Available online at www.sciencedirect.com
ScienceDirect
Protein Expression and Purification 57 (2008) 81–87
Protein Expression & Purification
www.elsevier.com/locate/yprep

A new Gateway® vector and expression protocol for fast and efficient recombinant protein expression in *Mycobacterium smegmatis*

Rachael M. Goldstone [a,b,1], Nicole J. Moreland [a,b,1], Ghader Bashiri [a,b], Edward N. Baker [a,b], J. Shaun Lott [a,c,*]

[a] *Maurice Wilkins Centre for Molecular Biodiscovery, School of Biological Sciences, Thomas Building, 3a Symonds Street, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand*
[b] *Laboratory of Structural Biology, School of Biological Sciences, Thomas Building, 3a Symonds Street, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand*
[c] *AgResearch Structural Biology Laboratory, School of Biological Sciences, Thomas Building, 3a Symonds Street, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand*

Received 25 June 2007, and in revised form 9 August 2007
Available online 14 September 2007

Abstract

A major obstacle associated with recombinant protein over-expression in *Escherichia coli* is the production of insoluble inclusion bodies, a problem particularly pronounced with *Mycobacterium tuberculosis* proteins. One strategy to overcome the formation of inclusion bodies is to use an expression host that is more closely related to the organism from which the proteins are derived. Here we describe methods for efficiently identifying *M. tuberculosis* proteins that express in soluble form in *Mycobacterium smegmatis*. We have adapted the *M. smegmatis* expression vector pYUB1049 to the Gateway® cloning system by the addition of *att* recombination recognition sequences. The resulting vector, designated pDESTsmg, is compatible with our in-house Gateway® methods for *E. coli* expression. A target can be subcloned into pDESTsmg by a simple LR reaction using an entry clone generated for *E. coli* expression, removing the need to design new primers and re-clone target DNA. Proteins are expressed by culturing the *M. smegmatis* strain mc^{2}4517 in autoinduction media supplemented with Tween 80. The media used are the same as those used for expression of proteins in *E. coli*, simplifying and reducing the cost of the switch to an alternative host. The methods have been applied to a set of *M. tuberculosis* proteins that form inclusion bodies when expressed in *E. coli*. We found that five of eight of these previously insoluble proteins become soluble when expressed in *M. smegmatis*, demonstrating that this is an efficient salvage strategy.

Keywords: Recombinant protein expression; Inclusion bodies; *Mycobacterium tuberculosis*; *Mycobacterium smegmatis*; Gateway; Autoinduction media

Introduction

One of the main obstacles encountered with recombinant protein over-expression in *Escherichia coli* is the production of insoluble inclusion bodies (IBs)[2]. In the context of Structural Genomics projects we and others have found insoluble protein expression to be a major bottleneck in obtaining suitable targets for structural studies [1–3]. This problem is particularly pronounced with *Mycobacterium tuberculosis* proteins. Recent statistics from the TB Structural Genomics Consortium show that across the 130 member laboratories,

* Corresponding author. Address: AgResearch Structural Biology Laboratory, School of Biological Sciences, Thomas Building, 3a Symonds Street, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. Fax: +64 0 9 3737414.
E-mail address: s.lott@auckland.ac.nz (J. Shaun Lott).
[1] These authors contributed equally to this work.

[2] *Abbreviations used:* IBs, inclusion bodies; ORFs, open reading frames; MALDI-TOF, matrix-assisted laser desorption/ionization reflection time-of-flight; LC-MS, liquid chromatography tandem mass spectrometry.

1046-5928/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2007.08.015

less than half of *M. tuberculosis* proteins expressed in *E. coli* are obtained in soluble form (see http://www.webtb.org). In our laboratory, researchers typically find that just three of each set of eight (37.5%) Gateway® cloned *M. tuberculosis* proteins are soluble when expressed in *E. coli* [2].

One salvage strategy to overcome the formation of IBs is using an expression host that is more closely related than *E. coli* to the organism from which the proteins are derived. However, such strategies are rarely used by Structural Genomics laboratories as the time taken to establish protocols for an alternative expression system, and to subclone targets into new expression vectors, is significant. The large numbers of *M. tuberculosis* proteins that express as IBs in *E. coli* represent a significant investment in cloning and expression studies. The development of a simple and rapid system to enable expression of *M. tuberculosis* proteins in an alternative host could rescue otherwise intractable proteins and increase success rates.

The non-pathogenic mycobacterium *M. smegmatis* is an attractive alternative host for parallel expression of *M. tuberculosis* proteins. It is relatively fast growing with a doubling time of approximately 3 h [4] and is likely to provide mycobacterium-specific chaperones which may aid correct folding of heterologously expressed proteins. Furthermore, we and others have reported successful expression of individual *M. tuberculosis* proteins using *M. smegmatis* as a host [5–7]. For example, the *M. tuberculosis* protein FGD1 was completely insoluble when expressed in *E. coli*, but when expressed in *M. smegmatis* using the shuttle vector pYUB1049, soluble protein was produced with a yield of 7 mg/L culture [5]. Based on our success with FGD1 we sought to expand our *M. smegmatis* expression capability and to assess its efficiency as a salvage strategy.

Here we describe a method for rapidly identifying *M. tuberculosis* proteins that express in soluble form in *M. smegmatis.* We have adapted the original *M. smegmatis* expression vector pYUB1049 to the Gateway® cloning system by the addition of *att* recombination recognition sequences. The resulting vector, designated pDESTsmg, is compatible with our in-house Gateway® methods for *E. coli* expression [2]. A target can be subcloned into pDESTsmg by a simple LR reaction using the Entry clone that was generated for *E. coli* expression, thereby removing the need to design new primers and re-clone target DNA. Proteins are expressed by culturing the *M. smegmatis* strain $mc^2$4517 in autoinduction media [8] supplemented with Tween 80, building on protocols recently developed in our laboratory [5]. Autoinduction media affords the advantage of decreased handling as cultures do not require monitoring of optical density and induction [8]. It lends itself well to parallel processing and is particularly useful in the case of *M. smegmatis* where cultures are grown for 4–5 days. The methods have been applied to a set of *M. tuberculosis* proteins that form IBs when expressed in *E. coli.* We found that >50% of these previously insoluble proteins are soluble when expressed in *M. smegmatis* using pDESTsmg, demonstrating that this is an efficient salvage strategy.

Materials and methods

Construction of pDESTsmg

A new hexahistidine (His_6) tag sequence was introduced into the *M. smegmatis* vector pYUB1049 (Fig. 1a; kindly provided by Professor W. R. Jacobs, Albert Einstein College of Medicine) to enable expression of an N-terminal His_6-tag following Gateway® conversion. The oligonucleotides YUB_his_1 (5′-CATGTCGCATCACCATCACCATCACCTCGAAT-CAC-3′) and YUB_his_2 (5′-CATGGTGATTCGAGGTGATGGTGATGGTGATGCGA-3′) were annealed and ligated into *Nco*I-cleaved pYUB1049 (Fig. 1b) to produce pYUB1049+his (Fig. 1c). Correct orientation of the YUB_his insert was confirmed by DNA sequencing.

pYUB1049+his was digested with the restriction enzymes *Nco*I and *Bam*HI to remove the original His_6-tag and thrombin cleavage site. The digested vector was treated with Klenow enzyme (Roche) to generate blunt ends and ligated with Gateway® cloning cassette B (Invitrogen) using T4 DNA ligase (Fig. 1d). The correct orientation of the Gateway® cassette was confirmed by digestion with the restriction enzyme *EcoRI* and DNA sequencing.

The insertion of Gateway® cassette B created a stop codon (TGA) prior to the *att*R1 recombination site (Fig. 1e). This was removed using whole plasmid inverse PCR using the Ex-Site Mutagenesis Kit (Stratagene) and the oligonucleotides YB4_Gate_mut_1 (5′-TCAACAAGTTTGTACAAAAAAGCTGAAGC-3′) and YB4_Gate_mut_2 (5′-TGGTGA-TTCGAGGTGATGGTGATGGTG-3′). Removal of the stop codon was confirmed by DNA sequencing and the resulting vector was designated pDESTsmg (Fig. 1f).

Cloning into pDESTsmg

Gateway® entry clones harbouring *M. tuberculosis* open reading frames (ORFs) were sub-cloned *via* an LR reaction into pDESTsmg for subsequent expression in *M. smegmatis*. The entry clones had been generated previously using published methods [2]. Briefly, nested PCR was performed on *M. tuberculosis* H37Rv genomic DNA to generate products in which the target ORFs are flanked by the *attB* sequences necessary for subsequent Gateway® cloning. BP reactions, where *attB*-flanked PCR products are recombined with an *attP* substrate (pDONR221; Invitrogen) using BP Clonase™ (Invitrogen), were performed to generate the entry clones. The entry clones, which contain *attL* sequences, were recombined with the *attR* substrate pDESTsmg in an LR reaction using LR Clonase™ (Invitrogen) to generate a *M. smegmatis* expression construct.

To verify that recombination with pDESTsmg had occurred the expression constructs were transformed into *E. coli* TOP10 cells and plated on low salt LB agar medium (yeast extract 5 g/L, tryptone 10 g/L, NaCl 5 g/L and agar