Received: 27 January 2020 | Revised: 22 September 2020 | Accepted: 29 September 2020

DOI: 10.1002/pro.3962

METHODS AND APPLICATIONS

THE PROTEIN SOCIETY WILEY

The pMy vector series: A versatile cloning platform for the recombinant production of mycobacterial proteins in *Mycobacterium smegmatis*

Katherine S. H. Beckham[1] | **Sonja Staack**[1] | **Matthias Wilmanns**[1,2] | **Annabel H. A. Parret**[1]

[1]Hamburg Unit, European Molecular Biology Laboratory, Hamburg, Germany

[2]University Hamburg Clinical Centre Hamburg-Eppendorf, Hamburg, Germany

Correspondence
Katherine S. H. Beckham, Hamburg Unit, European Molecular Biology Laboratory, Notkestrasse 85, 22607, Hamburg, Germany.
Email: kbeckham@embl-hamburg.de

Abstract

Structural and biophysical characterization of molecular mechanisms of disease-causing pathogens, such as *Mycobacterium tuberculosis*, often requires recombinant expression of large amounts highly pure protein. For the production of mycobacterial proteins, overexpression in the fast-growing and non-pathogenic species *Mycobacterium smegmatis* has several benefits over the standard *Escherichia coli* expression strains. However, unlike for *E. coli*, the range of expression vectors currently available is limited. Here we describe the development of the pMy vector series, a set of expression plasmids for recombinant production of single proteins and protein complexes in *M. smegmatis*. By incorporating an alternative selection marker, we show that these plasmids can also be used for co-expression studies. All vectors in the pMy vector series are available in the Addgene repository (www.addgene.com).

KEYWORDS

mycobacteria, *Mycobacterium smegmatis*, protein expression, recombinant proteins

1 | INTRODUCTION

The Gram-positive genus *Mycobacterium* includes several human pathogens, including *Mycobacterium tuberculosis* (*Mtb*). *Mtb* is listed by the World Health Organization as the leading cause of death from an infectious agent and led to 1.5 million deaths in 2019 alone.[1] The increase in multi-drug resistant strains of *Mtb* remains a public health crisis and the need for novel antibiotic therapies to treat *Mtb* is a priority. The growing importance of *Mtb* and other mycobacterial pathogens to human health has led to an intensive effort from several structural biology consortia to investigate the structure and function of mycobacterial proteins.[2,3] Despite these efforts, currently only approximately 15% of the *Mtb* proteome has been structurally characterized,[4] in part due to the challenge of *Mtb* protein production.

The production of large amounts of highly pure, properly folded and functional protein remains a bottle neck in the structural biology pipeline. *Escherichia coli* is typically the standard expression host for protein

Abbreviations: GFP, green fluorescent protein; His_6 tag, hexahistidine tag; hyg, hygromycin; IMAC, immobilized metal affinity chromatography; kan, kanamycin; SLiCE, seamless ligation cloning extract; TEV, tobacco etch virus.

production and there are a variety of modified strains available that are optimized for tackling challenging proteins.[5] However, studies have shown that for the production of mycobacterial proteins, expression in standard *E. coli* strains is only successful in one third of cases.[6,7] There are several factors that may limit the suitability of *E. coli* for the production of mycobacterial proteins, such as the mismatch in codon usage between mycobacterial genes and the *E. coli* translation machinery caused by the higher GC bias in *Mtb* genes.[8,9] In addition, the absence of key cofactors, post-translational modifications and chaperones in *E. coli* may further impede the production of mycobacterial proteins.[10] Owing to these deficiencies, many groups have turned to the fast-growing, non-pathogenic mycobacterial expression host, *Mycobacterium smegmatis*. The benefits of the use of *M. smegmatis* as an expression host, leading to an improved yield, solubility and functionality of purified proteins has been reported in several studies.[7,11,12]

Despite the advantages of mycobacterial protein production using *M. smegmatis*, the limited range of genetic tools compared to those available for *E. coli* has restricted the widespread use of this expression host. However, several groups continue to develop expression strains and optimize expression vectors with different features such as induction system, promoter strength and purification tags.[6,13–17] There are two principle *M. smegmatis* strains used for protein expression, *M. smegmatis* $mc^2$4517[17] and $mc^2$155 *groEL1ΔC*.[16] *M. smegmatis* $mc^2$4517 has been modified to allow the expression of T7-promoter based systems following the incorporation of the bacteriophage T7 RNA polymerase.[17] Several vector systems have been developed for use in this strain including, pYUB1062 and pYUB1049.[17] Further modified versions allow for a choice of N- or C- terminal hexahistidine (His_6) tag positioning (pYUB28b),[18] expression of a GFP fusion protein (pYUB1062-GFP)[19] or for the co-expression of two protein targets (pYUBDuet).[18] The $mc^2$155 *groEL1ΔC* strain has been modified to reduce the co-purification of GroEL1 chaperone protein following the deletion of its histidine-rich C-terminus[16] and can be used for expression of vectors carrying conditional promoter systems, such as acetamidase,[20] tetracycline[21] or arabinose.[22] The acetamidase promoter of *M. smegmatis* can be induced by the addition of acetamide, and several vector systems utilize this promoter to drive protein expression, including the pSD[15] and pMyNT/ pMyC vectors.[23] Several modified variants of these vectors exist, including the pMyCA vector, which contains a minimized acetamidase promoter.[13,14] Induction of the acetamidase promoter leads to a high level of protein expression, which in the case of toxic protein production may not be desirable. In *E. coli*, the arabinose-inducible promoter (P_{BAD}) enables tightly controlled and tunability of gene expression.[24] The arabinose promoter system is currently not widely used in *M. smegmatis,* but the tunability of this promoter prompted us to further explore its use for protein expression.

The aim of this work was to further expand the versatility of the pMyNT and pMyC vectors, which have been successfully used for both the production of single soluble proteins as well as protein complexes expressed from a single operon.[16,23,25] The pMyNT and pMyC vectors are shuttle vectors which can be propagated in *E. coli* cells for ease of manipulation, due to the presence of the *OriE* and the *OriM* origins of replication (*Ori*). These *Ori* are used for replication in *E. coli* and *M. smegmatis*, respectively.[23] Both vectors encode a His_6 tag at the N-terminus (pMyNT) or at the C-terminus (pMyC) and a hygromycin resistance marker for selection. Here we describe the modification of the pMyNT and pMyC vectors generating variants with an alternative selection marker and an arabinose-inducible promoter resulting in the "pMy vector series". Using fluorescent reporter proteins, we show that the pMy vectors can be used for the overexpression of a single protein and in combination for the production of multiple targets. In addition, we demonstrate the tunability of the P_{BAD} arabinose-based promoter, which may prove advantageous for the production of toxic proteins. The pMy vector series has been deposited with Addgene (www.addgene.com).

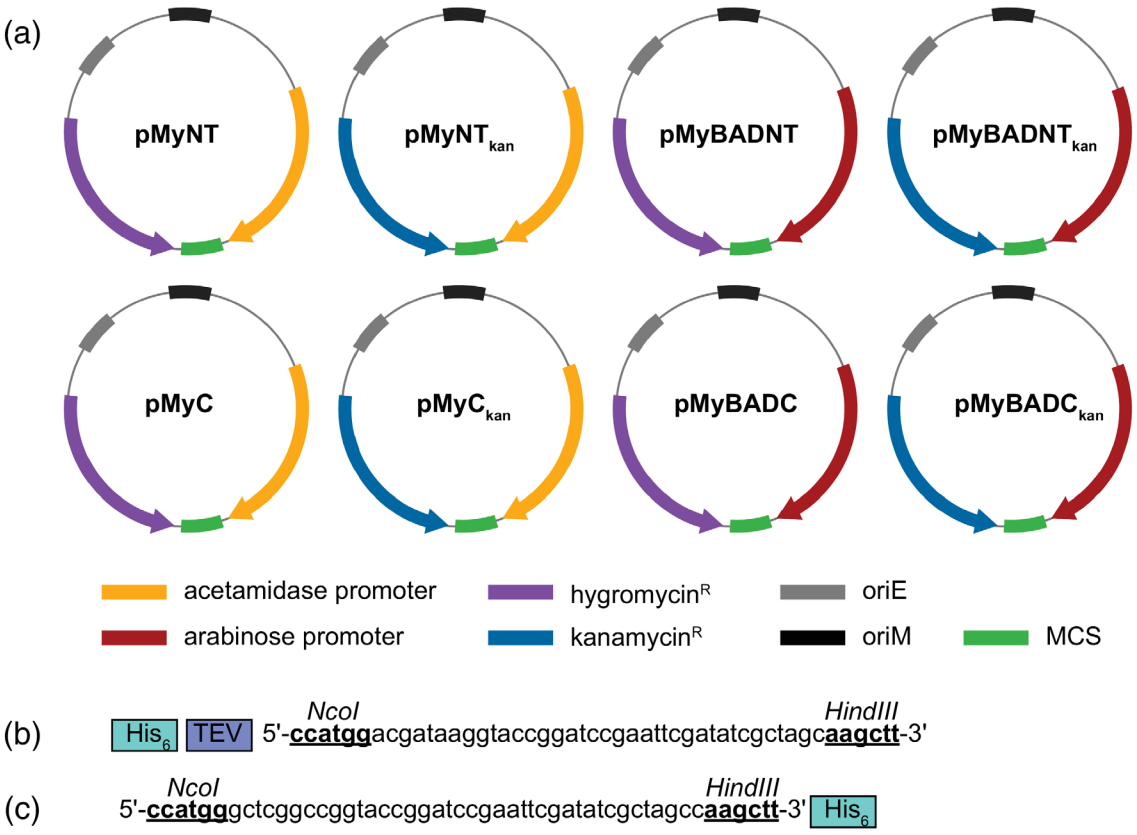
2 | RESULTS

2.1 | Construction of pMy vectors

To expand the existing repertoire of *M. smegmatis* expression vectors and create more tools for the production of mycobacterial vectors, we created variants of the existing pMyNT and pMyC vectors, which were previously generated by our group.[23] First we exchanged the existing acetamidase promoter present in both the pMyNT and pMyC for the arabinose-inducible promoter from the pBAD vector (P_{BAD}),[24] with the aim of producing a vector with tunable expression. Using Gibson cloning methods, the linearized fragments of the pMyC and pMyNT vectors without the acetamidase promoter were ligated with the P_{BAD} arabinose promoter producing the arabinose-inducible, hygromycin-resistant vectors with a N-terminal His_6 or C-terminal His_6 tag, pMyBADNT and pMyBADC, respectively. In addition, we extended the co-compatibility of the pMy vectors by including the *kanR* gene, which is widely used in other *M. smegmatis* vectors.[13] Utilizing Gibson cloning approaches again,

vector backbones of pMyNT, pMyC, pMyBADNT, and pMyBADC were amplified to omit the hygromycin resistance cassette, these backbone fragments were then ligated with the kanamycin resistance cassette, thus generating $pMyNT_{kan}$, $pMyC_{kan}$, $pMyBADNT_{kan}$, and $pMyBADC_{kan}$. An overview of the pMy vectors and their respective properties are outlined in Figure 1. All of the vectors produced in this study have been made available on Addgene (www.addgene.org) with their catalog numbers listed in Figure 1d.

As for the original pMyNT vector, all pMy vectors with N-terminal His_6 tags are preceded by a tobacco etch virus (TEV) cleavage site allowing the removal of the tag following immobilized metal affinity chromatography (IMAC). The pMy vectors derived from the original pMyC vector with a C-terminal His_6 tag do not include the TEV cleavage site due to the fact that following cleavage with TEV protease, five additional amino acids from the TEV recognition site remain,[26] comparable in length to the His_6 tag itself. The multiple cloning site (MCS) of



(d)

Name	Size (bp)	Fusion Tag	Promoter	Resistance	Addgene #	Source
pMyNT	6902	N-His6-TEV	Acetamidase	Hyg	42191	[23]
pMyC	6876	C-His6	Acetamidase	Hyg	42192	[23]
$pMyNT_{kan}$	6832	N-His6-TEV	Acetamidase	Kan	84693	This study
$pMyC_{kan}$	6811	C-His6	Acetamidase	Kan	84692	This study
pMyBADNT	5472	N-His6-TEV	araBAD	Hyg	84690	This study
pMyBADC	5474	C-His6	araBAD	Hyg	84688	This study
$pMyBADNT_{kan}$	5312	N-His6-TEV	araBAD	Kan	84691	This study
$pMyBADC_{kan}$	5409	C-His6	araBAD	Kan	84689	This study

FIGURE 1 The pMy vector series. (a) Overview of the features of the pMy vectors, highlighting the arrangement of the promoter and resistance genes (maps not shown to scale). (b) Multiple cloning site of the pMy vectors with an N-terminal hexahistidine tag (His_6) followed by a TEV cleavage site. (c) Multiple cloning site of the pMy vectors with a C-terminal hexahistidine tag (His_6). The unique restriction sites *Nco*I and *Hind*III are indicated. (d) Table summarizing the properties of the pMy vectors

the pMy vectors all contain the unique restriction sites *NcoI* and *HindIII* that can be used for linearizing the plasmids for ligation with a gene of interest using either restriction enzyme (RE) based cloning methods or ligation-independent cloning methods (Figure 1b,c). However, the limited number of unique restriction sites in the pMy vectors restricts the use of RE based cloning. Therefore, we primarily use ligation-independent cloning methods such as seamless ligation cloning methods (SliCE) or Gibson assembly methods for cloning genes into the pMy vectors. The recommended primer extension sequences for use with these methods for each of the pMy vectors are listed in Table S2.

2.2 | pMy vectors provide inducible protein expression in M. smegmatis

The level of protein expression from each of the pMy plasmid variants was evaluated using green fluorescent protein (GFP), which has been successfully used to monitor protein expression levels in other *M. smegmatis* vector systems.[19] The GFP gene was amplified using the primers listed in Table S3 and ligated into each of the pMy vectors using SliCE.[27] For protein production, we routinely use the *M. smegmatis* mc^{2}155 *groEL1ΔC*[16] strain that has been optimized for purification of proteins by IMAC methods, and therefore we tested the activity of the vectors in this strain. However as the pMyNT and pMyC vectors are compatible with other *M. smegmatis* strains,[28] it is likely that the new pMy variants produced in this study will also be compatible as the vector backbone is largely unchanged.

M. smegmatis cultures were grown to an OD_{600} of 1 in 7H9 expression medium before induction with either 1% acetamide or 1% arabinose. A concentration of 1% of the inducer molecule was chosen as higher concentrations of acetamide lead to an increase in cell aggregation reducing the accuracy of the fluorescence measurements. The amount of GFP produced in whole cells was measured using a plate reader following 18-hour induction with either arabinose or acetamide, as appropriate (Figure 2). The highest level of GFP expression was detected from the pMyNT and $pMyNT_{kan}$ vectors, which encode the acetamidase promoter and have the His_6 tag positioned at the N-terminus. In comparison to the pMyC and $pMyC_{kan}$ vectors (acetamidase promoter, C-terminal His_6 tag) the pMyNT and $pMyNT_{kan}$ vectors displayed approximately a three-fold higher level of GFP fluorescence. This difference was also observed for pMy vectors with arabinose inducible promoters when comparing GFP expression between the vectors with an N-terminal His_6 tag (pMyBADNT and $pMyBADNT_{kan}$) and the

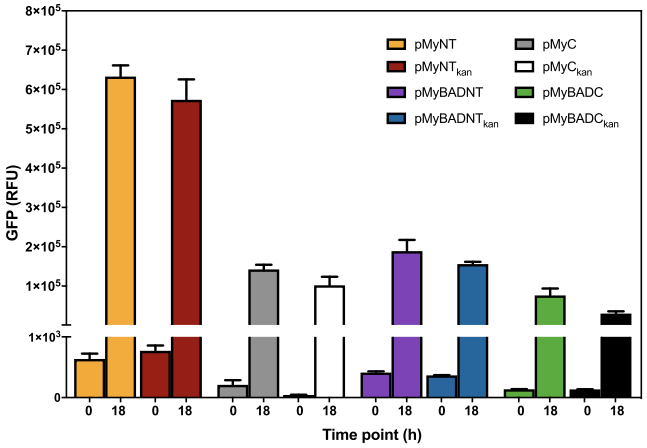


FIGURE 2 GFP expression using the pMy vectors in *M. smegmatis*. *M. smegmatis* cultures expressing pMy vectors encoding GFP2+ were induced at an OD_{600nm} of 1 with either 1% acetamide or arabinose, as appropriate. Determination of the GFP expression was calculated as relative fluorescence unit (RFU). All data were averaged from three independent samples of each time point. Samples were taken before and 24 hr after addition of inducer. Error bars depict standard deviation of three independent experiments

C-terminal His_6 tag (pMyBADC and $pMyBADC_{kan}$). This difference was significant for all vector variants ($p = .0001$) and suggests that the N-terminal position of the His_6 tag leads to more effective translation, which has been similarly observed in other systems.[29] When comparing the pMy vectors with the same affinity tag position (e.g., pMyNT vs pMyBADNT), the acetamidase-based vectors produce a significantly higher amount of GFP ($p < .001$). Additionally, there was no significant difference in the level of protein production between the hygromycin- and kanamycin-resistant variants, indicating that the choice of selection marker did not impact protein production.

As protein overexpression can be toxic to the host cell, an uninduced sample was included to monitor the level of background or "leaky" expression from the vectors. For all vectors the level of background expression was below 0.5% of the total amount of GFP being produced compared with the induced sample. Thus, even at high inducer concentrations both the acetamidase promoter and P_{BAD} promoter appear to be tightly regulated in *M. smegmatis* mc^{2}155 *groEL1ΔC*.

2.3 | pMyBAD vectors provide tunable protein expression

To investigate the tunability of the acetamidase (pMyNT) and P_{BAD} (pMyBADNT) promoters in *M. smegmatis,* we followed the expression of GFP over time using a range of

inducer concentrations (Figure 3). At the concentrations tested, induction of the acetamidase promoter leads to a rapid increase in GFP production, which does not appear to be dependent on the inducer concentration. While the highest concentration of acetamide used was 1%, decreasing the acetamide levels did not significantly reduce the level of GFP expression at the concentrations used in this study (Figure 3a). In contrast, increasing the concentration of arabinose proportionally increased the level of GFP expression from the pMyBADNT vector (Figure 3b). At the final 24 hr time point, the level of GFP produced from the different arabinose concentrations was significantly different ($p < .001$). Together these results indicate that P_{BAD} promoter is more tightly regulated than the acetamidase promoter in *M. smegmatis*. The tunability of the pBAD promoter in the pMy vectors could be exploited for the production of toxic proteins where unregulated levels of protein expression may lead to cell death.

2.4 | pMy constructs can be combined in co-expression studies

One of the aims of generating the pMy vector series with different antibiotic selection markers was to facilitate co-expression studies. To test whether a combination of pMy vectors could successfully express multiple proteins, we monitored the expression of GFP and mCHERRY simultaneously by using the corresponding excitation and emission wavelengths for each protein. mCHERRY was cloned into $pMyNT_{kan}$, $pMyBAD_{kan}$ using the primers listed in Table S3 with SliCE cloning methods as described above. Different combinations of pMy plasmids encoding either GFP or mCHERRY were co-transformed into *M. smegmatis* mc^2155 *groEL1ΔC* by electroporation and co-transformants were selected using hygromycin and kanamycin. To test the level of co-expression from two plasmids carrying the acetamidase promoter pMyNT-GFP and $pMyNT_{kan}$-mCHERRY were co-transformed (Figure 4a). Similarly, to test co-expression from two pMy plasmids carrying the arabinose promoter system pMyBAD-GFP was combined with $pMyBAD_{kan}$-mCHERRY (Figure 4b). Finally, pMyBAD-GFP was combined with $pMyNT_{kan}$-mCHERRY to test the co-expression from the two different protomer systems (Figure 4c). For all combinations fluorescent protein expression was monitored 18 hr following induction with 1% (v/v) acetamide and /or arabinose, as appropriate. For all vector combinations, the production of GFP and mCHERRY increased after induction showing that co-expression from two independent pMy vectors is possible. To compare the protein amounts produced during a co-expression experiment to the production from a single vector, the amount of GFP and mCHERRY produced during co-expression has been shown relative to the amount produced from the expression of the single protein from the corresponding vector. For example, the amount of GFP produced in the co-expression of pMyNT-GFP and $pMyNT_{kan}$-mCHERRY (Figure 4a) has been normalized to the amount of GFP produced by expressing pMyNT-GFP alone using the same conditions. Based on the normalized RFU readings the amount of GFP or mCHERRY produced in a co-expression experiment reduces by approximately 35%–55% compared to single expression. There was no significant difference between the level of GFP or mCHERRY

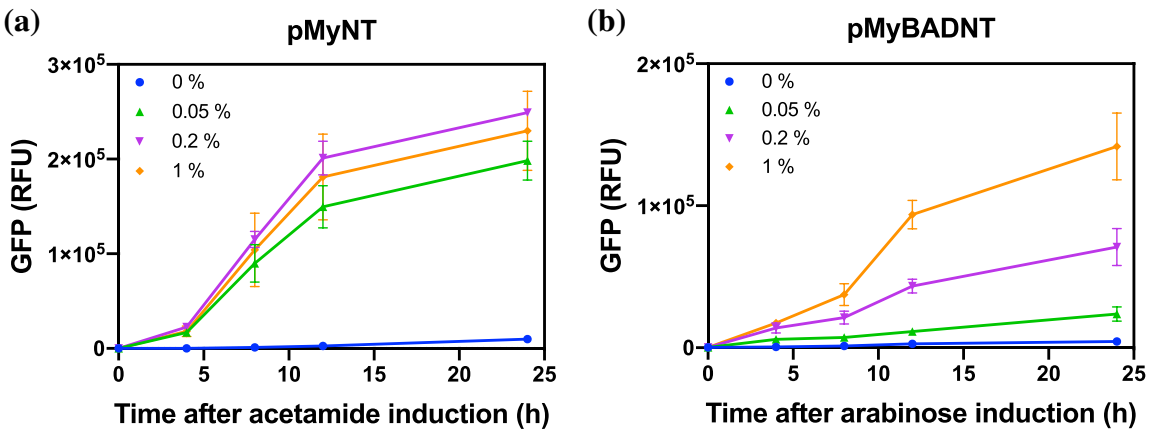


FIGURE 3 Time course of GFP expression in *M. smegmatis* following induction of the acetamidase (a) or arabinose promoters (b) with a range of inducer concentrations. Cultures of *M. smegmatis* transformed with either pMyNT-GFP (a) or pMyBADNT-GFP (b) were induced at an OD_{600nm} of 1 with acetamide or arabinose, respectively. The concentration (v/v) of inducer used varied between 0.05–1%, uninduced cultures (0%) indicate the level of unregulated background expression. GFP fluorescence was monitored at various time points after induction (0 hr). Error bars indicate the standard deviation of three independent experiments

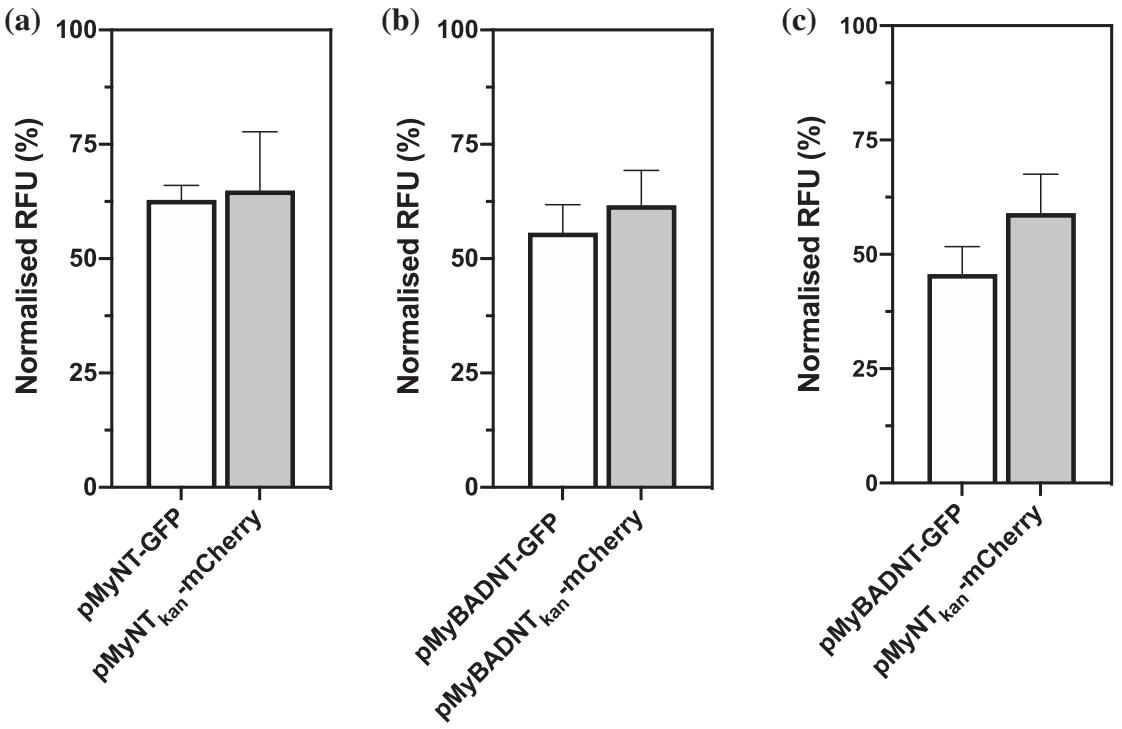


FIGURE 4 Co-expression of GFP and mCHERRY in *M. smegmatis* from the pMy vectors. *M. smegmatis* was co-transformed with different combinations of pMy vectors encoding GFP or mCHERRY fluorescent protein. Fluorescence of GFP and mCHERRY was measured 18 hr after the addition of the appropriate inducer to a final concentration of 1% (v/v). The mean fluorescence values have been normalized to the amount of GFP/mCHERRY produced following expression from a single vector under the same conditions, shown as relative fluorescence units (RFU). Error bars depict the standard deviation of three independent experiments. (a) Co-expression of two pMy vectors carrying an acetamidase promoter (pMyNT-GFP and pMyNT$_{kan}$-mCHERRY). (b) Co-expression of two pMy vectors carrying an arabinose promoter (pMyBADNT-GFP and pMyBADNT$_{kan}$-mCHERRY). (c) Combination of one pMy vector carrying an arabinose promoter (pMyBAD-GFP) and one pMy vector with an acetamidase promoter (pMyNT$_{kan}$-mCHERRY)

produced by the different vector combinations, indicating that none of the pMy plasmids were expressed preferentially over the other. In summary, the expression of the two fluorescent proteins from independent vectors demonstrates that pMy vectors encoding different antibiotic resistance markers and promoter systems can be combined for co-expression studies.

3 | DISCUSSION

The production of mycobacterial proteins for structural, functional and biochemical studies remains an important step in the drug discovery pipeline. The production of mycobacterial proteins in *M. smegmatis* is becoming more common owing to the advantages of using a more native expression host over the traditional *E. coli* strains.[7,11,12,25] The aim of this work was to further expand the tools available for recombinant protein expression in *M. smegmatis*.

The pMy vector series derives from the pMyNT and pMyC vectors that have been used in the mycobacterial field over the past decade.[23,30,31] Induction of protein expression from the acetamidase promoter is an established system in *M. smegmatis* and leads to high levels of target protein expression.[6] However, when producing proteins that are toxic to the cell such as membrane proteins, toxins or DNA binding proteins it can often be advantageous to regulate the level of protein overexpression. In *E. coli* the arabinose inducible promoter system, which includes the pBAD promoter and the AraC regulator has been successfully exploited to tightly regulate protein expression and its activity in *M. smegmatis* has also been demonstrated.[22] It is for this reason that we introduced the arabinose promoter system into the pMy vectors. By comparing the induction of GFP expression from both the acetamidase promoter (pMyNT) and the arabinose-inducible promoter (pMyBADNT), we showed that induction with different concentrations of arabinose is correlated with the level of GFP expression. This is in contrast to induction of the acetamidase promoter, where all concentrations tested resulted in a similar level of GFP fluorescence. The tuneability of the arabinose promoter in *M. smegmatis* could provide a useful alternative to the acetamidase

promoter for the production of toxic proteins and allow better control of the production of protein complexes where the correct stoichiometry of the different components may be essential for solubility.

To test the overall performance of pMy vectors we examined the level of GFP production from each of the vectors in the pMy series following induction with either acetamide, arabinose, or no inducer as a control. As expected, the acetamidase-based vectors (pMyNT, $pMyNT_{kan}$, pMyC and $pMyC_{kan}$) produced higher amounts of GFP compared to the arabinose-inducible vectors (pMyBADNT, $pMyBADMNT_{kan}$, pMyBADC and $pMyBADC_{kan}$). The significant reduction in the amount of GFP produced from the C-terminally tagged vectors was unexpected, but has been observed for other expression systems which show an increased production of target proteins with N-terminal fusions.[29] Recent work by Vergara *et al*., produced a modified pMyC variant with a reduced acetamidase regulon which increased protein production,[14] although whether this change resulted in a similar level of expression as from the pMyNT vectors was not tested in this work.

Using high concentrations of inducer, the pMyBAD vectors produced lower quantities of GFP compared to the acetamide-based vectors. This correlates with the previous reports of the low level of activity of the arabinose-inducible promoter system in *M. smegmatis*.[22] However, owing to the tunability of this promoter we propose that the pMyBAD vectors would be beneficial in cases where the target protein is toxic, as even low concentrations of acetamide leads to high levels of protein expression. Under these expression conditions, both the acetamidase- and arabinose-inducible systems have a low level of background expression following comparison of the induced to the uninduced samples. The low level of "leaky expression" contrasts with previous observations on the acetamide inducible systems using, for example, the pYUB1062[19] vectors in the *M. smegmatis* strain $mc^2$4517. This may reflect differences in the plasmid, *M. smegmatis* strain or growth media conditions.

In addition to creating expression vectors where the level of recombinant protein production can be as tightly regulated as for the pMyBAD vectors, we also aimed to facilitate co-expression studies by introducing an alternative selection marker, kanamycin. The kanamycin-resistant pMy variants show the same expression levels as the hygromycin-resistant variants, and as kanamycin is significantly cheaper than hygromycin it may be more accessible to some laboratories. Here we show that the co-expression of GFP and mCHERRY from the combination of pMy vectors with different promoter systems is possible. The combination of the tightly regulated pBAD promoter with the highly expressed acetamidase promoter may provide a useful strategy for the expression of toxin-antitoxin systems where the cytotoxic protein component be expressed at a lower level than is anti-toxin counterpart,[32] however this has not been tested as part of this work.

In summary, this work contributes additional tools for the production of recombinant proteins in *M. smegmatis*. As more tools become available, we hope to see further development of specific *M. smegmatis* expression strains for the production of recombinant proteins,[16] as for example the C41 strain for membrane protein production in *E. coli* (Lucigen). The pMy vector series could be further modified to include different purification tags to expand the applications of these vectors for the production of multi-protein complexes. The production of mycobacterial proteins for biochemical, biophysical and structural studies remains a key step in the search for novel anti-mycobacterial therapies. We anticipate that the expansion of the pMy vector series will be a useful tool for the community.

4 | MATERIALS AND METHODS

4.1 | Reagents, bacterial strains, and growth conditions

All cloning and plasmid propagation were conducted in *E. coli* DH5α-$T1^R$ (Life Technologies) using standard protocols. *E. coli* was transformed by conventional heat shock transformation and grown in Lennox Broth (LB) or LB agar plates (Carl Roth). The seamless ligation cloning extract (SLiCE) was produced from the *E. coli* DH10B-PPY strain.[27] In liquid cultures *M. smegmatis* $mc^2$155 *groEL1ΔC*[16] was grown in Middlebrook 7H9 medium (BD Biosciences) supplemented with 0.2% (w/v) glucose (Carl Roth), 342 mM NaCl), 0.05% (v/v) Tween-80 (Carl Roth) and 0.2% (v/v) glycerol (Carl Roth). Alternatively, the solid growth media used was 7H10 agar (Sigma-Aldrich) supplemented with 10% albumin–dextrose saline (ADS: 5% (w/v) BSA cold ethanol fraction, pH 5.2, ≥96% (Sigma-Aldrich), 2% (w/v) glucose (Carl Roth), 342 mM NaCl), 0.05% (v/v) Tween-80 (Carl Roth) and 0.2% (v/v) glycerol (Carl Roth). All bacterial strains were grown at 37°C. Where required the growth media was supplemented with 94 μM hygromycin B (Carl Roth) or 35 μM kanamycin (Sigma-Aldrich).

4.2 | Construction of pMy vectors

The expanded pMy vector series is based on the previously described pMyNT and pMyC vectors[23] which are derived from pSD31.[15] The sequences of the primers used in the construction of the pMy vectors are listed in

Table S1. PCR amplification was performed with Q5® High-Fidelity DNA Polymerase according to the manufacturer's instructions (New England Biolabs) and DNA fragments were purified with the Wizard SV Gel and PCR Clean-up System (Promega).

The different pMy vectors were created using Gibson assembly methods to generate variants encoding kanamycin resistance or the P_{BAD} promoter. The acetamidase promoter composed of the *amiC*, *amiA*, *amiD* and *amiS* genes encoded in in the pMyNT and pMyC vectors was replaced by the arabinose promoter (*araC* gene) from the pBAD/His (Thermo Fischer) vector using Gibson cloning producing pMyBADNT and pMyBADC. The parent plasmids pMyNT and pMyC were linearized by PCR to generate the "vector insert" using the primers listed in Table S1. The P_{BAD} promoter was amplified by PCR from the P_{BAD}/His (Thermo Fischer) using the corresponding "insert" primers listed in Table S1. The PCR products were purified and ligated using Gibson Assembly Master Mix (New England Biolabs) as per the manufacturer's instructions, using the recommended vector concentration of 50 ng and a two-fold molar excess of insert. Ligation mixtures were transformed into *E. coli* DH5α-T1^R (Life Technologies) by heat shock transformation and plated onto LB-agar plates containing hygromycin and incubated at 37°C overnight. The resultant plasmids, pMyBADNT and pMyBADC were sequence verified (Eurofins Genomics).

To make kanamycin resistant variants of the pMyNT, pMyC, pMyBADNT, and pMyBADC the hygromycin B (*hygR*) resistance marker was exchanged for the for the *kanR* marker. The *kanR* gene was amplified by PCR using the pMV306 vector[33] as template using the "insert" primers listed in Table S1 for each of the corresponding pMy vectors. Linear fragments of the pMyNT, pMyC, pMyBADNT, and pMyBADC vectors without the *hygR* gene were generated by PCR amplification using the "vector" primers listed in Table S1. Each of the vector backbone fragments was ligated with the complementary *kanR* product using the Gibson Assembly Master Mix (New England Biolabs) as described above. These reactions resulted in $pMyNT_{kan}$, $pMyC_{kan}$, $pMyBADNT_{kan}$, and $pMyBADC_{kan}$ plasmids which were all sequence verified (Eurofins Genomics). The pMy plasmids generated in this study, $pMyNT_{kan}$, $pMyC_{kan}$, $pMyBADNT_{kan}$, $pMyBADC_{kan,}$ pMyBADNT, and $pMyBADC_{kan}$ have been deposited on Addgene (www.addgene.com).

4.3 | Construction of pMy vectors encoding fluorescent reporter proteins

Expression constructs were generated using SliCE cloning methods with the SliCE ligation mix prepared as described by Zhang et al.[27] Vectors were linearized by restriction enzyme digestion with *Nco*I/HindIII followed by dephosphorylation using Antarctic phosphatase (New England Biolabs). Genes encoding the green fluorescent protein (GFPm2+) and red fluorescent protein (mCHERRY3) were synthesized and provided by Genscript using the sequences shown in Figure S1. The genes were amplified by PCR using Q5® High-Fidelity DNA Polymerase (New England Biolabs). All primers used for the generation of plasmids used in this study are listed in Table S3. DNA fragments were purified with the Wizard SV Gel and PCR Clean-up System (Promega). SliCE cloning reactions were performed using 50 ng linearized vector with a 5:1 (insert:vector) molar excess of purified insert. Ligation mixtures were transformed to *E. coli* DH5α-T1^R and transformants were selected on LB plates containing the appropriate antibiotic. Plasmid DNA was prepared using QIAprep Spin Miniprep kit (Qiagen) and sequence-verified with vector-specific primers (AP-328, 5'-CGCAGTTGTTCTCGCATACC-3′ and pMyNT-rev, 5′- TGGATCTCTCCGGCTTCAC-3′) before transformation to *M. smegmatis* mc^2155 *groEL1ΔC*. Electrocompetent *M. smegmatis* mc^2155 *groEL1ΔC* were prepared as previously described.[16]

4.4 | Monitoring the expression of fluorescent proteins expressed from pMy vectors in M. smegmatis

To monitor protein expression from the pMy vectors in *M. smegmatis*, the fluorescent reporter proteins GFP2+ or mCHERRY were used. Expression constructs outlined in Table S3 were transformed into electrocompetent *M. smegmatis* and selected on 7H10 agar supplemented with either kanamycin or hygromycin, as appropriate. In the case of co-expression studies, both plasmids were simultaneously transformed by electroporation into *M. smegmatis* and plated onto 7H10 agar plates supplemented with both kanamycin and hygromycin, (35 μM or 94 μM, respectively). Transformants were confirmed using colony PCR.

M. smegmatis starter cultures were cultivated from freshly streaked colonies or glycerol stocks for grown for 3 days at 37°C with orbital shaking at 120 rpm. For small-scale expression studies, a 1% volume of a starter culture was used to inoculate 50 ml of 7H9 expression medium. Cultures were grown to an OD_{600} of 1 and induced with varying concentrations of arabinose or acetamidase, as indicated. The first time point (0) was taken at the point of induction and several time points over 24 hr after induction, at each time point 200 μl of each culture was transferred from the 50 ml culture to a black FLUOTRAC flat bottomed 96 well plate (Greiner Bio-One). The

fluorescence of GFP or mCHERRY, or both in the case of co-expression studies, was measured using a TECAN infinite M1000 plate reader. GFP fluorescence was measured at an excitation wavelength of 490 nm and an emission wavelength of 509 nm. For mCHERRY detection, an excitation wavelength of 587 nm and an emission wavelength of 610 nm was used. A gain value of 100 was used for all measurements. Statistical analysis of expression levels was performed using a Student's *t*-test. For the analysis of the co-expression of GFP and RFP, the values have been normalized to the 18 hr timepoint level of fluorescence (RFU) for the respective reporter.

ACKNOWLEDGMENTS

K. S. H. Beckham was supported by a postdoctoral fellowship from the EMBL Interdisciplinary Postdoctoral (EIPOD) programme under Marie Curie Actions. This work was in part supported by a Joachim Herz Stiftung "Add-On Fellowship for Interdisciplinary Science" (awarded to K. S. H. B.) and by the Joachim-Herz-Stiftung Hamburg via the project Infectophysics. We would like to thank Dr Nabil Hanna for his contribution during the initial phase of the project. Open access funding enabled and organized by Projekt DEAL.

AUTHOR CONTRIBUTIONS

Katherine S. H. Beckham: Conceptualization; formal analysis; funding acquisition; investigation; methodology; supervision; writing-original draft; writing-review and editing. **Sonja Staack:** Investigation; methodology. **Matthias Wilmanns:** Conceptualization; funding acquisition; resources; supervision; writing-original draft; writing-review and editing. **Annabel H. A. Parret:** Conceptualization; methodology; project administration; supervision; writing-original draft; writing-review and editing.

ORCID

Katherine S. H. Beckham https://orcid.org/0000-0002-4929-6119
Matthias Wilmanns https://orcid.org/0000-0002-4643-5435
Annabel H. A. Parret https://orcid.org/0000-0003-0635-8890

REFERENCES

1. Global Tuberculosis Report 2019. Geneva, Switzerland: World Health Organization; 2019:1–297.
2. Terwilliger T, Park M, Waldo G, et al. The TB structural genomics consortium: A resource for *Mycobacterium tuberculosis* biology. Tuberculosis. 2003;83:223–249.
3. Holton SJ, Weiss MS, Tucker PA, Wilmanns M. Structure-based approaches to drug discovery against tuberculosis. Curr Protein Pept Sci. 2007;8:365–375.
4. RCSB Protein Data Bank. Available from: www.rcsb.org
5. Rosano GL, Morales ES, Ceccarelli EA. New tools for recombinant protein production in *Escherichia coli*: A 5-year update. Protein Sci. 2019;28:1412–1422.
6. Bashiri G, Baker EN. Production of recombinant proteins in mycobacterium smegmatis for structural and functional studies. Protein Sci. 2015;24:1–10.
7. Arbing MA, Chan S, Harris L, et al. Heterologous expression of mycobacterial Esx complexes in *Escherichia coli* for structural studies is facilitated by the use of maltose binding protein fusions. PLoS One. 2013;8:e81753.
8. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: Advances and challenges. Front Microbiol. 2014;5:1–17.
9. Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998;393:537–544.
10. Gasser B, Saloheimo M, Rinas U, et al. Protein folding and conformational stress in microbial cells producing recombinant proteins: A host comparative overview. Microb Cell Fact. 2008;7:1–18.
11. Bashiri G, Perkowski EF, Turner AP, Feltcher ME, Braunstein M, Baker EN. Tat–dependent translocation of an F420–binding protein of *Mycobacterium tuberculosis* Nigou J, sudarytojas. PLoS One. 2012;7:e45003.
12. Ahangar MS, Furze CM, Guy CS, et al. Structural and functional determination of homologs of the *Mycobacterium tuberculosis* N -acetylglucosamine-6-phosphate deacetylase (NagA). J Biol Chem. 2018;293:9770–9783.
13. Arnold FM, Hohl M, Remm S, et al. A uniform cloning platform for mycobacterial genetics and protein production. Sci Rep. 2018;8:9539.
14. Magaña Vergara C, Kallenberg CJL, Rogasch M, Hübner CG, Song YH. A versatile vector for mycobacterial protein production with a functional minimized acetamidase regulon. Protein Sci. 2017;26:2302–2311.
15. Daugelat S, Kowall J, Mattow J, et al. The RD1 proteins of *Mycobacterium tuberculosis*: Expression in mycobacterium smegmatis and biochemical characterization. Microbes Infect. 2003;5:1082–1095.
16. Noens EE, Williams C, Anandhakrishnan M, Poulsen C, Ehebauer MT, Wilmanns M. Improved mycobacterial protein production using a mycobacterium smegmatis groEL1ΔCexpression strain. BMC Biotechnol. 2011;11:27.
17. Wang F, Jain P, Gulten G, et al. *Mycobacterium tuberculosis* dihydrofolate reductase is not a target relevant to the antitubercular activity of isoniazid. Antimicrob Agents Chemother. 2010;54:3776–3782.
18. Bashiri G, Rehan AM, Greenwood DR, Dickson JMJ, Baker EN. Metabolic engineering of cofactor F420 production in mycobacterium smegmatis Pastore a, sudarytojas. PLoS One. 2010;5:e15803.
19. Radhakrishnan A, Furze CM, Ahangar MS, Fullam E. A GFP-strategy for efficient recombinant protein overexpression and purification in mycobacterium smegmatis. RSC Adv. 2018;8:33087–33095.
20. Triccas JA, Parish T, Britton WJ, Gicquel B. An inducible expression system permitting the effcient purifi cation of a recombinant antigen from mycobacterium smegmatis. FEMS Microbiol Lett. 1998;167:151–156.

21. Ehrt S, Guo XV, Hickey CM, et al. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. Nucleic Acids Res. 2005;33:1–11.
22. Carroll P, Brown AC, Hartridge AR, Parish T. Expression of *Mycobacterium tuberculosis* Rv1991c using an arabinose-inducible promoter demonstrates its role as a toxin. FEMS Microbiol Lett. 2007;274:73–82.
23. Poulsen C, Holton S, Geerlof A, Wilmanns M, Song Y. Stoichiometric protein complex formation and over-expression using the prokaryotic native operon structure. FEBS Lett. 2010;584:669–674.
24. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol. 1995;177:4121–4130.
25. Milewski MC, Broger T, Kirkpatrick J, et al. A standardized production pipeline for high profile targets from *Mycobacterium tuberculosis*. Proteomics Clin Appl. 2016;10:1049–1057.
26. Stevens RC. Design of high-throughput methods of protein production for structural biology. Structure. 2000;8:R177–R185.
27. Zhang Y, Werling U, Edelmann W. SLiCE: A novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res. 2012;40:e55.
28. Poulsen C, Akhter Y, Jeon AH, et al. Proteome-wide identification of mycobacterial pupylation targets. Mol Syst Biol. 2010;6:386.
29. Costa S, Almeida A, Castro A, Domingues L. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: The novel Fh8 system. Front Microbiol. 2014;5:1–20.
30. Delorme V, Diomandé SV, Dedieu L, et al. MmPPOX inhibits *Mycobacterium tuberculosis* lipolytic enzymes belonging to the hormone-sensitive lipase family and alters mycobacterial growth. PLoS One. 2012;7:e46493.
31. Lou Y, Rybniker J, Sala C, Cole ST. EspC forms a filamentous structure in the cell envelope of mycobacterium tuberculosis and impacts ESX-1 secretion. Mol Microbiol. 2017;103:26–38.
32. Freire DM, Gutierrez C, Garza-Garcia A, et al. An NAD+ phosphorylase toxin triggers *Mycobacterium tuberculosis* cell death. Mol Cell. 2019;73:1282–1291.
33. Andreu N, Zelmer A, Fletcher T, et al. Optimisation of bioluminescent reporters for use with mycobacteria. PLoS One. 2010;5:e10777.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Beckham KSH, Staack S, Wilmanns M, Parret AHA. The pMy vector series: A versatile cloning platform for the recombinant production of mycobacterial proteins in *Mycobacterium smegmatis*. *Protein Science*. 2020;29:2528–2537. https://doi.org/10.1002/pro.3962