

RESEARCH

Open Access

# A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in *Escherichia coli*

Simone Balzer<sup>1</sup>, Veronika Kucharova<sup>1</sup>, Judith Megerle<sup>2</sup>, Rahmi Lale<sup>1</sup>, Trygve Brautaset<sup>3</sup> and Svein Valla<sup>1\*</sup>

## Abstract

**Background:** Production of recombinant proteins in bacteria for academic and commercial purposes is a well established field; however the outcomes of process developments for specific proteins are still often unpredictable. One reason is the limited understanding of the performance of expression cassettes relative to each other due to different genetic contexts. Here we report the results of a systematic study aiming at exclusively comparing commonly used regulator/promoter systems by standardizing the designs of the replicon backbones.

**Results:** The vectors used in this study are based on either the RK2- or the pMB1- origin of replication and contain the regulator/promoter regions of *XylS/Pm* (wild-type), *XylS/Pm* ML1-17 (a *Pm* variant), *Lacl/P<sub>T7lac</sub>*, *Lacl/P<sub>trc</sub>* and *AraC/P<sub>BAD</sub>* to control expression of different proteins with various origins. Generally and not unexpected high expression levels correlate with high replicon copy number and the *Lacl/P<sub>T7lac</sub>* system generates more transcript than all the four other cassettes. However, this transcriptional feature does not always lead to a correspondingly more efficient protein production, particularly if protein functionality is considered. In most cases the *XylS/Pm* ML1-17 and *Lacl/P<sub>T7lac</sub>* systems gave rise to the highest amounts of functional protein production, and the *XylS/Pm* ML1-17 is the most flexible in the sense that it does not require any specific features of the host. The *AraC/P<sub>BAD</sub>* system is very good with respect to tightness, and a commonly used bioinformatics prediction tool (RBS calculator) suggested that it has the most translation-efficient UTR. Expression was also studied by flow cytometry in individual cells, and the results indicate that cell to cell heterogeneity is very relevant for understanding protein production at the population level.

**Conclusions:** The choice of expression system needs to be evaluated for each specific case, but we believe that the standardized vectors developed for this study can be used to more easily identify the nature of case-specific bottlenecks. By then taking into account the relevant characteristics of each expression cassette it will be easier to make the best choice with respect to the goal of achieving high levels of protein expression in functional or non-functional form.

**Keywords:** Recombinant expression, Regulator/promoter systems, *XylS/Pm*, *XylS/Pm* ML1-17, *Lacl/P<sub>T7lac</sub>*, *Lacl/P<sub>trc</sub>*, *AraC/P<sub>BAD</sub>*, Systematic comparison

\* Correspondence: svein.valla@ntnu.no

<sup>1</sup>Department of Biotechnology, NTNU, Sem Sælands vei 6, Trondheim 7491, Norway

Full list of author information is available at the end of the article

## Background

Parameters affecting recombinant protein expression in *Escherichia coli* have been studied extensively and numerous methods aiming at improving protein yields have been reported, usually involving genetic manipulations and/or production process optimization [1-4]. However, in spite of the large number of potentially useful approaches available there is still no guarantee that a satisfactory result will be obtained in each specific case, and trial and error is therefore currently an integrated part of development of new protein production processes. The work involved in this can become very laborious since many parameters such as choice of strains, vector construct designs, growth media and cultivation conditions can potentially have a big and unpredictable effect on the process. Steadily more promoter systems for regulated protein expression in *E. coli* ([1] and references therein, [2-6]) are being developed, increasing the complexity. The studies of those novel expression systems were commonly based on experiments involving vectors with different backbones [2,4,7,8]; typically commercially available and commonly used vectors from the pET [9], pTrc [10] or pBAD [11] series. More theoretical approaches have also been used [6,12]. However, expression is influenced by many parameters even within vectors, like the presence or absence of sequences of the 5' coding region encoding N-terminal fusion partners (His<sub>6</sub> tag [13], N-terminal signal peptides [14], and others), different origins of replication [15-17], different terminators [18] or selection markers. Penicillins for example are very frequently used for selection in spite of their known rapid degradation due to secreted  $\beta$ -lactamase [19]. A first step towards a more systematic, backbone-independent approach is described in a study performed by Tegel et al. [20] in which expression from three different IPTG-inducible promoters ( $P_{T7lac}$ ,  $P_{trc}$ ,  $P_{lac}$ ) is compared. These are all based on the negative regulator LacI, while positively regulated promoters such as  $P_{BAD}$  and  $P_{pm}$  have not been used in such comparative studies. The regulators of these two promoters (AraC and XylS, respectively) are both members of the same family of transcriptional activators [21]. The AraC/ $P_{BAD}$  system is quite extensively used and its characteristics have been reviewed [1]. The XylS/ $P_{pm}$  system was included because it has several beneficial traits for protein expression in general (reviewed by Brautaset et al. [21]), and in combination with RK2 minimal replicons it has been demonstrated to be capable of expressing proteins at industrial levels in high cell density cultivations [14,22]. We have used this system extensively in our laboratory as a model for studies of recombinant gene expression. Particular advantages of this system are that the levels of expression can be fine-tuned by various means [23-25], that it is not host-dependent in contrast to most other systems and that the inducer is cheap. Furthermore, expression

from the native system could be greatly improved by generating variants of the regulator protein XylS [26], the DNA region corresponding to the  $P_{pm}$  promoter region [27] as well as the region corresponding to the  $P_{pm}$  5'-untranslated region (5'-UTR) [28].

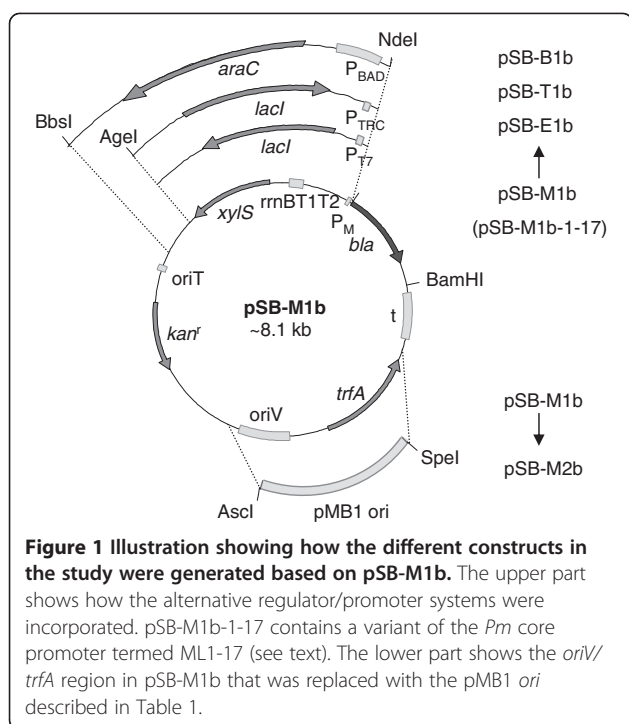
In this report we describe a systematic comparison of both positively and negatively regulated expression systems. Being aware of the influence of the 5' end of the coding region on expression [29,30], we intentionally chose to use model genes with native 5' ends as opposed to commonly used regions encoding N-terminal detection tags or solubility-enhancing fusion partners. The expression analyses were carried out at both the transcript and the protein level (activity assays and total protein), and we also included a flow cytometry based analysis of expression in individual cells. All comparisons were performed using identical vector backbones, a procedure we believe can be used generally as a diagnostic tool to identify bottlenecks in recombinant protein production processes.

## Results and discussion

### Construction of a set of plasmids specifically designed for comparative studies of commonly used expression systems in *E. coli*

To reduce potential effects on expression unrelated to the features of the regulator/promoter systems themselves all replicons used for comparisons were designed in such a way that the backbones were identical and the expression cassettes were in all cases integrated at the same location (Figure 1 and Table 1). The selected systems include XylS/ $P_{pm}$  (the native system; denoted in the figures as M); the high level expression variant  $P_{pm}$  ML1-17 (abbreviated by M-1-17) [27]; LacI/ $P_{T7lac}$  originating from the pET vector series (Novagen; denoted as E); the LacI/ $P_{trc}$  system from the pTRC series of vectors (Pharmacia; denoted as T); and finally the AraC/ $P_{BAD}$  system from the pBAD series of vectors (Invitrogen, abbreviated by B). Further details related to transcriptional start sites and 5'-UTR regions are described in the Methods section.

It is well known that gene dosage and expression levels often correlate, at least to some extent. In order to investigate any potential gene-specific effects related to this the cassettes were integrated into a mini-RK2 based replicon (pSB-M1b, 5-7 copies per cell [33]), and the pMB1 replicon (15-20 copies per cell [19], Novagen, Invitrogen) used in commercially available vectors such as pET and pBAD. In these two plasmid sets, genes coding for five different model proteins of varying biological origins were placed under control of the five promoters to cover a broad range of problems that may occur during recombinant protein production (Table 2). Note also that the use of one common N-terminal fusion tag for



all proteins was avoided to study the effect of the respective promoter-5'-UTR regions on different 5' coding sequences, as opposed to the study of Tegel et al. [20]. Specific gene sequence dependent parameters such as mRNA secondary structures and the presence of rare codons were taken into account by using optimized (for *E. coli*) synthetic genes. The corresponding genes were inserted into the two replicon types carrying the different expression cassettes, (Table 1). Not only can expression be directly compared from different regulator/promoter systems using these standardized vectors, but they can also be used more generally as tools to identify an appropriate expression system for the production of any selected target protein.

Due to the nature of the expression systems it was necessary to use two different *E. coli* strains as hosts. Strain ER2566 was chosen to compare expression from *LacI/P<sub>T7lac</sub>* with *XylS/P<sub>m</sub>* because it carries a chromosomal copy of the T7 polymerase integrated into the *lac* operon (NEB). Since the *LacI/P<sub>trc</sub>* system is also induced by IPTG, it was decided to study expression in the same host under the assumption that the expression of T7 polymerase does not affect expression from *LacI/P<sub>trc</sub>* due to the specificity of this polymerase for its cognate promoter [40]. Expression from *XylS/P<sub>m</sub>* compared to *AraC/P<sub>BAD</sub>* was performed in *E. coli* DH10B which is unable to catabolize L-arabinose, the inducer of the *AraC/P<sub>BAD</sub>* system.

### Protein production levels are generally stimulated by increased gene dosage, but none of the tested cassettes are superior for all genes

Three different genes, encoding luciferase, an antibody fragment fused in frame to alkaline phosphatase (scFv173-2-5-AP) and green fluorescent protein (GFP), respectively, were selected as models in the initial study of the performances of the various expression cassettes (Figure 2). The alkaline phosphatase fusion protein is translocated to the periplasm, while luciferase and GFP are cytoplasmic. The results were monitored as activities, meaning that only functional proteins were measured. The only parameter that gave a consistent response for all systems was not surprisingly gene dosage, as all cassettes gave rise to more activity when they were utilized in a high plasmid copy number context. However, the fold increase was heavily protein and expression cassette dependent, ranging from 1.6 for GFP (Figure 2, Panel C) to 10.4 for the alkaline phosphatase fusion in the *LacI/P<sub>T7lac</sub>* system (Figure 2, Panel B). We also observed that cell growth was strongly affected in several of the alkaline phosphatase fusion protein producing strains, and it was generally much more difficult to obtain reproducible data for this particular protein. We believe the reason for this is that the export of large amounts of protein is toxic to cell growth [41], in some cases also in the uninduced state. This potential toxic effect may even have resulted in accumulation of mutants that grow faster than the originally inoculated strain due to reduced scFv173-2-5-AP production. The maximal expression level is obviously very important in the context of recombinant protein production, and Figure 2 shows that in this respect none of the systems is superior for all proteins. Generally *XylS/P<sub>m</sub>* ML1-17 and *LacI/P<sub>T7lac</sub>* tended to produce most recombinant protein in the studies in strain ER2566 (Figure 2, Panels A-C). The mutations in the *P<sub>m</sub>* core region were of vital importance, as *XylS/P<sub>m</sub>* ML1-17 produced between 1.2- and 4.0- fold more active protein than the corresponding wild-type system. Note also that *AraC/P<sub>BAD</sub>* generated similar amounts of active protein compared to *XylS/P<sub>m</sub>* ML1-17 when the studies were done in an *ara* negative strain (DH10B; Figure 2, Panels D-F). *LacI/P<sub>T7lac</sub>* is generally known to be a very strong system because of the efficient transcription exerted by the T7 RNA polymerase [9,40], but the comparative analysis demonstrated that this system was not superior to *XylS/P<sub>m</sub>* ML1-17 for the genes studied here. Especially in the higher copy-number plasmids, up to four times more activity was detected in strains harboring *XylS/P<sub>m</sub>* ML1-17. We also noted in this and other related ongoing studies in our laboratory that to get stable expression from the *LacI/P<sub>T7lac</sub>* system, ER2566 cells needed to be freshly transformed prior to expression studies, as also recommended by Vethanayagam and Flower [42]. Similar observations were not made for any of the other three systems.

**Table 1 Plasmids used in this study<sup>a</sup>**

Name	Key features	Source
pTA16/pSB-M1b <sup>b</sup>	<i>m</i> -toluate- inducible <i>P<sub>m</sub></i> , <i>xylS</i> activator gene, RK2 replicon, <i>bla</i> reporter, Kan <sup>r</sup>	[31]
pET16b	IPTG-inducible <i>P<sub>T7lac</sub></i> , <i>lacI</i> repressor gene, Amp <sup>r</sup>	Novagen
pBAD/gIII_calmodulin	L-arabinose- inducible <i>P<sub>BAD</sub></i> , <i>araC</i> activator gene, Amp <sup>r</sup>	Invitrogen
pTrc99A	IPTG- inducible <i>P<sub>trc</sub></i> promoter, <i>lacI</i> repressor gene, Amp <sup>r</sup>	Pharmacia
pIB11- <i>lucS</i>	pIB11 [25] with <i>lucS</i> under control of <i>xylS/P<sub>m</sub></i> , Kan <sup>r</sup>	unpublished
pBAD24-GFP	pBAD24 with <i>gfpmut3</i> insert, Amp <sup>r</sup>	[32]
pHOG-173-2-5-AP	pHOG plasmid with <i>scFv173-2-5-phoA</i> fusion gene insert, provided by Affitech AS, Oslo, Amp <sup>r</sup>	unpublished
pMA-GH	pMA vector (GeneArt <sup>®</sup> , Invitrogen) with <i>GH1<sub>S</sub></i> insert, provided by Vectron Biosolutions AS, Trondheim, Amp <sup>r</sup>	unpublished
pMA-T-IL-1RA	pMA vector (GeneArt <sup>®</sup> , Invitrogen) with <i>IL1RN<sub>S</sub></i> insert, provided by Vectron Biosolutions AS, Trondheim, Amp <sup>r</sup>	unpublished
pSB-P0x	pSB-M1b variants with combinations of different features: P... regulator/promoter system M... <i>xylS/P<sub>m</sub></i> M-1-17... <i>xylS/P<sub>m</sub></i> variant ML1-17 E... <i>lacI/P<sub>T7lac</sub></i> (from pET) T... <i>lacI<sup>r</sup>/P<sub>trc</sub></i> (from pTrc) B... <i>araC/P<sub>BAD</sub></i> (pBAD) O... origin of replication 1... RK2 replicon 2... pMB1 replicon X... reporter gene b... <i>bla</i> l... <i>lucS</i> s... <i>scFv173-2-5-phoA</i> g... <i>gfpmut3</i> h... <i>GH1<sub>S</sub></i> r... <i>IL1RN<sub>S</sub></i>	This study
e.g. pSB-M2l	<i>m</i> -toluate- inducible <i>P<sub>m</sub></i> , <i>xylS</i> activator gene, pMB1 <i>ori</i> , <i>lucS</i> reporter, Kan <sup>r</sup>	This study

<sup>a</sup> *bla*: β- lactamase gene; *lucS*: synthetic luciferase gene; *scFv173-2-5-phoA*: single-chain antibody fragment 173-2-5 alkaline phosphatase fusion gene; *gfpmut3*: gene for the optimized green fluorescent protein mutant 3; *GH1<sub>S</sub>*: synthetic gene for human growth hormone, *IL1RN<sub>S</sub>*: synthetic gene for human interleukin 1 receptor antagonist.

<sup>b</sup> pTA16 was named pSB-M1g in this study for consistency purposes.

### The *LacI/P<sub>T7lac</sub>* system is unique by its generation of large amounts of transcript and insoluble protein

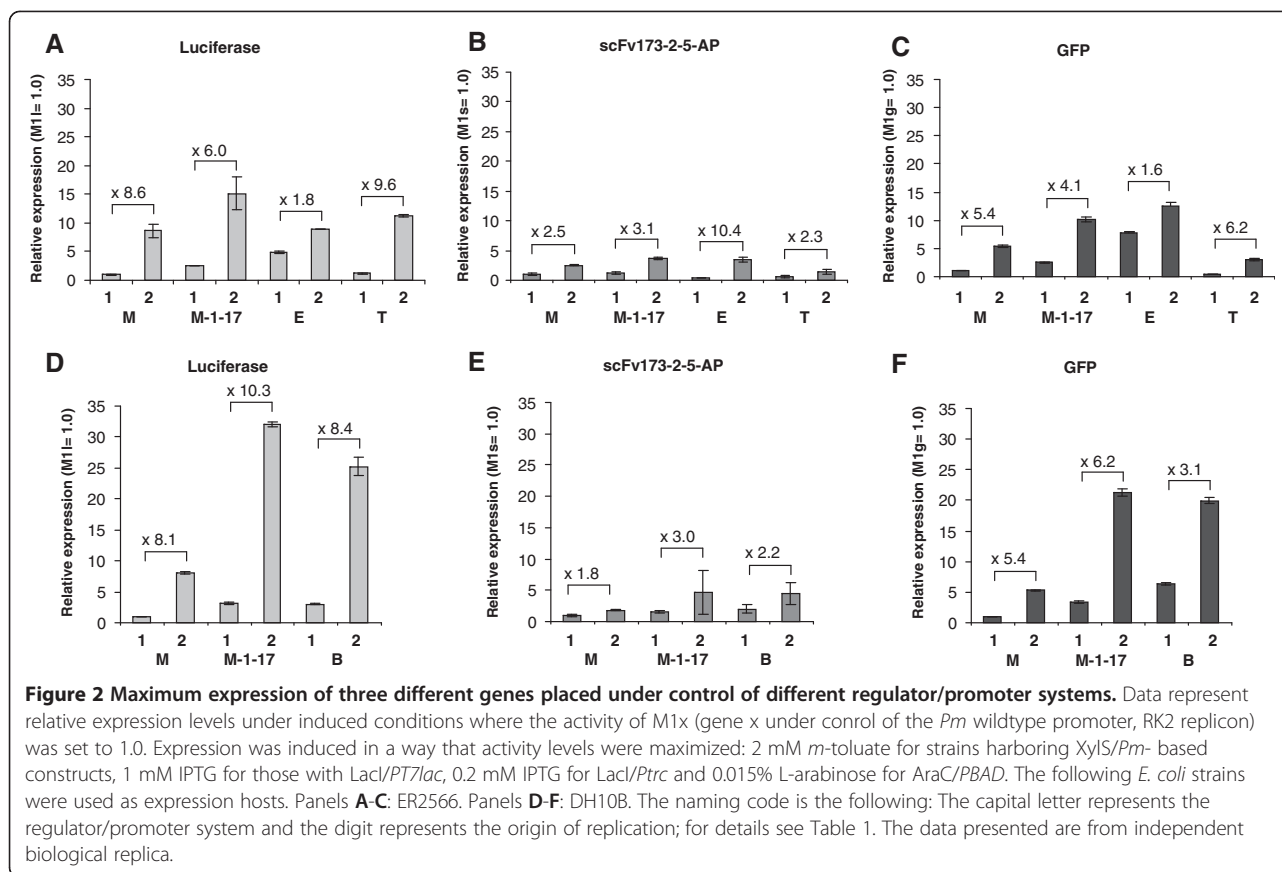
In the analyses described above only active protein was monitored, but potential big differences in target gene transcript accumulation or inactive (insoluble) protein production would not be discovered by such an analysis.

We therefore investigated to what extent total protein production is proportional to the amounts of transcript produced, which is not necessarily the case [43-45]. For this purpose, we included two additional proteins, the medically relevant human growth hormone (HGH) and interleukin-1RA (IL-1RA), see also (Table 2). The

**Table 2 Properties of the proteins selected as expression reporters**

Protein	Properties
Luciferase	reporter protein, ~ 60.8 kDa, cytoplasmic localization, generally low expression, rather easy to detect, very sensitive detection via bioluminescence assay
scFv173-2-5-AP	industrially relevant protein, ~77.2 kDa, fusion protein, disulfide bonds, translocated to the periplasm, detectable through AP <sup>a</sup> fusion, AP needs to be translocated to be active [34]
GFP	reporter protein, ~ 26.9 kDa, cytoplasmic localization, stable, known to be produced virtually only in its soluble form [35], very easy to detect by direct fluorometry
HGH	industrially relevant protein, ~25.1 kDa, cytoplasmic localization, usually expressed in <i>E. coli</i> as soluble protein [36,37]
IL-1RA	industrially relevant protein, ~20.1 kDa, cytoplasmic localization, usually expressed in <i>E. coli</i> as soluble protein [38,39]

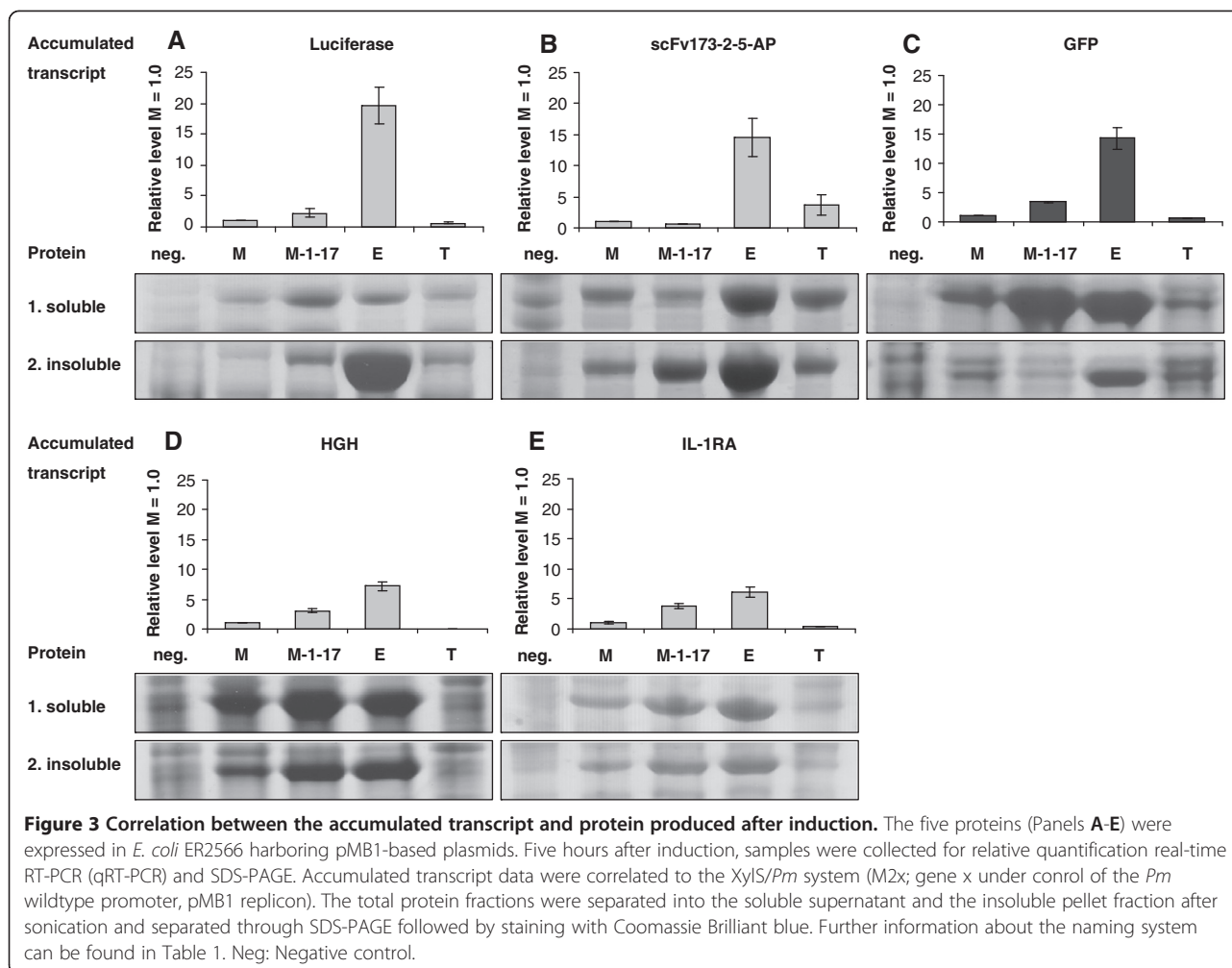
<sup>a</sup>Alkaline phosphatase.



comparison was carried out in strains harbouring plasmids with the pMB1 replicon, which as described above generally led to a higher level of protein production (measured as activity). One general conclusion following from these experiments was that the *Lacl/P<sub>T7lac</sub>* system generated much more transcript than *XylS/Pm* (between 6.2 and 20 times more) and *Lacl/P<sub>trc</sub>* (between 3.9 and 206 times more) for all the five tested genes. *XylS/Pm* ML1-17 generated more transcript than *Lacl/P<sub>trc</sub>* as well (Figure 3), ranging from 3.3 times for *luc<sub>S</sub>* mRNA and 88 times for *GHI<sub>S</sub>* mRNA, except for the special case with *scFv173-2-5-phoA*. Studies of *AraC/P<sub>BAD</sub>* was not included here since it required another host (DH10B) and since initial experiments indicated that this system (in contrast to T7) behaved very similar to *XylS/Pm* in the sense that transcript and protein amounts correlated well. At the total protein production level the analysis revealed more protein-specific effects compared to in the functional studies (Figure 2). In case of luciferase the amount of active protein was highest for *XylS/Pm* ML1-17 both according to activity measurements (see above) and deduced as soluble protein (Figure 3, Panel A). However, the very high level of transcription in the *Lacl/P<sub>T7lac</sub>* system resulted in a correspondingly big production of

insoluble and inactive luciferase protein, not seen to a comparable extent for any of the other systems.

For GFP and HGH (Panels C and D) production of soluble protein was very effective in both *XylS/Pm* ML1-17 and *Lacl/P<sub>T7lac</sub>*, and the final outcome at the protein level was more similar for these proteins than for luciferase. Generally, *Lacl/P<sub>T7lac</sub>* had an apparent advantage by its performance at the transcriptional level, but this potential was often not reflected at the translational level, such that the system often produced a vast amount of transcripts that were either translated into inactive protein or were not translated at all. Note also that the amounts of protein and transcript correlated well for *XylS/Pm* and *XylS/Pm* ML1-17 (except for *scFv173-2-5-AP*, Panel B), probably mainly because the amounts of transcript were generally much lower than for *Lacl/P<sub>T7lac</sub>* and therefore did not overload the translational machinery. It is also interesting to note that, in terms of both active and total protein produced, *XylS/Pm* ML1-17 and *Lacl/P<sub>T7lac</sub>* generally performed best. For *scFv173-2-5-AP* (Figure 3, Panel B) a more complex picture was observed, but this could be mainly related to the effects of toxic protein production on host growth or variability among the systems in the kinetics of induction [46].



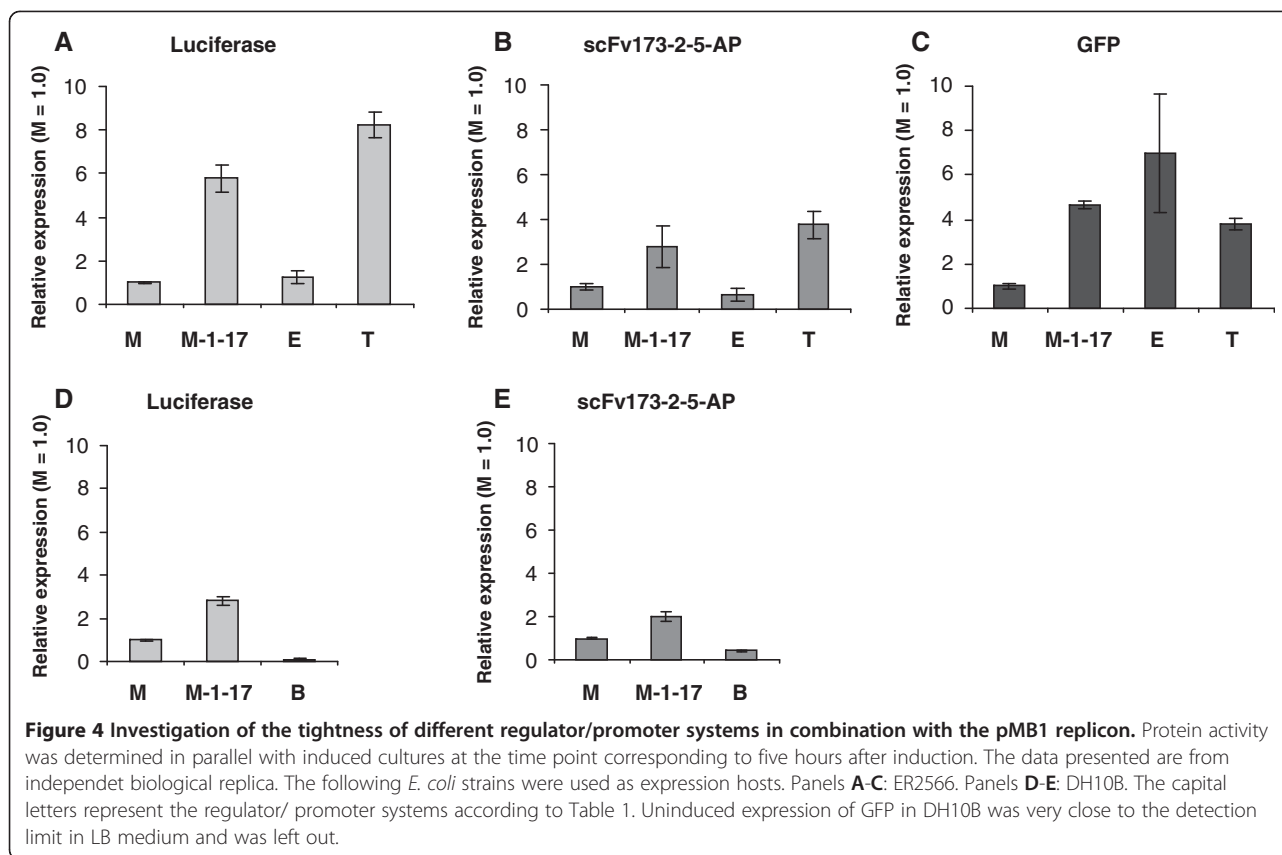
### Uninduced expression levels are highest for *Lacl/P<sub>trc</sub>* and lowest for *AraC/P<sub>BAD</sub>*

The tightness of the different regulator/promoter systems is another important feature, particularly for production of host-toxic proteins [47]. We studied this with the same set-up as for induced conditions, using luciferase, scFv173-2-5-AP and GFP, and as expected the background increased for all systems when the higher copy number vectors were used. The increase was in most cases approximately proportional to that of the plasmid copy number. Therefore, only findings collected from strains harboring pMB1-based plasmids are presented (Figure 4).

Generally, *Lacl/P<sub>trc</sub>* tended to be the leakiest system producing 3.8 to 8.2 times more active protein than *XylS/Pm* under uninduced conditions. Similarly, *XylS/Pm* ML1-17 displayed 2.8- to 5.8-fold higher background expression than the wild-type system. *AraC/P<sub>BAD</sub>* appeared to be, as expected, the tightest system giving rise to 0.1 and 0.4 times the background level for luciferase and scFv173-2-5-AP, respectively. *Lacl/P<sub>T7lac</sub>* was also quite

tightly regulated although it generated the highest background expression for GFP (Figure 4, Panel C).

The ratio between the induced and the uninduced expression levels was protein dependent with relatively small induction windows for svFv173-2-5-AP (1.2-25) and large for luciferase (60–3,000). In strain ER2566, *XylS/Pm* and *Lacl/P<sub>T7lac</sub>* displayed the highest induction windows, while *Lacl/P<sub>trc</sub>* was by far the least inducible system (0.1-0.2 times compared to *XylS/Pm*). In DH10B, induction ratios for *AraC/P<sub>BAD</sub>* were 1.3-27 times higher than the ratios of *XylS/Pm* and *XylS/Pm* ML1-17. These results are consistent with a previous report documenting that the induction ratio in the *AraC/P<sub>BAD</sub>* system can reach up to 1,200-fold when functionally compared for the *phoA* reporter gene [11]. As for *XylS/Pm* [24,25], the induction level can also be modulated over a wide concentration range by varying the inducer concentration. In addition, uninduced levels can be even further reduced by the presence of glucose, which represses the expression in this system [47]. The main disadvantage of the *AraC/P<sub>BAD</sub>* system is that the inducer can be metabolized in most strains of *E. coli*.



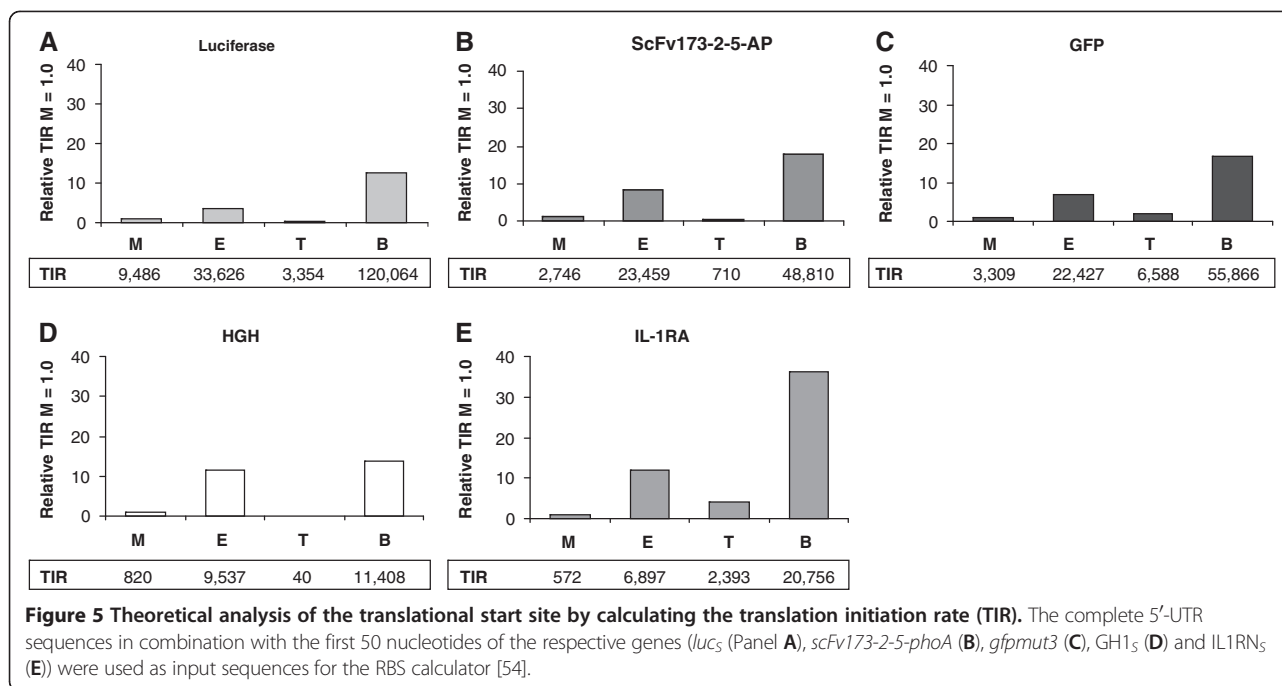
### The predicted translational efficiencies of the ribosomal binding sites vary over a wide range

The DNA region corresponding to the 5'-UTR plays a central role in regulation of gene expression [48-50]. It covers the untranslated nucleotides at the 5' end of the mRNA [51,52], including the ribosome binding site (RBS) that together with the translational start site influence expression [28,49,53]. One program frequently used to analyse the expected efficiency of these nucleotide sequences is the RBS calculator [54]. We applied its reverse engineering function on the various 5'-UTR-gene combinations used in the study and determined the translation initiation rate (TIR) values of the respective expression systems. The most striking finding was that the relative differences between the calculated TIRs of the four cognate RBSs were rather similar for all the five genes studied (Figure 5), although there were exceptions (see *LacI/P<sub>T7lac</sub>* for HGH and *AraC/P<sub>BAD</sub>* for IL-1RA). Generally, the calculator predicted that the TIR values of the *LacI/P<sub>T7lac</sub>* and the *AraC/P<sub>BAD</sub>* RBSs were higher than those of *XylS/P<sub>m</sub>* and *LacI/P<sub>trc</sub>* RBSs, suggesting a more efficient translation. The relative differences between the TIRs of the *LacI/P<sub>trc</sub>* and *XylS/P<sub>m</sub>* RBSs depended on the coding sequence.

To correlate the calculated TIR values with our experimental data is not straight forward because the total

protein levels are obviously also dependent on the efficiencies of the promoter sequences, which are not a part of the calculation of the TIR values. However, by comparing both transcript and protein amounts available from the data presented in Figure 3 such effects can at least partly be taken into account. The amounts of accumulated transcripts derived from *LacI/P<sub>T7lac</sub>* were generally highest and combined with a predicted more efficient TIR one might expect that this system would come out best at the protein level in all cases. However, this prediction was only in agreement with the luciferase data, and with the ScFv-173-2-5-AP and IL-1RA data to a more limited extent. In contrast, for GFP and HGH the experimental data did not support the prediction. It should also be remembered that efficient translation in itself may contribute to more accumulated transcript due to translation-mediated transcript stabilization [55,56]. For *XylS/P<sub>m</sub>* ML1-17 there appeared to be more protein per transcript compared to *LacI/P<sub>T7lac</sub>* and the total amounts of protein were at least equally good for this system, presumably indicating a better balance between the capacities of the transcriptional and translational systems. For *LacI/P<sub>trc</sub>* the calculator correctly predicted a very poor expression of HGH.

In general, it is possible to some extent to use the RBS calculator to predict which regulator/promoter system



would produce most protein. However, RBS function is just one among several parameters that affect the final protein production level. We have analyzed the previously reported very efficient UTR variants obtained by screening [28]. Despite the great stimulatory effect of these screened UTRs on protein expression (up to 20-fold), the calculator only predicted minor improvements relative to the wild-type sequence (between 1.5 and 3.6 times for the best variants).

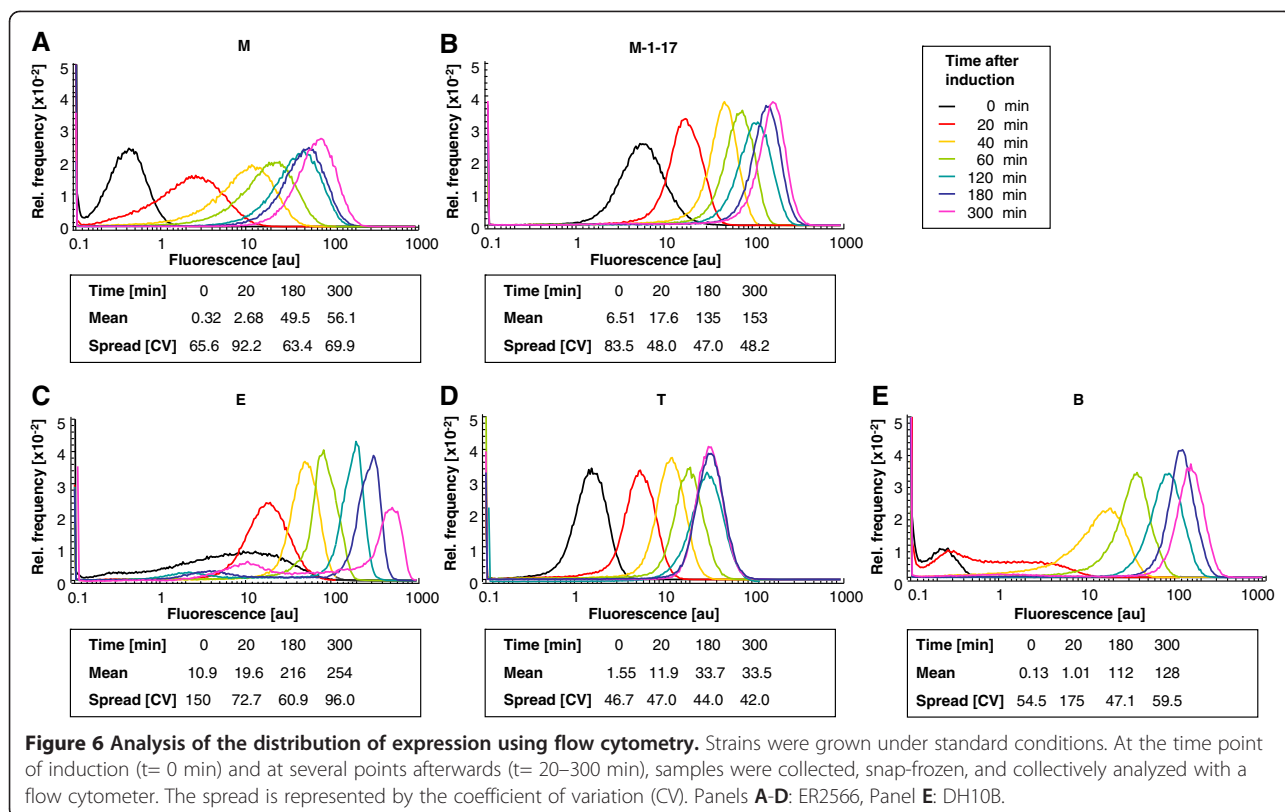
#### Flow cytometry analysis of GFP expression in individual cells revealed significant differences among the various regulator/promoter systems

Analyses of recombinant protein expression are mostly carried out at the level of cell populations, potentially masking significant differences in the level of expressed proteins between individual cells, which are known to occur [57,58]. If such heterogeneity exists it may represent another possibility for system improvement, e.g. by finding ways to reduce the fraction of cells with low expression level. This is also relevant in metabolic engineering projects involving metabolite flux control in biochemical pathways [59].

To analyze the level of homogeneity we used flow cytometry to quantitate GFP as it can be easily detected and because it was shown to be produced at high levels from the regulator/promoter systems used in this work, thus representing a relevant example in recombinant protein production. The fluorescence level, which reflects the number of GFP molecules, among the majority of cells harvested at a given time point typically varied

in a 5–10 fold range (Figure 6). In most cases, the fluorescence values fell within a signal peak, which moved to higher intensities with extended time after induction, as expected. The highest production levels were found in cells expressing GFP from *XylS/Pm* ML1-17, *LacI/P<sub>T7lac</sub>* and *AraC/P<sub>BAD</sub>* (where a different host strain was used), also consistent with what was observed at the population level. However, the analysis also revealed several new observations. For the two *XylS/Pm*-based systems the distributions were broader for the wild-type system (Figure 6, Panel A) than for *XylS/Pm* ML1-17 (Figure 6, Panel B), meaning that the promoter mutations improved culture homogeneity. The reasons for this are not clear but they might be related to differences in the efficiency of transcription initiation. Fluorescence distributions of cells expressing GFP from *LacI/P<sub>T7lac</sub>* (Figure 6, Panel C) were quite unique compared to those from the other systems. The expression profile at the time of induction is surprisingly broad in this system compared to the profiles of the remaining systems, possibly indicating low and varying (between individual cells) levels of T7 RNA polymerase production. Secondly, from two hours post induction onwards, two peaks became visible, one at rather low and one at rather high fluorescence values. The peak heights were also strongly reduced at the end. Most probably, the peak around lower fluorescence values late after induction reflects the formation of two subpopulations of cells as described by Zhao et al. [60], one being soluble GFP bearing and the other being dominated by inclusion bodies. Our findings also support those of a previous





report where GFP expression was studied from a pET vector context [4].

The  $\text{LacI}/P_{\text{trc}}$  system (Figure 6, Panel D) is characterized by a very even signal distribution throughout the entire induction period. Interestingly, the mean fluorescence remained constant already two hours after induction, possibly a consequence of a very fast activation of transcription after inducer addition in this system.

The  $\text{AraC}/P_{\text{BAD}}$  system, displayed a similar behaviour as  $\text{XylS}/P_{\text{m}}$  meaning that it takes an extended time until all cells are induced as reflected by a tail of the distribution towards low fluorescence values (Figure 6, Panel E). One hour after induction, the distribution fell into a single, rather narrow peak that was shifted towards higher fluorescence values over time.

The outcomes of the flow cytometry experiments showed that there is a quite big variation in GFP expression level among individual cells. By better understanding the factors controlling this variability it may become possible to improve expression at the population level. This conclusion is supported by the observation mutations in the  $P_{\text{m}}$  promoter region lead to more homogeneity.

## Conclusions

Development of efficient recombinant gene expression protocols is often based a lot on case-specific trial and error approaches, and the results reported here contribute to the

understanding of why. We have summarized the various observations (Table 3), and the  $\text{LacI}/P_{T7\text{lac}}$  system can be distinguished from all the other systems by its general tends to give rise to more transcript than all the other systems. The difference relative to the  $\text{XylS}/P_{\text{m}}$  system may be reduced or eliminated by incorporating more mutated control elements, but at the moment this would lead to very high levels of protein synthesis also in the absence of inducer [31]. Since we have shown before that even the wild-type  $\text{XylS}/P_{\text{m}}$  system can in some cases generate protein production at industrial levels it is clear that  $\text{LacI}/P_{T7\text{lac}}$  will only have an important advantage in those cases where the amount of transcript is the bottleneck. The experiments with GFP, HGH and IL-1RA illustrate cases where this has limited or no relevance (compared to  $\text{XylS}/P_{\text{m}}$  ML1-17). In contrast, for luciferase the amounts of transcript appears to be very important, but the potential in the  $\text{LacI}/P_{T7\text{lac}}$  system is in this case lost by the excessive production of inactive protein.  $\text{LacI}/P_{\text{trc}}$  generally has the advantage (for applications where this might be relevant) of a fast onset of protein production and a homogenous expression profile. However, both high levels of expression in the absence of inducer and comparatively low total production make it the least desirable if one is aiming at highest possible level of expression.  $\text{AraC}/P_{\text{BAD}}$  seems to be best with respect to tight regulation of the uninduced state coupled with high expression when induced. The  $\text{XylS}/P_{\text{m}}$  system has a big

**Table 3 Summary of the findings derived from the comparative expression study**

Category	Regulator/promoter system				References
	XylS/ <i>P<sub>m</sub></i> and <i>P<sub>m</sub></i> ML1-17	Lacl/ <i>P<sub>T7lac</sub></i>	Lacl/ <i>P<sub>trc</sub></i>	AraC/ <i>P<sub>BAD</sub></i>	
<b>Components</b>	XylS regulator <i>P<sub>m</sub></i> promoter (native or variant)	Lacl regulator <i>T7lac</i> promoter CAP binding site	Lacl regulator <i>trp/lac</i> hybrid promoter	AraC regulator <i>P<sub>BAD</sub></i> promoter CAP binding site	
<b>Strain requirements</b>	none	strain supplying T7 polymerase (and lysozyme) <sup>a</sup>	none	<i>araBADC</i> -/ <i>araEFGH</i> + strain	
<b>Medium requirements</b>	none	(glucose) <sup>b</sup>	none	(glucose) <sup>b</sup>	[3,10,11,42]
<b>Range of inducer</b>	0.001 - 2.0 mM	0.05 - 2.0 mM	0.05 - 2.0 mM	0.001% - 1%	[1,25]
<b>Expression level</b>	low - high	intermediate - high	low - intermediate	intermediate - high	This study
<b>Basal expression</b>	low - high	low - high	High	low	This study
<b>Induction ratio</b>	intermediate	intermediate-high	Low	high	This study
<b>Accumulated transcript</b>	low - intermediate	high	below detection - intermediate	intermediate	This study
<b>RBS strength</b>	weak - intermediate	intermediate - strong	weak - intermediate	strong	This study
<b>Homogeneity</b>	homogeneous populations	mixed populations	homogeneous populations	mixed populations	This study
<b>Recommended applications</b>	high level expression	high level expression	(high level expression) <sup>c</sup>	high level expression	
	expression of toxic proteins metabolic engineering	(expression of toxic proteins) <sup>c</sup>	(metabolic engineering) <sup>c</sup>	expression of toxic proteins (metabolic engineering) <sup>c</sup>	This study

<sup>a</sup> Expression of lysozyme, the natural inhibitor of T7 RNA polymerase, reduces the basal transcription from *P<sub>T7lac</sub>*.

<sup>b</sup> Supplementing glucose leads to catabolite repression which reduces basal transcription levels.

<sup>c</sup> Limited suitability. See 'Results and discussion' section for detailed information.

advantage of not being strain dependent in *E. coli*, and it is probably easiest to adapt to new bacterial hosts for cases where *E. coli* cannot be developed to perform in a satisfactory way. In summary we believe that the vectors developed for this study can be used as an efficient early test system for new proteins, perhaps by using XylS/*P<sub>m</sub>* ML1-17, Lacl/*P<sub>T7lac</sub>* and AraC/*P<sub>BAD</sub>*. The outcome of such a simple first experiment will probably often lead to an identification of the nature of the main bottleneck for this particular case, shortening the time from testing to development of a good production process. The further studies would involve a detailed analysis of parameters such as growth media composition, culture incubation temperature and host strain, which are known to affect recombinant protein expression at various levels.

## Methods

### Strains, standard DNA manipulations and growth conditions

*E. coli* DH5α (Bethesda Research Laboratories) was used for plasmid propagation during cloning steps. Recombinant DH5α strains were grown at 37°C in liquid Luria Bertani (LB) broth or on solid LB plates with appropriate

antibiotics (kanamycin 50 µg/mL; ampicillin 200 µg/mL). *E. coli* ER2566 (New England Biolabs, NEB) and *E. coli* DH10B (Invitrogen) served as expression hosts during the comparative studies. In comparison to the commonly used strain *E. coli* BL21(DE3), the former strain offers higher transformation efficiency for toxic clones and less background expression (NEB). All DNA manipulations were carried out according to standard procedures [61] or according to manufacturers' instructions. PCR was performed using the Expand High Fidelity PCR systems kit (Roche), and essential regions in PCR products were verified by sequencing. Functionality of the regulator/promoter systems was confirmed using *bla* as reporter gene determining the levels of resistance to ampicillin as described previously [62].

### Vector constructions

PCR primers used during various cloning steps are listed in Table 4. Plasmids used as templates or constructs that were generated in this study are listed in Table 1.

Construction of pSB-M2b: The region of pBAD\_gIII\_calmodulin containing the origin of replication from pMB1 was PCR amplified using primer pair Pwitw6\_badF

**Table 4 Oligonucleotides used in this study**

Name	Sequence (5' → 3')
a) PCR primers	
Pwitw4_AscI	AAAGTGAGGGCGCGCCGGTTGATGAGAG
Pwitw5_SpeI	ATCCACCCGGAAGTATGCCCTGCTC
Pwitw6_badF	AGACTAGTAAGCCCTCCCGTATCGTAGTTA
Pwitw6_badR	TGGCGCGCCAGATGCGTAAGGAGAAAA TACCG
ET_AgeI_fwd	GATGGCCCATATGATATCTCTCT
ET_NdeI_rev	GATCACCGTCCAGTGATCGAA
BAD_BbsI_fwd	GGCCTTTGCTCTCCCGCATCCGCTTA CAGACA
BAD_NdeI_rev	GACGCCCATATGTAATTCCTCCTGT AGCCCAAAAAACG
TRC_AgeI_fwd1	TGCATGTGTCACCGTTTTACCGTC
TRC_NdeI_rev1	GAGCTCGAATCATATGGTCTGTTTCCTG
pelB_fwd	AGTACATATGAAATACCTATTGCCTACG
Aphis_rev2	AGGATCCGAGCCTTTCTGTTTATTGATGC
b) qRT-PCR primers	
RT-synluc_fwd2	CCATGGCTTCGGCATGTT
RT-synluc_rev2	ACACGAAAGCCGCAATCA
gfpmut3_fwd1	CATGGCCAACACTGTCACT
gfpmut3_rev1	CTGCTTCATGTGATCTGGGTATCT
RT-hGH_fwd1	GCCTGTGTTTTAGCGAAAGCAT
RT-hGH_rev1	AGATTGCTTTTCTGCTGGGTTT
RT-IL-1-RA_fwd1	ATTGATGTGGTGCCGATTGA
RT-IL-1-RA_rev1	TCAGACACATTTTACCACCATGAA
scFv198_fwd	GAAGGGCCGGTTCACCAT
scFv255_rev	CATTTGCAGATACAGCGTGTCT
RT-16S-Fwd	ATTGACGTTACCCGAGAAGAA
RT-16S-Rev	GCTTGACCTCCGTATTACC

and Pwitw6\_badR. In parallel, pair Pwitw4\_AscI and Pwitw5\_SpeI was used to amplify pSB-M1b [31] excluding the RK2 *ori* (*trfA* coding region and the *oriV* origin of replication). After digestion with AscI and SpeI of both the amplified pMB1 *ori* and the pSB-M1b -resulting PCR product, the two fragments were ligated to each other resulting in plasmid pSB-M2b. The difference between copy-numbers of RK2- and pMB1-based plasmids was confirmed by agarose gel electrophoresis. Construction of pSB-P0b introducing different regulator/promoter systems: Three different regulator/promoter systems were chosen to substitute the region spanning *xylS/Pm* in pSB-M1b and pSB-M2b. The *lacI/P<sub>T7lac</sub>* region was amplified from pET16b using ET\_AgeI\_fwd and ET\_NdeI\_rev and inserted into the two depicted backbones using NdeI and AgeI, generating pSB-E1b and pSB-E2b. The *lacI<sup>f</sup>/P<sub>trc</sub>* region was amplified from pTrc99A using TRC\_AgeI\_fwd1 and TRC\_NdeI\_rev1 prior to insertion into pSB-M1b

and pSB-M2b using AgeI and NdeI, generating pSB-T1b and pSB-T2b. Finally, the PCR product covering the *araC/P<sub>BAD</sub>* region from pBAD/gIII\_calmodulin generated with the primers BAD\_BbsI\_fwd and BAD\_NdeI\_rev was inserted into the above mentioned backbones using BbsI and NdeI, creating pSB-B1b and pSB-B2b. In order to insert the *Pm* variant ML1-17 [27], pSB-M1b and pSB-M2b were digested with XbaI and PciI removing the *Pm* core promoter region which was replaced by two annealed oligonucleotides that constitute the double-stranded *Pm* ML1-17 fragment with XbaI and PciI compatible ends, creating pSB-M1b-1-17 and pSB-M2b-1-17. Introduction of other genes of interest: All pSB-P0b variants, except for pSB-B2b, were digested with NdeI and BamHI to excise the *bla* gene and to insert the *lucS* gene from pIB11-*lucS* instead, generating pSB-P0l variants. pSB-B2b and pSB-M1l were digested with NdeI and KpnI. The resulting DNA fragment corresponding to the pSB-B2 backbone and the *lucS* gene were ligated to each other to generate pSB-B2l. The *scFv173-2-5-phoA* gene was PCR cloned from pHOG-173-2-5-AP with primer pair pelB\_fwd and Aphis\_rev2. The enzyme combination NdeI and BamHI was used to replace the *bla* gene from pSB-M1b with the digested *scFv173-2-5-phoA* PCR product resulting in pSB-M1s. From there on NdeI and BamHI were used to generate all pSB-P0s variants, except for pSB-B2s. This construct was generated by digesting pSB-B2b and pSB-B1s with BamHI and ligating the pSB-B2 backbone with the *scFv173-2-5-phoA* BamHI digested insert from pSB-B1s. *gfpmut3* originating from pBAD24-GFP was inserted into the pSB-P0b variants using NdeI and BamHI with the exception of pSB-B2b. Instead, BamHI was used to excise the gene from pSB-B1g and to place it into pSB-B2 backbone (originating from pSB-B2l) to generate pSB-B2g. Genes *GHI<sub>S</sub>* and *IL1RN<sub>S</sub>* were excised from pMA-GH and pMA-T-IL-1RA with NdeI and BamHI and transferred to the pSB-P0b variants with the *Pm*, *Pm* ML1-17, *P<sub>T7lac</sub>* and *P<sub>trc</sub>* promoter using the same enzymes, resulting in pSB-P0h and pSB-P0r variants.

#### Growth conditions for comparative expression studies

The general cultivation protocol was based on recommendations published by the European Molecular Biology Laboratory (EMBL) [63]. For *E. coli* cultivations LB medium was chosen because it is widely used among molecular biologists and at the same time it was avoided to use media with glucose as a carbon source due to the influence of glucose on background expression from *P<sub>T7lac</sub>* and *P<sub>BAD</sub>* through catabolite repression [64]. A growth temperature of 30°C was applied for slowing down the growth rate of *E. coli*, as this generally leads to more soluble protein [65]. Initially the kinetics of protein accumulation was studied for all expression cassettes, using GFP (fluorescence) and luciferase (activity) as the main models.

The inducer concentrations and culture harvesting times post induction were varied and we found that five hours induction was sufficient to reach a plateau of accumulated protein per OD unit of cells. For GFP the accumulation rate was nearly constant (slightly lower from 3–5 hours) over this time-period. For most of the proteins it was complicated to follow the kinetics accurately since there was no quantitative method for measurement available, and in case of luciferase activity measurements may not necessarily correlate exactly with the accumulation kinetics of the insoluble fraction.

Recombinant *E. coli* ER2566 and DH10B strains were grown in 2 ml LB supplemented with 50 µg/ml kanamycin at 30°C over-night. Then 15 ml of LB with kanamycin in shake flasks were inoculated with the overnight culture to an initial OD<sub>600</sub> of 0.05. Following incubation at 200 rpm and 30°C expression was induced at OD<sub>600</sub> = 0.5–0.6 as follows: 2 mM *m*-toluate for strains harboring *P<sub>m</sub>*- based constructs, 1 mM IPTG for those with *P<sub>T7lac</sub>*, 0.2 mM IPTG for *P<sub>trc</sub>* and 0.015% L-arabinose for *P<sub>BAD</sub>*. Growth was continued for 5 more hours at 30°C.

#### Transcript analysis by qRT-PCR

At harvest, 0.5 ml of culture was stabilized with RNA protect (Qiagen) prior to freezing. The subsequent total RNA isolation, cDNA synthesis and relative transcript quantification by qRT-PCR was performed as described previously [28]. Primer pairs used during amplification are listed in Table 4. Transcript generated from the 16S rRNA gene was used for normalization.

#### Activity measurements of the different reporters

The luciferase assay was performed using the Luciferase assay System (Promega). At harvest, the cell culture was normalized to an OD<sub>600</sub> of 0.5. 90 µL of this mixture was supplemented with 10 µL of K<sub>2</sub>HPO<sub>4</sub>, pH 7.8, 20 mM EDTA prior to lysis with the Luciferase Cell Culture Lysis Reagent (CCLR, Promega). The remaining steps of the protocol were carried out according to the manufacturer's instructions except that the luciferase activities were determined from 10 µL lysed culture mixed with 50 µL of substrate. The alkaline phosphatase assay was performed as described previously [66]. Fluorescence measurements of strains expressing GFP were performed with the FLUOstar Omega instrument (BMG Labtech) together with the corresponding Omega Software. Fluorescence intensity was determined directly from the cultures using an appropriate filter set (excitation: 485 nm; emission: 520 nm). Values were normalized against the optical density. Data were acquired from three biological and thereof three technical replica.

#### Protein analysis by SDS-PAGE

For SDS-PAGE analysis 50 ml culture volume was used. Because of impaired growth of recombinant strains expressing scFv173-2-5-AP, 3xLB was used to get sufficient cell mass for analysis. The general growth conditions were as described above for the comparative expression studies. At harvest, bacterial pellets were washed with 0.9% NaCl and 100 mg pellet (wet weight) was frozen until further processing. Pellets were resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl, 8 mM MgCl<sub>2</sub>). The solution was sonicated using a Branson Sonifier DSM tip (sonication for 3.5 minutes on ice, duty cycle 35% and output control 3.0). Soluble and insoluble fractions were separated by centrifugation and treated with 62.5 U/ml Benzonase nuclease (Merck). Protein gels were run under denaturing conditions using ClearPAGE 10% gels and ClearPAGE SDS-R Run buffer (C.B.S. Scientific) followed by staining with Coomassie Brilliant blue R-250 (Merck).

#### Flow-cytometry

Cultures were grown essentially as described for SDS-PAGE analysis. At various time points after induction, 1 ml of culture was collected, supplemented with glycerol to 10% and snap-frozen in liquid nitrogen until further analysis. For single-cell fluorescence measurements, samples were thawed on ice and diluted in PBS. Flow cytometry was performed using the CyFlow<sup>®</sup> Space flow cytometer (Partec) equipped with a 488 nm blue solid state laser (200 mW) and a 536/40 nm band pass filter. For each sample, 150,000 events were collected at a rate between 800 and 2,000 events per second. Data were analysed with the Windows<sup>™</sup> XP FloMax(R) software (Quantum Analysis). The mean and spread (coefficient of variation (CV)) of the distributions were calculated over all collected values after gating.

#### Abbreviations

5'-UTR: 5'-untranslated region; qRT-PCR: Relative quantification real-time RT-PCR; scFv173-2-5-AP: Single-chain antibody fragment 173-2-5 alkaline phosphatase fusion protein; GFP: Green fluorescent protein; HGH: Human growth hormone; IL-1RA: Human interleukin 1 receptor antagonist; EMBL: European Molecular Biology Laboratory; SDS-PAGE: Sodium dodecyl sulfate- polyacrylamide gel electrophoresis; RBS: Ribosome binding site; TIR: Translation initiation rate; CV: Coefficient of variation; LB: Luria Bertani; NEB: New England Biolabs; IPTG: Isopropyl β-D-1-thiogalactopyranoside.

#### Competing interests

The authors declare no competing interests.

#### Authors' contributions

SB prepared all genetic constructs and strains, performed all experimental work on luciferase, GFP and HGH, participated in the design of the study and wrote the paper. VK performed the experimental work on scFv173-2-5-AP and IL-1RA and assisted in editing the paper. JM performed the analysis of the flow cytometry data. RL and SV conceived of the study. In addition, SV and TB participated in its design, coordinated the work and critically edited the paper. All authors read and approved the final manuscript.

## Acknowledgements

We are very grateful to Laila Berg for the gift of pIB11-luc<sub>s</sub>, Affitech, AS, Oslo for providing pHOG-173-2-5-AP and Vectron Biosolutions AS, Trondheim, for the gift of pMA-GH and pMA-T-IL-1RA. We thank Andrea Ebert for assisting with the collection of the flow cytometry data. This work was funded by a stipend that SB obtained from the Faculty of Natural Sciences and Technology, NTNU, Trondheim.

## Author details

<sup>1</sup>Department of Biotechnology, NTNU, Sem Sælands vei 6, Trondheim 7491, Norway. <sup>2</sup>Faculty of Physics and CeNS, LMU Munich, Munich 80539, Germany. <sup>3</sup>Department of Biotechnology, SINTEF Materials and Chemistry, Trondheim 7465, Norway.

Received: 9 November 2012 Accepted: 1 March 2013

Published: 18 March 2013

## References

1. Terpe K: Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 2006, **72**(2):211–222.
2. Lee SK, Keasling JD: A propionate-inducible expression system for enteric bacteria. *Appl Environ Microbiol* 2005, **71**(11):6856–6862.
3. Blatny JM, Brautaset T, Winther-Larsen HC, Haugan K, Valla S: Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. *Appl Environ Microbiol* 1997, **63**(2):370–379.
4. Choi YJ, Morel L, Le François T, Bourque D, Bourget L, Groleau D, Massie B, Míguez CB: Novel, versatile, and tightly regulated expression system for *Escherichia coli* strains. *Appl Environ Microbiol* 2010, **76**(15):5058–5066.
5. Menzella HG, Gramajo HC: Recombinant protein production in high cell density cultures of *Escherichia coli* with galactose as a gratuitous inducer. *Biotechnol Prog* 2004, **20**(4):1263–1266.
6. Makart S, Heinemann M, Panke S: Characterization of the AlkS/P<sub>alkB</sub>-expression system as an efficient tool for the production of recombinant proteins in *Escherichia coli* fed-batch fermentations. *Biotechnol Bioeng* 2007, **96**(2):326–336.
7. Giacalone MJ, Gentile AM, Lovitt BT, Berkley NL, Gunderson CW, Surber MW: Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system. *Biotechniques* 2006, **40**(3):355–364.
8. Wu H, Pei J, Jiang Y, Song X, Shao W: pHsh vectors, a novel expression system of *Escherichia coli* for the large-scale production of recombinant enzymes. *Biotechnol Lett* 2010, **32**(6):795–801.
9. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW: Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* 1990, **185**:60–89.
10. Amann E, Ochs B, Abel KJ: Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* 1988, **69**(2):301–315.
11. Guzman LM, Belin D, Carson MJ, Beckwith J: Tight regulation, modulation, and high-level expression by vectors containing the arabinose P<sub>BAD</sub> promoter. *J Bacteriol* 1995, **177**(14):4121–4130.
12. Schofield DA, Westwater C, Dolan JW, Norris JS, Schmidt MG: Tight regulation and modulation via a C1-regulated promoter in *Escherichia coli* and *Pseudomonas aeruginosa*. *Curr Microbiol* 2002, **44**(6):425–430.
13. Woestenenk EA, Hammarström M, van den Berg S, Härd T, Berglund H: His tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors. *J Struct Funct Genomics* 2004, **5**(3):217–229.
14. Sletta H, Tøndervik A, Hakvåg S, Aune TE, Nedal A, Aune R, Evensen G, Valla S, Ellingsen TE, Brautaset T: The presence of N-terminal secretion signal sequences leads to strong stimulation of the total expression levels of three tested medically important proteins during high-cell-density cultivations of *Escherichia coli*. *Appl Environ Microbiol* 2007, **73**(3):906–912.
15. Carrier T, Jones KL, Keasling JD: mRNA stability and plasmid copy number effects on gene expression from an inducible promoter system. *Biotechnol Bioeng* 1998, **59**(6):666–672.
16. Wong Ng J, Chatenay D, Robert J: Plasmid copy number noise in monoclonal populations of bacteria. *Phys Rev E* 2010, **81**(1):8.
17. Glick BR: Metabolic load and heterologous gene expression. *Biotechnol Adv* 1995, **13**(2):247–261.
18. Makrides SC: Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev* 1996, **60**(3):512–538.
19. Sørensen HP, Mortensen KK: Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol* 2005, **115**(2):113–128.
20. Tegel H, Ottosson J, Hober S: Enhancing the protein production levels in *Escherichia coli* with a strong promoter. *FEBS J* 2011, **278**(5):729–739.
21. Brautaset T, Lale R, Valla S: Positively regulated bacterial expression systems. *Microb Biotechnol* 2009, **2**(1):15–30.
22. Sletta H, Nedal A, Aune TE, Hellebust H, Hakvåg S, Aune R, Ellingsen TE, Valla S, Brautaset T: Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host-toxic single-chain antibody fragment in *Escherichia coli*. *Appl Environ Microbiol* 2004, **70**(12):7033–7039.
23. Blatny JM, Brautaset T, Winther-Larsen HC, Karunakaran P, Valla S: Improved broad-host-range RK2 vectors useful for high and low regulated gene expression levels in gram-negative bacteria. *Plasmid* 1997, **38**(1):35–51.
24. Winther-Larsen HC, Josefsen KD, Brautaset T, Valla S: Parameters affecting gene expression from the P<sub>m</sub> promoter in gram-negative bacteria. *Metab Eng* 2000, **2**(2):79–91.
25. Lale R, Berg L, Stüttgen F, Netzer R, Staßnes M, Brautaset T, Vee Aune TE, Valla S: Continuous control of the flow in biochemical pathways through 5′ untranslated region sequence modifications in mRNA expressed from the broad-host-range promoter P<sub>m</sub>. *Appl Environ Microbiol* 2011, **77**(8):2648–2655.
26. Vee Aune TE, Bakke I, Drabløs F, Lale R, Brautaset T, Valla S: Directed evolution of the transcription factor XylS for development of improved expression systems. *Microb Biotechnol* 2009, **3**(1):38–47.
27. Bakke I, Berg L, Aune TE, Brautaset T, Sletta H, Tøndervik A, Valla S: Random mutagenesis of the P<sub>m</sub> promoter as a powerful strategy for improvement of recombinant gene expression. *Appl Environ Microbiol* 2009, **75**(7):2002–2011.
28. Berg L, Lale R, Bakke I, Burroughs N, Valla S: The expression of recombinant genes in *Escherichia coli* can be strongly stimulated at the transcript production level by mutating the DNA-region corresponding to the 5′-untranslated part of mRNA. *Microb Biotechnol* 2009, **2**(3):379–389.
29. Tuller T, Waldman YY, Kupiec M, Ruppin E: Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci USA* 2010, **107**(8):3645–3650.
30. Loh E, Memarpour F, Vaitkevicius K, Kallipolitis BH, Johansson J, Sondén B: An unstructured 5′-coding region of the *prfA* mRNA is required for efficient translation. *Nucl Acids Res* 2012, **40**(4):1818–1827.
31. Zwick F, Lale R, Valla S: Strong stimulation of recombinant protein production in *Escherichia coli* by combining stimulatory control elements in the expression cassette. *Microb Cell Fact* 2012, **11**(1):133.
32. Megerle JA, Fritz G, Gerland U, Jung K, Rädler JO: Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys J* 2008, **95**(4):2103–2115.
33. Figurski DH, Meyer RJ, Helinski DR: Suppression of ColE1 replication properties by the Inc P-1 plasmid RK2 in hybrid plasmids constructed *in vitro*. *J Mol Biol* 1979, **133**(3):295–318.
34. Hoffman CS, Wright A: Fusions of secreted proteins to alkaline phosphatase: an approach for studying protein secretion. *Proc Natl Acad Sci USA* 1985, **82**(15):5107–5111.
35. Cormack BP, Valdivia RH, Falkow S: FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 1996, **173**(1 Spec No):33–38.
36. Shin NK, Kim DY, Shin CS, Hong MS, Lee J, Shin HC: High-level production of human growth hormone in *Escherichia coli* by a simple recombinant process. *J Biotechnol* 1998, **62**(2):143–151.
37. Goeddel DV, Heyneker HL, Hozumi T, Arentzen R, Itakura K, Yansura DG, Ross MJ, Miozzari G, Crea R, Seeburg PH: Direct expression in *Escherichia coli* of a DNA sequence coding for human growth hormone. *Nature* 1979, **281**(5732):544–548.
38. Schreuder HA, Rondeau JM, Tardif C, Soffientini A, Sarubbi E, Akeson A, Bowlin TL, Yanofsky S, Barrett RW: Refined crystal structure of the interleukin-1 receptor antagonist: Presence of a disulfide link and a *cis*-proline. *Eur J Biochem* 1995, **227**(3):838–847.
39. Steinkasserer A, Solari R, Mott HR, Aplin RT, Robinson CC, Willis AC, Sim RB: Human interleukin-1 receptor antagonist: High yield expression in *E. coli* and examination of cysteine residues. *FEBS Lett* 1992, **310**(1):63–65.
40. Studier FW, Moffatt BA: Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 1986, **189**(1):113–130.

41. Schlessinger MJ: Secretion of alkaline phosphatase subunits by spheroplasts of *Escherichia coli*. *J Bacteriol* 1968, **96**(3):727–733.
42. Vethanayagam JGG, Flower AM: Decreased gene expression from T7 promoters may be due to impaired production of active T7 RNA polymerase. *Microb Cell Fact* 2005, **4**(1):113–130.
43. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS: Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 2010, **329**(5991):533–538.
44. Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN: Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci USA* 2002, **99**(15):9697–9702.
45. Laursen BS, Sørensen HP, Mortensen KK, Sperling-Petersen HU: Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev* 2005, **69**(1):101–123.
46. Lim HK, Jung KH, Park DH, Chung SI: Production characteristics of interferon- $\alpha$  using an L-arabinose promoter system in a high-cell-density culture. *Appl Microbiol Biotechnol* 2000, **53**(2):201–208.
47. Saida F, Uzan M, Odaert B, Bontems F: Expression of highly toxic genes in *E. coli*: special strategies and genetic tools. *Curr Protein Pept Sci* 2006, **7**(1):47–56.
48. Chen H, Bjercknes M, Kumar R, Jay E: Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res* 1994, **22**(23):4953–4957.
49. de Smit MH, van Duin J: Translational initiation on structured messengers. Another role for the Shine-Dalgarno interaction. *J Mol Biol* 1994, **235**(1):173–184.
50. Barrick D, Villanueva K, Childs J, Kalil R, Schneider TD, Lawrence CE, Gold L, Stormo GD: Quantitative analysis of ribosome binding sites in *E. coli*. *Nucleic Acids Res* 1994, **22**(7):1287–1295.
51. Walker KA, Osuna R: Factors affecting start site selection at the *Escherichia coli* *fis* promoter. *J Bacteriol* 2002, **184**(17):4783–4791.
52. Lewis DEA, Adhya S: Axiom of determining transcription start points by RNA polymerase in *Escherichia coli*. *Mol Microbiol* 2004, **54**(3):692–701.
53. Vellanoweth RL, Rabinowitz JC: The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* *in vivo*. *Mol Microbiol* 1992, **6**(9):1105–1114.
54. Salis HM, Mirsky EA, Voigt CA: Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 2009, **27**(10):946–950.
55. Komarova AV, Tchufistova LS, Dreyfus M, Boni IV: AU-rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. *J Bacteriol* 2005, **187**(4):1344–1349.
56. Osterman IA, Evfratov SA, Sergiev PV, Dontsova OA: Comparison of mRNA features affecting translation initiation and reinitiation. *Nucleic Acids Res* 2012, **41**(1):474–486.
57. Raj A, van Oudenaarden A: Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 2008, **135**(2):216–226.
58. Eldar A, Elowitz MB: Functional roles for noise in genetic circuits. *Nature* 2010, **467**(7312):167–173.
59. Keasling JD: Gene-expression tools for the metabolic engineering of bacteria. *Trends Biotechnol* 1999, **17**(11):452–460.
60. Zhao Y, He W, Liu WF, Liu CC, Feng LK, Sun L, Yan YB, Hang HY: Two distinct states of *Escherichia coli* cells that overexpress recombinant heterogeneous beta-galactosidase. *J Biol Chem* 2012, **287**(12):9259–9268.
61. Sambrook J, Fritsch EF, Maniatis T: *Molecular cloning: a laboratory manual*. 2nd edition. New York: Cold Spring Harbor Laboratory Press; 1989.
62. Aune TEV, Bakke I, Drabløs F, Lale R, Brautaset T, Valla S: Directed evolution of the transcription factor XylS for development of improved expression systems. *Microb Biotechnol* 2009, **3**(1):38–47.
63. EMBL Protein Expression and Purification Core Facility. [www.embl.de/pepcore/](http://www.embl.de/pepcore/).
64. Miyada CG, Stoltzfus L, Wilcox G: Regulation of the *araC* gene of *Escherichia coli*: catabolite repression, autoregulation, and effect on *araBAD* expression. *Proc Natl Acad Sci USA* 1984, **81**(13):4120–4124.
65. Schein CH, Noteborn MHM: Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperature. *Nat Biotechnol* 1988, **6**(3):291–294.
66. Winther-Larsen HC, Blatny JM, Valand B, Brautaset T, Valla S: *Pm* promoter expression mutants and their use in broad-host-range RK2 plasmid vectors. *Metab Eng* 2000, **2**(2):92–103.

doi:10.1186/1475-2859-12-26

Cite this article as: Balzer et al.: A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in *Escherichia coli*. *Microbial Cell Factories* 2013 **12**:26.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

