Protein Expression and Purification 65 (2009) 230–237

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

A novel bicistronic vector for overexpressing *Mycobacterium tuberculosis* proteins in *Escherichia coli*

Yin Guo, Susan S. Wallace *, Viswanath Bandaru [1]

Department of Microbiology and Molecular Genetics, The Markey Center for Molecular Genetics, University of Vermont, Stafford Hall, 95 Carrigan Dr. Burlington, VT 05405-0068, USA

ARTICLE INFO

Article history:
Received 17 November 2008
and in revised form 19 December 2008
Available online 30 December 2008

Keywords:
Codon usage
mRNA secondary structure
Translation initiation
Bicistronic vector

ABSTRACT

A putative DNA glycosylase encoded by the *Rv3297* gene (MtuNei2) has been identified in *Mycobacterium tuberculosis*. Our efforts to express this gene in *Escherichia coli* either by supplementing tRNAs for rare codons or optimizing the gene with preferred codons for *E. coli* resulted in little or no expression. On the other hand, high-level expression was observed using a bicistronic expression vector in which the target gene was translationally coupled to an upstream leader sequence. Further comparison of the predicted mRNA secondary structures supported the hypothesis that mRNA secondary structure(s) surrounding the translation initiation region (TIR), rather than codon usage, played the dominant role in influencing translation efficiency, although manipulation of codon usage or tRNA supplementation did further enhance expression in the bicistronic vector. Addition of a cleavable N-terminal tag also facilitated gene expression in *E. coli*, possibly through a similar mechanism. However, since cleavage of N-terminal tags is determined by the amino acid at the P_1' position downstream of the protease recognition sequence and results in the addition of an extra amino acid in front of the N-terminus of the protein, this strategy is not particularly amenable to Fpg/Nei family DNA glycosylases which carry the catalytic proline residue at the P_1' position and require a free N-terminus. On the other hand, the bicistronic vector constructed here is potentially valuable particularly when expressing proteins from G/C rich organisms and when the proteins carry proline residues at the N-terminus in their native form. Thus the bicistronic expression system can be used to improve translation efficiency of mRNAs and achieve high-level expression of mycobacterial genes in *E. coli*.

Introduction

Escherichia coli remains a common host for high-level expression of heterologous genes, however, this often depends on the source of the target genes. A number of factors that significantly influence gene expression at the translation level have been identified with codon usage and mRNA secondary structures being the major concerns.

There are marked differences in codon usage from one organism to another. Significant variation in codon usage patterns among genes in one organism appears to be associated with their expression levels. Genes with a high proportion of optimal codons are highly expressed, whereas those with rare codons are poorly expressed [1,2]. Moreover, the presence of rare codons can cause ribosome stalling, slow translation, pre-mature translation termination and translation errors and therefore inhibit proper protein synthesis and even cell growth [3–7]. To avoid the potential expression problems resulting from rare codons, one can either optimize codon usage in the target gene by silent mutations for expression in *E. coli* or expand the intracellular tRNA pool of rare codons by introducing a plasmid which encodes these tRNAs [3]. Both strategies have been successfully used to enhance the expression of heterologous genes with rare codons in *E. coli* [2,8]. However, negative results have been reported indicating that factors other than codon usage can affect protein expression [8,9].

It has been widely accepted that secondary structure of mRNA in the translation initiation region (TIR)[2] plays a crucial role in controlling translation efficiency. A quantitative analysis revealed a strict correlation between the translational efficiency and the stability of the local secondary structure [10]. Since the 30S ribosomal subunit most likely binds to single-stranded regions of mRNA and slides into place while unfolding the TIR, the secondary structure(s) by sequestering either the ribosome binding site (RBS) and/or the initiation codon from ribosome binding will block translation initiation thereby inhibiting translation [11–13].

A series of two-cistron plasmids (bicistronic vectors) have been

* Corresponding author. Fax: +1 802 656 8749.
E-mail address: susan.wallace@uvm.edu (S.S. Wallace).

[1] Present address: Mylan Pharmaceuticals, 3711 Collins Ferry Rd., Morgantown, WV 26505, USA.

[2] *Abbreviations used:* TIR, translation initiation region; RBS, ribosome binding site; BER, base excision repair; LB, Luria Broth medium; ORF, open reading frame; MBP, maltose binding protein.

1046-5928/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2008.12.013

successfully used to overcome translational inhibition of mRNAs with stable secondary structures [14–18]. In such a system, the first/upstream cistron is generally an A/T-rich sequence, which can minimize local secondary structure(s) thereby allowing efficient translation initiation [15,16], while the second/downstream cistron containing the coding sequence of the target gene is translationally coupled to the first/upstream cistron. Therefore, protein production from a two-cistron system is theoretically dependent upon the efficiency of translation of the first/upstream cistron [19]. Two models have been proposed in terms of the mechanism of translational coupling [19–21]. One is that a ribosome translating the upstream cistron disrupts the inhibitory secondary structure thereby making the RBS of the downstream cistron accessible to other ribosomes which can initiate translation of the coupled downstream gene. The other is that the same translating ribosome can re-initiate and continue to translate the downstream cistron.

Mycobacterium tuberculosis, the causative agent of tuberculosis, is a high-G/C gram-positive bacterium with a genomic G/C content of approximately 65%. The high-G/C content results in a markedly different pattern of codon usage from that of *E. coli*. A number of rare codons for *E. coli*, widely used in the genome of *M. tuberculosis*, might be one reason for the poor expression of mycobacterial genes in *E. coli*, even in the presence of strong *E. coli* promoters [1,22,23]. Moreover, stable secondary structures of mRNA surrounding the TIR can be formed due to the high-G/C content and may be responsible for the poor expression. A number of expression systems have been developed for expressing mycobacterial proteins in bacteria that are phylogenetically closer to *M. tuberculosis*, particularly *Mycobacterium smegmatis* [24–28]. However, the success of these expression systems is still limited to a few genes [24–28]. Thus a versatile efficient expression system is still an urgent need.

The sequence analysis of the genome of *M. tuberculosis* [29] allowed us to identify three putative DNA glycosylase genes of the Fpg/Nei family, *Rv2464c*, *Rv2924c* and *Rv3297*. The *Rv2924c* gene encodes a 32.0 kDa formamidopyrimidine (Fpg) DNA glycosylase (MtuFpg1). The *Rv2464c* and *Rv3297* genes encode a 29.7 kDa endonuclease VIII (MtuNei1) and a 28.5 kDa endonuclease VIII (MtuNei2) respectively. The DNA glycosylases of the Fpg/Nei family recognize and remove oxidized DNA bases as a first step in the base excision repair (BER) process responsible for removing endogenous oxidative damages from the DNA [30]. As shown in *Salmonella typhimurium* [31] and *Helicobacter pylori* [32], the BER pathway may be involved in pathogen proliferation or colonization conferring a virulence advantage for the microorganisms, and if this is true for *M. tuberculosis*, it could provide a target for future therapy.

The potential role of codon usage and mRNA secondary structure(s) in regulating gene expression was tested using the mycobacterial *Rv3297* gene as an example. Although both are regarded as potential causes of poor expression of *M. tuberculosis* genes in *E. coli*, our results suggest that mRNA secondary structure(s), rather than codon usage, is the primary determinant in influencing translation efficiency. To our knowledge, this is the first report where a mycobacterial gene has been overexpressed in high-levels in *E. coli*, and the bicistronic vector designed here should be generally applicable to improve translation efficiency of mRNAs and to achieve high-level gene expression of heterologous genes in *E. coli*.

Materials and methods

Materials

The genomic DNA of *M. tuberculosis* H37Rv was kindly provided by Dr. Karin Eiglmeier (*Unité de Génétique Moléculaire Bactérienne*, Paris, France). The DNA sequences were retrieved from GenBank for *M. tuberculosis* H37Rv (*Rv2924c*: gi|15610061; *Rv2464c*: gi|15609601; *Rv3297*: gi|1877352). The primers were chemically synthesized by Operon Biotechnologies, Inc. (Germantown, MD) and Midland Certified Reagent Company, Inc. (Midland, TX). Restriction enzymes were purchased from New England Biolabs (NEB, Beverly, MA). Cloned Pfu DNA polymerase was purchased from Stratagene (Cedar Creek, TX). Expression vectors pET30a (Novagen, Madison, WI) and the bicistronic pET vector (pET30a/ORF6) constructed in this paper were used to overexpress the three mycobacterial genes in *E. coli*. Recombinant plasmids were amplified in *E. coli* TOP10 cells (Invitrogen, Carlsbad, CA) and plasmid DNA was purified using Wizard Plus Midiprep DNA Purification System (Promega, Madison, WI). BL21-Gold (DE3) (Stratagene) was used as the host strain for protein expression in *E. coli*.

Construction of the bicistronic vector

A 74-mer oligonucleotide (5′GGAATTCCATATGAAAATCGAAGCAGGTAAACTGGTACAG**aaggag**ATTAACT*GATATC*GGATCCCTCGAGCGG) and its complementary strand were designed and chemically synthesized by Midland Certified Reagent Company, Inc that included NdeI and XhoI restriction sites (underlined) at the N and C-termini, an EcoRV restriction site (italics) for cloning the target gene and a RBS (lowercase/bold) for the target gene. An equal amount of the oligonucleotide and its complementary strand were annealed in NEBuffer 4 (50 mM potassium acetate, 20 mM Tris acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9) and then cleaved with NdeI and XhoI. The resulting fragment was purified from a 1% agarose gel with β-agarase (NEB) and then cloned between the NdeI and XhoI sites of pET30a vector to create the pET30a-ORF6 vector.

Cloning of the target Rv3297 gene into the pET30a vector

The nucleotide sequence of the *Rv3297* gene was optimized with *E. coli* preferred codons and chemically synthesized by GeneScript Corporation (Piscataway, NJ) resulting in the synthesized *Rv3297* (*sRv3297*) gene. Using two primers: sRv3297-Fwd1 (5′GATCCAATCATATGCCGGAAGGTGATACCGTGTGGC) and sRv3297-Rev1 (5′ TGGCTCGAGACGCTGGCACGCCGGGCACCAATAGC), the gene was amplified and cloned between NdeI and XhoI sites in pET30a vector as described below resulting in pET30a/sRv3297-His expression vector.

The *Rv3297* gene was also cloned from the genomic DNA using two primers: Rv3297-F (5′AGATATACATATGCCGGAGGGCGACACCGTCTGGCAC) and pETRv3297Rev-2 (5′CCGCTCGAGGCGCTGGCAGGCCGGGCACCAATACC) to yield the cloned *Rv3297* (*cRv3297*) gene. The PCR reaction was carried out with cloned Pfu DNA polymerase in an Air Thermo-Cycler (Idaho Technology, Salt Lake City, UT) for an initial denaturation for 5 min at 94 °C, 40 cycles with 94 °C for 15 s, 60 °C for 45 s and 72 °C for 2.5 min and an additional extension for 5 min at 72 °C. The PCR product was cleaved with NdeI and XhoI restriction enzymes, purified from a 1% agarose gel with β-agarase and then inserted between the NdeI and XhoI sites of pET30a vector to yield pET30a/cRv3297-His vector.

Additionally, in order to express MtuNei2 fused to a 37 amino acid N-terminal His/thrombin/S-tag, the *cRv3297* gene was subcloned into pET30a vector between the KpnI and XhoI sites using a forward primer Fwd-5 (5′GAAGGAGATGGTACCATGCCGGAGGGCGACACCG) and pETRv3297Rev-2 to yield pET30a/His-cRv3297.

Cloning of the genes into the bicistronic pET30a-ORF6 vector

To improve expression of the *Rv3297* in *E. coli*, the target gene was amplified and cloned into the biscistronic pET30a-ORF6 vector as follows. Forward primers, Rv3297-F (5′GCCGGAGGGCGACAC