

## Alternative approach to express *Mycobacterium tuberculosis* proteins in *Escherichia coli*

Sandeep Mukherjee, Nada Daifalla, Chao Liu, and Antonio Campos-Neto  
*Infectious Disease Research Institute, Seattle, WA, USA*

*BioTechniques* 35:34-40 (July 2003)

The rigor of our understanding of the various aspects of gene expression in *Escherichia coli* has allowed geneticists not only to engineer increasingly sophisticated expression vectors to maximize gene expression but also to create a growing number of host strains for optimizing the expression of gene products. The advent of inducible expression systems, especially in the bacteriophage T7 RNA polymerase-based system (1–3), allows nearly routine overexpression at levels from 2% to as high as 50% of the cell protein. However, for unknown reasons, several proteins (usually of low molecular weight)

are very difficult to express in *E. coli* host cells; typical examples are several proteins of *Mycobacterium tuberculosis* and *Pseudomonas aeruginosa* (4–7). The use of low copy number T7 expression vectors has improved the expression of some proteins of both microorganisms in *E. coli* (8). Another alternative to circumvent this constraint uses the standard artifice of engineering the genes encoding such proteins “fused” with genes encoding proteins that are readily overexpressed in *E. coli*. In addition, these proteins contain a specific binding motif to facilitate the purification and detection of the recombinant fusion proteins. Some common examples are calmodulin-binding peptide, glutathione-S-transferase (GST), maltose-binding protein, and thioredoxin (9). However, a serious limitation of this procedure is the obvious fact that the recombinant protein is “contaminated” with an unrelated molecule, which is a situation that is in direct opposition to the very concept of producing recombinant proteins to generate highly purified molecules. Despite the fact that most of these systems offer the option of specific digestion of the junction of the two fused proteins (by engineering specific enzymatic sites at the junction), followed by purification of the protein of interest, this procedure is cumbersome and often does not work properly. Moreover, many proteins are not expressed well even as fusion proteins.

Here we report the results of a simple, alternative procedure to express *M. tuberculosis* problem proteins in *E. coli*. In contrast to engineering the genes of these proteins with a heterolo-

gous “carrier” gene, we constructed “homo-fusions” (i.e., fusions of two or three copies of the same gene in tandem expressed as dimers or trimers). This simple procedure resulted in the excellent expression of the immunologically reactive recombinant proteins.

This alternative was initially used to express the gene of a small molecular weight protein of *M. tuberculosis*, which we could not otherwise express. The gene encoding this 6.4-kDa protein (which we named U2) [GenBank® accession no. CAB08507; open reading frame (ORF) Rv0909] has been under investigation in our laboratory as a potential diagnostics and vaccine candidate. Our initial approach, involving PCR amplification followed by cloning into the pET-17b vector and transformation into *E. coli* BL-21 (DE3) pLysS host cells (both from Novagen, Madison, WI, USA), was disappointing. Upon induction with IPTG, no significant expression of the protein could be obtained; in fact, a region of diffused band of lower molecular weight than expected was observed. No significant expression of this molecule could be achieved either with the BL-21(DE3) or the BL-21(DE3) pLysS. Before testing the expression of U2 in conventional and commercially available systems involving fusion proteins (hetero-fusions), we tested the possibility of facilitating expression by simply increasing the molecular weight of the protein using the artifice of constructing homofusions instead. Three sets of oligonucleotide PCR primers were used separately to amplify the full-length ORF of U2 from genomic DNA of the virulent Erdman strain of *M. tuberculosis*. Table 1 depicts the sequences of these three sets of primers. The resultant PCR products were double-digested with *Nde*I and *Hind*III, *Hind*III and *Bam*HI, and *Bam*HI and *Eco*RV, respectively. The three fragments were ligated (T4 DNA ligase; Stratagene, La Jolla, CA, USA) and cloned into the pET-17b vector in tandem, which was also double-digested with *Nde*I and *Hind*III, *Hind*III and *Bam*HI, and *Bam*HI and *Eco*RV, respectively, in successive rounds of digestion, ligation, transformation, and plasmid preparation. The result was three clones: one, the original copy of the

**Table 1.** Sets of Forward (5') and Reverse (3') Primers Used to Engineer the Homo-Trimer U2 Gene

Set 1	5'	5'-CAATTACAT <u>ATG</u> <b>CATCAC</b> ACCACATGGGAATCCTGGACAAG-3'
		<i>NdeI</i>
	3'	5'-CAATTAA <u>AGCTT</u> GCTCTGCTGGTCGCTCAT-3'
		<i>HindIII</i>
Set 2	5'	5'-CAATTAA <u>AGCTT</u> ATGGGAATCCTGGACAAG-3'
		<i>HindIII</i>
	3'	5'-CAATTAG <u>GATCC</u> GCTCTGCTGGTCGCTCAT-3'
		<i>BamHI</i>
Set 3	5'	5'-CAATTAG <u>GATCC</u> ATGGGAATCCTGGACAAG-3'
		<i>BamHI</i>
	3'	5'-CAATTAG <u>AGATAC</u> <b>CTAG</b> GCTCTGCTGGTCGCTCAT-3'
		<i>EcoRV</i>

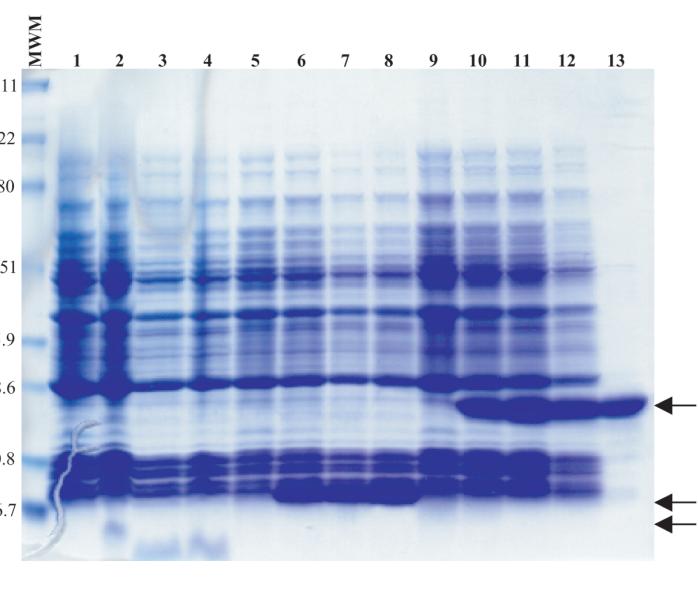
Set 1 contains at the 5' end an *NdeI* restriction site and the ATG initiation codon together (underlined), followed by a sequence encoding six histidines (bold) and sequence derived from the U2 gene (italic). The 3' end contains the N-terminal protein-coding sequence followed by a *HindIII* restriction site (underlined). Set 2 contains at the 5' end the *HindIII* restriction site, followed by the gene-derived sequence, and the 3', the gene coding sequence (italic), followed by a *BamHI* restriction site (underlined). Set 3 contains at the 5' end the *BamHI* restriction site (underlined), followed by the gene-derived sequence, and the 3', the gene coding sequence (italic), followed by the stop codon CTA (bold/italic) and by an *EcoRV* restriction site (underlined).

gene in pET-17b; one in which two copies of the gene were fused (expressed as a homo-dimer); and one in which three copies were fused (expressed as a homo-trimer). The ligation products were initially transformed into *E. coli* XL1-Blue competent cells (Stratagene) and were subsequently transformed into *E. coli* BL-21(DE3) pLysS for expression.

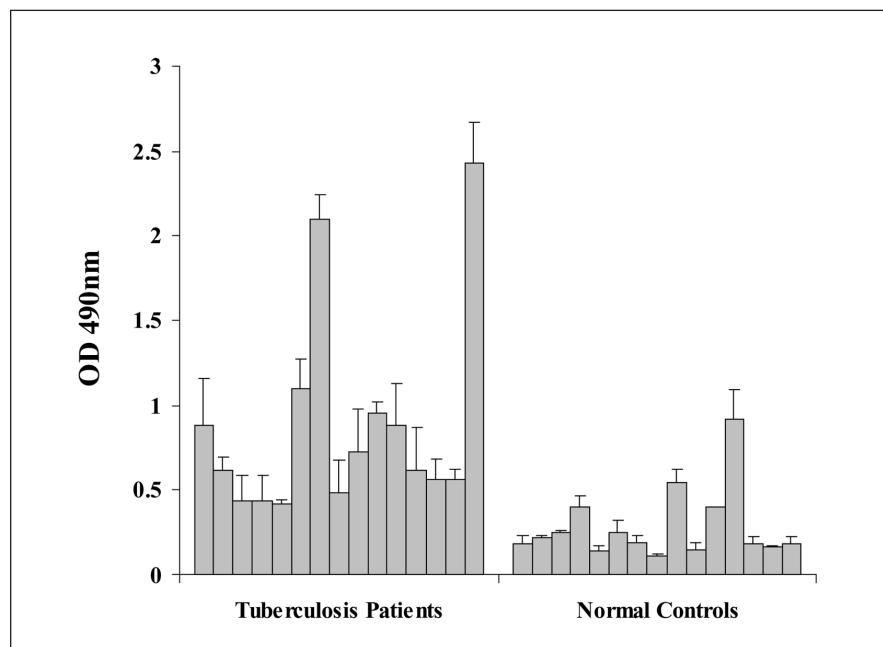
The genes encoding the recombinant dimer and trimer fusion proteins were induced in LB medium with 0.25, 0.5, 1.0, and 2.0 mM IPTG for 3 h at 37°C. Before induction, bacteria were expanded in LB medium for approximately 18 h at 37°C. Regardless of the IPTG concentration, similar yields of the recombinant proteins (8–10 mg/L *E. coli* culture) were achieved consistently.

Coomassie® staining of polyacrylamide gels performed under reduction conditions clearly revealed the presence of distinct induced proteins, migrating at positions that matched the expected molecular weights of both the homo-dimer and homo-trimer of the native U2 (Figure 1) induced with 1 mM IPTG. Western blot analyses with anti-His-tag antibody confirmed that the overexpressed proteins were the correct proteins (data not shown). The recombinant trimer was found in the insoluble inclusion body and was easily purified from 1 L IPTG-induced batch cultures by affinity chromatography using the one-step QIAexpress® Ni-NTA agarose matrix (Qiagen, Valencia, CA, USA) in the presence of 8 M urea, as previously described (10). The yield of recombinant protein varied from 8 to 10 mg/L induced bacterial culture. The purity of the purified recombinant protein was assessed by SDS-PAGE, followed by Coomassie Blue staining (Figure 1), and N-terminal sequencing using Edman chemistry with a Procise® 494 protein sequencer (Applied Biosystems, Foster City, CA, USA) confirmed the correct sequence (data not shown).

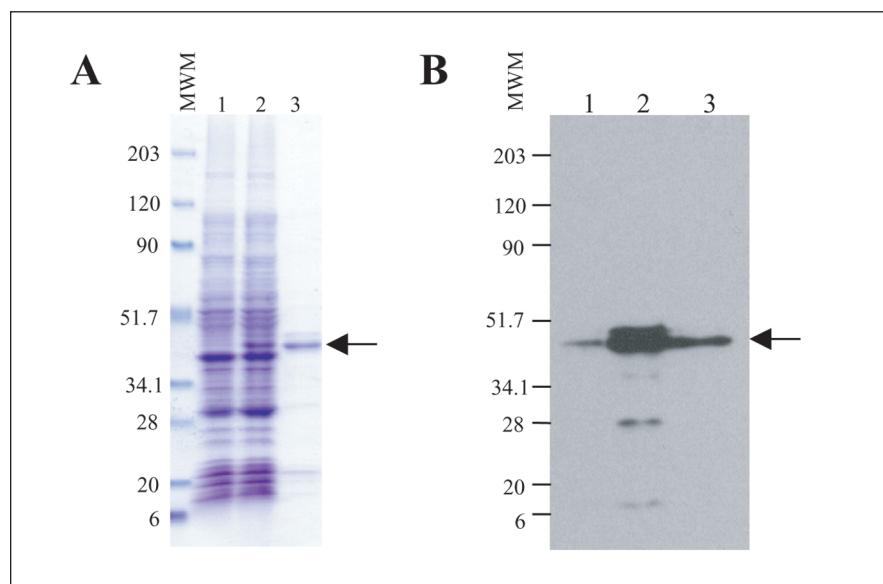
To ascertain that the expressed and



**Figure 1.** Expression of U2 as homo-dimer and homo-trimer recombinant proteins. *E. coli* BL21(DE3)pLysS transformed with expression vector pET-17b containing the genes encoding U2 monomer, U2 dimer, and U2 trimer, respectively. The gel was stained with Coomassie Blue. U2 monomer (lanes 1–4); U2 dimer (lanes 5–8); U2 trimer (lanes 9–12); purified U2 trimer (lane 13). U2 trimer was purified by affinity chromatography with one-step QIAexpress Ni-NTA agarose matrix. Non-induced *E. coli* lysate (lanes 1, 5, and 9), *E. coli* lysate from induction with 1 mM IPTG for 1 h (lanes 2, 6, and 10), 2 h (lanes 3, 7, and 11), and 3 h (lanes 4, 8, and 12). The numbers on the left indicate molecular weights of the markers in kilodaltons (prestained SDS-PAGE marker, broad range; Bio-Rad Laboratories, Hercules, CA, USA). The numbers on the right indicate the relative molecular weights of expressed trimer (23 kDa) and dimer (13 kDa) proteins and the position of the expected migration of the monomer molecule (6.4 kDa). MWM, molecular weight marker.



**Figure 2. Serological reactivity of U2 with sera from tuberculosis patients and normal individuals.** Sera from 15 patients with pulmonary tuberculosis and no signs of co-infection with HIV and from 15 healthy purified protein derivative-negative individuals (Normal) were evaluated for reactivity with U2 using standard sandwich ELISA. The results are plotted as OD<sub>490</sub>, given by the sera diluted at 1/25. The serological reactivity of U2 with these two panels of sera was performed three times (one month apart). Bars represent the standard deviation of the absorbance readings obtained with these three determinations.



**Figure 3. Expression of the homo-trimer DPPD gene as a recombinant protein.** *E. coli* BL21(DE3)pLysS transformed with expression vector pET-17b, containing the homo-trimer DPPD gene, was grown and induced with IPTG. The cells were lysed, and the DPPD protein was purified by affinity chromatography using the one-step QIAexpress Ni-NTA agarose matrix. Expression and purification were evaluated by SDS-PAGE (gradient gel 20–4%) under reducing conditions, and gel was stained with Coomassie Blue. (A) Lane 1, noninduced *E. coli* lysate; lane 2, induced *E. coli* lysate; lane 3, purified recombinant DPPD protein. (B) Western blot analysis of DPPD. A replica of the gel described in panel A was transferred to a nitrocellulose membrane. Proteins were identified using a mouse anti-His-tag monoclonal antibody. Reactivity was detected with peroxidase-labeled goat anti-mouse immunoglobulin and developed using ECL® chemoluminescent reagent (Western blot detection system; Amersham Biosciences). The numbers on the left side of panels A and B indicate the molecular weights of the markers. Arrows point to the homo-trimer DPPD. MWM, molecular weight marker.

purified recombinant U2 homo-fusion proteins had preserved epitopes, both ELISA and Western blot analyses were performed using human sera from patients with pulmonary tuberculosis and from healthy persons from areas where tuberculosis is not an endemic disease (USA) as a control. Figure 2 shows the results and clearly demonstrates the recognition of the recombinant protein by the sera from patients with tuberculosis and not by the control sera, thus indicating the preservation of the tuberculosis-specific epitopes in the trimerized format of the U2 molecule.

To evaluate the potential of engineering homo-fusions to overexpress problem proteins in *E. coli*, we tested another small molecular weight *M. tuberculosis* protein (9 kDa), which we have previously described (11). This recombinant molecule, named DPPD (based on sequence of the first four amino acids of its N-terminus), is a very useful tool for the diagnosis of tuberculosis (accession no. NP\_334475; ORF CDC1551). Unfortunately, thus far, we have been unable to express DPPD either as a single or fusion recombinant protein using the following systems: pET vector (Novagen), Rosetta™ BL-21(DE3) systems (Novagen), pQ30 (Qiagen), pET 32A (Novagen), pThioHis (Invitrogen, Carlsbad, CA, USA), and pGEX-2T (Amersham Biosciences, Uppsala, Sweden). To overcome this problem, we engineered DPPD as a chimerical construct with a highly expressing 12-kDa *M. tuberculosis* protein named Ra12 (11,12). This construct yielded a DPPD fusion protein that could be purified and tested. However, because the Ra12 gene, in contrast to DPPD, is broadly distributed among the *Mycobacterium* genus, this fusion protein introduced an undesired property to the recombinant DPPD. To circumvent this problem, three DPPD gene copies were contiguously fused and cloned into the pET plasmid vector (pET-17b), followed by the transformation of *E. coli* BL-21(DE3) pLysS for high-level expression. Exactly the same step-wise ligation and cloning strategies we used to express U2 were used to engineer the DPPD homo-fusion. Figure 3A illustrates the results and clearly shows that the tri-fusion DPPD

## BENCHMARKS

was easily expressed and purified. Western blot analyses revealed that the expressed and purified molecule was the correct His-tag recombinant protein (Figure 3B). Moreover, amino acid sequencing using Edman degradation confirmed the correct N-terminal sequence of the molecule (data not shown). In addition, the immunological reactivity of the purified trimer [elicitation of delayed type hypersensitivity (DTH)] was tested in *M. tuberculosis*-infected guinea pigs. Five infected and three noninfected control animals were skin-tested with 5 µg purified recombinant tri-fusion DPPD, and the development of DTH was read 24 h later. All five infected guinea pigs showed typical DTH reactions, with diameters ranging from 8 to 12 mm induration. No reaction was observed in any of the noninfected controls (data not shown).

At this point, in the absence of further mechanistic studies, it is difficult to explain how the act of fusing genes to each other facilitates gene expression. Further studies involving RNA stabilization, for example, are needed to evaluate the mechanism involved in protein overexpression by fusing several copies of the same gene in tandem.

Finally, although the validation of this alternative approach needs to be tested for proteins other than those encoded by *M. tuberculosis* genes, this system already represents an attractive and significant achievement for the production of the recombinant proteins of this important pathogen.

### ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health grant no. AI 43528 to A.C.-N.

### REFERENCES

1. Studier, F.W. and B.A. Moffatt. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189:113-130.
2. Studier, F.W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* 219:37-44.
3. Studier, F.W., J. Rosenberg, J. Dunn, and J. Dubendorff. 2003. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* 185:60-89.
4. Barker, L.P., S.F. Porcella, R.G. Wyatt, and P.L. Small. 1999. The *Mycobacterium marinum* G13 promoter is a strong sigma 70-like promoter that is expressed in *Escherichia coli* and mycobacteria species. *FEMS Microbiol. Lett.* 175:79-85.
5. Harth, G., B.Y. Lee, J. Wang, D.L. Clemens, and M.A. Horwitz. 1996. Novel insights into the genetics, biochemistry, and immunocytochemistry of the 30-kilodalton major extracellular protein of *Mycobacterium tuberculosis*. *Infect. Immun.* 64:3038-3047.
6. Harth, G., B.Y. Lee, and M.A. Horwitz. 1997. High-level heterologous expression and secretion in rapidly growing nonpathogenic mycobacteria of four major *Mycobacterium tuberculosis* extracellular proteins considered to be leading vaccine candidates and drug targets. *Infect. Immun.* 65:2321-2328.
7. Roche, P.W., N. Winter, J.A. Triccas, C.G. Feng, and W.J. Britton. 1996. Expression of *Mycobacterium tuberculosis* MPT64 in recombinant *Mycobacterium smegmatis*: purification, immunogenicity and application to skin tests for tuberculosis. *Clin. Exp. Immunol.* 103:226-232.
8. Hoang, T.T., Y. Ma, R.J. Stern, M.R. McNeil, and H.P. Schweizer. 1999. Construction and use of low-copy number T7 expression vectors for purification of problem proteins: purification of *Mycobacterium tuberculosis* RmlD and *Pseudomonas aeruginosa* LasI and RhII proteins, and functional analysis of purified RhII. *Gene* 237:361-371.
9. Terpe, K. 2003. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 60:523-533.
10. Skeiky, Y.A., M. Kennedy, D. Kaufman, M.M. Borges, J.A. Guderian, J.K. Scholler, P.J. Ovendale, K.S. Picha, et al. 1998. LeIF: a recombinant *Leishmania* protein that induces an IL-12-mediated Th1 cytokine profile. *J. Immunol.* 161:6171-6179.
11. Campos-Neto, A., V. Rodrigues-Junior, D.B. Pedral-Sampaio, E.M. Netto, P.J. Ovendale, R.N. Coler, Y.A. Skeiky, R. Badaro, et al. 2001. Evaluation of DPPD, a single recombinant *Mycobacterium tuberculosis* protein as an alternative antigen for the Mantoux test. *Tuberculosis (Edinb.)* 81:353-358.
12. Coler, R.N., Y.A. Skeiky, P.J. Ovendale, T.S. Vedvick, L. Gervassi, J. Guderian, S. Jen, S.G. Reed, et al. 2000. Cloning of a *Mycobacterium tuberculosis* gene encoding a purified protein derivative protein that elicits strong tuberculosis-specific delayed-type hypersensitivity. *J. Infect. Dis.* 182:224-233.

Received 20 February 2003; accepted 9 April 2003.

Address correspondence to Antonio Campos-Neto, Infectious Disease Research Institute, 1124 Columbia Street, Suite 600, Seattle, WA 98104, USA. e-mail: acampos@idri.org