 through 8, and the overexpressed band persists in lanes 11 and 13 relative to empty vector control (lane 10).

pRE-ACCD induced *E. coli* shown in lanes 3 through 5 to total protein from induced *E. coli* carrying empty pRE vector (lane 1) does not reveal any apparent ACCD overexpression at the expected 30 kDa molecular weight. However, the same comparison for pRE-GroEL(191–345)-ACCD induced *E. coli* (lanes 6 through 8) with pRE-GroEL(191–345) (lane 2) shows overexpression with an expected fusion protein size of \approx 45 kDa in total protein samples. Further, much of the overexpressed GroEL(191–345)-ACCD fusion protein from Mt1625 and Mlp1399 (lanes 11 and 13) can be found in the soluble fraction after cell lysis; however, GroEL(191–345)-ACCD from Mlp0201 remained in inclusion bodies despite fusion as can be seen by comparing the soluble fraction (lane 12) with total protein (lane 7).

Persistent solubility of the protein-of-interest after cleavage from a GroEL(191–345) fusion supports the idea that GroEL(191–345) aids in the folding of its fusion partner instead of simply acting as a solubility enhancing partner or tag. This is demonstrated by enterokinase cleavage

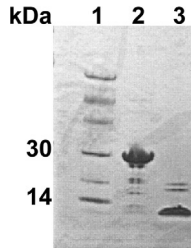


Fig. 3 GroEL(191–345)–CAMLP fusion and cleaved CAMLP product. 8% to 25% gradient SDS-PAGE analysis with Lane 1 containing molecular weight marker. GroEL(191–345)–CAMLP fusion protein can be purified in soluble form as shown in Lane 2. After cleavage of the fusion protein with enterokinase and further purification, CAMLP remains soluble (Lane 3). This figure is reprinted with modification from [Reddy et al. \(2003\)](#).

of a GroEL(191–345)–CAMLP fusion with results shown in [Fig. 3](#). CAMLP is a 55 AA protein whose overexpression was not apparent on SDS-PAGE when cloned into the parental pRE vector (data not shown). However, when expressed as a GroEL(191–345) fusion, CAMLP overexpresses and is soluble both in the purified fusion form (lane 2) as well as after cleavage and size-exclusion purification to remove the GroEL(191–345) fragment (lane 3).



7. Advantages

The pRE-GroEL(191–345) fusion expression vector combines the advantages of tightly regulated expression for toxic proteins of interest with the apparent chaperone activity of the GroEL(191–345) minichaperone for proteins prone to insolubility upon *E. coli* overexpression. When transformed into a host *E. coli* strain bearing one (or more) genomic copies of the temperature sensitive variant of lambda repressor, λ cI857, inducible expression can be achieved by temperature shift using the pRE:GroEL vector alone with no additional helper plasmids required.



8. Limitations

The temperature sensitive lambda repressor variant λ cI857 requires culture maintenance at 28 °C for maximal repression and at 42 °C for maximal transcription from the P_L promoter ([Remaut et al., 1981](#)). (Sufficient repression can generally be maintained at 32 °C as described in the method.)

Such temperatures can negatively impact the growth rate and/or fitness of *E. coli*; however, λ cI857 can partially derepress at temperatures as low as 39°C (Reddy et al., 1989) or even 37°C (George, Watson, Harbrecht, & DeLorbe, 1987). In addition, the use of *E. coli* strains harboring temperate λ or λ 857 prophage for subcloning and/or expression incurs the risk of phage induction, particularly at the elevated 42°C temperature required to derepress P_L. Selecting an *E. coli* host harboring a defective prophage or transforming with a second plasmid expressing λ cI (subcloning) or λ cI857 (expression) can avoid the risk of phage induction.

Disclaimer

Certain commercial equipment, instruments or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

References

- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, *393*, 537–544. <https://doi.org/10.1038/31159>.
- Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., et al. (2001). Massive gene decay in the leprosy bacillus. *Nature*, *409*, 1007–1011. <https://doi.org/10.1038/35059006>.
- Ellis, K. J., & Hartl, F.-U. (1996). Protein folding in the cell: Competing models of chaperonin function. *The FASEB Journal*, *10*(1), 20–26. <https://doi.org/10.1096/fasebj.10.1.8566542>.
- George, H. J., Watson, R. J., Harbrecht, D. F., & DeLorbe, W. J. (1987). A bacteriophage λ cI857 cassette controls λ P_L expression vectors at physiologic temperatures. *Bio/Technology*, *5*(5), 600–603. <https://doi.org/10.1038/NBT0687-600>.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology*, *166*(4), 557–580. [https://doi.org/10.1016/S0022-2836\(83\)80284-8](https://doi.org/10.1016/S0022-2836(83)80284-8).
- Isaacs, L. N., Echols, H., & Sly, W. S. (1965). Control of lambda messenger RNA by the C₁-immunity region. *Journal of Molecular Biology*, *13*(3), 963–967. [https://doi.org/10.1016/S0022-2836\(65\)80163-2](https://doi.org/10.1016/S0022-2836(65)80163-2).
- Jaenicke, R. (1998). Protein self-organization *in vitro* and *in vivo*: Partitioning between physical biochemistry and cell biology. *Biological Chemistry*, *379*(3), 237–243. <https://doi.org/10.1515/bchm.1998.379.3.237>.
- Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M. K., & Hartl, F. U. (1992). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature*, *356*, 683–689. <https://doi.org/10.1038/356683a0>.
- Lieb, M. (1966). Studies of heat-inducible lambda bacteriophage: I. Order of genetic sites and properties of mutant prophages. *J. Mol. Biol.*, *16*(1), 149–163. [https://doi.org/10.1016/S0022-2836\(66\)80269-3](https://doi.org/10.1016/S0022-2836(66)80269-3).
- Reddy, P., Peterkofsky, A., & Mckenney, K. (1989). Hyperexpression and purification of *Escherichia coli* adenylate cyclase using a vector designed for expression of lethal gene products. *Nucleic Acids Research*, *17*(24), 10473–10488. <https://doi.org/10.1093/nar/17.24.10473>.

- Reddy, P. T., Prasad, C. R., Reddy, P. H., Reeder, D., McKenney, K., Jaffé, H., et al. (2003). Cloning and expression of the gene for a novel protein from *Mycobacterium smegmatis* with functional similarity to eukaryotic calmodulin. *Journal of Bacteriology*, *185*(17), 5263–5268. <https://doi.org/10.1128/JB.185.17.5263-5268.2003>.
- Remaut, E., Stanssens, P., & Fiers, W. (1981). Plasmid vectors for high-efficiency expression controlled by the PL promoter of coliphage lambda. *Gene*, *15*(1), 81–93. [https://doi.org/10.1016/0378-1119\(81\)90106-2](https://doi.org/10.1016/0378-1119(81)90106-2).
- Schein, C. H. (1989). Production of soluble recombinant proteins in bacteria. *Nature Biotechnology*, *7*, 1141–1149. <https://doi.org/10.1038/nbt1189-1141>.
- Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification*, *41*(1), 207–234. <https://doi.org/10.1016/j.pep.2005.01.016>.
- Zahn, R., Buckle, A. M., Perrett, S., Johnson, C. M., Corrales, F. J., Golbik, R., et al. (1996). Chaperone activity and structure of monomeric polypeptide binding domains of GroEL. *Proceedings of the National Academy of Sciences*, *93*(26), 15024–15029. <https://doi.org/10.1073/pnas.93.26.15024>.