Published in final edited form as:

Protein Expr Purif. 2016 July ; 123: 75–82. doi:10.1016/j.pep.2016.04.003.

Optimized expression and purification of biophysical quantities of the Lac repressor and Lac repressor regulatory domain

Matthew A. Stetz, Marie V. Carter, and A. Joshua Wand*

Johnson Research Foundation and Department of Biochemistry & Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104 USA

Abstract

The recombinant production of Lac repressor (LacI) in *E. coli* is complicated by its ubiquitous use as a regulatory element in commercially-available expression vectors and host strains. While LacI-regulated expression systems are often used to produce recombinant LacI, the product can be heterogeneous and unsuitable for some studies. Alternative approaches include using unregulated vectors which typically suffer from low yield or vectors with promoters induced by metabolically active sugars which can dilute isotope labels necessary for certain biophysical studies. Here, an optimized expression system and isolation protocol for producing various constructs of LacI is introduced which eliminates these complications. The expression vector is an adaptation of the pASK backbone wherein expression of the *lacI* gene is regulated by an anhydrotetracycline inducible *tetA* promoter and the host strain lacks the *lacI* gene. Typical yields in highly deuterated minimal medium are nearly 2-fold greater than those previously reported. Notably, the new expression system is also able to produce the isolated regulatory domain of LacI without co-expression of the full-length protein and without any defects in cell viability, eliminating the inconvenient requirement for strict monitoring of cell densities during pre-culturing. Typical yields in highly deuterated minimal medium are significantly greater than those previously reported. Characterization by solution NMR shows that LacI constructs produced using this expression system are highly homogenous and functionally active.

Keywords

Lac repressor; NMR spectroscopy; *tetA* promoter; protein deuteration

1. Introduction

The Lac repressor (LacI) is a transcription factor that regulates expression of the *lac* operon in bacteria. The operon consists of three genes which encode proteins that either transport or metabolize lactose. LacI represses transcription of the operon by binding with very high affinity to a specific operator sequence located upstream of the genes but downstream of the promoter. When lactose enters the cell, it is converted to allolactose that can bind

*Correspondence to: Professor A. Joshua Wand, Department of Biochemistry & Biophysics, University of Pennsylvania Perelman School of Medicine, 905 Stellar-Chance Laboratories, 422 Curie Blvd, Philadelphia, Pennsylvania 19104-6059, telephone: 215-573-7288, facsimile: 215-573-7290, ; Email: wand@upenn.edu

specifically to LacI. Allolactose binding induces a change in LacI such that its affinity for the operator drops by over three orders of magnitude [1]. In this “induced” state, LacI is easily outcompeted off the operator, resulting in transcription initiation. For decades, LacI has been used as a canonical model for protein-DNA interactions and so-called “molecular switches” [2]. However, the mechanistic details that underlie the transition from the repressed state to the induced state have yet to be fully elucidated. Early, low-resolution crystallographic studies showed that LacI is a modular protein that tetramerizes into an unusual V-shaped dimer of dimers [3]. Each dimer binds one individual operator sequence. Monomers consist of a DNA-binding domain (residues 1-45), an extended linker region commonly referred to as the “hinge” (residues 46-62), a regulatory domain (residues 63-331) which contains a sugar-binding pocket and dimerization interface, and a helical tetramerization domain (residues 332-360). The sugar-binding pocket is located approximately 40 Å from the DNA-binding domain, suggesting that induction is as an allosteric process. While a complete structure of LacI bound to operator was solved, only a partial structure of LacI bound to the synthetic allolactose analog, isopropyl β-D-1-thiogalactopyranoside (IPTG), could be obtained. The IPTG-bound state of LacI was crystallized in the absence of operator and no electron density was observed for the DNA-binding domain or hinge. The low resolution of these structures impeded an atomic-level delineation of the changes in the regulatory domain that accompany induction.

In order to obtain higher-resolution structural data, a dimeric construct of LacI, which we will refer to as LacI*, was crystallized [4,5]. LacI dimers can be formed by simply deleting the tetramerization domain and these dimers are functionally equivalent to the tetramer [6,7]. Similar to previous structural studies, a complete structure of LacI* bound to operator was solved and only a partial structure of LacI* bound to IPTG in the absence of operator could be obtained. Though these structures provided critical, atomic-level insight into how the regulatory domain responds to inducer binding, the lack of electron density for the DNA-binding domain left the allosteric mechanism unexplained.

In principle, solution NMR could elucidate the structural properties of LacI unobtainable by crystallography and potentially illuminate the allosteric mechanism of induction. However, the high molecular weight of both the tetrameric and dimeric form of LacI (~154 kDa and ~71 kDa, respectively) has limited solution NMR studies to the isolated DNA-binding domain (~7 kDa) [8-10] and an engineered dimeric form of the DNA-binding domain (~14 kDa) [11-15]. While studies of the isolated DNA-binding domain have provided key information regarding the structural details underlying operator recognition, no insight into allostery or induction could be obtained directly.

Recent advances in solution NMR methodology have made characterization of high molecular weight proteins feasible [16,17]. In order to properly exploit these advances, proteins need to be prepared in highly deuterated conditions (minimal medium, typically > 90% D₂O v/v). Accordingly, high levels of expression are desired in order to minimize sample preparation costs. Moreover, NMR typically requires highly homogeneous sample preparations for accurate characterization. Unfortunately, LacI is utilized widely in commercially-available expression systems (in both vectors and host strains) and using such systems to express LacI itself usually results in sample heterogeneity via heterodimerization

of recombinant LacI and regulatory LacI. This issue becomes acute when the production of alternative constructs of LacI is desired. Additionally, inducing LacI-regulated expression vectors requires the introduction of high concentrations of IPTG into the growth medium which co-purifies with the recombinant product [18]. Alternative approaches that avoid contamination include using constitutive or unregulated expression systems and vectors with promoters induced by other sugars such as arabinose [19] (pBAD, ThermoFisher Scientific/Invitrogen) or rhamnose [20] (pRham, Lucigen). Each of these alternatives has its own associated limitations here. Though constitutive expression is widely used to express LacI in highly rich media (e.g. 2xYT), typical biochemical studies require 12 L of culture [21]. An unregulated expression system was recently introduced for expressing a thermostable mutant of LacI* in highly deuterated minimal medium which utilized a vector featuring a T7 promoter and the BL21(DE3) host strain [18]. This vector was a pET derivative which lacked the *lacI* gene and *lac* operator. To avoid adding IPTG, expression levels of T7 RNA polymerase were limited to the inherent leakiness of its *lacUV5* promoter. While homogenous LacI* could be obtained using this expression system, yields from growths in highly deuterated minimal medium were relatively low. Arabinose-inducible and rhamnose-inducible promoters require high concentrations of sugar (1-2 g/L) which are metabolized by *E. coli* once glucose reserves are depleted. This raises issues with dilution of ¹³C isotope via glucose that are required for a variety of NMR studies, particularly the critical experiments necessary for chemical shift assignment. Increasing total glucose content in the growth medium is often not economically feasible due to the relatively high cost of isotopically labeled glucose. Moreover, both pBAD and pRham promoters are repressed by glucose [20,22].

Here a novel vector/host strain combination for producing recombinant LacI* is evaluated along with a facile isolation procedure. Though the focus of this report is on producing the dimeric construct of LacI, the methods presented are also applicable to producing the tetrameric protein. The vector is a derivative of the pASK backbone [23] which employs an anhydrotetracycline-inducible promoter regulated by the Tetracycline repressor (TetR). The host is the *lacI*-free BLIM strain [24]. No copies of the *lacI* gene are present in the vector or host strain which eliminates the need to add IPTG to the growth medium and ensures homogenous preparations of LacI*. Additionally, this system produces LacI* in yields nearly 2-fold greater than those previously reported for unregulated expression from a T7 promoter. An added benefit of this expression system is its ability to produce the isolated regulatory domain of LacI* which was previously reported to be toxic [18]. Past attempts to recombinantly produce the isolated regulatory domain necessitated co-expression with full-length LacI and strict monitoring of cell densities during pre-culturing. It is shown that the pASK vector/BLIM host strain expression system produces isolated regulatory domain with no noticeable growth defects or restrictions. Total yields of isolated regulatory domain are about 2-fold greater than those previously reported and are comparable to the yields obtained for full length LacI*.

2. Materials and Methods

2.1 Expression vector construction

An expression vector containing the gene encoding LacI* under control of an inducible *tetA* promoter regulated by TetR was constructed. The gene encoding wild type *E. coli* LacI* (residues 1-331 of LacI, 109A isoform from strain K-12, Swiss-Prot entry: P03023) was amplified by PCR from an arabinose-inducible vector (pBAD Gateway, ThermoFisher Scientific/Invitrogen) containing the *lacI** gene in its expression cassette (gift from Prof. Mitchell Lewis, University of Pennsylvania). A Tobacco Etch Virus (TEV) protease cleavage site and EcoRI restriction site were added to the 5' end of the sequence and an XhoI restriction site was added to the 3' end of the sequence by primer extension. The gel-purified PCR product was ligated into a high copy number, linearized pGEM-T Easy vector (Promega) and transformed into XL1-Blue cells (Agilent). Cells harboring vectors which contained the insert were identified by blue-white screening by plating cells on X-gal. Vectors were amplified in rich media and subsequently isolated by miniprep. Vectors were then digested using EcoRI and XhoI restriction enzymes (NEB). The insert was isolated by gel purification and then subcloned into a pASK-IBA35plus vector backbone (IBA) using the corresponding restriction sites. The pASK vector backbone confers ampicillin resistance. Ligated vectors were transformed into *E. coli* DH5 α cells (ThermoFisher Scientific/Invitrogen), amplified in rich media, and isolated by miniprep. Ligation of the correct insert was confirmed by DNA sequencing. A vector harboring the gene for the isolated regulatory domain of LacI* (residues 60-331) was produced using identical procedures. Site-directed mutagenesis was performed using the QuikChange II kit (Agilent) according to the manufacturer's protocol.

Recombinant LacI* constructs expressed from this vector possess an N-terminal 6x-His tag and TEV-protease cleavage site. Following treatment with TEV-protease, a single non-native glycine residue is left at the N-terminus. For both constructs full length LacI* and LacI* regulatory domain, the first few N-terminal residues are unstructured. The additional glycine is therefore not expected to affect functionality.

2.2 Host strain

The pASK vector utilizes an *E. coli* promoter and so there are no restrictions placed on the choice of host strain. The BLIM strain [24] was chosen as the host strain for the pASK vector. BLIM cells are derived from the BL26 Blue strain (Novagen), a derivative of BL21 cells from which the *lac* operon has been deleted and an F' episome carrying *lacI* under control of the *lacP* promoter has been introduced. The BLIM strain was originally made by curing BL26 Blue cells of the F' episome, resulting in a strain completely devoid of the *lacI* gene. BLIM cells were obtained from Prof. Kathleen Matthews' laboratory (Rice University) via Addgene (Bacterial strain #35609) and the absence of the *lacI* gene was confirmed by colony PCR using the same primers employed to amplify the *lacI** gene.

2.3 Expression of LacI* and LacI* regulatory domain

All growth media were supplemented with ampicillin. Chemically competent BLIM cells were transformed with either the vector harboring the gene encoding LacI* or the regulatory

domain of LacI* and selection was performed using LB plates following an overnight (< 16 hour) incubation period at 37°C. For proteins expressed in rich or H₂O-M9 minimal medium, a single colony was first used to inoculate a 50 mL culture of LB medium and grown in a shake flask at 37°C overnight to stationary phase (OD₆₀₀ ~ 3-4). The next morning, 1 L cultures of either LB or H₂O-M9 minimal medium were inoculated to an initial OD₆₀₀ of 0.05 and left to grow at 37°C in shake flasks until an OD₆₀₀ of 0.5 (H₂O-M9 minimal medium) or 1.0 (LB medium) was reached. At this point, cultures were induced with 100 µL of a 2 mg/mL stock of anhydrotetracycline (IBA) prepared in ethanol. The temperature was then reduced to 25°C and cells were left to grow overnight (< 16 hours). Cells were harvested by centrifugation and frozen at -80°C as 30 mL suspensions in 50 mM sodium phosphate, 500 mM sodium chloride, pH = 7.8.

Preparation of highly deuterated LacI* followed a procedure similar to that previously published for preparing highly deuterated malate synthase G [25]. A single colony was selected from LB plates and used to inoculate a 5 mL LB culture. The culture was grown at 37°C until cells reached an OD₆₀₀ ~0.8-1.0. At this point, cells were spun down and re-suspended in 20 mL of H₂O-M9 minimal medium and left to grow at 37°C until an OD₆₀₀ of 0.5 was reached. Cells were then spun down and re-suspended in 75 mL D₂O-M9 minimal medium (> 90% D₂O v/v). Cultures were left to grow overnight at 37°C. The next morning, cells were harvested and used to inoculate a 900 mL cultures of D₂O-M9 minimal medium to an initial OD₆₀₀ of 0.05. Cells were grown at 37°C until an OD₆₀₀ of 0.4-0.5 was reached. Cultures were induced with 100 µL of a 2 mg/mL stock of anhydrotetracycline prepared in ethanol. Induced cultures were left to grow overnight at 30°C (< 16 hours). Cells were harvested by centrifugation and frozen at -80°C as 30 mL suspensions in 50 mM sodium phosphate, 500 mM sodium chloride, pH = 7.8.

2.4 Purification of LacI* and LacI* regulatory domain

Frozen cell suspensions were thawed in a room-temperature water bath and then transferred to ice. MgCl₂ was added to a final concentration of 10 mM, CaCl₂ was added to a final concentration of 5 mM, Triton-X 100 was added to a final concentration of 0.2% (v/v), and fresh PMSF was added to a final concentration of 1 mM. One “cOmplete Mini” EDTA-free, broad range protease inhibitor tablet (Roche), 5 mg of hen egg white lysozyme, and 3.5 mg of bovine pancreas DNaseI were also added. The suspension was left to rock gently at 4°C for one hour. Sonication was then performed on ice using a repeated series of 15 pulses followed by resting on ice. Cell debris was pelleted by centrifugation at 15,000 rpm for 30 minutes at 4°C and the lysate was further clarified by passage through a 0.4 µm syringe filter.

All purification steps utilized an Äkta Prime Plus FPLC system operating at 4°C with the exception of the final gel filtration step which was sometimes performed at room temperature. Protein was detected by UV/Vis absorbance at 280 nm. Immobilized metal affinity chromatography (IMAC) was performed using a 3-5mL nickel-IDA column (His60 Ni Superflow, Clontech) and two buffers: buffer A (50 mM sodium phosphate, 500 mM sodium chloride, and 3 mM fresh BME, pH = 7.8) and buffer B (50 mM sodium phosphate, 500 mM sodium chloride, 250 mM imidazole, and 3 mM fresh BME, pH = 7.8). The lysate

was loaded onto the column using a 50 mL superloop and the loaded column was washed with buffer A then 16% buffer B until the detector baseline was flat. Protein was then eluted isocratically using 100% buffer B. The eluted fractions were analyzed by SDS-PAGE, pooled, and dialyzed against 4 L of buffer A overnight at 4°C in the presence of 1 mg of recombinant 6x-His tagged TEV protease [26].

Protease-treated protein solutions were then subject to a second round of IMAC using the same nickel IDA column equilibrated with buffer A at 4°C. The flow through was collected and purity was assessed by SDS-PAGE. Pure fractions were then pooled, concentrated to about 0.2-0.3 mM (for both LacI* and LacI* regulatory domain) via spin concentration using a 10 kDa molecular weight cutoff Amicon Ultra centrifugal filter (Millipore), then subject to gel filtration using either a ~120 mL Superdex 75 (at room temperature) or ~65 mL Superdex 200 (at 4°C) column (GE Healthcare) equilibrated with NMR buffer: either 20 mM sodium phosphate, 150 mM sodium chloride, 3 mM DTT, pH = 7.4 (LacI*) or 20 mM sodium phosphate, 50 mM sodium chloride, 3 mM DTT, pH = 7.4 (LacI* regulatory domain) at 4°C. In some cases HEPES was used instead of sodium phosphate. The final purity of LacI* or LacI* regulatory domain was typically > 99% as determined by quantification of SDS-PAGE results using the program ImageJ [27].

2.5 Solution NMR spectroscopy

NMR samples were typically 0.1-0.2 mM protein (for both LacI* and LacI* regulatory domain samples) in NMR buffer with 0.1 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and 0.02% w/v sodium azide added. TROSY ¹H-¹⁵N HSQC [16] experiments were collected at 298 K on a Bruker Avance III spectrometer equipped with cryoprobe operating at 750 MHz ¹H Larmor frequency.

3. Results and Discussion

3.1 Expression Vector Construction

Because LacI is a homo-oligomer, expressing a recombinant copy of *lacI* in the presence of other copies of the *lacI* gene could result in hetero-oligomerization, particularly if the dimerization and/or tetramerization domains are intact. In order for significant hetero-oligomerization to occur, the expression levels of each copy of the *lacI* gene need to be somewhat comparable. The native *lacI* gene in *E. coli* is expressed at very low levels and only a few copies of LacI exist in a single *E. coli* cell at any given time [28]. Over-expressing a recombinant copy of *lacI* in this background is unlikely to result in detectable heterodimerization. However, popular commercial *E. coli* host strains used in T7 expression systems (such as BL21(DE3) and derivatives thereof) contain an additional copy of the *lacI* gene on the λ-DE3 lysogen which is used to regulate the expression of T7 RNA polymerase [29]. Additionally, the most commonly used vectors possessing LacI-regulated T7 promoters (for example, the pET series, Novagen) possess their own copy of the *lacI* gene. Most commercially-available pET vectors have a pBR322 origin of replication which yields copy numbers of 15-20 per cell. The potential for heterodimerization becomes even greater when *lacI* promoter mutations, such as the popular *lacP* [30] or *lacP*¹ [31] mutations, are used to increase expression levels of LacI (10-100 fold) for tighter regulation. Despite this, such

expression systems are sometimes used to express recombinant LacI and IPTG is used to induce recombinant expression [4,18,32,33]. Since the K_d of IPTG binding for LacI and LacI* is on the order of 1-2 μ M [7], contaminating IPTG can be retained throughout purification. Indeed, recombinant LacI* produced from an IPTG-inducible expression system has been shown to be a mixture of apo and IPTG-bound states, even though the concentration of IPTG used in that particular study was only 0.5 mM [18].

While procedures could be devised to homogenize LacI, such efforts are laborious and will almost invariably involve some degree of denaturation followed by refolding due to the tight dissociation constants for IPTG binding and dimerization [34]. An alternative approach is presented here which eliminates these complications by simply removing all instances of the *lacI* gene from the expression system. Fig. 1 shows a diagram of the LacI* constructs and expression vector used in this work. The backbone is based on the inducible pASK vector [23]. The LacI* gene with TEV-protease recognition sequence is inserted in frame with an N-terminal 6x-His tag. Recombinant expression is regulated by the native *E. coli tetA* promoter from transposon 10 (Tn10) which is repressed by the TetR protein. Induction is initiated by introducing a small amount of anhydrotetracycline, a non-toxic analog of tetracycline. The use of an *E. coli* promoter eliminates the need for host strains lysogenized with the *lacI*-containing λ -DE3 element. Here, a *lacF*⁻ B-strain derivative (BLIM) is used as a host [24].

3.2 Expression and purification of LacI*

The high molecular weight of LacI* necessitates extensive perdeuteration in order to enable characterization by solution NMR. In order to express highly deuterated LacI*, cells were only minimally adapted to D₂O using slightly modified protocols previously described [25]. Briefly, the procedure involved an initial growth in LB medium, followed by transfer to H₂O-M9 minimal medium, then immediate transfer to D₂O-M9 minimal medium (> 90% D₂O v/v). This method was favored over serial adaptation since repeated passages over the course of days in an *endA*⁺/*recA*⁺ host could result in plasmid instability. In order to qualitatively assess the expression level of LacI* in highly deuterated minimal media, several small scale expression tests were performed. Small cultures (~2 mL) of cells that had been transferred to highly deuterated minimal medium were left to grow at 37°C until densities reached an OD₆₀₀ ~0.4. At this point, a pre-induction sample was taken and the remaining cells were induced. Whole cell pellets, harvested before and after induction were then subjected to SDS-PAGE which is shown in Fig. 2. Prior to induction, no noticeable bands are evident above background, confirming previous reports that the *tetA* promoter is tightly regulated by TetR, despite being a native *E. coli* promoter [23]. Following induction with anhydrotetracycline, a prominent band appears at about 40 kDa, indicating LacI* over-expression. OD₆₀₀ values for cultures in stationary phase following induction were typically ~1.0 for D₂O-M9 minimal medium. These results confirmed that the expression system can tolerate growth in highly deuterated minimal medium and is suitable for the large scale production of highly deuterated LacI* necessary for biophysical studies.

LacI* produced from 1 L cultures was subject to three purification steps. Protein purity after each step was monitored by SDS-PAGE. The first step was IMAC using a nickel IDA resin

which results in highly pure protein as shown in Fig. 3a. Fractions containing significant amounts of LacI* were pooled, treated with His-tagged TEV protease, then subject to IMAC again in order to separate TEV-protease, digested His tags, and any residual LacI* that still possessed a His tag (Fig. 3b). Tag-less LacI* elutes in the flow through but is also retained on the column until washed with 16% buffer B. This is because LacI has an inherent affinity for nickel resins [35]. Samples eluted in the flow through and after the 16% wash with buffer B were structurally identical as judged by NMR. The final purification step was gel filtration on a Superdex 75 column (Fig. 4). The elution profile reveals a single, narrow peak that is consistent with a highly homogeneous species. Typical yields of LacI* are ~13-14 mg/L of culture in highly deuterated minimal medium. These yields are nearly 2-fold higher than those previously reported for a thermostable mutant of LacI* (K84M) expressed from an unregulated vector [18].

3.3 Recombinant production of LacI* regulatory domain

A common approach to studying large, multi-domain proteins by solution NMR is to parse the protein into its constituent domains and characterize each domain individually to simplify spectra or improve the sensitivity of experiments [36]. This “divide and conquer” method is particularly useful for assigning chemical shifts since the required experiments are often limited by sensitivity and become difficult when the number of observed resonances is large due to increased spectral overlap. A previous report showed that recombinant production of a thermostable mutant of the isolated regulatory domain of LacI* (K84M), was toxic to three common *E. coli* host strains [18]. The toxicity was so pronounced that cells could not even be successfully transformed. It was observed that the toxicity was reduced only when full length LacI was introduced into the expression system at appreciable levels. As such, an IPTG-inducible expression vector was created to co-express full length LacI with the regulatory domain of LacI* (K84M). Three major limitations to using this expression system were noted: 1) Heterodimerization between full length LacI and LacI* (K84M) regulatory domain; 2) IPTG co-purification; 3) The requirement to keep starter cultures at low cell densities. While the first two issues could be addressed by purification and a special IPTG-removal step, the latter issue seemed unavoidable and was attributed to leaky expression of the toxic regulatory domain. This was perhaps surprising considering the host strain used, Rosetta(DE3) pLysS, utilizes a *lacUV5* promoter and expression of T7 lysozyme to regulate the expression and function of T7 RNA polymerase. The *lacUV5* promoter should impart very tight regulation since it is insensitive to decreases in catabolite repression as cell densities increase and T7 lysozyme inhibits T7 RNA polymerase [37]. This would suggest that the toxicity of LacI* (K84M) is particularly acute.

Recombinant production of the isolated regulatory domain was attempted using the pASK vector/BLIM strain combination. Fig. 5a shows the growth curve for BLIM cells carrying the vector for wild type LacI* regulatory domain in highly deuterated minimal medium. All measurements were made prior to inducing over-expression. The growth curve for BLIM cells carrying the vector for full length LacI* is also shown for reference. It is clear that cells carrying the LacI* regulatory domain expression vector grow robustly at a rate nearly identical to that for cells carrying the full length LacI* expression vector. To see if the previously observed toxicity stemmed specifically from the use of the thermostable K84M

mutation, growths were monitored for BLIM cells transformed with a vector harboring LacI*(K84M) regulatory domain and a similar mutant, K84L, which exhibits an even stronger degree of thermostabilization [38]. The growth curves are shown in Fig. 5b and it is apparent that neither vector is toxic to BLIM cells.

The effect of over-expressing LacI* regulatory was then examined. Fig. 6a-c shows an expression test for wild type, K84L, and K84M LacI* regulatory domains in highly deuterated minimal medium, respectively. It is clear that all three constructs exhibit high levels of over-expression. Final OD₆₀₀ values for induced 1L cultures in stationary phase were ~2.0 for wild type regulatory domain and the K84L mutants and ~1.5 for the K84M mutant. Yields obtained from 1 L growths in highly deuterated minimal medium were ~15 mg for wild type, ~10 mg for K84L, and ~9 mg for K84M. The slightly reduced yields for the K84 mutants may not necessarily reflect enhanced toxicity, since it was observed the His tag cleavage was markedly less efficient for these mutants (data not shown). This is perhaps expected considering that the crystal structure of the K84L mutant exhibits a more tightly packed monomer-monomer interface relative to the structure of wild type [39]. This likely reduces flexibility and, in turn, access to the His tag. Regardless, the observed yields are all over 2-fold higher than the previously reported yield [18]. Ostensibly, this increase may not appear that drastic. However, it should be noted that *E. coli* RNA polymerase is five-fold less efficient than T7 RNA polymerase [22], implying the improvement in yield is actually quite considerable.

3.4 Expression toxicity

The previously observed toxicity associated with expressing LacI* regulatory domain from a T7 promoter was hypothesized to originate from disruption of regulatory LacI functionality through the formation of non-functional LacI-LacI* regulatory domain heterodimers [18]. This would presumably result in high levels of basal T7 RNA polymerase expression which would increase basal expression of LacI* regulatory domain. This increased basal expression of LacI* regulatory domain would only result in significant toxicity if the LacI* regulatory domain had an inherent level of toxicity. The results above, however, suggest that LacI* regulatory domain is not inherently toxic (at least not prohibitively so) and that full length LacI is not required for expression when the pASK vector/BLIM strain expression system is used. The striking differences in toxicity between the previous report and what is presented here would suggest that the toxicity is specific to the expression system used and not the LacI* regulatory domain protein. The lack of significant LacI* regulatory domain toxicity in the pASK vector/BLIM host strain expression system eliminates the inconvenient need to keep cell densities low at all stages during pre-culturing.

3.5 LacI* and LacI* regulatory domain are functional

The functionality of LacI* and LacI* regulatory domain was assessed using NMR spectroscopy. Fig. 7 a-b. shows TROSY ¹H-¹⁵N HSQC spectra of LacI* in the apo and IPTG-bound states, respectively. Because binding of IPTG is slow on the NMR-timescale, free and bound-states yield a separate set of resonances which facilitates rapid detection of inactive/contaminated protein. The data shown in Fig. 7 show that each state of LacI* yields a single set of resonances, which confirms that the expression system presented produces

highly homogenous samples. Fig. 7c-d shows TROSY ^1H - ^{15}N HSQC spectra of wild type LacI* regulatory domain in the apo and IPTG-bound states, respectively. Again, a single set of resonances is observed for each state. It is clearly apparent that the spectra of the isolated regulatory domain map to those of the full length protein, suggesting the feasibility of transferring chemical shift assignments. This confirms that the isolated regulatory domain will be a useful tool for solution NMR studies of LacI*.

Conclusions

Biophysical studies of LacI*, particularly solution NMR studies, require an expression system that produces homogenous protein in high yield, even under high-stress growth conditions. While several expression systems for producing LacI* have been introduced over the years, they have all exhibited specific limitations which include sample heterogeneity, low-yield, and intolerance to producing the isolated regulatory domain. The expression system introduced here which utilizes an inducible *tetA* promoter and *lacI* deficient host strain ensures a highly homogenous product without the possibility of hetero-oligomerization or IPTG co-purification. Importantly, LacI* yields from highly deuterated minimal medium were higher than previously reported for other expression systems. Perhaps the most beneficial aspect of the expression system is its ability to produce recombinant LacI* regulatory domain without toxicity or the need for co-expression of the full length LacI protein. This greatly simplifies isolation and increases yields by over 2-fold. While the expression system and procedures outlined here are likely to be most beneficial to the production of recombinant LacI and constructs thereof, the approaches and considerations may be general to any scenario where multiple copies of a gene of interest complicate protein production. Moreover, the widespread use of LacI as both a model system and general tool in biology (other than regulating recombinant expression) suggest the expression system presented here will be useful to studies across many disciplines.

Acknowledgement

This work was supported by NIH GM102447 awarded to A.J.W. M.A.S. is an NIH predoctoral trainee (GM008275). We gratefully acknowledge Prof. Mitchell Lewis and Dr. Leslie Milk for providing the gene encoding LacI* and expert advice throughout the course of this work. We would like to thank Dr. Igor Dodevski and Dr. Yu-San Huoh for helpful discussion regarding molecular cloning and protein purification. We also thank Dr. Kathleen Valentine and Dr. Sabrina Bedard for technical assistance.

References

- [1]. Swint-Kruse L, Matthews KS. Allostery in the LacI/GaIR family: Variations on a theme. *Cur. Opin. Microbiol.* 2009; 12:129–137. 10.1016/j.mib.2009.01.009.
- [2]. Lewis M. A tale of two repressors. *J. Mol. Biol.* 2011; 409:14–27. 10.1016/j.jmb.2011.02.023. [PubMed: 21392509]
- [3]. Lewis M, et al. Crystal structure of the Lactose operon repressor and its complexes with DNA and inducer. *Science.* 1996; 271:1247–1254. 10.1126/science.271.5253.1247. [PubMed: 8638105]
- [4]. Daber R, Stayrook S, Rosenberg A, Lewis M. Structural analysis of Lac repressor bound to allosteric effectors. *J. Mol. Biol.* 2007; 370:609–619. 10.1016/j.jmb.2007.04.028. [PubMed: 17543986]
- [5]. Bell CE, Lewis M. A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Biol.* 2000; 7:209–214. [PubMed: 10700279]

- [6]. Chen J, Matthews KS. Deletion of Lactose repressor carboxyl-terminal domain affects tetramer formation. *J. Biol. Chem.* 1992; 267:13843–13850. [PubMed: 1629185]
- [7]. Chen J, Surendran R, Lee JC, Matthews KS. Construction of a dimeric repressor - dissection of subunit interfaces in Lac repressor. *Biochemistry.* 1994; 33:1234–1241. 10.1021/bi00171a025. [PubMed: 8110756]
- [8]. Kaptein R, Zuiderweg ERP, Scheek RM, Boelens R, Vangunsteren WF. A protein structure from nuclear magnetic resonance data. Lac repressor headpiece. *J. Mol. Biol.* 1985; 182:179–182. 10.1016/0022-2836(85)90036-1. [PubMed: 3889346]
- [9]. Spronk C, Slijper M, vanBoom JH, Kaptein R, Boelens R. Formation of the hinge helix in the Lac repressor is induced upon binding to the lac operator. *Nat. Struct. Biol.* 1996; 3:916–919. 10.1038/nsb1196-916. [PubMed: 8901866]
- [10]. Spronk C, et al. The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator. *Structure.* 1999; 7:1483–1492. 10.1016/s0969-2126(00)88339-2. [PubMed: 10647179]
- [11]. Kalodimos CG, Folkers GE, Boelens R, Kaptein R. Strong DNA binding by covalently linked dimeric Lac headpiece: Evidence for the crucial role of the hinge helices. *Proc. Natl Acad. Sci. USA.* 2001; 98:6039–6044. 10.1073/pnas.101129898. [PubMed: 11353825]
- [12]. Kalodimos CG, Boelens R, Kaptein R. Toward an integrated model of protein-DNA recognition as inferred from NMR studies on the Lac repressor system. *Chem. Rev.* 2004; 104:3567–3586. 10.1021/cr0304065. [PubMed: 15303828]
- [13]. Kalodimos CG, et al. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science.* 2004; 305:386–389. 10.1126/science.1097064. [PubMed: 15256668]
- [14]. Romanuka J, et al. Specificity and affinity of Lac repressor for the auxiliary operators O2 and O3 are explained by the structures of their protein-DNA complexes. *J. Mol. Biol.* 2009; 390:478–489. 10.1016/j.jmb.2009.05.022. [PubMed: 19450607]
- [15]. Kalodimos CG, et al. Plasticity in protein-DNA recognition: Lac repressor interacts with its natural operator O1 through alternative conformations of its DNA-binding domain. *EMBO J.* 2002; 21:2866–2876. 10.1093/emboj/cdf318. [PubMed: 12065400]
- [16]. Pervushin K, Riek R, Wider G, Wuthrich K. Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl Acad. Sci. USA.* 1997; 94:12366–12371. 10.1073/pnas.94.23.12366. [PubMed: 9356455]
- [17]. Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. Cross-correlated relaxation enhanced H-1-C-13 NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* 2003; 125:10420–10428. 10.1021/ja030153x. [PubMed: 12926967]
- [18]. Romanuka J, van den Bulke H, Kaptein R, Boelens R, Folkers GE. Novel strategies to overcome expression problems encountered with toxic proteins: Application to the production of Lac repressor proteins for NMR studies. *Protein Express. Purif.* 2009; 67:104–112. 10.1016/j.pep.2009.05.008.
- [19]. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose p-BAD promoter. *J. Bacteriol.* 1995; 177:4121–4130. [PubMed: 7608087]
- [20]. Giacalone MJ, et al. Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system. *Biotechniques.* 2006; 40:355–364. 10.2144/000112112. [PubMed: 16568824]
- [21]. Zhan H, Sun Z, Matthews KS. Functional impact of polar and acidic substitutions in the Lactose repressor hydrophobic monomer-monomer interface with a buried lysine. *Biochemistry.* 2009; 48:1305–1314. 10.1021/bi801357f. [PubMed: 19166325]
- [22]. Boomersshine WP, Raj MLS, Gopalan V, Foster MP. Preparation of uniformly labeled NMR samples in *Escherichia coli* under the tight control of the araBAD promoter: Expression of an archaeal homolog of the RNase P Rpp29 protein. *Protein Express. Purif.* 2003; 28:246–251. 10.1016/s1046-5928(02)00707-6.

- [23]. Skerra A. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. *Gene*. 1994; 151:131–135. 10.1016/0378-1119(94)90643-2. [PubMed: 7828861]
- [24]. Wycuff DR, Matthews KS. Generation of an AraC-araBAD promoter-regulated T7 expression system. *Anal. Biochem*. 2000; 277:67–73. 10.1006/abio.1999.4385. [PubMed: 10610690]
- [25]. Tugarinov V, Kanelis V, Kay LE. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nature Protocols*. 2006; 1:749–754. 10.1038/nprot.2006.101. [PubMed: 17406304]
- [26]. Tropea JE, Cherry S, Waugh DS. Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol. Biol. (Clifton, N.J.)*. 2009; 498:297–307. 10.1007/978-1-59745-196-3_19.
- [27]. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*. 2012; 9:671–675. 10.1038/nmeth.2089. [PubMed: 22930834]
- [28]. Gilbert W, Mullerhi B. Isolation of Lac repressor. *Proc. Natl Acad. Sci. USA*. 1966; 56:1891–&. 10.1073/pnas.56.6.1891. [PubMed: 16591435]
- [29]. Studier FW, Moffatt BA. Use of bacteriophage-T7 RNA-polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol*. 1986; 189:113–130. 10.1016/0022-2836(86)90385-2. [PubMed: 3537305]
- [30]. Mullerhi B, Crapo L, Gilbert W. Mutants that make more Lac repressor. *Proc. Natl Acad. Sci. USA*. 1968; 59:1259–&. 10.1073/pnas.59.4.1259. [PubMed: 4870861]
- [31]. Glascock CB, Weickert MJ. Using chromosomal lacI(Q1) to control expression of genes on high-copy-number plasmids in *Escherichia coli*. *Gene*. 1998; 223:221–231. 10.1016/S0378-1119(98)00240-6. [PubMed: 9858738]
- [32]. Kim S, et al. Probing allostery through DNA. *Science*. 2013; 339:816–819. 10.1126/science.1229223. [PubMed: 23413354]
- [33]. Haeusler AR, et al. FRET studies of a landscape of Lac repressor-mediated DNA loops. *Nucleic Acids Res*. 2012; 40:4432–4445. 10.1093/nar/gks019. [PubMed: 22307389]
- [34]. Chen J, Matthews KS. Subunit dissociation affects DNA-binding in a dimeric Lac repressor produced by C-terminal deletion. *Biochemistry*. 1994; 33:8728–8735. 10.1021/bi00195a014. [PubMed: 8038163]
- [35]. Owens RM, Grant A, Davies N, O'Connor CD. Copurification of the lac repressor with polyhistidine-tagged proteins in immobilized metal affinity chromatography. *Protein Express. Purif*. 2001; 21:352–360. 10.1006/prev.2000.1384.
- [36]. Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature*. 2007; 445:618–622. 10.1038/nature05512. [PubMed: 17237764]
- [37]. Wanner BL, Kodaira R, Neidhardt FC. Regulation of lac operon expression - reappraisal of theory of catabolite repression. *J. Bacteriol*. 1978; 136:947–954. [PubMed: 214424]
- [38]. Nichols JC, Matthews KS. Combinatorial mutations of lac repressor. Stability of monomer-monomer interface is increased by apolar substitution at position 84. *J Biol Chem*. 1997; 272:18550–18557. [PubMed: 9228020]
- [39]. Bell CE, Barry J, Matthews KS, Lewis M. Structure of a variant of Lac repressor with increased thermostability and decreased affinity for operator. *J. Mol. Biol*. 2001; 313:99–109. 10.1006/jmbi.2001.5041. [PubMed: 11601849]

Fig. 1.

Map of the expression vectors used in this study. The backbone is the pASK35plus vector which possesses an inducible tetracycline promoter (*ptet*) from *E. coli*, regulated by the TetR protein. The vector confers ampicillin resistance (*bla*) and is medium copy number (*colE1* origin). The multiple cloning site (MCS) follows a 6x-His tag. Genes encoding either LacI* or the isolated LacI* regulatory domain (*lacI** RD) with N-terminal TEV protease recognition sites were subcloned into the backbone in frame with the His tag.

Fig. 2.

LacI* expression test in BLIM cells cultured in highly deuterated M9 minimal medium. Whole cell pellets were subject to SDS-PAGE. The results for three separate small scale (~2 mL) cultures are shown. The left most lane is a molecular weight marker sample. “Pre-Ind” lanes were pellets harvested just prior to induction ($OD_{600} \sim 0.4$) and “Ind” lanes were pellets harvested following overnight induction at 30°C ($OD_{600} \sim 1.0$). The red arrow corresponds to the expected molecular weight of the tagged LacI* gene product (38.6 kDa).

Fig. 3. SDS-PAGE illustrating the effectiveness of the LacI* purification protocol. The left most lane for each panel is a molecular weight marker sample. (A) Eluted fractions containing tagged LacI* from the first IMAC step. The elution volume is relative to the start of the application of 100% buffer B. (B) Flow through containing tag-less LacI* from the second IMAC step. The last lane of (B) marked “E” is the eluted fraction obtained from washing the IMAC column with 100% buffer B and contains a small amount of tagged LacI*, untagged LacI*, TEV (two bands presumably due to TEV protease autodigestion) and the cleaved tag. The red arrow corresponds to the expected molecular weight of the tag-less LacI* product (~35.5 kDa).

Fig. 4. Size exclusion chromatography elution profile for LacI*. The inset shows the SDS-PAGE results for the eluted fractions where the left most lane is a molecular weight marker sample.

Fig. 5.

Cell growth as monitored by OD₆₀₀. (A) BLIM cells carrying the full length (FL) LacI* expression vector (black circles) and BLIM cells carrying the LacI* isolated regulatory domain (RD) vector (red circles). (B) BLIM cells carrying the LacI* (K84M) isolated regulatory domain vector (black triangles) and BLIM cells carrying the LacI* isolated regulatory domain vector (red triangles). All cells were cultured in highly deuterated M9 minimal medium and grown at 37°C in shake flasks.

Fig. 6.

Expression profiles for (A) wild type LacI* regulatory domain, (B) LacI*(K84L) regulatory domain, and (C) LacI*(K84M) regulatory domain. Whole cell pellets were subject to SDS-PAGE. For each panel, left most lanes are molecular weight marker samples, center lanes are pre-induction (PI) samples ($OD_{600} \sim 0.4$) and right most lanes are post-induction (I) samples ($OD_{600} \sim 2.0$ for wild type and K84L, ~ 1.5 for K84M). The red arrow corresponds to the expected molecular weight of the tagged regulatory domain (~ 32.1 kDa).

Fig. 7.
TROSY ^1H - ^{15}N HSQC spectra of (A) LacI*, (B) LacI* + IPTG, (C) LacI* RD, (D) LacI* RD + IPTG. Data were collected at 298K at 750 MHz ^1H Larmor frequency and processed identically.