Simone Balzer

Bacterial Recombinant Expression by the Positively Regulated XylS/*Pm* Promoter System - Comparative, Genetic and Physiological Studies on Modulation of Protein Production Levels

Thesis for the degree of Philosophiae Doctor

Trondheim, April 2014

Norwegian University of Science and Technology
Faculty of Natural Sciences and Technology
Department of Biotechnology

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Natural Sciences and Technology
Department of Biotechnology

ISBN 978-82-326-0152-3 (printed ver.)
ISBN 978-82-326-0153-0 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2014:114

Printed by NTNU-trykk

Acknowledgements

The scientific work presented in this thesis was carried out at the Department of Biotechnology at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. The work was mainly funded by a personal stipend from the Faculty of Natural Sciences (NTNU, Trondheim); some direct costs were paid by the three Norwegian Research Council-funded projects called 'A Combinatorial Mutagenesis Approach to Improve Microbial Expression Systems', ''Development of versatile bacterial expression systems for use in recombinant protein production, metabolic engineering, and systems biology' and 'Molecules for the future - Novel enzyme activities from environmental libraries'.

I am very grateful to have worked under guidance of Svein Valla, a supervisor with great research ideas who constantly helped out with constructive discussions, with his positive spirit and his ability to solve both scientific problems and personal conflicts. Thank you also for your encouragement to present my work at the two PEGS conferences.

I also thank my co-supervisor Research Director at SINTEF Trygve Brautaset for introducing me to the XylS/*Pm* expression tool, for believing in my abilities and for giving me opportunities along the way.

At this point, it feels like being the last successor in a row of people performing great research on mutagenizing the control elements of the XylS/*Pm* system and I could not have performed this work without their groundwork before my time at the Molgen group and their work achieved while we shared the labs. My special thanks goes to Rahmi Lale who stepped in as a co-supervisor in the last stages of my thesis and provided helpful advice during the whole thesis. I am also grateful to have collaborated with Veronika Kucharova, Laila Berg, Friederike Zwick, Hanne Jørgensen and Jon Andreas Lorentzen here at the Molgen group as well as Anne Tøndervik, Tone Haugen and my other (current) colleagues at SINTEF. Thanks also to Jørgen Skancke for valuable input on bioinformatics issues. Help provided by the staff at the Department of Biotechnology was also much appreciated.

Trondheim, January 2014
Simone Balzer

“You are worth far more than you think. Your work and presence on this Earth are important, even though you may not think so. Of course, thinking in this way, you might have many problems because you are breaking the Law of Jante – but don’t feel intimidated, go on living without fear and in the end you will win." – anti-Law of Jante (Paulo Coelho's Blog)

Abstract

Genetically engineered bacteria, especially *Escherichia coli*, find applications in both university and industry-based research; one important purpose being heterologous production of proteins. Experiences from empiric genetic engineering strategies in pursuit of maximizing or, generally speaking, modulating protein production levels have demonstrated that results are often unpredictable. Even after several decades of research in this field, the complex interplay of the many different genetic and physiological parameters that affect protein production levels are still not fully understood. Among the various expression vectors, mini-RK2 replicons containing the positively regulated *xylS*/*Pm* system can be used in *E. coli* and other Gram-negative bacteria. Using recombinant *E. coli* strains harboring these vectors as a basis, the influence of different genetic and physiological parameters on recombinant gene expression was systematically compared during this PhD project.

The first parameter studied was how different regulated promoter systems influence expression. Interestingly, a survey of the scientific literature indicated that the performance of commonly used expression systems such as LacI/P_{T7lac}, LacI/P_{trc}, and AraC/P_{BAD}, (and also XylS/*Pm*) had to a very limited extent been properly and systematically compared to each other. Such a comparison was selected as one objective of this PhD project. Given a common vector backbone and expression system insertion points, this comparison made it possible to understand the influence of the systems on production of different selected proteins. This study confirmed that no system was superior to meet all requirements an ideal system should have. However, it was possible to recommend certain systems for different expression purposes. The LacI/P_{T7lac} system, for example, was still best suited for achieving most total production judged by the vast amount of accumulated transcript and total recombinant protein. As for production of soluble (and active) protein, however, use of a variant of XylS/*Pm* (XylS/*Pm* ML1-17) or AraC/P_{BAD} were more beneficial for some proteins. The plasmid vector tools developed for this study may be used in future studies to analyze limiting factors for production of in principle any specific protein, including production in an active form.

The second genetic parameter included in the systematic studies was the DNA region corresponding to the 5′-UTR of mRNA. My findings demonstrate that 5′-UTR DNA sequences with strong RBS (like the P_{T7lac} UTR), could enhance protein (here: β-lactamase) production from XylS/*Pm* (7-fold) compared to the *Pm* 5′-UTR. Still, *Pm* 5′-UTR DNA

sequences that were obtained by combinatorial mutagenesis and screening approaches gave rise to even higher protein amounts (up to 20-fold). Also, the relative changes could not be predicted using current sequence analysis tools. Prior to this study, it was already shown that a 5′-UTR DNA region plays a key role in expression due to its involvement in transcription, translation and transcript stability. Therefore, it was hypothesized that one cannot study the effect of 5′-UTR DNA regions on protein production based on parameters influencing translation only (e.g. length of a Shine-Dalgarno sequence). To determine sequence features that influence transcription and to distinguish them from features influencing translation, combinatorial mutagenesis and screening using two efficient vector tools was applied. The new 5′-UTR DNA regions identified with these tools led to primarily stimulated transcript accumulation or protein production indeed, but it was not possible to identify positional hot-spots for mutations that specifically influenced either process. It was however possible to combine a 5′-UTR DNA region carrying mutations that primarily stimulated transcript accumulation with a 5′-UTR DNA region whose mutations primarily stimulated translation The total improvement achieved by this strategy was an impressive 170- fold compared to the native 5′-UTR. It can also be possible to adjust this 5′-UTR DNA region to other promoters and to other coding regions in the future. Hopefully, this strategy will enable more rational design in recombinant protein expression.

Among the remaining parameters that were varied, it was not surprisingly growth temperature, inducer concentration and plasmid (and indirectly gene) copy number that influenced recombinant protein production. This latter parameter also influenced culture heterogeneity at the single-cell level.

The last parameter that this work focused on was the qualitative influence of the expression host on the final product. Certain proteins are prone to be problematic for functional expression in *E. coli* even independent of the expression levels, the growth conditions or the protein engineering strategies. This was demonstrated by the example of the two difficult-to-express Norwegian Salmonid alphavirus E1 and E2 capsid proteins. Therefore, exploring the use of alternative bacterial hosts was considered; *Pseudomonas putida*, the organism XylS/*Pm* originates from, as well as cold-adapted members of the genus *Pseudomonas*. Vectors developed in this project were suitable to be directly transferred to these hosts; and at least for one protein (mCherry), even higher protein production levels could be achieved in *P. putida* compared to *E. coli*. The cold-adapted *Pseudomonas* strains also showed potential to be used as expression hosts for certain

proteins, however, slow growth and low expression levels demonstrated the need to engineer these novel hosts further in the future.

The findings presented in this work not only expand our current understanding of gene expression in general, but also help to approach the goal to optimize bacterial recombinant protein production in a more rational manner than is currently possible.

List of papers

Paper I [a] BALZER, S., KUCHAROVA, V., MEGERLE, J., LALE, R., BRAUTASET, T. and VALLA, S. (2013) A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in *Escherichia coli*. Microb Cell Fact, 12, 26.

Paper II [b] BALZER, S., LORENTZEN, J.A. VALLA, S. and LALE, R. Improving bacterial recombinant gene expression by separate optimization and combination of two functional parts of 5′-UTR sequences. (draft manuscript)

Paper III [c] TØNDERVIK, A.*, BALZER, S.*, HAUGEN, T., SLETTA, H., RODE, M., LINDMO, K., ELLINGSEN, T. E. and BRAUTASET, T. (2013) High production of recombinant Norwegian salmonid alphavirus E1 and E2 proteins in *Escherichia coli* by fusion to secretion signal sequences and removal of hydrophobic domains. Biotechnology and Bioprocess Engineering, 18, 742-750.

* joint first authors

Contributions

[a] As the first author, I contributed to this article by participating in the design of the story, planning of the experiments and methodology design, by carrying out most of the experiments (except for work on scFv173-2-5-AP and IL-1RA, VK), by participating in data analysis and in the interpretation of the results, by writing the first draft of the manuscript and by editing the final version.

[b] As the first author, I contributed to this article by participating in the design of the story, planning of the experiments, by carrying out most of the experiments (except for construction and validation of the artificial operon constructs (RL) and work on Figure 2 which was shared between JAL and myself), by participating in data analysis and in the interpretation of the results, by writing the first draft of the manuscript and by editing the final version.

[c] As joint first author, I designed the genetic constructs, planned and performed the experiments in shake-flasks including protein purification, participated in the interpretation of the results, wrote the first draft of the manuscript and participated in editing the final version.

Abbreviations

API	active pharmaceutical ingredient
CHO	Chinese hamster ovary
CoFi blot	colony filtration blot
CRP	cAMP receptor protein
CSP	consensus signal peptide
CTD	C-terminal domain
EC	elongation complex
EF	elongation factor
EMEA	European Medicines Agency
FDA	Food and Drug Administration
FP	fusion partner
GFP	green fluorescent protein
GOI	gene of interest
GST	glutathione S-transferase
HCDC	high-cell-density cultivation
HTS	high-throughput screening
IB	inclusion body
IC	initiation complex
IF	initiation factor
Inc group	plasmid incompatibility group
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITS	initially transcribed sequence
MBP	maltose-binding protein
MCS	multiple cloning site
MRS	multimer resolution system
NSAV	Norwegian Salmonid alphavirus
NTD	N-terminal domain
Nus	N-utilization substance
oriT	origin of transfer
oriV	origin of vegetative DNA replication
PIC	preinitiation complex
PTM	posttranslational modification
qPCR	relative quantification real-time reverse transcription PCR
RBS	ribosome binding site (contains the SD sequence and the start codon)
Rep protein	replication initiation protein
RF	release factor
RNAP	RNA polymerase
Rop protein	repressor of primer protein
SD	Shine-Dalgarno sequence (part of an RBS)
SEVA	Standard European Vector Architecture
SLIC	sequence- and ligase- independent cloning
SUMO	small ubiquitin-like modifier
TF	transcription factor
TIR	translation initiation rate
TM	transmembrane domain
TrfA	trans-acting replication function
tRNA	transfer RNA
UP	upstream promoter

UTR	untranslated region
xylS/*Pm*	*xylS* gene and *Pm* promoter region; specifically used to denote DNA region as part of an expression vector
XylS/*Pm*	XylS protein and *Pm* promoter region; general term for the expression system

Table of contents

1 Introduction

Bacterial recombinant gene expression is a very general term and implies the use cloning and transformation techniques to insert foreign DNA into a heterologous bacterial host which is thereby enabled to produce the corresponding recombinant product (mostly protein). Since the birth of recombinant DNA technology in the 1970th [1] this field has spread into many different research areas studying gene expression in species across the whole kingdom. Genetic engineering with the aim to modify a cell's biosynthetic machinery is a well-established practice. Motivating factors for further improvements of existing and development of novel expression platforms come from both academic research and different industries covering the biopharmaceutical, industrial biotechnology, agricultural, environmental, chemical and bioenergy sectors, with typical products ranging from monoclonal antibodies [2], laundry detergents [3], improved cellulose-degrading microorganisms [4] and biosensors [5] to antioxidant pigments [6] and food supplements such as amino acids [7]. In the following chapter, some of the most important applications of recombinant gene expression will be presented and special attention will be paid to classic use in overexpression of single genes as well as applications in other modern research disciplines. Afterwards, an overview over the current biological and computational tools used to engineer bacterial host strains will be given. In the second chapter, the focus lies on regulation of gene expression and how different genetic parts are combined to achieve certain expression outputs highlighting the role of the 5′-untranslated region (5'-UTR) due to its multiple functions in control of gene expression. This chapter will also explain some of the challenges associated with design and predictability of novel gene expression tools.

1.1 Applications of bacterial recombinant gene expression

1.1.1 Recombinant protein production

Many biomolecules that are of interest to the industry or academic research like active pharmaceutical ingredients (APIs) in pharmaceutical drugs or restriction enzymes for cloning are only available in insufficient quantities in nature and they needed to be extracted and purified from large amounts of starting material in the past. The development of recombinant DNA technology, however, enabled researchers to turn

bacteria [8] (and over time also species from all kingdoms [9,10]) into production factories for heterologous proteins and other biomolecules instead. As an example taken from academic research, structural biologists require up to several milligrams of pure proteins to resolve their three-dimensional structure by X-ray crystallography or NMR spectroscopy [11]. Independent of the final use, the basis is a suitable production host that is usually developed using genetic engineering strategies. The five most common expression hosts (based on number of registered publications) are the bacterium *Escherichia coli*, insect-cell lines derived from *Spodoptera frugiperda*, the yeast *Pichia pastoris*, Chinese hamster ovary (CHO) cell lines and the yeast *Saccharomyces cerevisiae* [12]. Looking at the pharmaceutical market in the year 2009, when this thesis was initiated, ~39% of the protein-based recombinant pharmaceuticals licensed up by the Food and Drug administration (FDA) and European Medicines Agency (EMEA) were produced by mammalian cells, ~30% by *E. coli*, ~19% by *S. cerevisiae*, ~11% by hybridoma cells leaving ~1% to transgenic goat milk and insect cells [13]. Over half of the industrial enzymes were made by yeasts and molds (e.g. *Kluyveromyces lactis*, *Saccharomyces cerevisiae, Aspergillus niger* and *Trichoderma reesei)*, with bacteria (e.g. *E. coli* and different *Bacillus* systems) producing about 30%. Animals provided 8% and plants 4% [14]. Finding an appropriate host for a protein production process is a multi-factorial challenge and the choice can be based on the requirement of the final product for example. The requirements can include protein yield, ability to secrete proteins, ability to correctly process the protein as well as time and effort required in the upstream and downstream processes [14]. For this work, the use of expression tools restricted itself to the important production host *E. coli*, and, as alternative hosts, both mesophilic and cold-adapted members of the genus *Pseudomonas*. Therefore important applications using mainly *E. coli* and other Gram-negative bacteria will be addressed.

What makes *E. coli* and other bacteria the preferred hosts are their rapid growth rates, their potential to be easily modified by well-characterized genetic tools, low media costs and potentials to achieve high protein production levels [10]. There are examples described where the accumulated heterologous protein accounted for a high fraction (up to 50%) of the total cellular protein [15,16]. These advantages are also reflected by the high fraction of publications and pharmaceutical proteins produced using this host. Maximization strategies for recombinant protein production at the DNA level typically involve gene design [17], choice of a regulated promoter [18], a suitable ribosome binding site (RBS) [19], a fusion partner (FP) for translocation [20,21], solubility enhancement [22],

specific detection or affinity purification [23] and adjustment of the gene dosage [24] amongst others [21]. Even though there are many tools available, advanced efforts can go into the direction of further improvements of these parameters like optimizing an RBS [25,26], improving the promoter [27,28], generating more efficient transcriptional regulators [29,30] and stabilizing the transcript [31]. With all these potential strategies in mind, one can envision the complexity behind the task to maximize production of recombinant proteins. And these efforts do not even include protein engineering, host strain engineering, optimization of the fermentation conditions and downstream processing efforts. Process developments often start broad from the parameter perspective and might even involve screening of mutant libraries. Current efforts for overproduction of proteins either thrive to make early efforts in process development amenable to high throughput screening (HTS) efforts and/or to make the early stages less complex by narrowing down the number of variants to test in an expression cassette.

1.1.2 Metabolic engineering

In metabolic engineering single enzymes, partial or even complete metabolic pathways are modified in an organism or even transferred from one organism to another to metabolize a novel compound, to create a novel product or to change the metabolic flow within a cell through modification of existing pathways [32]. The engineering part of this field is to predict which modifications to perform and to suggest a strategy to control the desired output. One important goal for example is to achieve a certain ratio between the numbers of enzymes present in a cell. This is desired in order to facilitate a balanced flow of metabolites through a metabolic pathway. As opposed to protein overproduction, metabolic engineering depends on production of adequate, not maximal, amounts of proteins. An important reason for this is that overexpression of certain genes within a pathway can cause metabolite drainage. In addition, bottlenecks can be caused for example when an enzyme is present in low abundance causing an intermediate product to accumulate (for a review see [33]). This might lead to a feedback response within the metabolic pathway, toxicity caused by an intermediate product or other undesired effects that can influence growth or the yield of the final product. To avoid imbalances, one of the most common strategies is to adjust the protein production levels by changing the regulatory elements of the expression cassette. As an example, genes coding for enzymes that act consecutively in a metabolic pathway can be arranged in synthetic operons and their expression levels can be controlled by adjusting the RBS, RNase sites

and the mRNA secondary structures within the intergenic regions [34,35]. Another successful strategy that has been applied was the coupling of different enzymes in appropriate ratios to a protein scaffold [36]. Whichever strategy might be chosen, this area of research will also benefit from advances in genetic engineering to make pathway engineering more rapid, less costly and more predictable in the future [37].

1.1.3 Synthetic biology

Current synthetic biology applications focus on the adaptation of engineering principles to biological systems to understand them better, to improve cellular functions or to create functions that are new to nature. One approach to understand biological systems is to develop genetic circuits that resemble electronic networks. In these circuits, signals might come from the extracellular environment or even from the cell's own metabolism in which case circuits were designed to form oscillators [38]. A good example for implying a circuit in a simple manner is one with a sensor function in which a cell is enabled to detect a signal from the environment and converts this event into a cellular response like expression of a fluorescent reporter gene. One more complex example that has caught attention in the scientific community is the recombinase example by Siuti et al. [39] in which logic and biologic memory were combined. In this case, two different input signals in form of two different molecules activate expression of two different recombinases. The respective recombinase recognition sites were placed around the genetic elements of an expression cassette consisting of a promoter, a terminator and a *gfp* reporter gene. In presence of the signaling molecules, the recombinases either excised or switched orientation of the genetic elements and thereby created a lasting effect at the DNA level. By arranging these elements in certain orders, different GFP readings could be observed depending on the logic functions implied in this gate. Despite the use of engineering terms, the basis is still recombinant gene expression where promoters, transcriptional regulators and reporter genes are connected to fulfill a logic function. Genetic engineering for synthetic biology application can aid in generation of more precise and reliable circuits by designing effector- and DNA-binding specificities of transcription factors, by modifying protein-protein interactions or by creating novel transcription factors with multiple functions.

1.2 The modern genetic engineering tool box with special emphasis on the particular tools used in this thesis

Over the past decades molecular biologists have developed a range of tools that represent the basis for modern genetic engineering purposes. Generally, bacteria harboring plasmids that contain an expression cassette (**Figure 1.1**) are used as platforms to express heterologous genes although integration of the expression cassette into the chromosome or cell-free expression systems without the cellular environment are possible alternatives [40,41].

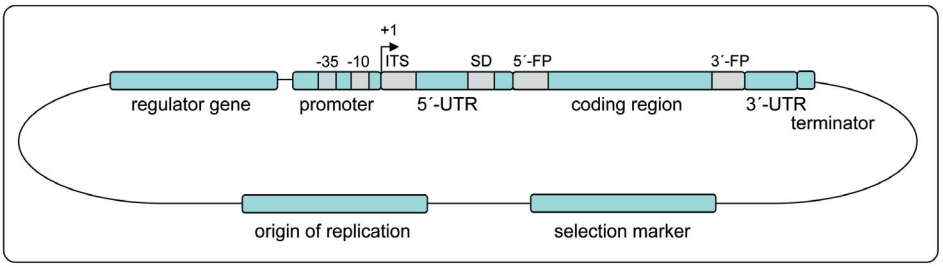

Figure 1.1: Overview of the arrangements of the most relevant genetic elements on a plasmid that can be modified to change expression of a gene of interest. Commonly the use of a regulated promoter is preferred. The product of the regulator gene (with its own promoter, 5′-UTR, 3′-UTR and terminator) might act positively (transcriptional activator) or negatively (transcriptional repressor) on transcription from the promoter. It binds to its own recognition sites within the promoter, upstream of the promoter or within the 5′-UTR. A transcript is formed which consists of the 5′- UTR - including the initial transcribed sequence (ITS) and typically a Shine-Dalgarno (SD) sequence - , the coding region of a gene of interest and the 3′-UTR. The translated target protein might have an N-terminal and/or C-terminal fusion partner (FP). The origin of replication determines the plasmid copy number and thereby the number of copies of the gene of interest. In bacteria, selection of the plasmid is usually accomplished by antibiotic resistance markers.

I would like to stress at this point that during this thesis the term ‘ribosome binding site’ (RBS) will be distinguished from the term ‘Shine-Dalgarno’ (SD) sequence. RBS will be used to describe a region at the mRNA level which is covered by the ribosome during translation initiation including the SD sequence and the start codon as well as nts within the 5′-UTR and the 5′ proximal end of the coding region.

The common arrangement of a regulated promoter, a 5′-UTR, a coding region and a terminator can be expanded by inserting multiple coding regions under control of the promoter. These so-called synthetic operons are beneficial in co-expressing several genes in accordance with naturally occurring bacterial operons. Another extension could

be to subject the promoter or regulator to a more complex control mechanism like the combination of the T7 promoter with the lac operator in pET vectors. Further options are to include a second origin of replication for plasmid maintenance in other hosts (in shuttle vectors) or to insert elements that facilitate transfer of an expression cassette to the chromosome. In order to understand the function of the different genetic elements, an overview over the most relevant replicon types, regulated promoters and modifications of coding regions of a gene of interest for this thesis will be presented.

1.2.1 Replicon types

Propagation of plasmids within a bacterial cell is facilitated by an origin of replication found on the plasmid, a replication initiation (Rep) protein which is mostly encoded by the plasmid as well as additional factors supplied by the plasmid and the host [42]. Which type of plasmid replication system is used determines the host range [43], the gene dosage [44], the compatibility with other plasmids [45], the stability within a bacterial population [46] and the heterogeneity by which the plasmid molecules are distributed amongst the daughter cells [47]. Concerning the host-range, there are replicon types that greatly rely on host enzymes during initiation and thereby often replicate in only a few hosts. Plasmids whose replication is initiated by host independent initiation factors on the other hand have a rather broad host range [42,48].

The mechanism by which the number of plasmids per cell is maintained directly determines the number of DNA copies of the gene of interest per cell which in turn affects the protein production level. This number depends both on the type of replicon and the host [49]. Interesting approaches to control the copy-number are not to use plasmids with a fixed range of copies per cell, but rather maintaining the plasmid at a single copy per cell until protein production is initiated and the copy-number is elevated [50].

Maintenance of two or more types of plasmid is not uncommon in recombinant gene expression. Coresidence of several plasmids is desired for example for coexpression of chaperones [51], rare tRNAs [52] or even other genes of interest [53]. Essential for co-existence of several types of replicons in one cell is that these belong to different incompatibility (Inc) groups [54].

Plasmid loss is a phenomenon associated with overproduction of recombinant proteins from plasmids with high copy-numbers [46]. Simultaneous maintenance of a plasmid and overexpression of the antibiotic resistance gene and the gene of interest

(GOI) impose a metabolic burden on the cell redirecting cellular resources towards the maintenance and processing of foreign DNA [55]. A consequence of this might be segregational instability leading to accumulation of plasmid-free cells in a bacterial community that often gain a growth advantage over plasmid-containing cells. In addition to selection markers, plasmid stabilization elements such as the *cer* locus or post-segregational killing systems like *hok*/*sok* can be inserted to ensure successful plasmid propagation to all daughter cells or killing of plasmid-free cells [56,57].

RK2

RK2 [58] and its derivatives belong to the IncPα group of plasmids which are known to replicate in a broad range of bacteria [59]. In order to make the replicon applicable for recombinant gene expression, the original RK2 plasmid was dramatically reduced in size. Derivatives of RK2, the so-called RK2-mini replicons [60], contain the essential elements for replication. These elements comprise the *trans*-acting replication function *trfA* gene encoding the replication initiation protein TrfA and *oriV*, the *cis*-acting origin of vegetative replication [61]. In addition these plasmids harbor the origin of transfer *oriT* which allows for the conjugative transfer of the plasmid to other microorganisms [62]. Host factors required for RK2 plasmid propagation in *E. coli* are the host initiation factor DnaA, the accessory HU protein, the helicase DnaB, the helicase accessory protein DnaC, the single stranded DNA-binding protein SSB, the primase DnaG, the DNA polymerase III and DNA gyrase [63,64]. The minimal replication origin *oriV* comprises a DNA region consisting of four DnaA binding boxes followed by five TrfA binding iterons, an AT-rich region and a GC-rich region. In order for replication to take place, TrfA and DnaA must bind to their respective binding sites for strand-opening at the AT-rich site to happen. A DnaB-DnaC complex is recruited to *oriV* leading to template unwinding (for more information, see [65]). Further details on formation of the replication complex and DNA synthesis are not known. TrfA has got a second important role besides replication initiation in which its ability to form dimers plays an important role. The handcuffing mechanism serves as a model to explain plasmid copy-number control. According to this model two RK2-based plasmids are coupled by direct interaction of iteron-bound TrfA molecules [61]. Thereby, further replication is inhibited through steric hindrance [65]. The copy-number of the original mini-RK2 plasmids is four - seven in *E. coli* and two - three in *P. putida* [66]. However, copy-up variants of *trfA* that are less sensitive to intermolecular coupling were identified [67] among which *trfA* cop271C (17-19 copies per cell [40,68]),

cop251M (31-33 copies per cell [68]) and cop254C (77-89 copies per cell [69]) were tested in our laboratories.

Besides the essential regions for replication, RK2 contains further control regions. These elements enable the plasmids to co-localize in specific subcellular positions, resolve multimers that are formed during replication and ensure that each daughter cell receives at least one plasmid copy [70,71]. Active partitioning systems encoded by *E. coli* plasmids make plasmids localize in characteristic mid- and quarter-cell positions of the bacterial cell while deletion of the corresponding DNA region results in random clustering of plasmids at the cell poles or the mid-cell position outside of the nucleoid [72,73]. The corresponding locus in RK2 comprises the O_{B1}*incC korB* region of RK2 [70]. This region is able to re-establish the quarter- and mid-cell positions when inserted into mini-RK2 replicons whereas only a marginal effect on plasmid stability can be observed. In contrast to RK2 namely, its minimal replicon derivatives are not stable. Stabilizing elements of RK2 are located in the *par* locus consisting of two operons [71]. The *parCBA* operon codes for a plasmid multimer resolution system while a post-segregational killing system is encoded by *parDE*. Many currently used mini-RK2 plasmids [27,29,40,74] neither contain an active partitioning system nor the original multimer resolution system known to affect plasmid stabilization [75]. Under normal laboratory conditions (shake-flask cultivation, small inserts) in presence of antibiotic selection markers, plasmid loss has not been a problem. However, when cells were cultivated under high-cell density conditions in absence of antibiotic selection or when large DNA fragments where inserted, plasmid instability was observed. Therefore, insertion of the *hok*/*sok* or *parDE* post-segregational killing systems in some our mini-RK2 plasmids was performed [68,76] and plasmid stability was regained. It has even been hypothesized that plasmid localization might affect protein expression [69].

ColE1-like replicons

The origin of replication found in plasmid pMB1 is closely related to the ColE1 replicon. It is also found in the well-known pBR322 plasmid [77] and commercially available plasmids from the pET (Novagen) and pBAD [78] series and is maintained at a copy-number of 15-20 per cell. Plasmids with a pMB1-based replicon can only be propagated in *E. coli* and other closely related *Enterobacteriaceae*. The only plasmid-encoded factors for replication of these plasmids are the primer RNA molecule RNAII, the incompatibility-mediating small RNA molecule RNAI and the small repressor of primer

(Rop) protein with additional inhibitory functions [79]. This means that plasmid replication purely depends on proteins supplied by the host cell. The plasmid-encoded RNAII is the only one of the three factors that plays a direct role in replication. It binds to a DNA region close to the origin of replication and this DNA-RNA hybrid serves as a substrate for RNaseH. RNaseH in turn cleaves RNAII and thereby forms a mature primer molecule for DNA polymerase I action. The remaining steps of initiation, melting and unwinding of the dsDNA, protection of single stranded DNA, elongation from the RNAII-based primer and termination are carried out by the host replication machinery. RNAI has a role in regulating the frequency of plasmid replication by hybridizing with the RNAII primer precursor and thereby blocking the RNAII-DNA binding. Rop influences the formation of RNAI-RNAII hybrids and thereby also functions as a copy-number control factor. Plasmids with a pBR322-derived replicon are believed to be distributed by random partitioning and were lost to a great extent under non-selective conditions [80,81].

1.2.2 Regulated expression systems

To be able to influence the timing and amount of expression, a heterologous gene is often placed under control of a regulated promoter. There are many different systems available in order to express genes in bacteria. Due to the many different expression purposes (section **1.1**), there is no ideal expression system on the market that has got all the desired qualities. Among the number of available systems, however, there are a few that were of particular interest for this work and these will be described in the following sections. The focus here is on the promoter and how expression from the promoter is controlled. Characteristics of the 5′-UTR and typical ribosome binding sites will be addressed later.

XylS/*Pm*

Pm is a positively regulated promoter that originates from the *P. putida* pWWO plasmid. This plasmid encodes proteins involved in the catabolism of aromatic hydrocarbons [82] and *Pm* controls transcription of one of two operons found on this plasmid. Inducibility of transcription from *Pm* is mediated by the transcription factor (TF) XylS. In nature, expression of *xylS* is controlled at two levels; constitutive low-level expression from the weak *Ps2* promoter and hyperexpression from the strong and regulated *Ps1*. The *Ps1* promoter is activated by aromatic hydrocarbons which bind to the TF XylR, the second regulator of the benzoate degradation pathway. This protein

dimerizes in presence of inducer leading to activation of transcription from *Ps2* [82]. In most of our in-house vectors, however, regulation of *xylS* expression was simplified by deleting the *Ps1* promoter meaning that *xylS* expression is not further regulated [60] (**Figure 1.2**).

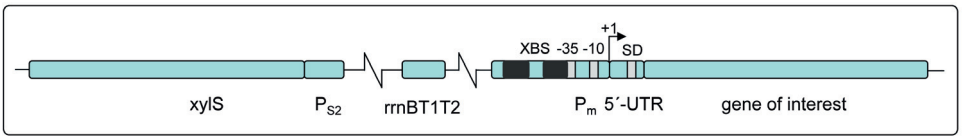

Figure 1.2: Schematic representation of the key DNA elements of the XylS/*Pm* expression system as found in the in-house set of vectors. The gene coding for the positive regulator XylS is constitutively expressed from the weak *Ps2* promoter. XylS can dimerize upon inducer binding and is thereby enabled to bind to two adjacent XylS-binding sites (XBS). The -10 and -35 *Pm* core promoter elements are recognized by σ^{32} and σ^{38} factors. The *Pm* 5′-UTR region includes a SD sequence involved in translation initiation. The bidirectional transcription terminator *rrnBT1T2* prevents any influence on transcription from *Pm* caused by transcription from other upstream promoters.

The XylS N-terminal domain (NTD) is responsible for inducer binding, establishing physical contact with the RNA polymerase α-subunit and facilitating dimerization [83,84,85,86] while the XylS C-terminal domain (CTD) makes contact with the DNA binding sites one of which overlaps with the -35 element of *Pm* [87]. The second layer of regulation of gene expression from *Pm* is accomplished by the involvement of different sigma factors depending on the growth phase. σ^{32} (σ^{H}) binds to *Pm* in the early exponential phase and σ^{38} (σ^{S}) in the late exponential or stationary phase [88]. The XylS/*Pm* system has been used successfully for different expression purposes ranging from overproduction of medically relevant proteins [20,68] and metabolic pathway engineering [89] to identification of novel FPs for the target protein [90,91]. Parts of this system have also been subjected to random mutagenesis and these comprise the regulator gene *xylS* [29], the *Pm* promoter core region [27] and the DNA region corresponding to the 5′-UTR [25,74,89] (**Figure 1.3**). The underlying screening procedure to identify the *xylS*, *Pm* and *Pm* 5′-UTR variants was based on expression of the *bla* gene (coding for β-lactamase) [92]. Varying expression of this gene leads to different amounts of the β-lactamase product and thereby causes the host cell to tolerate different amounts of the antibiotic ampicillin. It could be demonstrated that *bla* expression levels correlate quite well with the ampicillin concentration the cells tolerate [25,90]. *E. coli* strains harboring the *xylS*, *Pm* and *Pm* 5′-UTR mutant libraries were grown on agar plates with

increasing ampicillin concentrations and clones with for example the highest tolerance were picked and subjected to further analysis.

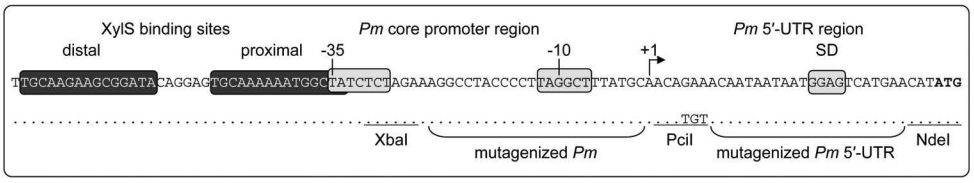

Figure 1.3: Sequence of the DNA region covering the *Pm* core promoter with its -10 and -35 elements, the distal and proximal XylS binding sites and the DNA region corresponding to the 5′-UTR. In the in-house vectors, a unique restriction site (PciI) was introduced at the 5′-end of the UTR DNA region to facilitate exchange of the *Pm*- and *Pm* 5′-UTR DNA fragments that were randomly mutagenized. *Pm* and *Pm* 5′-UTR variant elements can be conveniently combined using XbaI plus PciI and PciI plus NdeI, respectively. The transcriptional start site is marked with +1, the SD sequence is highlighted with a box and the nts making up the start codon at the mRNA level are typed in bold (derived from [25,27,87]).

In case of *xylS*, the gene was mutagenized using an error-prone PCR strategy followed by random shuffling of selected mutations. Target mutants stimulated transcription from *Pm* and thereby increased expression of the downstream gene *bla*. The novel XylS variants all contained substitutions in the NTD of the protein and two of them (XylS-StEP11 and –StEP13) could stimulate ampicillin tolerance 9.5 times compared to the wild-type in presence of inducer accompanied with a 4-times increase of the uninduced tolerance level. *Pm* promoter variants were identified using a different mutagenization strategy. A 24-bp region around the -10 element was replaced with an oligonucleotide mixture with a high doping frequency of the wild-type nt and low frequencies for the other three nts [25,74,93]. Screening for high ampicillin tolerance led to the identification of *Pm* variants that could enhance β-lactamase production by a factor of 14 [27]. A corresponding method was applied to randomly mutagenize the original *Pm* 5′-UTR. Expression of *bla* could be increased up to 7-fold at the transcript level and up to 20-fold at the functional protein product level [25]. At the other end of the scale, *Pm* 5′-UTR variants leading to reduced (down to 1.5% of the wild-type) β-lactamase production were identified [89]. A further way to increase expression from XylS/*Pm* is to place *xylS* under control of a different, stronger promoter and to supply the *xylS* gene *in trans* on a separate plasmid [94].

Advantages of this system are that the inducer is cheap, that inducer uptake is passive, that expression from XylS/*Pm* is activated in an inducer dose-dependent manner that this system can function in several bacterial species and that background expression is low when wild-type elements are used [95]. However, substantial leakiness can be observed when variant elements of *Pm*, the *Pm* 5′-UTR, *xylS* or combinations of these are used [40].

AraC/P_{BAD}

The AraC/P_{BAD} system is one of the most commonly used systems in recombinant gene expression and the pBAD series of vectors is commercially available (Invitrogen; [78]). The AraC protein acts both as an activator and repressor. Just like XylS, AraC consists of two domains, one responsible for DNA binding and one for effector binding and dimerization [96,97]. In absence of inducer, an AraC dimer is formed which connects the O_2 binding site found within the *araC* gene and the I_1 site found in the P_{BAD} promoter region (**Figure 1.4**) thereby forming a DNA loop [98,99]. AraC also regulates its own expression [100] at two levels; one comprises AraC binding to two O_1 half-sites within P_C which prevents transcription from this promoter and the other the above mentioned DNA looping process. AraC works as an activator protein in presence of L-arabinose, the signal which leads to rearrangement of the AraC dimer towards I_1 and I_2 binding and subsequent transcription from P_{BAD}.

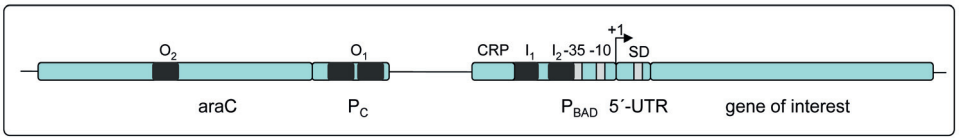

Figure 1.4: Overview over the DNA regions that make up the AraC/P_{BAD} system for L-arabinose- induced expression of a gene of interest. Transcription of the *araC* gene is controlled by the P_C promoter (derived from [97,101], Invitrogen). Genes under control of AraC/P_{BAD} are transcribed by the host-cell RNAP. Control of the system is achieved by AraC binding to different recognition sites; I_1 and I_2 in close proximity to the P_{BAD} core promoter -35 element, two O_1 half-sites in P_C and one O_2 half site located within the *araC* gene.

The third role of AraC is the regulation of expression of L-arabinose uptake systems [97] which makes regulation of AraC/P_{BAD} even more complex compared to XylS/*Pm*. L-arabinose enters the cell by active transport catalyzed by two transporters, one high-capacity, low-affinity L-arabinose transporter [102] and one high-affinity

transporter [103]. The glucose concentration in the growth medium adds a second level of control [104] which is mediated by cAMP and its receptor protein CRP. The CRP recognition site overlaps with the I_1 half-site. Low glucose levels coincide with high cAMP-CRP levels the complex of which can interact with the CRP DNA binding site to help rearrange AraC-binding to promote expression from P_{BAD}. Moreover, CRP positively stimulates P_C.

Consequences of the complex regulation of AraC/P_{BAD} reveal themselves at different levels. On the one hand tightness of the system is ensured which makes this system popular for expression experiments especially when leakage and subsequent toxicity would impair high production levels. At the same time, the promoter is strong enough to be used in protein overproduction and is even precisely adjustable when controlled expression for improved soluble expression is desired [78]. Disadvantages of the system are connected to catabolite repression and inducer uptake. In the former case, the important carbon source glucose, commonly used in growth media affects expression from AraC/P_{BAD}. In the latter case, the necessity for active inducer uptake leads to the all-or-nothing induction phenomenon which causes extensive heterogeneity of expression in cell populations. Adjustments of the expression system to render it more homogenous included L-arabinose transporter engineering [102,105]. Other modifications of the system were performed [106,107], but not as systematic as for XylS/*Pm* with exception of a more recent pursued strategy applied by a Danish iGEM team in 2011 [108].

LacI/P_{T7lac}

Even more heavily used among the expression systems is LacI/P_{T7lac} found for example in pET vectors (Novagen) [21]. It contains elements from the T7 system (T7 polymerase and its promoter) and elements from the *lac* promoter system (*lacI* repressor gene and its promoter, *lac* operator sequence). Supplying the T7 polymerase gene is usually accomplished by using an *E. coli* host strain with a chromosomally integrated T7 gene 1 (encoding the polymerase) for example under control of the L8-UV5 *lac* promoter (*E. coli* BL21(DE3) cells, Invitrogen) or the P_{BAD} promoter (*E. coli* BL21-AI™ cells, Invitrogen). LacI serves as a repressor by forming homotetramers that make contact with the operator sequence *lacO 1* in the 5′-UTR DNA region preventing T7 polymerase from binding to its promoter (**Figure 1.5**). In presence of an inducer like IPTG however, LacI dissociates from the *lac* operator making room for T7 RNAP. The mechanism of T7 RNAP binding to its promoter differs from the way *E. coli* RNAP established contact with a

promoter. In contrast to *E. coli* RNAP which consists of six subunits the T7 enzyme consists of only a single unit that recognizes a stretch of 17 bps without the requirement of host σ factors [109].

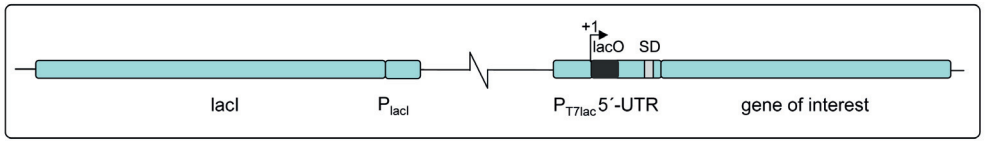

Figure 1.5: Depiction of the plasmid-borne genetic elements of the LacI/P_{T7lac} system. To make the system complete, a T7 polymerase gene needs to be supplied *in trans* from a further plasmid, a phage or the bacterial chromosome. Note the absence of typical *E. coli* -10 and -35 core promoter elements. The T7 RNAP recognizes a promoter sequence that is rather distinct from the *E. coli* consensus promoter making transcription from P_{T7lac} very selective.

This system is probably the one which underwent most development due to the many known problems associated with protein overproduction using it. A major drawback is the high leakiness which can be attributed to the removal of regulatory elements found in the natural *lac* system like the CRP binding site in direct proximity to the promoter controlling expression of the target genes or further *lacO* sites [110]. Reduction of leakiness was addressed by introducing T7-lysozyme-encoding plasmids to the cells, the inhibitor of T7 RNAP [111]. Slight reduction of basal expression can also be achieved when using *E. coli* strains with λDE3 lysogens due to the remaining responsiveness to catabolite repression despite the introduction of mutations within the L8-UV5 *lac* promoter that led to decreased sensitivity to glucose [112]. Another disadvantage is the high cost of the inducer IPTG. Costs can be reduced by using auto-induction media [113]. In spite of the efforts, some problems remain, namely the decoupling of transcription and translation [114], mutation of the T7 RNAP gene in the host strain [115] or limitation in cell growth associated with the high productivity of this system.

LacI/P_{trc}

LacI/P_{trc} is a second example for an expression system containing a *lac* promoter derivative (**Figure 1.6**). For once, both *lac* core promoter elements have been modified. The -10 regions in the P_{trc} promoter stems from the *lac* UV5 promoter which distinguishes itself from the natural *lac* promoter -10 region by two nts [110] making it a strong TATAAT site. The -35 region was taken from the *trp* promoter [116]. One further modification was

included in pTrc vectors, namely a mutation within the *lacI* promoter (*lacI*^{Q}) leading to increased *lacI* transcripts and thereby stronger production from P_{trc}.

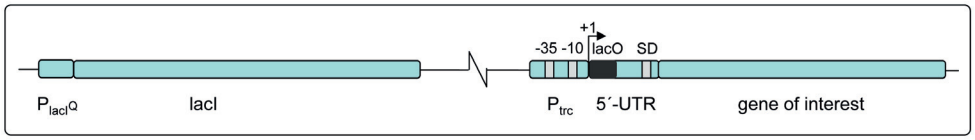

Figure 1.6: Schematic representation of the DNA regions forming the LacI/P_{trc} expression system. Most control elements from the natural *lac* promoter (CRP binding site, two additional *lacO* sites) have been removed.

Using this system, *E. coli* cells could accumulate up to 30% of the total cellular protein and it is generally regarded as a strong system [117]. However, leakiness is a drawback of this system which again can be attributed to lack of control mechanisms on top of repression via $LacI_4$-lacO. It is therefore not recommended to use LacI/P_{trc} for expression of toxic genes. The second disadvantage, namely the high costs of the IPTG inducer, can be circumvented by using temperature-sensitive mutants of LacI. Effects of the heat-shock response on the bacterial cell caused by elevated temperatures though needs to be kept in mind.

1.2.3 Genes of interest and their encoded protein products

Proteins represent the most important products of recombinant expression projects [12]. On top of regulation of gene expression by using certain promoter systems, the coding region, the amino acid sequence and the encoded function need to be considered to find a good production strategy. An often employed means is gene design which has become widely applicable today due to reduced costs and which serves different purposes [118]. Gene design is a term used to describe the modification of the coding region of a protein using synonymous mutations in favor of better expression output. Factors that are considered in gene design make use of the degeneracy of the genetic code and change the nt sequence in favor of the host's codon preference or GC content, for removal of unwanted sequences like restriction sites, repeats or promoter-resembling sequences and for reducing mRNA secondary structures [17]. More details on control of translation initiation and elongation will follow in section **1.3.2**. A strategy that involves

modification of the final amino acid sequence is the use of FPs at the target protein's N- or C-terminus. FPs can lead to solubility enhancement and/or tagging for simplified detection and affinity purification and typical examples are small ubiquitin-like modifier (SUMO), maltose-binding protein (MBP), glutathione-S-transferase (GST), thioredoxin, N-utilizing substance A (NusA), histidine tags of different lengths and the c-myc epitope tag. A further use of an N-terminal tag is general expression enhancement for example by displacing unfavorable secondary structure elements in the ribosome binding site [91,119]. Third, FPs for secretion of recombinant proteins to the periplasm are commonly used for minimization of inclusion body (IB) formation, reducing chances for protease degradation or improved disulfide bond formation. Yet another advantage of using a FP is the correct processing of the N-terminus by either automatic removal of the partner via the translocation machinery or use of specific protease recognition sites [120]. More drastic approaches are the redesign of the protein like deletion of complete domains [121].

1.2.4 Genetic tools for DNA modification

Current advances in cloning technologies speed-up the upstream stage of a recombinant protein production process by saving hands-on work, by enabling simultaneous connection of several DNA elements, by avoiding dependence on restriction enzyme recognition sites and expensive enzymes or combinations of these. Although traditional restriction cloning using type II restriction enzymes is still common (e.g. BioBricks [122], BglBricks [123] and Standard European Vector Architecture (SEVA; [124])) other methods have gained interest to standardize, parallelize and simplify connection of several different DNA fragments. Examples for a current cloning method includes use of type IIs restriction enzymes as in Golden Gate cloning [125]. This technique relies on single-cutters that cut outside the recognition sequence. Techniques that do not require restriction enzymes were also refined within the past years and include Gateway cloning using λ recombination [126] or the ligation-independent cloning (LIC) techniques Circular Polymerase Extension cloning (CPEC; [127]) and Isothermal Assembly (Gibson method; [128]) amongst others. Another very versatile, quick and easy LIC method is one-step sequence- and ligase- independent cloning (SLIC) [129] (**Figure 1.7**) which was also relevant for this thesis.

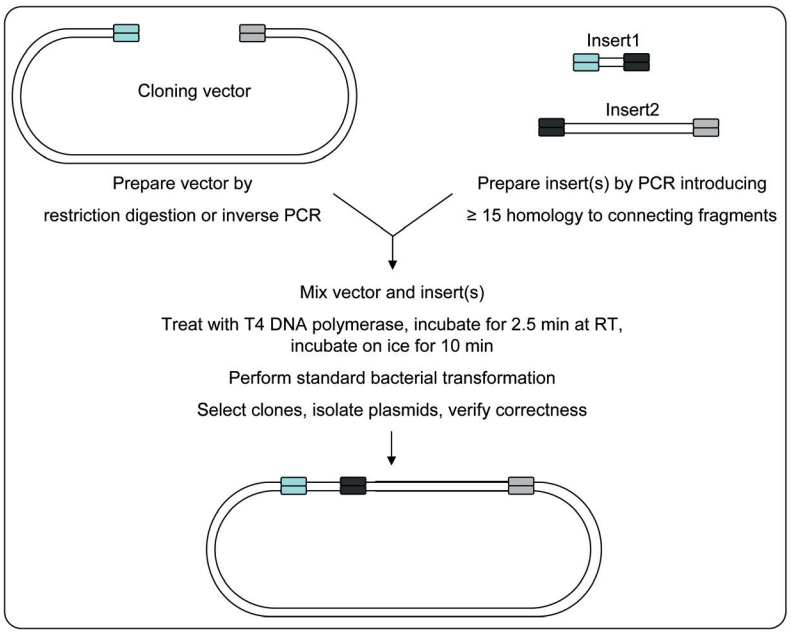

Figure 1.7: Schematic representation of the one-step sequence- and ligase-independent cloning (SLIC) method to connect several DNA fragments. Derived from Jeong et al [129].

1.2.5 Bacterial hosts

The bacterial cell creates the environment for the process of recombinant gene expression and it thereby influences the yield and quality of the protein product. *E. coli* is still one of the dominant production platforms [130]. *E. coli* strains have been engineered to meet certain requirements for recombinant protein (over)production. These cover combinations of characteristics for plasmids maintenance, expression and protein product quality where changes have been made either via chromosomal modifications or the addition of plasmids. Strains for modified expression in general include those that are deficient in Lon and OmpT proteases, those that co-express chaperones like DnaK, DnaJ, GroES, GroEL, Skp or ClpB which aid in folding of the target proteins and prevention of aggregation, those that create a reducing cytoplasmic environment for disulfide bond formation using glutathione reductase- (*gor*), glutathione synthetase- (*gshA*) and thioredoxin reductase- (*trxB*) deficiency and those that co-express tRNAs for rare codons

for improved translation efficiency (reviewed by Makino et al. [131]). Particular efforts on strain development have been performed for use of pET expression vectors and include integration of the T7 RNAP gene into the chromosome under control of different promoters (section **1.2.2**), addition of a gene encoding for T7 lysozyme or mutant variants of the LacY transporter for improved dose-response to IPTG [21]. Besides these classic strains, more modern approaches have been pursued for *E. coli*. Some highlights were the integration of post-translational modification machineries *via* insertion of pathways for N-glycosylation of eukaryotic proteins [132] or an enzyme responsible for acetylation [133] and the insertion of non-natural amino acids into proteins for expansion of the genetic code [134]. Other interesting engineering strategies were the generation of RNAseE-mutant strains for increased mRNA longevity (Invitrogen) or global strains engineering approaches [135].

A second bacterial host that was used in this thesis to study expression from XylS/*Pm* was *P. putida* KT2440, a strain cured for the TOL plasmid. Besides overproduction of proteins [136], this strain is utilized for bioremediation, production of fine-chemicals, bioplastics [137] and is currently studied to further exploit its metabolic and biotechnological capacities (Victor de Lorenzo's lab at CNB, Madrid, Spain [138]). A great advantage of this host is the compatibility with our in-house vectors (expression system and replicon-type), ease of genetic manipulation and biosafety.

1.3 Regulation of gene expression

Once a genetic construct is made and transferred to the host strain more complex regulatory mechanisms than encoded by the expression system itself act on expression of the gene of interest. To put it simple, the flow of genetic information during gene expression starts with transcription of the DNA region between the promoter and the transcriptional terminator into mRNA. The flow continues with translation of information encoded by the coding region into a protein which then maturates into its active state to fulfill its function. The process ceases with degradation of both transcript and protein. However, it is important to remember that for prokaryotic organisms such as *E. coli* the translation process is initiated as soon as the newly formed transcript 5′-end protrudes from the RNA polymerase. Also there is direct physical contact between the elements of transcription and translation involved. Furthermore, several RNAP complexes can transcribe one coding region and a multitude of ribosomes translate the information from

one transcript into several proteins simultaneously. At the same time, mRNA degradation might occur while translation takes place. All of these steps are highly regulated and there is no linear relationship between the number of genes of interest, mRNA and protein in the cell at a given time point due to complex regulatory mechanisms [139]. However, there is no doubt that changing one of the processes also has an impact on the following processes. In the following sections, the fundamental processes of transcription, translation and mRNA degradation are explained. These give a foundation for discussing challenges and pitfalls in genetic engineering approaches.

1.3.1 The process of transcription and its regulation

The transcription process can be divided into three major steps. During the first step, RNA polymerase binds to the promoter and initiates transcript formation. The second step involves processive elongation of the newly formed mRNA while termination is the third and last step. Among these, initiation of transcription is regarded as the key step in regulation of gene expression due to the possibility to prevent excessive transcript formation at an early stage and thereby limiting the use of cellular resources at subsequent stages [140]. The first prerequisite for transcription initiation to occur is the correct positioning of the core RNA polymerase at the core promoter -10 and -35 elements (and occasionally an extended -10 or an upstream promoter (UP) element) facilitated by σ-factors [141]. Depending on the promoter, different σ-factors (σ^{70}, σ^{38}, σ^{32}, σ^{28}, σ^{24}, σ^{19} (σ^{70} family) and σ^{54} (σ^{54} family)) are involved in promoter recognition [142] among which σ^{70} is the one responsible for transcription of house-keeping genes and σ^{38} and σ^{32} for stress-induced genes such as genes under control of *Pm* (see section 1.2.2). In this context, cellular stress can refer to entering the stationary growth phase (σ^{38}), heat-shock and subsequent accumulation of unfolded proteins (σ^{32}) or presence of certain chemicals like benzoic acid derivatives. As far as *Pm* is concerned the σ-factor recognition sites are not specific for σ^{38} or $\sigma^{32}$and transcription can take place in several growth phases [88]. In addition to the essential σ-factors, transcription factors belonging to the two general classes of activators or repressors determine whether transcription takes place or not. Action of these classes of DNA binding proteins is facilitated by small ligands which might be referred to as environmental stimuli in general or inducers in recombinant gene expression [143,144]. Another level of complexity is added when TFs possess dual functions which can occur naturally such as in AraC (section **1.2.2**) or artificially such as in

engineered LacI variants [145]. One further way of RNAP action modulation occurs via proteins that establish direct contact with RNAP, but lack DNA binding domains [146]. Following the first contact of the RNAP holoenzyme with a specific promoter sequence, an open complex is formed [147]. This happens spontaneously at σ^{70}–dependent promoters, while the action of enhancer-binding proteins is required for σ^{54}–dependent promoters [143]. The process of open complex formation is characterized by opening and unwinding of the double-stranded DNA around the promoter and up to +2 relative to the transcriptional start site followed by repetitive synthesis and release of abortive transcripts with a length of up to 15 nts [148,149]. The part of the DNA region corresponding to the 5′-UTR that matches the length of the abortive transcripts is called initially transcribed sequence (ITS). During the initiation phase, RNAP unwinds the DNA template while still being attached to the promoter which requires a pulling action, a phenomenon known under the term of DNA scrunching [150]. However, at some point, RNAP escapes the promoter and transcription enters the elongation phase. It is believed that σ-factors (at least σ^{70}) remain attached to the transcription machinery. How many full-length transcripts are formed relative to the amount of abortive transcripts is termed promoter escape efficiency. Termination of transcription is mediated in one of two ways. One way is protein-factor independent and involves intrinsic terminators which are hairpin structures that form in GC-rich DNA stretches which are followed by a stretch of U residues [151]. The other way is factor-dependent and requires a protein like the major termination factor Rho and its recognition sites (see [152] for more information).

Points of transcriptional control

Sequence-specific determinants for transcription initiation efficiency are not only found in the promoter region itself, but also reach further downstream. Naturally, promoter core elements, extended promoter recognition regions (extended -10 and UP elements) and other transcription factor binding sites themselves play a role in transcription initiation because their nt composition determines the affinity by which DNA binding domains of the TFs make contact with their specific recognition sites [153,154,155]. This is often referred to as basal promoter strength [143]. On top of that the distance between these sites is of importance as well as the composition of nts surrounding them. During transcription initiation, DNA bending takes place which is part of the correct positioning of the elements of the transcription initiation complex with respect to each other [121,156]. Changing the nt composition affects the flexibility of the DNA region and thereby affects the arrangement of the protein factors. Second, transcription involves unwinding and melting

of the double-stranded DNA which affords different amounts of energy depending on the nt composition [157,158]. Upon release of the initially formed transcript from its template, binding energies between the DNA:RNA hybrid might play a role (Jørgen Skancke, personal communication) just like in the elongation phase [140]. Not only is the DNA region in direct proximity to the promoter important for transcription initiation, but also the part of the DNA region corresponding to the 5′-UTR which lies beyond the initially transcribed sequence. The mechanism for this is unkown [25].

Other control mechanisms influence the elongation stage. For once the composition of nts in the DNA region corresponding to the mRNA determines the evenness of the polymerization speed [159]. The course of elongation is namely prone to compete with alternative pathways [140]. One alternative to elongation is pausing caused by transcriptional arrest signals encoded in the DNA sequence [160,161]. Biological functions for transcriptional pausing include the control of the overall rate of mRNA formation, the interaction with external factors and coordination of transcription and translation (more details follow in section **1.3.4**) [161,162,163]. Sequence elements that are involved in pausing include promoter-like regions that are recognized by σ^{70} [164,165,166]. A second and third alternative is editing and pyrophosphorolysis, processes which allow for correction of mis-incorporated nts through removal of a short oligonucleotide stretch or a single nt, respectively [140]. However, the rate of transcription elongation as a whole is believed to depend on the initiation efficiency [167,168]. Several RNAP complexes can transcribe from the same promoter and act positively on one another by pushing backtracked RNAPs forward given that transcription initiation is efficient enough.

Besides the DNA sequence, the polymerase itself determines the rate of transcription. Elongation happens at a speed ranging from 20-200 nts/s. More efficient transcription can be achieved by using a foreign RNAP. A prominent example is the T7 RNAP which catalyzes the coupling of ribonucleotides at a rate five- eight times faster than the *E. coli* host cell RNAP [21,114].

1.3.2 The translation process and its regulation

Translation of the information encoded within an mRNA occurs in the four steps of initiation, elongation, termination and ribosome recycling all of which are well-understood in bacteria thanks to decades of genetic and biochemical studies, including

crystallographic studies [18,169,170,171]. Initiation involves the assembly of the preinitiation complex (PIC) consisting of the mRNA, the 30S ribosomal subunit, initiation factors (IF1, IF2 and IF3) and the initiator tRNA fMet-tRNA^{fMet}. The first codon-anticodon interaction between the start codon (e.g. ATG) and the fMet-tRNA^{fMet} takes place. The 50S ribosomal subunit then loosely connects with the PIC, but GTP hydrolysis mediated by IF2 establishes a more stable 70S initiation complex (IC). One by one IF dissociates from the 70S IC and the elongation phases commences (summary derived from [170]) involving the two elongation factors (EF-G and EF-Tu) and the non-canonical factor EF4 (LepA) [172]. During elongation a constant of exchange of tRNAs within the ribosome takes place. Each tRNA is transferred from the acceptor A site in the ribosome to the peptidyl P site and the exit E site while new peptide bonds are formed. In this way, the transcript slides through the ribosome until the stop codon is reached. Translation termination is mediated by three release factors (RF1, RF2 and RF3) and dissociation of the ribosome requires additional proteins (ribosome release factor and IF3) [173]. The factors involved in translation are recycled at this stage and used in subsequent translation events.

Regulation of translation

Gene expression is highly regulated at the post-transcriptional level [174]. As for transcription, the initiation stage of translation is also regarded as the rate-limiting step. By changing translation initiation for example, protein production levels could be varied over two orders of magnitude [175]. The mRNA features that influence translation initiation have been summarized in an experimental review [176]. Obvious sequence features that affect translation are the kind of start codon, the length of the SD sequence, the distance between the SD sequence and the start codon, the position and strength of secondary structure elements in form of hairpins and the presence of an A/U-rich enhancer sequence. Sequence features that result in high expression levels are the use of an ‘AUG’ or ‘GUG’ start codon and stem-loop structures upstream of the SD-sequence or downstream of the start codon. Concerning length of the SD sequence and position relative to the start codon (measured as the distance between the center of the SD sequence and the first nt of the start codon), choice of a 8/7 or 6/10 SD sequence (length SD/distance to start codon) is of benefit for an efficient translation initiation.

Another perspective to our current understanding of transcription initiation involves initiation of translation in absence of the SD sequence. Only 57.1% of the *E. coli* genes

possess an SD sequence [177]. One way to explain why translation initiation in absence of SD sequences is possible is that absence of secondary structures around the start codon is sufficient to support ribosome binding. Initiation sites that promote translation in absence of an SD sequence contain an A/U-rich stretch devoid of stable secondary structures [169]. Moreover, presence of upstream-hairpins can have a positive effect by redirecting base-pairing motifs away from the initiation site. However, a conflicting theory exists which suggests that this unstructured region close to the start codon alone is not sufficient for translation initiation. Instead, it is believed that an SD sequence further upstream is required which is brought in proximity to the start codon *via* folding of the 5′-UTR [178]. Additional factors that regulate translation at the initiation stage are small RNAs like in the *hok*/*sok* post-segregational killing systems, metabolites found in riboswitches and temperature. Riboswitches are defined as elements within a 5′-UTR that are able to bind diverse metabolites and thereby repress or activate gene expression at the transcriptional and translational level [179].

In the elongation phase, pausing can be observed which is either attributed to pairing of internal SD sequences with the anti-SD sequence [180] or simply reduction of the translocation rates due to presence of rare codons [181]. By making use of the degeneracy of the genetic code, the nts in the coding sequence can be changed in favor of high translational efficiencies. SD-like sequences can be removed by using synonymous codons that interrupt this region. Mostly, a coding sequence is changed by introducing more frequently used codons, but it has been shown that introduction of rare codons at the 5′-end of the coding region (ramp theory) might lead to a more even distribution of ribosomes across the transcript by prevention of jamming [182].

1.3.3 Mechanisms of mRNA degradation and its regulation

At the same time as information encoded by an mRNA is translated into protein, mRNA degradation takes place. RNA degradation involves the action of ribonucleases (RNases) which can either cut internally (endonucleases) or attack the RNA from the termini (5′- or 3′- exonucleases) as well as accessory proteins performing 5′- and 3′-end modifications. The enzymes involved in mRNA degradation are often organized in a multiprotein complex called degradosome. Important enzymes found in *E. coli* that act mainly from the 3′-end are RNaseE, an endonuclease that targets A/U rich single stranded regions and commonly initiates mRNA decay, polynucleotide phosphorylase (PNPase),

RNase R and RNase II, three different 3′-exonucleases that require single stranded RNA and work in response to RNaseE action as well as poly(A) polymerase which couples adenines to the 3′-end of a stem-loop and thereby promotes its resolution and subsequent degradation [183,184,185]. The enzyme complex acting from the 5′-terminus contains RNA pyrophosphohydrolase which cleaves off pyrophosphate and which physically interacts with RNaseE and thereby enhances this enzyme's catalytic activity to cut the mRNA further downstream [186].

Generally speaking, the number of transcripts per cell affects the number of encoded proteins. Therefore, regulation of mRNA degradation is crucial for the fate of a transcript. Half-lives of bacterial mRNAs range from s to several min [187,188] making it a labile molecule. However, the turnover can be modulated in favor of a higher stability. Bacterial mRNAs have some intrinsic properties that ensure certain stability. For once, the triphosphate at the 5′-end stabilizes a transcript. Furthermore, translating ribosomes simply act by sterically hindering RNases from binding to their target. In addition, self-interaction of mRNA plays a vital role in particular at the 5′-end of the transcript and both existing as well as willingly introduced stable stem- loop structures can protect the mRNA from being degraded to a certain degree. Another major stem-loop structure can often be found at the 3′-end. Further control points involve the interplay between translation and transcription which will be explained below.

1.3.4 Transcription:translation coupling and its impact on transcript turnover

Several translation events can occur on one transcript, as has already been visualized by Miller et al. in 1970 [189] and translation is initiated co-transcriptionally. This means that changes affecting transcription also affect translation, especially in the rate-determining initiation step. Besides their connection to mRNA, direct physical contact between the transcription and translation machinery is established through protein interactions. The NusG TF is considered to be the central coupling factor between transcription and translation [190,191]. This protein is able to interact both with the RNA polymerase *via* its amino-terminus and the TF NusE *via* its carboxy-terminus (identical to the ribosomal S10 protein found in the 30S subunit). NusG has an important role in transcription elongation where it suppresses transcriptional pausing, involves in transcription antitermination, prevents backtracking of RNAP and consequently increases the elongation rate [161,192,193]. These functions are observed in translated transcripts.

In absence of translating ribosomes, however, NusG is able to bind to the Rho protein by which it exerts its role in transcription termination [194]. Another Nus factor with a dual function is NusA. Together with other Nus-factors this protein is involved in formation of stable ECs. Its second and major involvement, however, is in intrinsic termination due to the ability to recognize pausing and termination signals.

Understanding the structural basis for transcription:translation coupling gave rise to an updated version of the traditional schematic illustration of the translation process. The traditional view was that the transcription rate determines the speed by which ribosomes travel across the transcript. According to an updated model the speed of the ribosomes determines the processivity of transcription elongation not *vice versa* [195,196]. Moving ribosomes can accelerate RNAP. Absence of translating ribosomes on the other hand e.g. caused by rare codons stalls the TEC waiting for the ribosomes to connect with the RNAP again [197]. Alternatively, premature termination is induced in case translation is abolished e.g. by amino acid starvation. Considerations to why the two processes are coupled in bacteria are for once that the coupling prevents that non-functional transcripts accumulate in the cytoplasm [198]. Secondly, it is hypothesized that tightly coupled transcription and translation minimizes the chances for *de novo* DNA:RNA hybrid formation, so called R-loops, during transcription. However, uncoupling of the two processes can be observed [114].

A second important relation is the impact of translation on transcript stability. Iost and Dreyfus could demonstrate experimentally that efficient translation stabilizes a transcript [199]. This study supports earlier studies in which antibiotics targeting translation were utilized to study how translation inhibition affects mRNA stability [200,201]. In general, reducing translation initiation efficiency by introducing mutations around or within the SD site and thereby reducing the affinity of a ribosome to the translation initiation site destabilizes a transcript [202] and *vice versa* [203,204]. One accepted explanation is that translating ribosomes mask RNase sites, in particular within the 5′-UTR [205]. Interestingly, factors involved in translation and mRNA degradation share similar recognition sites. An A/U rich stretch around the SD sequence both leads to an unstructured region that promotes ribosome binding [206,207] and is associated with RNase E binding [208] implying that the binding affinity of the two factors has an impact on the fate of an mRNA [209]. Not only is ribosome binding around the SD sequence important for mRNA stability, but also the general distribution of translating ribosomes across the coding region [210].

How efficiently a protein is produced also depends on the amounts of σ factors, RNAP, nts, ribosomes, tRNAs, amino acids, energy equivalents amongst others available at the location of gene expression within the bacterium. Due to limited availability of the mentioned resources and high packing density of the molecules in the bacterial cytoplasm, excessive expression of a gene can lead to local depletion of resources [211,212] restricting total protein production. It has been hypothesized that better distribution of gene copies in a cell e.g. accomplished by high copy number plasmids that are randomly spread in the cytoplasm makes use of cellular resources more efficient which in turn leads to better protein production per gene copy [69].

1.4 Challenges during combination of genetic elements with focus on the central role of the 5′-UTR

A central goal of genetic engineering is to influence expression of a gene and thereby change production of the encoded protein product. Although engineering of the host-cell RNAP and other global changes of the transcription machinery to generate different cellular phenotypes are possible [213,214], more specific approaches to focus on a gene or a set of genes are desired. Over the past decades, it has become clear that gene expression involves a sequence of complex and highly regulated steps and that in order to gain control over the outcome of gene expression more than simply combining a promoter, a regulator, a 5′-UTR, a coding region and a terminator is necessary. This is of particular importance when even more complex tasks than expression of a single gene are pursued such as engineering of whole metabolic pathways (section **1.1.2**) or generation of regulatory circuits (section **1.1.3**). The DNA region corresponding to the 5′-UTR is of special importance due to its involvement in transcript formation, transcript degradation and translation. Currently, single genetic parts for design or redesign of systems are available which are often combined in a ‘plug-and-play’ manner to study gene expression [215]. Owing to complex regulatory mechanisms, however, these single parts can not be treated as context-free meaning that the function of the connected parts can not be purely derived from the sequence of the single parts alone. This leads to unpredictable effects on expression.

Table 1.1: The complex role of the 5′-UTR in gene expression. Positions of the nts are given relative to transcriptional start site (+) or to the translation start codon (-).

Process	Part of 5′-UTR	Involvement	Source
Transcription	nt +1 and +2	Initial melting during initiation	[141]
	nt +1 to +15	Promoter escape	[148,149]
	nt +10 to +24	Unknown	[74]
	variable	Repression of transcription (e.g. *lacO*/ LacI interaction)	
	nt -20 to -1	Backwards effect from translation to transcription	[196]
	whole sequence	5′-UTR secondary structure formation: - Premature termination mediated by metabolite binding	[216]
Translation	nt -20 to -1 incl.:	Translation initiation:	[217]
	- SD sequence	- Pairing with 16S rRNA	
	- Spacing SD-start codon	- Positioning of 30S complex	
	- secondary structure elements	- Availability of RBS (SD, start codon)	[218]
	whole sequence	5′-UTR secondary structure formation in riboswitches: - translation inhibition by metabolite binding (RNA, small molecules)[a]	[216]
Transcript degradation	RBS	Indirect effect on transcript stability through variable translation efficiency	[206]
	A/U-rich regions	Degradation mediated by RNase E	[208]
	5′-PPP	Protection from direct 5′-> 3′ degradation	[219]
	secondary structure elements	Stem loops at 5′-end protect from degradation	[203,220]

[a] A special role of the 5′-UTR is found in riboswitches and aptamer technology where external signals such as temperature, metabolites or proteins apart from the general expression machinery act on expression of a gene by influencing transcription or translation [216].

The influence that the parts have on each other has earlier been defined as 'context dependency' [221] and 'part-junction-interference' [222]. In case of the 5′-UTR, its influence can be seen from the perspective of its position within a stretch of DNA. To be more specific it forms the direct linear link between a promoter and a coding region (compositional context). But it is also subjected to the cellular environment with *in trans*-influence from external factors (host context; e.g. growth phase, strain) whose metabolism is influenced by the environmental context (e.g. temperature, pH, [221]). Some of the most relevant roles of the 5′-UTR in regulation of expression were already mentioned in the previous paragraphs. However, since this region was a topic of particular interest in this study, the key roles are summarized here (**Table 1.1**). Because of this complex role, choosing one particular 5′-UTR for expression of one gene probably does not have the same effect on expression of another gene due to the interference with the 5′-proximal coding region [218,223]. It has been proposed to physically separate the RBS from the remaining part of the 5′-UTR by insulators [222,224] that contain elements that mediate cleavage of the 5′-UTR to get a better control on the expression output. And indeed, these RNA-processing strategies led to the design of more predictable systems since relative strengths of promoters could be maintained for different genes by adjusting the RBS. To make this adjustment, both computational methods and experimental strategies are available.

Computational tools for prediction of expression levels and design of novel 5′-UTR elements are the RBS calculator [225], the RBS designer [226] and the UTR designer [227]. On the other hand, screening of mutant libraries is still a common approach [74,175]. Generally, strategies to improve expression by focusing on translation initiation are successful and their usefulness can be backed-up by the latest findings on transcription:translation coupling (section **1.3.4**). However, they neither cover the complete regulatory space nor lead to complete reliability of expression output. That is why more complex set-ups to address control of transcription and translation were generated in form of combining regulatory element libraries including different promoters and RBS [228]. These examples show that there still is groundwork to be done by molecular biologists to get a deeper understanding of gene expression that can be used in more targeted rational design of biological systems.

2 Aims of the study

The XylS/*Pm* system has been studied extensively over a long period of time in our research group. At the onset of this PhD work we knew that at least some proteins could be expressed at very high (industrial) levels with this system and that it in addition had certain unusual attractive properties such as broad host range and fine-tuning capacity. The group has tried to exploit this potential both for maximization of recombinant protein production and for metabolic engineering purposes, however, just like with other expression systems, results were often unexpected. In this PhD work the focus was to study the influence of different genetic and physiological parameters on recombinant protein production and to contribute knowledge on how to achieve a more controlled expression output. The specific goals were as follows:

- Many different bacterial expression cassettes are used in parallel in academia and industry, but we felt that there was no way of objectively comparing the performances of these systems based on the available scientific literature. It was therefore decided to set up a study which would allow a comparison of the core elements of commonly used expression cassettes, namely their regulated promoters, and systematically comparing their features directly in the same genetic background (replicon and host).
- The DNA regions corresponding to the 5′-UTR downstream of the various promoters are very different, but they can in principle be substituted in any promoter system. Since it is well established that 5′-UTR DNA sequences can have a profound effect on the total expression outcome we wanted to exploit the potential of designing 5′-UTR sequences that would work efficiently in the XylS/*Pm* system. The 5′-UTR design would be based on the idea that these sequences are important both for transcription and translation, and that these sequence features could be separated and optimized individually.
- A further genetic parameter that is known to affect recombinant protein production is the plasmid backbone. Characteristics of XylS/*Pm* would be determined by variation of the replicon type, plasmid copy-number and presence of plasmid stabilization elements.
- The XylS/*Pm* system would be utilized to control production of toxic proteins to analyze the effects of different host growth conditions (physiological parameters) on recombinant protein production.

- The last influence studied during this project would be the choice of bacterial host strain and the nature of the chosen mini-RK2 plasmid backbone containing the *xylS*/*Pm* system allows for the direct transfer of the vectors to different hosts like *Pseudomonas* in cases where *E. coli* might not be the suitable expression host.

This in-depth analysis of the promoter systems would hopefully lead to more general findings that could be applied for future recombinant expression strategies both to maximize protein production and to control expression at any desired level.

3 Summary of results and discussion

The core of this thesis is represented by the work on the positively regulated *Pm* promoter and its associated 5′-UTR region (**Figure 3.1**). Many findings are published in **Papers I-III**, but this section also supplies further (unpublished) data and discusses them with respect to recent developments published in scientific journals.

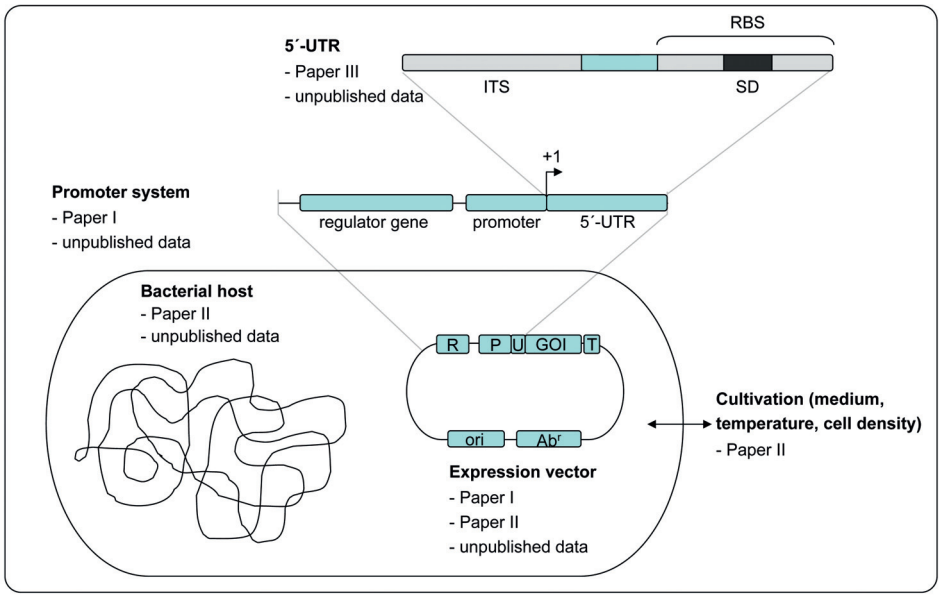

Figure 3.1: Visual representation of the topics covered by this thesis. The work in this thesis focused on recombinant gene expression in Gram-negative bacteria mainly using XylS/*Pm*, but also alternative regulated promoter systems with the goal to minimize trial-and-error attempts in the development of a protein production process. The abbreviations stand for: ITS - initially transcribed sequence, SD - Shine-Dalgarno sequence, RBS - ribosome binding site, R - regulator gene, P - promoter, U - 5′-UTR, GOI -gene of interest, T - terminator, ori - origin of replication, Ab^r - antibiotic resistance marker. Notice that for this thesis, RBS is defined as a stretch of ribonucleotides which is covered by the ribosome during translation initiation meaning that it includes the SD sequence, the start codon and also nts around these regions both within the 5′-UTR and the 5′ proximal end of the gene.

Within the following chapters, the most relevant results covering different approaches will be described and discussed. The first subchapter focuses on the results from a systematic comparison of features of commonly used regulated promoter systems at the transcript and protein level performed in *E. coli*. In the following chapter, work on the DNA

region corresponding to the 5′-UTR is summarized. This work involved random mutagenesis, rational and generic design approaches to study the complex role of this region at the transcriptional and translational level. Next, the work in this thesis also included studies on the influence of the plasmid backbone and the growth conditions (physiological parameters) on recombinant gene expression in *E. coli* (third and fourth subchapter). At last, the use of alternative bacterial hosts for recombinant expression under control of XylS/*Pm* was evaluated, namely *P. putida* and cold-adapted *Pseudomonas* strains (fifth subchapter).

3.1 A systematic study on the influence of different regulated promoter systems on recombinant protein production

Regulated promoters represent one important genetic key element to control gene expression. Despite the large number of such promoters available on the market, there are a few that are heavily used in heterologous expression in Gram-negative bacteria. These all have certain advantages and disadvantages making it impossible to name a system that is best for any purpose. Each of these promoter systems is often part of a certain expression vector like LacI/P_{T7lac} in pET vectors, XylS/*Pm* in pJB vectors, AraC/P_{BAD} in pBAD vectors or LacI/P_{trc} in pTrc vectors. Certain promoters even require a specific bacterial host like LacI/P_{T7lac} which requires an *E. coli* strain supplying a T7 polymerase gene. It is known that the genetic context (like the plasmid copy-number), the bacterial host or transcriptional terminators influence expression which means that a comparison of the features of regulated promoters from different vector backbones and hosts is not suitable to truly evaluate the characteristics of one promoter with respect to another. Surprisingly, comparisons among different promoter systems had only been performed using different vector backbones or describing features at a theoretical level when this work was initiated. And during the course of this work, only one systematic study had been published using LacI-regulated systems [229]. In this study the performances of different regulated promoters were compared by using carefully designed vectors that functioned as test systems for early stages of protein production (**Paper I**).

3.1.1 Design of a set of expression vectors enabling comparison of bacterial promoter systems

The first step of the vector development was to define the 'regulated promoter' DNA region and it was decided to extract the regions corresponding to the promoter, the 5′-UTR, the regulator gene and its promoter as well as the region between the regulator and the promoter including additional control elements (like CRP in pBAD) from commercial vectors and transfer this device to a common vector backbone. It would have been possible to also pick a common 5′-UTR region or at least a common RBS within the 5′-UTR for all promoters. However, it is known that there is a context dependency between the promoter and the 5′-UTR DNA region meaning that variations within the DNA regions corresponding to the ITS of the 5′-UTR DNA region can affect transcriptional start site selection and level of active target protein [25,230]. To ensure that the correct transcriptional start site was maintained and that the *lacO* region in two of the promoter systems (*lacI*/P_{T7lac} and *lacI*^{Q}/P_{trc}) was unchanged with respect to the promoter and the remaining downstream 5′-UTR DNA region, it was decided to keep the 5′-UTR DNA regions that were found in the original vectors. I also made sure that all coding regions started with 'ATG' and that gene sequences optimized for *E. coli* were used ensuring that expression was not limited by an unfavorable coding region (e.g. presence of rare codons, transcriptional pausing signals, strong mRNA secondary structures). For some commercial vectors like pET or pBAD, it is possible to insert genes of interest via a multiple cloning site (MCS). By using alternative restriction sites within an MCS different RBS are created changing the spacing and sequence composition between a SD sequence and a start codon. However, these sequence modifications can affect translation initiation and indirectly affect protein production [170,218]. Due to this, the default RBS present in pET16b, pBAD/gIII_calmodulin, pTrc99A and pTA16, respectively, was picked (see **Paper I**; Methods or **3.2.1** below). Common for all four 5′-UTR DNA regions was a slight modification just upstream of the start codon where the ‘CATATG’ site (for NdeI) was introduced. The well-known context dependency between a 5′-UTR and the 5′end of a coding region was not part of this comparison and was analyzed in separate studies (**Paper II** and unpublished data).

Another important consideration was with respect to which variant of the XylS/*Pm* system to choose. In our group, mutant variants of the regulator XylS [29], the *Pm* promoter core region [27] and its cognate 5′-UTR region [25] had been identified earlier leading to improved gene expression. One important reason why a promoter variant of

Pm, not a XylS or a *Pm* 5′-UTR variant, was chosen in this comparative expression study of regulated promoter systems was that uninduced expression levels were only increased by a factor of ~2 (vs. ~4.3 (XylS) or ~7.5 (5′-UTR)). Another reason was that enhancement achieved by *Pm* promoter variants was less gene-dependent than achieved by a 5′-UTR variant. In total, the DNA regions from the regulated promoters of *lacI*/P_{T7lac}, *lacI*^{Q}/P_{trc}, *xylS*/*Pm* and *araC*/P_{BAD} as well as a high level expression variant of *Pm*, *Pm* ML-1-17, were utilized and a set of 60 vectors was constructed (five promoter systems, two replicons, six GOI). Comparative expression studies using vectors containing the *lacI*/P_{T7lac}, *lacI*^{Q}/P_{trc}, *xylS*/*Pm* and *xylS*/*Pm* ML-1-17 regions were performed in *E. coli* ER2566 (T7 polymerase 1 gene positive) and studies using vectors with the *araC*/P_{BAD}, *xylS*/*Pm* and *xylS*/*Pm* ML-1-17 systems in *E. coli* DH10B (*ara* negative). For more details see **Paper I**.

3.1.2 Systematic study on the influence of the chosen regulated promoter systems on accumulated transcript and protein levels

Part of a systematic comparison was to find a good basis for evaluating the characteristics of the regulated promoters with respect to each other. Within recombinant expression, several steps during biosynthesis of proteins are important and these cover transcript accumulation, translation into recombinant protein in total as well as the fraction of soluble and active protein produced. These aspects differ from one another such that not all transcripts might be translated into protein [231] and that different forms of the protein might be present in the cell (folded vs. unfolded, processed vs. unprocessed, active vs. inactive protein). In recombinant protein production, it is generally desired to maximize the amount of folded, soluble, processed and active target protein which could be measured directly in case of reporter proteins or purified and used for more complex functional studies. Achieving this goal does not necessarily mean that expression should be pushed to the possible maximum, but requires a certain balance between efficiency of expression and ability of the protein to achieve its folded and active state instead [232]. One way to influence expression levels is to use different regulated promoter systems and the systematic comparison of some selected ones led to the following findings.

First, LacI/P_{T7lac} led to accumulation of the highest amount of transcript among the systems which was rather expected due to the very efficient T7 RNA polymerase [233]. However, the apparent advantage of this promoter system was not reflected at the protein activity level. When relative protein activity was compared among the different promoter

systems namely, it became evident that the XylS/*Pm* ML1-17 system was nearly as strong as or even slightly stronger than LacI/P_{T7lac}.

A second important observation was made when transcript data were correlated to the total amount of recombinant protein produced (soluble and insoluble protein fractions from protein gels together). According to the data, the excess of transcript produced by LacI/P_{T7lac} was in fact translated, but a distinct or even major fraction was accumulated as insoluble protein. This was especially evident for luciferase. For all the proteins studied, LacI/P_{T7lac} led to the highest accumulation of insoluble protein, but was also the system which was most likely to produce most total recombinant protein (**Paper I**, Figure 3). In some occasions, it might even be preferred to continue to work with the insoluble fraction in the downstream process, especially when this fraction makes up the major part of recombinant protein for example or when a protein is prone to aggregate and it is possible to refold the protein (**Paper III**). XylS/*Pm* and LacI/P_{trc} generally led to the least amount of total protein. However, the ratio between soluble and insoluble protein was higher compared to the two stronger promoters XylS/*Pm* ML1-17 and LacI/$P_{T7lac.}$

The interesting finding that luciferase tended to form insoluble aggregates when overexpressed led to the question whether the P_{BAD} promoter would have a similar effect as LacI/P_{T7lac}. Characteristics of the AraC/P_{BAD} system in relation to XylS/*Pm* and XylS/*Pm* ML1-17 at the transcriptional and translational level were performed in a different *E. coli* strain, DH10B, which is unable to catabolize L-arabinose (**Figure 3.2**).

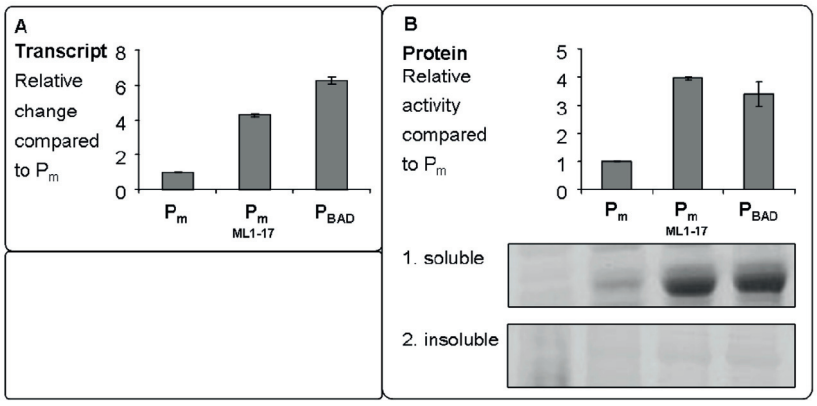

Figure 3.2: Luciferase production analysis in *E. coli* DH10B (unpublished data). Cell lysates from recombinant strains were subjected to qPCR (**A**), protein activity measurements (bar diagrams in **B**) and protein gel analysis (gel pictures of soluble and insoluble fractions in **B**). All data were related to the XylS/*Pm* system (values arbitrarily set to 1).

Data collected in this *E. coli* strain generally confirmed that it is more beneficial to use XylS/*Pm* ML1-17 compared to XylS/*Pm* which became evident from accumulated transcript data, activity data and the protein gel analysis. Although AraC/P_{BAD} led to a slightly higher amount of accumulated transcript compared to XylS/*Pm* ML1-17, similar amounts of active and soluble luciferase were detected. Virtually all recombinant luciferase was present in the soluble protein fraction.

3.1.3 Recommended applications of the regulated promoter systems in heterologous gene expression

To evaluate which of the promoter systems was the most suitable for expressing heterologous genes depends on the scope. If the goal is to produce as much protein in general, LacI/P_{T7lac} would be the system of choice. In this case, the efficient transcription is of advantage. In some cases, like shown for scFv173-2-5-AP, expression might be limited at the transcription stage and an efficient T7 polymerase was able to compensate for this limitation. For this protein efficient transcription was also beneficial for the soluble fraction of recombinant protein. Not coincidental, the LacI/P_{T7lac} system is the dominant promoter system used for recombinant protein production e.g. based on the total number of publications in the worldwide Protein Data Bank (wwPBD) data set, in which use of pET vectors is described [12]. On the other hand, if the amount of active protein is of higher importance, XylS/*Pm* ML-1-17 or AraC/P_{BAD} might as well be used. It is known that transcription and translation are coupled in bacteria [198] and my results indicate that these two systems maintain a better balance between transcript accumulation and translation indicating a better use of cellular resources. In contrast, the excess of transcript produced by LacI/P_{T7lac} can be interpreted by an uncoupling of the two processes. A third criterion would be the ratio between soluble and insoluble protein. Judged from the protein gels, a ratio in favor of soluble protein was achieved using XylS/*Pm*, AraC/P_{BAD} and LacI/P_{trc}, in some occasions also XylS/*Pm* ML1-17. In a previous study, expression from different negatively regulated promoters was compared in a direct manner, similar to this study [229]. Data collected by this group suggested that the weakest promoter tested (the *lacUV* promoter) produced the highest fraction of soluble protein, but that LacI/P_{T7lac} sill produced the highest amount of soluble protein in total followed by LacI/P_{trc}. This finding could be confirmed. At the same time more data were contributed which indicate that the use of the *Pm* promoter variant *Pm* ML1-17 or

AraC/P_{BAD} would yield similar amounts without producing excessive amounts of transcript. An alternative to a regulated promoter system that can be used to fine-tune expression levels is the rhamnose-inducible RhaR-RhaS*/rhaBAD* system [95]. RhaR-RhaS*/rhaBAD* is either utilized to control expression of T7 lysozyme to serve as additional control element for the LacI/P_{T7lac} system (Lemo system; Xbrane Bioscience AB), but also for recombinant protein production directly ([234,235]; Lucigen). This trend supports the idea of incorporating several regulated promoters in a portfolio of bacterial expression systems in order to maximize chances to achieve high expression levels; at least one system like XylS/*Pm* for proteins that require careful adjustment of expression levels (see **Paper III**) and one system like LacI/P_{T7lac} for easy to express proteins which can comprise a high fraction of the total cellular protein without forming insoluble aggregates or affecting cell physiology noticeably. Further findings and recommendations for expression purposes beyond maximization of recombinant protein production are summarized in **Paper I**, Table 3.

3.2 Demonstration of the complex role of the 5′-UTR DNA region in bacterial recombinant expression

Several genetic parts of the expression systems influence efficiency of transcription and translation and thereby the amount of target protein. Besides promoter strength and RNA polymerase efficiency, the nt composition of the 5′-UTR DNA region influences the processes of transcription and translation. A tool to evaluate the sequence of a 5′-UTR DNA region with respect to translation initiation efficiency is the RBS calculator [218]. This tool is based on a biophysical thermodynamic model for translation initiation. It takes into account the region around the SD sequence and the 5′ proximal coding region and calculates translation initiation rates (TIRs). TIRs were calculated for the regions around the 5′-UTRs of the different promoter systems and the 5′ proximal coding region transitions of the five different genes used in order to compare strengths of the RBS (**Paper I**, Figure 5). Due to the complexity of the regulation of the steps involved in protein biosynthesis, it was challenging to evaluate the actual effects of the calculated TIRs on expression of the different genes in most cases. That is why it was decided to extract the 5′-UTR DNA region as a part and to study effects of the region alone on gene expression to get a deeper understanding of its multiple roles and to open up new possibilities for rational design.

A body of scientific evidence has been collected in the past decades demonstrating the versatile and complex role the 5′-UTR in recombinant gene expression (section **1.4**). Previous findings in our group showed that creation of mutant libraries covering the 5′-UTR and screening made it possible to identify *Pm* 5′-UTR expression variants which either led to decreased or increased final protein product levels compared to the wild-type 5′-UTR sequence. In other words, mutagenization and screening could be applied to cover a big expression window simply by mutating this DNA region. This was possible because the mutations that were inserted into the *Pm* 5′-UTR DNA region affected both transcription and translation [25,74,89]. However, screening combinatorial libraries had certain disadvantages. For once, the effects the mutations had on gene expression were quite gene-dependent. As an example, the LII-10 *Pm* 5′-UTR variant led to ~20-fold increased β-lactamase production, but only ~1.5-fold increased phosphoglucomutase and luciferase production. Also, mutations were spread over the entire 5′-UTR DNA region which made it impossible to identify specific mutations that either led to increased transcript accumulation or translation efficiency. For future genetic engineering approaches, however, a more reliable, less gene-dependent and less time-consuming method to adjust a 5′-UTR DNA region to improve expression levels would be desired. Using the *Pm* promoter and its cognate 5′-UTR DNA region as a model system, the goal of designing a sophisticated 5′-UTR region was approached in three steps during the PhD project. The first was to verify the actual need to design better 5′-UTRs for expression by comparing different strategies to improve a 5′-UTR DNA region (**3.2.1**). Second, a successful attempt was made in identifying *Pm* 5′-UTR DNA sequences that either affected transcription or translation (**3.2.2**). Third, a novel 5′-UTR region was designed in which the knowledge from previous 5′-UTR modification attempts was applied (**3.2.3**).

3.2.1 Evaluation of different strategies to increase gene expression by changing the 5′-UTR DNA region

Despite the knowledge about the role of the 5′-UTR region in gene expression, it is still common practice to clone a gene behind a promoter and its associated 5′-UTR or a 5′-UTR with generally good features with respect to translation initiation without making adjustments in favor of a good 5′-UTR - 5′-coding region context. In order to demonstrate the need to make more rational choices when it comes to connecting 5′-UTRs and coding regions, it was decided to compare different strategies to improve a 5′-UTR with respect to

expression levels. We hypothesized that 5′-UTR DNA sequences identified by screening would lead to higher expression levels than simply using 5′-UTRs with strong RBS. To test this hypothesis the 5′-UTR regions from the regulated promoter systems studied in the previous section (**Table 3.1B**) were extracted and transferred to mini-RK2 plasmids containing *xylS*/*Pm* replacing the original *Pm* 5′-UTR DNA region.

Table 3.1: Sequences of the 5′-UTR variants that were used to compare their effects on *bla* expression. The SD sequences are written in bold face, the PciI (= BspLU11I) site at the 5′-end (not present in **B**) and the NdeI site at the 3′-end at underlined. The nts corresponding to the start codon are double underlined and the nts deviating from the *Pm* wt 5′-UTR DNA sequence are highlighted in grey, except for the sequences depicted in **B** which differ both in length and sequence composition.

A *Pm* wt 5′-UTR DNA sequence

Pm wt	AACATGTACAATAATAAT**GGAG**TCATGAACATATG

B 5′-UTR DNA sequences present in commercially available promoter systems (**Paper I**)

P_{T7lac}	GGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTT TGTTTAACTTTAAG**AAGGAG**ATATCATATG
P_{trc}	AATTGTGAGCGGATAACAATTTCACAC**AGGA**AACAGACCATATG
P_{BAD}	ATACCCGTTTTTTGGGCTAAC**AGGAGG**AATTACATATG

C *Pm* 5′-UTR up variants that mainly act on transcription (and translation to a lesser extent) [25]

Pm LV-1	AACATGTACCATTATAAC**GGAG**TAATGAACATATG
Pm LV-2	AACATGTACCATAATAC**AGGAG**TTATGAACATATG

D *Pm* 5′-UTR up variants that on transcription and translation [25,40]

Pm LII-10	AACATGTACCACAATAAT**GGAG**TTTTGAACATATG
Pm LII-11	AACATGTTACACAATAAT**GGAG**TAATGAACATATG
Pm H39	AACATGTACCATAATAAT**GGAG**TCTTGAACATATG

E *Pm* 5′-UTR down variants that on translation causing decreased expression [89]

Pm DI-3	AACATGTGGCATAATAAT**GGAG**TTATGCACATATG
Pm DI-7	AACATGTACAATGATTAT**GGAG**TCATGTACATATG
Pm DI-8	AACATGTCCCATAATAAT**GGAG**TCATGAACATATG

F *In silico*-designed 5′-UTR DNA sequences with maximal TIR (same length as *Pm* 5′-UTR) (**Paper II**)

dIB1	AACATGTTCGTCTTCACGC**TAAGGAGGT**ACATATG
dIB2	AACATGTTACTTATACG**AGGAGGT**TACAGCATATG
dIB3	AACATGTACCGTTCTTTCTAAGC**GAGGTT**CATATG

Two of the three regulated promoter systems (LacI/P_{T7lac} and AraC/P_{BAD}) contained stronger RBS than the XylS/*Pm* system independent of the coding sequence while the one present in the LacI/P_{trc} 5′-UTR was either only slightly stronger or even weaker than the RBS present in the *Pm* 5′-UTR dependent on the 5′-proximal end of the coding region (**Paper I**, Figure 5). Another way to improve translation initiation is to apply rational design tools to find the 5′-UTR DNA sequence that allows for the most efficient translation initiation given a certain coding region. The different strategies to improve a 5′-UTR were tested using *bla* as reporter gene due to the ability to use ampicillin tolerance of a cell as direct output for *bla* expression (section **1.2.2**). Using this gene also has the advantage that expression levels can be varied over two orders of magnitude [25,89]. Previously, it could be shown that some combinations of *Pm* promoter and *Pm* 5′-UTR variants could have an additive effect on each other [40]. This time, the *Pm* 5′-UTR in plasmids with the *Pm* wild-type or *Pm* ML1-17 promoter was replaced with $P_{T7lac,}$ P_{trc} and P_{BAD} 5′-UTRs and resulting ampicillin tolerance levels were determined. Data analysis revealed that for the *Pm* wild-type, 5′-UTRs from the three alternative regulated promoter systems caused higher ampicillin tolerance than the *Pm* 5′-UTR (**Figure 3.3A**). However, for the *Pm* ML1-17 promoter, it was different. The *Pm* promoter variant enhanced ampicillin tolerance five times. This increase was only exceeded by the effect of the P_{T7lac} UTR (7-fold), not the P_{trc} and P_{BAD} 5′-UTRs (four- and two- fold, respectively). Replacement of the *Pm* 5′-UTR with so-called UTR up-variants (**Table 3.1C and D**) led to even further increase of *bla* expression which exceeded 13-times improvement compared to the wt *Pm* 5′-UTR (**Figure 3.3B**). The DI UTR variants (**Table 3.1E**) on the other hand lowered ampicillin tolerance by a factor of 10 to 33. As could be assumed, the designed UTR variants with optimal features with respect to translation initiation (**Table 3.1F**) did improve *bla* expression. However, with only four- to eight- fold improvement of ampicillin tolerance compared to the *Pm* 5′-UTR, it was concluded that the screened up variants of the *Pm* 5′-UTR still were the best choice to achieve the highest *bla* expression improvement based on changing the 5′-UTR.

A different way of interpreting the data was to look at whether expression levels could have been predicted based on sequence analysis (**Figure 3.3C**). For the *Pm* 5′-UTR DI variants, lowered TIR values compared to the wild-type sequence coincided with lowered tolerated ampicillin concentrations. This supported previous experimental data which revealed that these UTR variants down-regulate translation [89]. Despite this it

would not have been possible to predict expression levels based on sequence analysis. This is especially evident for LV-1 and LII-11, two *Pm* 5′-UTR variants that only had a slightly increased TIR value compared to the wild-type (1.3- and 1.5- fold higher), but which could increase ampicillin tolerance by a factor of ~12. The mutations within these *Pm* 5′-UTR variants are known to affect transcript accumulation in addition to translation.

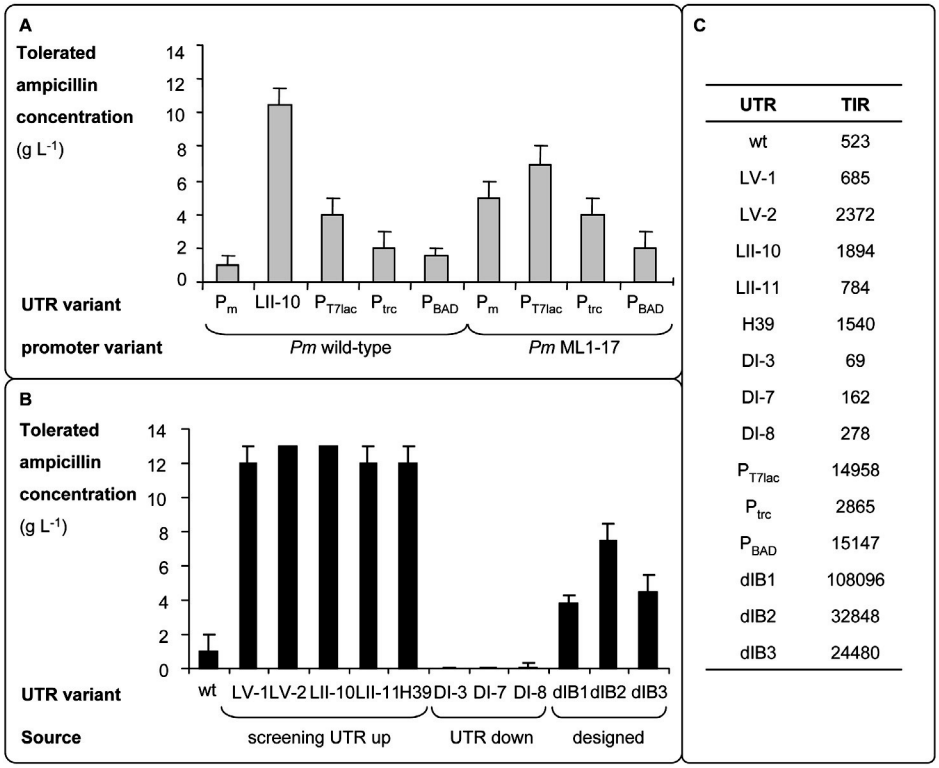

Figure 3.3: Analysis of the effects of different 5′-UTR DNA sequences on *bla* expression based on ampicillin tolerance testing and theoretical analysis of the RBS. A and **B** Results from ampicillin tolerance testing of strains expressing *bla* from *Pm* (or *Pm* ML1-17) in dependence of different 5′-UTR DNA sequences. Expression was induced with 2 mM *m*-toluate. Values represent the highest ampicillin concentration at which growth was observed. Error bars point to the next highest ampicillin concentration that inhibited growth. 13 g L^{-1} was the highest concentration used. The different *Pm* 5′-UTR variants were published before: LV-1, LV-2, LII-10 and LII-11 [25]; H39 [40]; DI-3, -7, -8 [89]. The three latter sequences led to tolerance of 0.03, 0.06 and 0.10 g L^{-1} ampicillin, respectively, under the given conditions. dIB1-3 were designed by the RBS calculator [218]. Ampicillin tolerance levels of LII-10 vary because two slightly different plasmid backbones were used: pTA16=pSB-M1b (**A**) and pIB11 (**B**) (difference in two restriction sites around the *xylS* gene; otherwise same sequence around *Pm*, *Pm* 5′-UTR DNA sequences and GOI. **C** Calculated TIRs obtained by analyzing UTR-*bla* DNA regions with the reverse engineering function of the RBS calculator. (**Paper II** and unpublished data)

Another example is designed 5′-UTR dIB1 which has a 207-fold higher TIR value than the *Pm* 5′-UTR, but which led to only 3.8-fold improvement of ampicillin tolerance. This analysis demonstrated the need to understand the sequence-determinants that affect transcription and translation better in order to be able to choose or re-design a 5′-UTR in a more predictable manner for example by utilizing RBS design tools.

3.2.2 Attempts to identify specific mutations within *Pm* 5′-UTR DNA sequences that influence transcription or translation

The comparison between the different 5′-UTR DNA sequences described above highlighted that screening was the preferred method to identify such sequences with respect to high expression. The analysis also revealed that the high expression levels achieved by the *Pm* 5′-UTR variants could not have been predicted based on sequence analysis. This could be attributed to the fact that mutations within these variants lead to increased transcript production and translation efficiency at the same time. This means that a prediction method purely based on translation-influence was not sufficient to evaluate the potential of 5′-UTR DNA sequences to enhance protein production levels. The extent to which the mutations found in the 5′-UTR variants stimulate one or the other process differs. Mutations in LII *Pm* 5′-UTR variants for example are believed to mainly act on translation while mutations in LV *Pm* 5′-UTR variants probably confer mainly increased transcript production [25]. Earlier, two artificial operon constructs had been created, pAO-Tn and pAO-Tr [236]. These were applied to generate mutant libraries to get a better insight into which mutations lead to stimulation of either of the processes, not a mixture (**Paper II**).

Novel artificial operon constructs to identify *Pm* 5′-UTR DNA sequences that either lead to increased transcript accumulation or protein production

The short description of how the artificial operon libraries were made and screened in order to identify novel *Pm* 5′-UTR variants is the following. First, a doped synthetic oligonucleotide mixture covering the central *Pm* 5′-UTR region (see **Table 3.1A**) was inserted into pAO-Tr and pAO-Tn using the PciI and NdeI sites. By not randomizing the 5′-UTR DNA core region completely, but by using a low substitution frequency for each position instead (14.4% substitution by a different nt versus 85.6% wild-type nt) a certain

resemblance to the original *Pm* 5′-UTR was assured. This rendered the mutated 5′-UTR DNA sequences variants of the *Pm* 5′-UTR instead of completely new 5′-UTRs. In previous 5′-UTR random mutagenesis studies, the SD was also included in the randomized region. However, the mutations causing high expression did not affect the original four nt long SD sequence [25] which is why the SD sequence was kept intact in the mutant library. The resulting Tr- and Tn-UTR libraries (UTR libraries derived from pAO-Tr and pAO-Tn, respectively) where then transferred to *E. coli* and transformants were selected on agar plates supplemented with kanamycin. Several 100,000 clones were pooled and subjected to agar plates containing *m*-toluate and different ampicillin concentrations (for details, see **Paper II**). Strains growing on high ampicillin concentrations were picked and tested phenotypically again (ampicillin tolerance). Plasmids were isolated from strains tolerating high concentrations and transformed back into *E. coli*. 5′-UTR DNA regions found in strains whose phenotype could be confirmed were then sequenced. The 5′-UTR DNA regions were re-synthesized as pairs of complementary oligonucleotides and cloned back into pAO-Tr or pAO-Tn. Finally, the constructs were transferred to *E. coli* one last time for ampicillin tolerance testing.

Among the five novel UTR variants identified from the Tr-UTR library, there were three (r31, r36 and r50) which could lead to the same level of ampicillin tolerance as caused by mutations contained in LV-1 and LV-2. Mutations in the previously identified LV-1 and LV-2 *Pm* 5′-UTR variants could lead to increased tolerance to ampicillin (about 4-fold) when inserted upstream of *celB*. This could be attributed to increased transcription of the operon because increased translation of *celB* would not influence translation of *bla* in a direct manner. This effect was expected because these two *Pm* 5′-UTR variants were proven to mainly cause transcript accumulation of a target gene [25]. The 5′-UTR variants all differed in their sequence. Three variants had a different total length (r28, r31 and r50). However, the number of novel Tr-UTR candidates was generally low and three rounds of screening needed to be performed. None of the Tr-UTR variants led to higher ampicillin tolerance than the previously identified variants. Although more clones with mutations in the Tr-UTR variant were identified in the initial screen, they could not be confirmed in later stages suggesting that mutations in other places than the *Pm* 5′-UTR DNA region led to increased ampicillin tolerance. Indeed, during the second screening round mutations within the second codon of *celB* (CCC-> T/UCC) were detected in more than half of the sequenced constructs and additional mutations were detected within the PciI and NdeI restriction sites. And these were only the mutations that could easily be identified. Similar

to the novel *Pm* 5′-UTR variants that could finally be confirmed, LV-1 and LV-2 were also identified in the same genetic context meaning that the *Pm* 5′-UTR library was placed directly preceding the *celB* gene. The only difference was that *celB* and *bla* were translationally fused by an overlapping start and stop codon. The *celB* gene was previously shown to be well-expressed in presence of the wild-type *Pm* 5′-UTR already. This means that both screening tools (the one described here and the one developed by Berg et al. [74] potentially led to identification of the mutations within the *Pm* 5′-UTR with the strongest effect on *celB* expression. No clear positional hot-spot was observed and there was not any mathematical tool available to analyze the sequence composition of a 5′-UTR DNA region with respect to transcription initiation available. That is why analysis of the 5′-UTR variants was performed at the transcript and protein level (see below).

Compared to the Tr-UTR library, screening of the Tn-UTR library led to the identification of more *Pm* 5′-UTR variants and in two screening rounds 22 variants were confirmed in the end. These caused up to 25 times increased ampicillin tolerance compared to the wild-type. During screening of the Tn-UTR library, some strains with even higher ampicillin tolerance were found. But when the DNA region between the *Pm* promoter and the *celB* was sequenced, it turned out that three *Pm* 5′-UTR DNA fragments had been inserted; two fragments in regular orientations interspersed with a third in the inversed orientation. Due to the presence of triple insertions of *Pm* 5′-UTR DNA fragments in strains tolerating high ampicillin concentrations, a colony PCR was performed prior to sequencing reactions to eliminate those constructs which contained more than a single *Pm* 5′-UTR DNA region. My work benefitted from this observation at a later stage (see **3.2.3**). Looking at the sequence composition, 19 out of 22 carried a mutation that interrupted the extra ATG sequence found between the SD sequence and the actual start codon DNA sequence (**Table 3.1A**). Three out of 22 (n24, n35 and n59) increased the length of the SD by one nt and three by two nts (n44, n47 and n58). These are good indicators for improvement of translation because the SD-antiSD interaction is important during translation initiation. In the *Pm* 5′-UTR sequence, the distance between the SD region and the ATG is relatively long (12 nts from the center of the SD sequence gGag to the ATG start codon) and with four nts in length the SD is relatively short. Osterman et al. [176] could show that for such big distances, a longer SD sequence is beneficial for efficient translation.

Characterization of Tr- and Tn-UTR DNA sequences at the accumulated transcript and protein level

After having identified the Tr- and Tn-UTR variants, different analyses were performed using the three strongest Tr-UTR variants together with the LV-2 control *Pm* 5′-UTR as well as four Tn-UTRs. The latter were chosen based on sequence composition (elongation of the SD region by one or two nts, change of the internal ATG or not, total number of mutations) to give a variation of sequence features. Effects of general sequence features within the 5′-UTR affecting translation have been systematically investigated before [176]. To evaluate many criteria for good translation efficiency at the same time, computational tools such as the UTR designer [227] or the RBS calculator [218] can be applied. Data obtained from analyses with the RBS calculator are found in **Paper II** while data from the most important UTR variants studied in this thesis are shown below (**Figure 3.4**). Another way to study effects of 5′-UTR variants on gene expression is to look at accumulated transcript and protein levels. For this purpose, the Tr- and Tn-UTR variants were transferred to pIB11, a monocistronic construct in which *Pm* and its cognate 5′-UTR control expression of the *bla* gene alone. Strains harboring pIB11 constructs with Tr- and Tn-UTR variants were subjected to qPCR and protein analyses using an enzymatic assay.

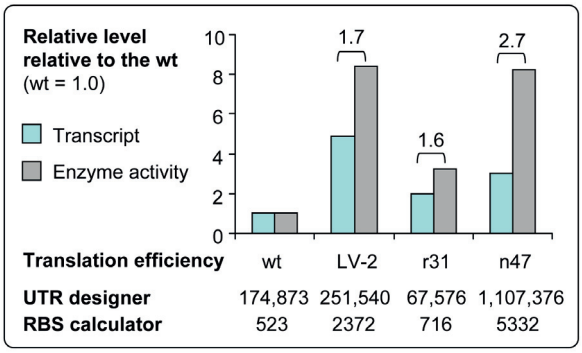

Figure 3.4: Analysis of the effect of selected 5′-UTR variants on accumulated transcript levels and enzymatic activity as well as a theoretical analysis of the translation efficiency. Recombinant *E. coli* strains harboring pIB11 constructs with different UTR variants were induced with 0.5 mM *m*-toluate. Expression of *bla* was analyzed from samples taken 5 hours post induction using qPCR and a β-lactamase assay. Results from one representative experiment are shown. The data were later confirmed by master student Jon A. Lorentzen. Below the bar diagram, the calculated translation efficiencies from two different softwares are shown. (**unpublished data**)

Protein-to-transcript ratios of the Tr- and Tn-UTRs were in the range of 1.1-1.4 and 1.2-2.6, respectively, under full induction (2 mM *m*-toluate) and slightly higher under partial induction (0.5 mM *m*-toluate). Data for some selected UTR DNA sequences are shown in **Figure 3.4.** Interestingly, it could be demonstrated that mutations within the LV-2 variant led to a significant improvement of protein production levels on top of increased accumulated transcript levels which was supported by the computational analysis. In contrast, the two important UTR variants identified in the study described in **Paper II** (SIII-r31 and SII-n47) displayed more distinct characteristics with respect to transcription and translation stimulation. Overall, TIR values (RBS calculator) of the three Tr-UTRs were slightly (1.2-2.7 times) higher than for the wild-type *Pm* 5′-UTR while the TIR values for the four Tn-UTRs were increased by a factor of 9.3-14.0. Based on these two criteria and a third analysis (change of position of the Tr- and Tn-UTRs in the artificial operon constructs (**Paper II**, Figure 2)), it was concluded that, most likely, Tn-UTRs mainly act on translation efficiency while Tr-UTR mainly, but not purely act on transcript accumulation.

3.2.3 Development of a novel 5′-UTR DNA sequence design to increase protein production levels

The 5′-UTR variants that were identified using the two artificial operons were shown to display distinct characteristics with respect to influencing relative accumulated transcript amounts and enzyme activity. However, generally the overall improvement of β-lactamase production was lower than achieved by previously identified *Pm* 5′-UTR variants (maximum 11-fold increase (n47; data not shown) vs. 20-fold increase at induction with 2 mM *m*-toluate; [25]). This could be explained by the fact that the mutations found in Tr- and Tn-UTRs mainly stimulate one of the processes instead of both at the same time. As a consequence, it was of great interest to combine mutations of different Tr-and Tn-UTRs to form a superior *Pm* 5′-UTR variant in which both transcript accumulation and protein product formation were enhanced. Different potential design strategies were evaluated on how mutations that influence transcription or translation could be combined in a single 5′-UTR DNA sequence.

The first idea was to merge mutations of a Tr-UTR variant with mutations from a Tn-UTR variant and generate a new 5′-UTR with the same length as the original *Pm* 5′-UTR. But the UTR variants all differed in sequence composition. As an example, the mutations in the r31 Tr-UTR were only located in the stretch of DNA located upstream of

the SD sequence. However, the six mutations within the n47 Tn-UTR were spread across the entire mutated region and thereby overlapped with the mutated region of the r31 Tr-UTR and the other Tr-UTRs. Therefore, it was not possible to merge the sequences or to piece together a *Pm* 5′-UTR out of two segments, one from the Tr-UTR and another from the Tn-UTR.

The second idea was to combine the entire Tr-UTR with an entire Tn-UTR and separate the units with a spacer sequence. This second option had the advantage that effects caused by the mutations within a Tr-UTR variant would probably not interfere with ribosome binding and translation initiation due to the physical distance generated by the spacer sequence. At the same time, mutations around the ribosome binding site (contained within the Tn-UTR) would not affect transcription. The spacer design was based on the above described screening artifact where three *Pm* 5′-UTR DNA sequences had been inserted simultaneously leading to high expression levels. At first, several parallel designs were tested; each of which contained three wild-type 5′-UTR DNA sequences; one in the regular orientation, one in the inversed orientation and one additional in the regular orientation. However, the sequence of the inversed 5′-UTR DNA sequence was modified in a way that the additional PciI and NdeI sites connecting the three sequences were removed and that one of several unique restriction sites was introduced in the center of this 5′-UTR DNA sequence. This would ensure that the 5′-DNA region and the 3′-DNA region of the novel 5′-UTR DNA region could be replaced using annealed oligonucleotides flanked by PciI plus enzyme X and enzyme X plus NdeI sites. Finally, one design was chosen with a SacI site in the center which gave rise to an ampicillin resistance level of 0.25 g L^{-1} at full induction (2 mM *m*-toluate). This pIB11-derived construct was called pDUTR and the particular 5′-UTR DNA region between the *Pm* promoter and the *bla* gene was termed dualUTR (**Paper II**).

3.2.4 Generation of dualUTR DNA sequences that primarily enhance transcription, translation or a mixture of both

Based on the above described 5′-UTR design, several mutations derived from Tr- and Tn-UTR DNA sequences were introduced in the 5′-UTR DNA region and variants of this region were denoted as pairs of Tr- and Tn-UTR variants; e.g. wtwt for a dualUTR composed of two wild-type units. First, several Tr- and Tn-UTR DNA sequences were combined and the effect of the mutations was tested based on resulting ampicillin tolerance of the respective recombinant *E. coli* strains. Among the different combinations

tested, it was the mutations contained in Tr-UTR r31 which were able to exert the strongest increase in *bla* expression on top of the effect caused by mutations contained in the different Tn-UTR variants. Among the Tn-UTR variants, it was n47 which displayed the strongest stimulating ability on *bla* expression. In order to analyze the contributions of the different dualUTR sequences on *bla* expression more precisely, recombinant *E. coli* strains harboring wtwt, r31wt, wtn47 and r31n47 dualUTR variants were subjected to qPCR analysis and an enzymatic assay. One important observation was that the mutations in the three different dualUTR variants did lead to higher accumulated transcript levels compared to the wtwt combination. However, the protein product levels did not increase proportionally. Calculated ratios between relative protein and transcript amounts were 0.6 for r31wt, 4.1 for wtn47 and 3.6 for r31n47 (based on **Paper II**, Figure 4A).

These data support two hypotheses derived by Laila Berg ([237]; Figure 14). The first stated that, theoretically, there exist 5′-UTR variants with maximized transcriptional stimulation ability but which are not optimal for translation. These 5′-UTR variants would lead to stronger relative transcript accumulation than protein production with a protein-to-transcript ratio below one. The r31wt dualUTR is an actual, practical example to support hypothesis 1 because it displays a protein-to-transcript ratio of ~0.6. This means that transcriptional features exerted by the *Pm* promoter region in combination with the mutations in the r31 Tr-region of the dualUTR are more dominant than translational features of the wt Tn-part which in turn could be interpreted by an uncoupling of the two processes. The second hypothesis states that there exist 5′-UTR variants with maximized translation efficiency, but reduced transcript production rate. This would cause a protein-to-transcript ratio well above one. Again, a 5′-UTR variant, wtn47, could be generated which fulfils the criteria to prove hypothesis 2. Combining the r31 and n47 DNA regions into one 5′-UTR causes a protein-to-transcript ratio of 3.6 which is just below the ratio caused by wtn47 (4.1). This indicates that transcriptional stimulation and translation efficiency are stimulated to a similar extent. However, an even lower ratio could be envisioned. This could be further exploited in the future by adjusting the Tr-DNA unit of the dualUTR to balance transcript accumulation even better with translation stimulation. According to the current understanding of the coupling of the transcription and translation machineries (section **1.3.4**), the optimal ratio would be around one. This is why I hereby suggest a revised interpretation of the effects of the LV-2, namely that the mutations within this *Pm* 5′-UTR variant lead to equally increased accumulated transcript and protein product levels. This can be supported by the theoretical findings (extended SD sequence

and higher TIR values which indicate a positive effect on translation initiation) as well as experimental data (protein-to-transcript ratio just above 1).

A parallel combinatorial engineering approach was performed in our research group [40]. In this study, it was not 5′-UTR DNA sequences that act on transcription or translation, but variant elements of the XylS/*Pm* system acting on the two processes that were combined. In brief, the major finding of this story was that combination of variant elements that act on different steps of expression act in a stimulating manner. As an example, a *xylS* variant or a *Pm* variant which both stimulate transcription enhanced the positive effect of a 5′-UTR variant which in turn mainly acts at the translational level. The total increase achieved by a combination of a *xylS* and a 5′-UTR variant was around 50-fold and the increase from a *Pm* and 5′-UTR variant combined 25-fold (vs. 15 times increase compared to the wt by a 5′-UTR variant alone). All three variant elements combined increased expression by a factor of 75.

3.2.5 Application of a rational RBS design tool to change the Tn-UTR part of the dualUTR

Previous mutagenesis studies on the *Pm* 5′-UTR demonstrated that there is a context dependency between the 5′-UTR and the 5′-proximal coding region. This means that 5′-UTR variants identified in proximity to a certain gene would not exert the same effect on a second gene. The Tn-UTR variants tested in the dualUTR design (identified in proximity to the *bla* gene could enhance mCherry production by an impressive factor of 67 (**Paper II**). This effect could have been predicted since the n47 variant and the 5′-end of the mCherry gene form an RBS with a high TIR value compared to the wild-type (2,309 vs. 12,766). Still, it is questionable whether the 5′-UTR sequences in this study that were identified to give rise to high protein production levels of β-lactamase (and mCherry) would work equally well for different genes. In order to avoid a new mutant library approach for every gene in order to find a suitable 5′-UTR DNA sequence that would increase translation initiation in a predictive manner, an attempt was made to apply the RBS calculator again. The reason to apply this tool for the dualUTR constructs (now in the forward engineering mode) under the assumption that more reliable protein product levels could be achieved is based on the physical separation of the Tr- and Tn-UTR DNA elements. Since mutations in the Tn-part of the dualUTR are most likely too far away from the transcriptional start site to affect transcript accumulation, it was reasonable to assume that the Tn-part of the dualUTR could be optimized based on translation-enhancing

properties only without affecting transcript accumulation at the DNA level. As exemplified for β-lactamase the use of three different designed Tn-UTR sequences (called dTn1-3 for designed Tn-dualUTR) resulted in the identification of at least one sequence which displayed increased protein production abilities equally well or better than the screened n47 Tn-UTR variant (**Figure 3.5**). This finding supports the hypothesis that a 5′-UTR DNA sequence cannot be optimized based on translation-affecting properties only. Separation of the region of a 5′-UTR DNA sequence which is part of the RBS from the ITS region, instead, opens up for the possibility to apply rational design tool to improve translation initiation.

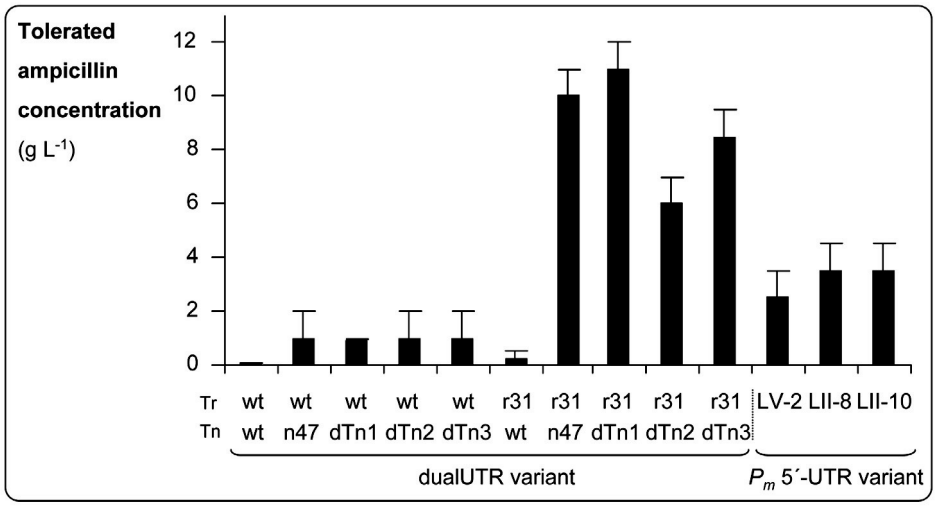

Figure 3.5: Comparative analysis of the effect of different 5′-UTR DNA sequences on *bla* expression measured as ampicillin tolerance. Strains harboring pDUTR- or pIB11-based plasmids with different dualUTR or *Pm* 5′-UTR DNA sequences, respectively, were transferred to agar plates containing 0.1 mM *m*-toluate and increasing amounts of ampicillin. The highest concentration tolerated is plotted and error bars point to the next highest concentration that restricted growth. (derived from **Paper II**, Figure 6A).

In general, the data obtained in this study agree well with the current literature in the field of synthetic biology. For example, the issue of context dependency or junction interference is currently taken up. Different solutions to prevent unforeseen effects of combining promoter, RBS and genes are presented. Two additional articles describe the use of an endoRNase site [224] and a ribozyme encoding region [222] within the 5′-UTR DNA region to cleave off the 5′-part of the UTR in order to achieve a more predictable effect on translation initiation caused by a rationally designed RBS. Another study on the

other hand describes the benefits of applying a synthetic library approach to combine promoters and 5′-UTR DNA sequences to achieve an optimal context between the promoter, the RBS and the gene and to be independent of predictions [228]. In general, the novel dualUTR described here could be applied for both strategies and it remains to be shown in the future which strategy gains more acceptance.

3.3 Analysis of the effects of different plasmid backbones on gene expression

Heterologous gene expression can be directly affected by the choice of the promoter system (**Paper I**) and the 5′-UTR (**Paper II**). In a broader perspective, e.g. at the systems level (expression vector); more parameters come into play which are able to exert an effect on protein production levels. These include the plasmid copy-number regulated by the origin of replication and associated plasmid control elements. In contrast to the first parameters, these would likely affect gene expression in a general, not gene-dependent manner.

3.3.1 Influence of plasmid stabilization elements on protein production

In previous studies the *hok*/*sok* system was used in pJB vectors to prevent plasmid loss under HCDC [20,68]. Another control region was used on mini-RK2 vectors, namely the *par* locus (Par) [76]. This locus is involved in postsegregational killing and plasmid multimer resolution. Both systems were able to stably maintain plasmids in cell populations in absence of selection. Besides the *par* locus, RK2 contains the O_{B1}*incC korB* partitioning (Inc) region which is involved in spatial organization of the plasmid within a bacterial cell (see **1.2.1**). However, it has not been shown before whether the use of such control regions would have a direct effect on heterologous protein production. Expression could be studied at the population level or at the level of individual cells using methods like flow cytometry. The outcome at the population level would then be the average sum of the production in each individual cell. Here, GFP expression was used to exemplify the influence of different elements contained in the plasmid backbone on protein production. Mini-RK2 plasmids were constructed based on pSB-M1g (**Paper I**) in which the Inc and Par loci (**Figure 3.6A and B**) were inserted downstream of the *xylS* gene.

GFP production was studied in *E. coli* MG1655[1] (**Figure 3.6C-E**). At the population level, both regions had a significant impact on GFP production increasing product levels by around 80% and 60% compared to when plasmids without control regions were used (**Figure 3.6C**).

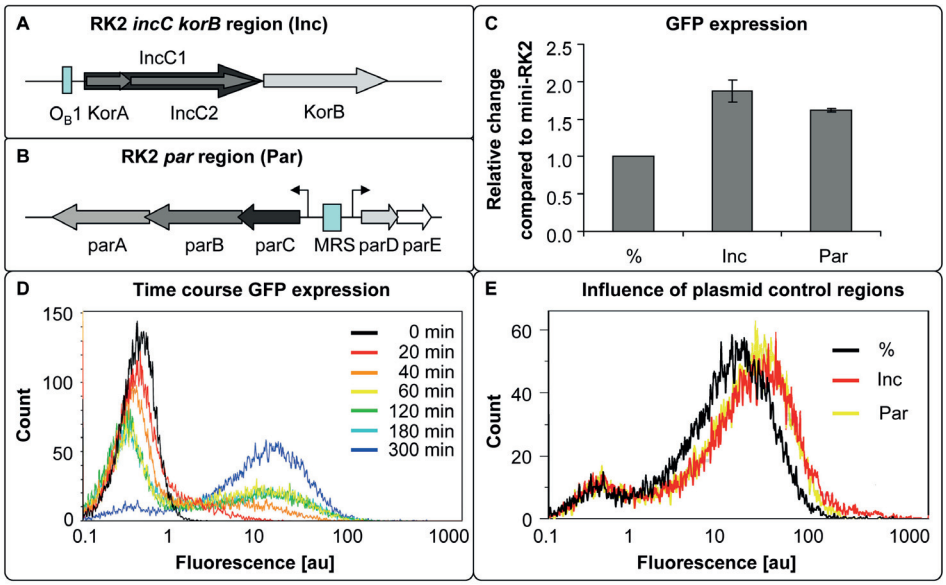

Figure 3.6: Influence of plasmid control regions on GFP expression profiles of recombinant *E. coli* cells harboring pSB-M1g. A Genetic organization of the O_{B1}*incC korB* partitioning region of RK2 that is known to shift the location of mini-RK2 replicons towards the quarter- and mid-cell positions (derived from Verheust and Helinski [70]). **B** Organization of the *par* locus of RK2 containing the postsegregational killing system encoded by the *parDE* operon and the multimer resolution system encoded by *parCBA*. MRS multimer resolution site (derived from Sobecky et al. [75]) Both the O_{B1}*incC korB* and the *par* locus were inserted into pSB-M1g (**Paper I**) downstream of the *xylS* gene and the resulting plasmids were called pSB-M1g Inc and pSB-M1g Par, respectively. **C** Recombinant *E. coli* MG1655 cells harboring pSB-M1g (Ø), pSB-M1g Inc (Inc) or pSB-M1g Par (Par) were grown in LB medium according to the protocol described in **Paper I** and fluorescence values were determined 300 min post induction. After normalization, data were related to pSB-M1g. **D** During growth of *E. coli* MG1655 (pSB-M1g), samples were collected at different time points after induction and analyzed by flow cytometry. This time, a Becton Dickinson LSR flow cytometer and the Kaluza 1.1 software (courtesy of Physics Department, NTNU) were used. **E** In parallel, samples from *E. coli* MG1655 (pSB-M1g Inc) and *E. coli* MG1655 (pSB-M1g Par) cultures were collected and distribution of cells collected 300 min after induction are presented.

[1] This strain was used by our collaboration partner Judith Megerle at LMU Munich to study culture heterogeneity.

Flow cytometry data were collected as well, following GFP production in individual cells over a period of five hours. This analysis revealed that onset of GFP production from mini-RK2 plasmids was generally slow. A significant population of cells that produced more GFP than the background level only formed one hour after induction (**Figure 3.6D**). Still, many cells remained uninduced as derived from the peak at low fluorescence values. Five hours post induction, most cells were induced. Distributions of cells harboring plasmids with the Inc and Par regions were strikingly similar to the distributions of cells harboring plasmids devoid any control region (**Figure 3.6E**). The only difference was that the major peak derived from induced cells was shifted towards higher fluorescence values. This is consistent with the data collected at the population level.

These results were somewhat surprising. It would have been possible that the presence of a multimer resolution system and a partitioning region would lead to a more homogenous distribution of plasmid between the cells in a population leading to a more homogenous expression. The data disprove this hypothesis. However, the presence of the Inc and Par regions was beneficial for the overall production level of GFP. This finding may be interpreted by assuming that a better distribution of plasmids within cells leads to a more favorable use of cellular resources (nts, amino acids, etc.). It is known that mRNA does not freely diffuse in a cell [238] due to a crowded cytoplasmic environment. Assumingly, transcripts stay in proximity to the plasmids. If plasmids are distributed at more locations within a cell, nutrients used for transcription and translation of heterologous genes can be used more efficiently [69]. This in turn might explain why more GFP can be produced from plasmid containing elements that lead to a more favorable positioning inside the cell.

3.3.2 Influence of different origins of replication on protein production

A different way to change the distribution of plasmids within the cell is the use of different origins of replication leading to elevated plasmid copy-numbers. Therefore the effect of a pMB1-based plasmid backbone on GFP production was studied (**Paper I** and unpublished data). Overall, GFP production levels were increased, not only for plasmids containing the XylS/*Pm* system, but also XylS/*Pm* ML1-17, the LacI/P_{T7lac} system, the LacI/P_{trc} system and the AraC/P_{BAD} system (**Paper I**, Figure 2). Increasing the copy-number from four- seven (mini-RK2) to 15-20 (pMB1) led to 1.6- to 6.2-fold higher GFP levels depending on the promoter system. The relatively low increase for the LacI/P_{T7lac}

system could be attributed to the accumulation of insoluble protein (**Paper I**, Figure 3) which was not detected by fluorescence measurements.

In a second analysis, populations of GFP producing cells were subjected to flow cytometry analysis. Surprisingly, the use of a pMB1-based backbone led to more homogenous populations compared to the use of RK2-based backbones as exemplified by GFP expression profiles of strains harboring the XylS/*Pm* system (**Figure 3.7A**). Already 20 min after induction, a single peak distinct from the uninduced peak was observed suggesting that all cells were induced and that these produced GFP. The peak migrated towards higher fluorescence values in agreement with fluorescence values at the population level (data not shown). Equally homogenous distributions were observed for the system with the *Pm* promoter variant *Pm* ML1-17 (**Figure 3.7B**). These data confirm that the use of a different plasmid backbone not only influences heterologous protein production in general, but also the homogeneity of a cell population. It can be assumed that this finding is based on how plasmids are distributed in a cell. However, a study visualizing plasmid distributions would be needed to confirm this hypothesis.

3.3.3 Influence of different regulated promoter systems on protein production studied at the single-cell level

Another parameter influencing homogeneity of expression is the promoter type. The comparative analysis of different regulated promoter systems revealed that distributions of fluorescing cells differed quite significanty from one another. Populations of cells harboring plasmids with LacI/P_{T7lac} were heterogeneous prior to induction and at later stages during expression (e.g. 5 hours post induction) (**Figure 3.7C**). The LacI/P_{trc} system caused homogenous distributions throughout the time course (**Figure 3.7D**). However, production of GFP did not increase further from one hour post induction on. Finally, the AraC/P_{BAD} system was characterized by a slower onset of GFP production, but distributions remained homogenous from one hour post induction on (**Figure 3.7E**).

A known issue associated with AraC/P_{BAD} is the so-called all-or-nothing phenomenon. This is based on the finding that two populations of cells are observed in flow cytometry studies after induction with sub-saturating inducer concentrations [105]. In order to demonstrate this phenomenon also in the novel vector context and to compare the behavior of XylS/*Pm* ML1-17 to AraC/P_{BAD} at lower inducer concentrations, a second experiment was conducted using 10-fold less inducer than in the full induction experiment (**Figure 3.7F** and **G**). The analysis revealed that GFP expression driven from the *Pm*

ML1-17 promoter variant led to formation of homogenous cell populations. Meanwhile, a significant portion of cells harboring constructs in which GFP production was regulated by the P_{BAD} promoter displayed a behavior similar to uninduced cells.

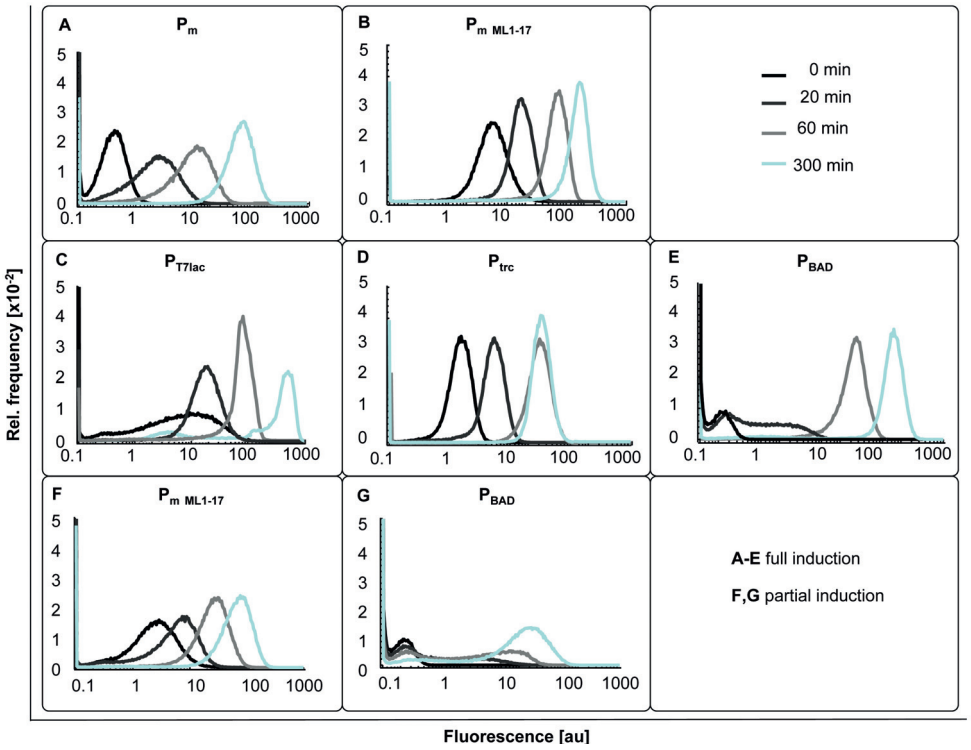

Figure 3.7: Distribution of cells producing GFPmut3. Recombinant cells were grown and expression was analyzed according to the protocol described in Paper I. A *E. coli* ER2566 pSB-M2g **B** *E. coli* ER2566 pSB-M2g-1-17 **C** *E. coli* ER2566 pSB-E2g **D** *E. coli* ER2566 pSB-T2g **E** *E. coli* DH10B pSB-B2g. In addition to studies at full induction, *gfpmut3* expression was also induced with a 10-fold lower inducer concentration for some selected strains: **F** *E. coli* DH10B pSB-M2g-1-17 induced with 0.2 mM *m*-toluate **G** *E. coli* DH10B pSB-B2g induced with 0.0015 % L-arabinose (derived from **Paper I** and unpublished data).

These findings can be related to inducer uptake. L-arabinose, the inducer of the AraC/P_{BAD} system, namely, undergoes active transport by one high-capacity, low-affinity L-arabinose transporter and one high-affinity transporter (see **1.2.2**). By making regulation of the production of the L-arabinose transporters independent of the inducer, homogenous cell populations could be achieved [102]. The XylS/*Pm* system whose inducer molecules enter

the cell by passive diffusion did not need to be modified (other than the replicon on the plasmid backbone) in order to achieve homogenous populations. This represents a great advantage of XylS/*Pm* over regulated promoter systems that are induced by molecules requiring an uptake system (e.g. AraC/P_{BAD}).

3.4 Analysis of the effects of different growth conditions on protein production

The focus of this PhD project as described up to this point has been on the genetic elements of expression cassettes only. However, it is well known that the surrounding environment of an expression vector (e.g. cell cytoplasm and culture medium) affects recombinant gene expression. Under laboratory scale experiments, a range of parameters can be tested with respect to the yield of a heterologous protein product. A selection was tested during this PhD project using XylS/*Pm* to express genes coding for two capsid proteins of the Norwegian Salmonid Alphavirus (NSAV) called E1 and E2. The results are presented in the following two subchapters.

3.4.1 Production of Norwegian Salmonid alphavirus E1 and E2 capsid proteins in *E. coli*

Analyses so far in addition to several previous studies revealed that the XylS/*Pm* system possesses beneficial traits for various recombinant expression purposes, amongst them the ability to produce industrial levels of heterologous proteins in a fermentor [20,68]. These traits were again put to the test by trying to produce medically relevant proteins. In a KMB[2] project (which this PhD project was indirectly associated with), Pharmaq, one of the four industry partners, supplied genes encoding NSAV E1 and E2 to be expressed within the project (**Paper III**). The goal was to produce the target proteins as antigens to make it possible to derive antibodies against them for use as vaccines later on. In order to produce antigens, a simple bacterial host such as *E. coli* can be used. A literature study and bioinformatics analysis suggested that these proteins were likely to be difficult to produce due to their origin (fish virus), the predicted presence of transmembrane helices and disulfide bonds. This led to a rational choice of an expression system. In **Paper I** it

[2] A Combinatorial Mutagenesis Approach to Improve Microbial Expression Systems (Norwegian Research Council project; Project nr: 182672/I40)

could be shown that the XylS/*Pm* system in combination with a replicon leading to an intermediate plasmid number could yield intermediate, yet detectable protein production levels, a good ratio between soluble and insoluble protein on top of the ability to fine-tune expression levels without the need of making new genetic constructs, for example by varying the inducer concentration. It was decided to use the system in its wild-type form as a basis to have an intermediate starting point with respect to production capacity. Using a too strong promoter for toxic proteins as an example could stress the cells which increases the probability to accumulate IBs. As a plasmid backbone, mini-RK2 replicons containing *trfA* variant cop271C and the *hok*/*sok* suicide system were chosen based on the pJB series of plasmids. Starting broad, full-length genes encoding E1 and E2 as well as truncated versions devoid of the transmembrane and hydrophobic interaction domains were cloned into the pJB vectors. This protein engineering strategy was pursued to increase chances to achieve production of soluble target proteins. It could be justified with the fact that potential epitopes would not be located in the parts of the proteins that are buried inside the membrane or inside regions that form interaction surfaces with associated proteins. Another choice that was made early on was the use FPs. N-terminal FPs comprised CSP and OmpA, two signal peptides which accomplish translocation of the target proteins to the periplasm. Both were previously shown to facilitate dramatic improvement of protein production besides being properly cleaved off [20,68]. Also, combinations of the *Pm* 5′-UTR DNA region and the coding regions for the CSP and OmpA signal peptides formed RBS with TIRs of 3,621 and 5,190, respectively, which are well in the range of TIRs described earlier. Therefore, no particular 5′-UTR modification strategy was applied. C-terminal FPs included the c-myc and His6 tags for detection and affinity purification. A schematic representation of the arrangements of the genetic elements that were coupled together during this study can be found here: **Paper III**, Figure 1. Further decisions regarding the choice of *E. coli* strains (RV308 and BL21-CodonPlus(DE3)-RIPL) and growth medium (Hi+YE) can also be found in this article.

3.4.2 Recombinant production of the NSAV E1 and E2 capsid proteins in small-scale under influence of different growth conditions

A common understanding among researchers working in the field of recombinant expression is that it is useful to approach laboratory-scale experiments from a broad parameter perspective due to many unforeseen effects in the upstream process. This is the reason why the influence of different parameters on production of the NSAV E1 and

E2 proteins was tested. These included protein engineering, choice of genetic context (see **3.4.1**) and physiological conditions. The primary focus was to produce as much soluble virus protein as possible. Recombinant *E. coli* strains producing different variants of the E1 and E2 proteins with or without N-terminally fused signal peptides were grown at 30 °C and expression was induced with 0.5 mM *m*-toluate. After the first round of experiments, it could be concluded that presence of the additional tRNAs produced by the *E. coli* BL21-CodonPlus(DE3)-RIPL strain did not improve soluble E1 and E2 production. Due to better growth characteristics *E. coli* RV308 was selected for use in these experiments (data not shown).

Furthermore, it could be shown that production of the virus proteins affected bacterial growth which is why it was decided to induce expression from XylS/*Pm* in the middle of the logarithmic growth phase (corresponding to OD600 = 8-10 in Hi+YE medium) (**Figure 3.8A-D**). Especially the use of a signal sequence led to reduction of growth after induction an effect which became most evident for the E2 protein in its full-length form (**Figure 3.8B**).

Growth of strains producing the truncated form of E1 (E1ΔIDTM) was only slightly impaired in the first 2 hours after induction. However, production of these proteins appeared to be host-toxic from then on, a conclusion drawn from the significant reduction of OD_{600} from ~20 down to ~15 (**Figure 3.8C**). Growth of recombinant *E. coli* RV308 strains producing E2ΔIDTM was only marginally impaired for the ones in which the protein was fused to CSP (**Figure 3.8D**). In the next step, periplasmic extracts were prepared and a Western blot analysis was performed to study the fraction of soluble and insoluble recombinant protein (**Figure 3.8E** and **F**). Comparison of the different strategies performed revealed that truncation had the most predominant effect on increasing both the soluble and insoluble fraction while fusion to signal sequences had a minor contribution. Truncation also rendered production of the viral proteins less toxic to the cells which could be shown for E1 to a certain extent and to a greater extent for E2.

From this point on, strategies to improve the soluble fraction were pursued. The first was temperature reduction to 16 °C after induction, the next decrease of the inducer concentration and the third, use of the Chaperone Plasmid Set from the TaKaRa. Western blot analysis revealed that the functionality of XylS/*Pm* was maintained at low temperatures such that fine-tuning of the inducer concentration (here use of 0.1 mM) increased the soluble fraction compared to when 0.5 mM was used. The use of

chaperones, however, was not suitable to significantly increase the soluble protein fraction (data not shown).

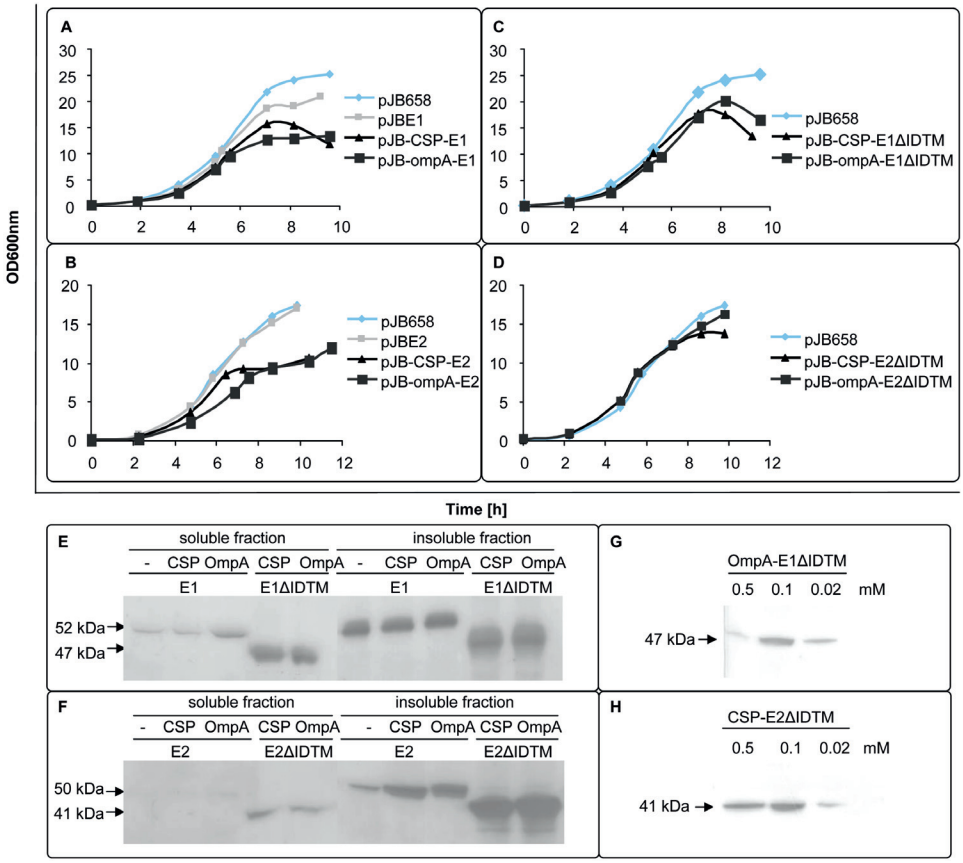

Figure 3.8: Study of Norwegian Salmonid Alphavirus E1 and E2 protein production regulated by the XylS/*Pm* system. Growth curves of recombinant *E. coli* RV308 strains producing different variants of E1 (**A**: full-length proteins and **B**: truncated proteins) and E2 (**C**: full-length proteins and **D**: truncated proteins). pJB658 is a mini-RK2-based plasmid without a gene of interest and confers Amp^R. For details see Paper III. Cell lysates were subjected to Western blot analysis. **E** Lysates obtained from E1-producing strains grown at 30 °C (0.5 mM induction) **F** Lysates obtained from strains grown at 16 °C; induction of OmpA-E1ΔIDTM production with decreasing *m*-toluate concentrations. **G** Same as **E**, but with E2-producing strains. **H** Same as **F**, but with E2-producing strains. (**Paper III** and **unpublished data**).

In this study, several classic approaches to address solubility were assessed like reduction of growth temperature and inducer concentration as well as use of chaperons to aid the folding process. Further strategies could have been applied like the use of

solubility-enhancing FPs (e.g. NusA or MBP) [239] or the addition of additives like L-arginine, glycine betaine or mannitol) [240]. However, since early refolding attempts of the insoluble aggregates followed by Ni-NTA affinity chromatography were successful (**Paper III**, Figure 3), alternative strategies for increasing the soluble protein fraction were not pursued.

3.5 Investigations on how the bacterial host affects expression of heterologous genes that are under control of XylS/*Pm*

E. coli is the predominant bacterium used in recombinant expression [18]. However, limitations of using this host can be seen compared to other bacteria with respect to safety (production of endotoxin) and production of insoluble aggregates (see **Paper I** and **Paper III**). Therefore, the use of alternative bacteria was considered. In the following two subchapters, some unpublished findings will be described involving the use of *P. putida* and cold-adapted members of the genus *Pseudomonas*. These strains were used in very preliminary experiments to test their potential to be used as alternative expression hosts and the the majority of the results have not been published, yet.

3.5.1 Exploring *P. putida* as a host to produce recombinant proteins from XylS/*Pm*

P. putida is the bacterium from which the pWWO plasmid was isolated. This plasmid is the source of the *xylS* gene and the *Pm* promoter meaning that the XylS/*Pm* promoter system is able to function for recombinant gene expression purposes in this bacterium. Moreover, mini-RK2 plasmids are known to be maintained in this organism [65]. Due to these two prerequisites, it was decided to directly transfer some of the newly constructed vectors into *P. putida* KT2440, a strain cured for the pWWO plasmid. In the line of systematic studies performed during this PhD project, production of a target protein was directly compared between *E. coli* and *P. putida* using the exact same plasmids, growth media, growth temperatures and inducer concentrations. The constructs that were chosen contained the gene coding for the red-fluorescent mCherry protein [241] (gene coding region optimized for expression in *E. coli*) under control of XylS/*Pm* connected to different *Pm* 5′-UTR DNA variants (**Figure 3.9**). The first important observation was that in *E. coli* the relative improvement of mCherry production achieved by (dual)UTR variants

compared to the wtwt dualUTR was lower than the relative improvement of β-lactamase production achieved by different (dual)UTR variants (**Figure 3.9A**).

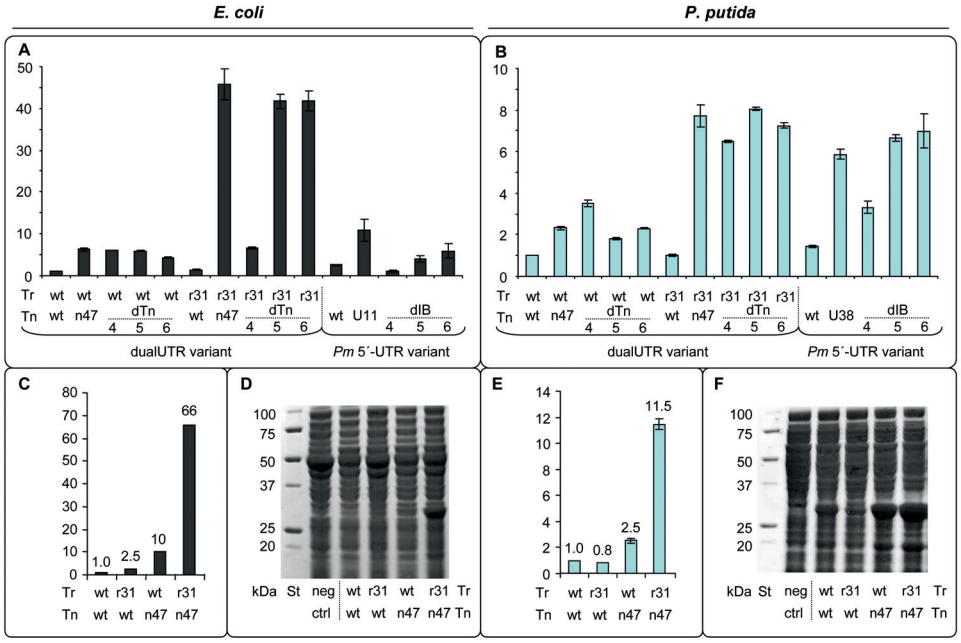

Figure 3.9: Comparative analysis of mCherry production in *E. coli* and *P. putida* under influence of different 5′-UTR DNA sequences. Recombinant *E. coli* DH5α (**A**) and *P. putida* KT2440 (**B**) strains producing mCherry from XylS/*Pm* were grown in 96-well plates at 30 °C and induction was accomplished by addition of 1 mM *m*-toluate after which growth was continued (5 h). Fluorescence values (in arbitrary units) were determined, normalized against OD_{600} and related to data obtained for strains harboring plasmids with the wtwt dualUTR (arbitrarily set to 1.0). Data represent averages and standard deviations from three biological replicas. U11 and U38 are the best 5′-UTR variants that have been identified by Rahmi Lale and Friederike Zwick when screening the same 5′-UTR library for high mCherry expression in *E. coli* or *P. putida*, respectively.[3] 'dTn' variants (length of the Tn-UTR part of the dualUTR) and 'dIB' variants (*Pm* 5′-UTR length) represent 5′-UTR DNA regions that were specifically designed by the RBS calculator to give maximal TIRs for mCherry. The sequences are the same for *E. coli* and *P. putida* due to the shared anti-SD sequence found in the 16S rRNA of these two bacteria. Production of mCherry under influence of different dualUTR variants was also compared between the two bacteria grown in shake flasks. **C** Relative normalized fluorescence values obtained from *E. coli* cultures. **D** SDS-PAGE analysis of the soluble fraction of the total protein obtained from *E. coli.* **E** Relative normalized fluorescence values obtained from *P. putida* cultures. **F** SDS-PAGE analysis of the soluble fraction of the total protein obtained from *P. putida*. Molecular weight of mCherry: 28.8 kDa; St: Precision Plus DualColor (BioRad). (**Paper II** and **unpublished data**)

[3] Part of the project: 'Development of versatile bacterial expression systems for use in recombinant protein production, metabolic engineering, and systems biology'

The relative improvement was even less in *P. putida* (**Figure 3.9B**). As an example, the best UTR DNA sequence (dualUTR r31n47) stimulated mCherry production about 45 times in *E. coli* and only about 8 times in *P. putida*.

The next important finding was that the synergistic effect of combining an improved Tr-UTR DNA sequence with an improved Tn-UTR DNA sequence in the dualUTR context was also true for mCherry in both bacteria (**Figure 3.9A-C** and **E**).

Third, the use of a dualUTR variant gave rise to higher relative mCherry fluorescence compared to a *Pm* 5′-UTR DNA variant. This effect was more evident for *E. coli* than for *P. putida*.

The fourth finding was that application of the RBS calculator to design novel Tn-UTR or *Pm* 5′-UTR DNA sequences generally led to similar fluorescence values than caused by Tn-UTR or *Pm* 5′-UTR DNA sequences identified by screening. An exception was that the dIB4-6 UTR variants tested in *E. coli* led to less improvement of mCherry production than caused by UTR variant U11. This supports the hypothesis from above (**3.2.1**) that a short UTR DNA sequence such as the *Pm* 5′-UTR cannot be optimized for high expression based on improving translation initiation only. When comparing total soluble production of mCherry between the two bacteria (**Figure 3.9D** and **F**), it became evident that production in *P. putida* was clearly higher than in *E. coli*. This observation can be interpreted with a generally higher efficiency of the XylS/*Pm* system in this host. Since this system originates from *P. putida*, one can assume that the action of the host factors (e.g. σ-factors, RNA polymerase etc.) leads to a more efficient transcription from *Pm*. This could explain the finding that the r31wt dualUTR DNA sequence does not increase mCherry production compared to the wtwt sequence. Just like in Zwick's study [40], there is a limit in how far protein production can be stimulated by just enhancing transcription. In her study, a XylS variant in combination with a *Pm* promoter variant (both acting on transcription) could not improve *bla* expression further than achieved by the variants alone. In this case, when we assume that the *Pm* promoter is stronger in *P. putida* than in *E. coli*, an additional effect on transcription exerted by the r31 UTR variant would not be evident on the protein product level. In fact, r31 even decreased mCherry production compared to the wt Tr-UTR.

To summarize this analysis it can be concluded that *P. putida* KT2440 is a well-suited host for expression of mCherry. The mini-RK2 plasmids containing XylS/*Pm* can be used as expression system in this host making it possible to explore the influence of different genetic elements even further. Also, the XylS/*Pm* system appears to be stronger

in this host compared to *E. coli*. The use of constructs with artificial dualUTR variants was proven to be useful for this host as well. For example, it could be demonstrated that it is possible to adjust the Tn-UTR part of the dualUTR to a certain coding region. Generally, genetic manipulation is simple and standard media such us LB are proven to be suitable. Based on this and the findings described above it can be recommended to test expression of other genes in this host as well.

3.5.2 Use of cold-adapted *Pseudomonas* hosts for recombinant protein production

Besides relying on existing expression platforms, efforts are made on finding novel hosts for recombinant protein production. Sources for novel hosts can be environments with extreme temperatures for example. Especially organisms adapted to cold environments are often explored for their potential to be used as expression hosts. These might circumvent common problems associated with mesophilic hosts like *E. coli* and *P. putida* such that growth of the cold-adapted bacteria would actually be promoted at low temperatures and that thermally labile proteins or proteins prone to aggregation and protease degradation might be functionally expressed [242]. In the MARZymes project[4] which was ongoing during the period of this PhD project, marine arctic microorganisms isolated from environmental samples and proteins selected based on data mining of metagenomic libraries were sent to NTNU for expression analyses. After careful selection, four strains that were shown to be cold-adapted members of the genus *Pseudomonas* (based on 16S rRNA analysis) and 11 genes encoding enzymes with putative functions as proteases, carbohydrases and nucleases were handed over to me. These proteins were mainly produced as insoluble aggregates in *E. coli* (Rahmi Lale/ UiT, personal communication). The rationale was to construct vectors for expression of these 11 genes based on XylS/*Pm* and the mini-RK2 replicon and to explore whether the target proteins could be produced in their soluble form using the four cold-adapted *Pseudomonas* hosts.

To achieve this goal two methods were used:

(1) One-step SLIC (**1.2.4**) to fuse the open reading frame to His6 tag coding regions to the 5′- or the 3′- end of the gene and transfer of these gene fusions to the plasmid backbone in one step.

[4] Molecules for the future - Novel enzyme activities from environmental libraries, UiT, NTNU, Umeå

(2) Colony filtration (CoFi) blot, a medium-scale method to evaluate production of the soluble fraction of the different fusion proteins in different host strains under different growth conditions simultaneously (**Figure 3.10A**).

Application of sequence- and ligase- independent cloning (SLIC)

One-step SLIC was carried out using three DNA fragments. The first fragment comprised the vector backbone. Basically, the backbone covered the region between the NdeI and BamHI sites of pSB-M1x (**Paper I**) where the *trfA* gene was replaced with the *trfA* cop271C variant (made by Hanne Jørgensen). The His6 coding region was generated by annealing two overlapping oligonucleotides which also contained complementary regions towards the vector backbone. The third fragment consisted of a PCR product covering the coding region of the target protein, a complementary region towards the backbone at one end and a complementary region to the His6 coding region at the other end. The SLIC reaction was performed according to the protocol described by Jeong et al. [129] and the plasmids were transferred to *E. coli*. After confirming the correct insertion of the coding region by sequencing, the correct constructs were transferred to *P. putida* KT2440 (control strain) by electroporation and to the four cold-adapted *Pseudomonas* strains L11, H12, H26 and H32 by conjugation (the latter step performed by Rahmi Lale). Out of 22 possible variants (11 target genes, two positions of His6 coding regions), 13 were successfully transferred to *P. putida* KT2440, 13 to L11, 10 to H12, 11 to H26 and 11 to H32.

Establishment and application of the Colony filtration blot method

In parallel to making the genetic constructs, the CoFi method [243,244] was tested. CoFi works as follows: Recombinant strains producing certain proteins which can be detected using specific antibodies are grown on agar plates. When colonies have formed, they are lifted up from the agar plate using a Durapore® membrane (Millipore) which is in turn placed on top of a new agar plate containing an appropriate inducer, colony side facing up. This new plate is then incubated at the desired induction temperature for a certain time (e.g. 30 °C, 5 hours for *E. coli*). Next, a filter sandwich is assembled according to **Figure 3.10A** and covered in lysis buffer. Upon cell lysis, the soluble proteins diffuse through the Durapore® membrane onto the nitrocellulose membrane. This nitrocellulose membrane is then subjected to immunoblotting to visualize the proteins. With this method, one can screen libraries of small and intermediate size (several

thousand clones) for soluble protein expression. The great advantages are that this method is independent of reporter protein FPs which might interfere with the structure of the target protein and that the throughput is higher compared to approaches in 96-well format using filtration or centrifugation to separate the soluble and insoluble protein fractions [245].

This method was first tested using two *E. coli* RV308 strains; one expressing hGH (**Paper I**) which was predominantly expressed in its soluble form and PelB-IFN-α2B [74] which was only detectable in the insoluble protein fraction (**Figure 3.10B**). CoFi results were consistent with previous findings judged from the strong signal for hGH and the weak signal for PelB-IFN-α2B.

Next, the CoFi method was applied to evaluate the cold-adapted *Pseudomonas* strains as novel expression hosts based on their ability to express soluble target proteins (**Figure 3.10C-F**). *P. putida* KT2440, *Pseudomonas* L11, H12, H26 and H32 were grown at 5 °C, 10 °C and 20 °C during the induction period to compare different temperatures. This comparative analysis revealed that signals derived from strains subjected to 10 °C were generally stronger than signals obtained from strains subjected to 5 °C or 20 °C. Moreover, signal intensity was higher for the cold-adapted strains than for *P. putida*. Among the cold-adapted strains, it was H32 and H26 that gave rise to the most intense signals. These results looked very promising because they indicated that the cold-adapted *Pseudomonas* strains were able to produce the target proteins in their soluble form.

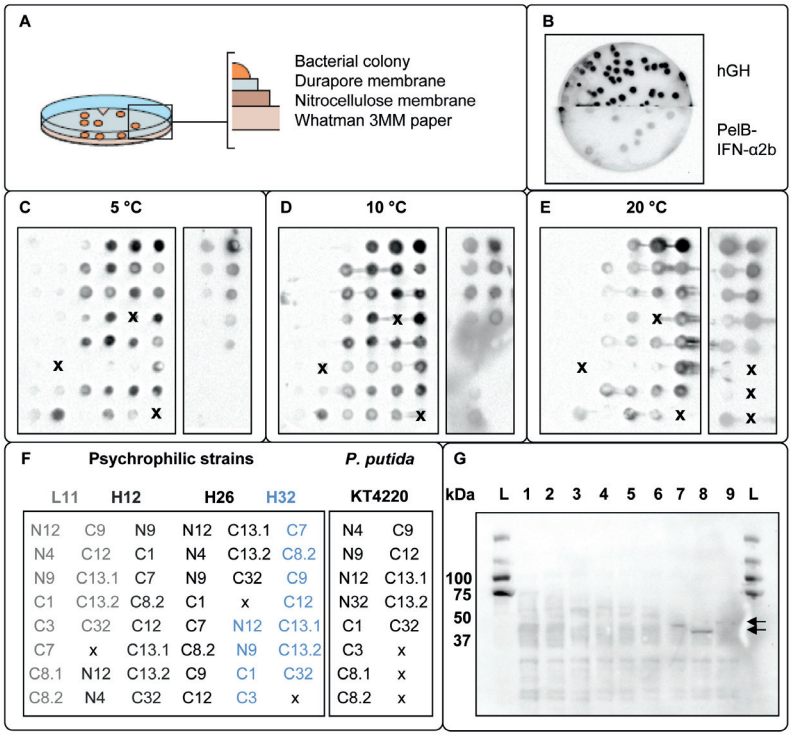

Figure 3.10: Application of the Colony filtration blot (CoFi) method to evaluate cold-adapted *Pseudomonas* strains expressing genes originating from a metagenomic library under different growth conditions. A Schematic representation of the CoFi filter sandwich. First, a Whatman filter is placed into a petri dish followed by a nitrocellulose membrane. These two layers are soaked in lysis buffer containing lysozyme amongst others. A Durapore® membrane with bacterial colonies attached is then placed on top of the nitrocellulose membrane. Cells are lyzed enzymatically and by repeated freeze-and-thaw cycles. The soluble protein fraction diffuses through the Durapore® membrane and is blotted onto the nitrocellulose membrane. After disassembling the filter sandwich, proteins can be detected via specific antibodies [244]. **B** Results from a CoFi trial experiment. *E. coli* RV308 cells were transformed with pSB-M2h harboring the gene encoding human growth hormone which was mainly expressed in its soluble form (**Paper I**) or pVK1 harboring the gene coding for IFN-α2b with an N-terminal PelB FP and C-terminal c-myc and His6 tags [74]. The latter was mainly expressed in its insoluble form. Cell suspensions were spread on agar plates, colonies were lifted up with a Durapore® membrane which was transferred to new agar plates containing 1 mM *m*-toluate and grown for 5h at 30 °C. **C-E** Four different cold-adapted *Pseudomonas* strains (L11, H12, H26 and H32) in addition to *P. putida* KT2440 were transformed with plasmids containing XylS/*Pm* to control expression of interesting genes from a metagenomic library encoding different proteins (No 1,3,4,7,8 (version 1 and 2), 9,12,13 (version 1 and 2) and 32) with N-terminal (N) or C-terminal (C) His_6 tags. x denotes an empty position. Strains were generally grown at 20 °C, but growth was continued at different temperatures after induction (1 mM *m*-toluate). Proteins were detected using the HisProbe™-HRP (ThermoScientific) and the Pierce ECL Western blot substrate. **F** Arrangement of the recombinant strains on agar plates used in **C-E**. **G** Western blot results of recombinant H32 strains expressing different proteins from the metagenomic library. Strains were grown in LB medium at 20 °C until OD_{600} = 0.4-0.5. Then 1 mM *m*-toluate was added and growth was continued at 10 °C over night. A Western blot was performed and the His-

tagged proteins were detected using HisProbe ^{TM}-HRP (1:2,000) The order is: L (Precision Plus Dual Color, BioRad);1 - negative control (H32 without any plasmid); 2 - N12; 3 - C7; 4 - C8; 5 - C9; 6 - C12; 7 - C13.1; 8 - C13.2; 9 - C32 Arrows point to 43 and 36 kDa, the expected molecular weights of proteins C13.1 plus C32 and C13.2, respectively.

To confirm the CoFi results, some strains were selected for experiments in shake-flasks (**Figure 3.10G**). Western blot analysis revealed that expression of the target proteins was generally close to the detection limit. Especially when analyzing the soluble protein fraction, it became evident that the CoFi results and the Western blot result did not match well. The protein no 7 with a C-terminal His-tag for example gave rise to the strongest signal using the CoFi procedure (**Figure 3.10D**). In contrast, no signal was detected using Western blot analysis from samples taken after growth in shake flasks (Lane 3 in **Figure 3.10G**). Only the two variants of protein No 13 gave rise to a weak signal (Lanes 7 and 8 in **Figure 3.10G**). Inspection of the Western blot membrane derived from samples taken from the insoluble protein fraction revealed that several more proteins could be detected. However, others were not detected at all (data not shown). These data led to the conclusion that the cold-adapted *Pseudomonas* strains have a certain potential to be used as expression hosts for heterologous proteins. However, more work would be needed to find more suitable growth conditions (media, shaking, etc.) to be able to transfer results obtained from agar plates to liquid cultures. And although the CoFi method is not 100% reliable [243], it still represents a simple, easy-to-adopt and relatively fast method to select among several conditions and strains simultaneously. The more detailed characterization of the strains of course is left to experiments in liquid cultures.

4 Concluding remarks

The work presented in this thesis addressed different issues related to bacterial recombinant expression with special focus on increasing recombinant protein production. The work mainly dealt with genetic and physiological parameters like regulated promoters (such as XylS/*Pm*), variants of the *Pm* 5′-untranslated region and bacterial host strains (*E. coli* and others) and did not focus on the coding region of the target protein.

One important conclusion is that the systematic comparison of the performance of different regulated promoters in recombinant protein production was necessary to unravel direct effects of these on different steps involved in gene expression. Despite theoretical knowledge about the features of certain regulated promoters, it would not have been possible before to truly evaluate the use of the chosen promoter systems for certain expression purposes, especially XylS/*Pm* and AraC/P_{BAD} which had not been included in systematic comparisons before. This work also confirms the advantage of starting with a small portfolio of bacterial expression platforms including different promoters to identify bottlenecks in expression.

In addition, several strategies are presented how to deal with certain bottlenecks. For once, it could be demonstrated that adjustment of the *Pm* 5′-UTR DNA region might help to prevent that limitations associated with the initiation phase of transcription and translation occur. An especially important contribution was the development of the versatile artificial dualUTR tool which makes it possible to increase protein production in a less gene-dependent, less laborious and more rational manner compared to previous tools. For the future, it would be of great interest to adjust the artificial dualUTR tool to other regulated promoter systems to design an even more flexible set of vectors for early stages in a protein production process.

On top of that, more general strategies to improve protein production are presented which include the incorporation of plasmid stabilization elements in the expression vectors and the use of several bacterial hosts. Further adjustments regarding the *Pm* promoter core region, the regulator protein XylS or the 5′-end of the coding region were already performed by my former colleagues and it would be of great value to combine the genetic engineering strategies of XylS/*Pm* presented here with strategies presented in related work in the future.

Although incorporation of the different presented parameters would mean that a range of strains needs to be constructed and tested, use of modern cloning and screening

technologies can facilitate higher throughput than achieved by constructing and testing strains one by one.

References

1. Cohen SN, Chang AC, Boyer HW, Helling RB (1973) Construction of biologically functional bacterial plasmids *in vitro*. Proc Natl Acad Sci U S A 70: 3240-3244.

2. Donzeau M, Knappik A (2007) Recombinant monoclonal antibodies. Methods Mol Biol 378: 14-31.

3. Rao MB, Tanksale AM, Ghatge MS, Deshpande VV (1998) Molecular and biotechnological aspects of microbial proteases. Microbiol Mol Biol Rev 62: 597-635.

4. Lin L, Xu J (2013) Dissecting and engineering metabolic and regulatory networks of thermophilic bacteria for biofuel production. Biotechnol Adv 31: 827-837.

5. Morris MC (2010) Fluorescent biosensors of intracellular targets from genetically encoded reporters to modular polypeptide probes. Cell Biochem Biophys 56: 19-37.

6. Netzer R, Stafsnes MH, Andreassen T, Goksøyr A, Bruheim P, et al. (2010) Biosynthetic pathway for γ-cyclic sarcinaxanthin in *Micrococcus luteus*: heterologous expression and evidence for diverse and multiple catalytic functions of C_{50} carotenoid cyclases. J Bacteriol 192: 5688-5699.

7. Jakobsen OM, Brautaset T, Degnes KF, Heggeset TM, Balzer S, et al. (2009) Overexpression of wild-type aspartokinase increases L-lysine production in the thermotolerant methylotrophic bacterium *Bacillus methanolicus*. Appl Environ Microbiol 75: 652-661.

8. Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, et al. (1977) Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science 198: 1056-1063.

9. Frommer WB, Ninnemann O (1995) Heterologous Expression of Genes in Bacterial, Fungal, Animal, and Plant Cells. Annu Rev Plant Physiol Plant Mol Biol 46: 419-444.

10. Yin J, Li G, Ren X, Herrler G (2007) Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. J Biotechnol 127: 335-347.

11. Kim Y, Bigelow L, Borovilos M, Dementieva I, Duggan E, et al. (2008) Chapter 3. High-throughput protein purification for x-ray crystallography and NMR. Adv Protein Chem Struct Biol 75: 85-105.

12. Broadway N (2012) Recombinant Protein Expression: Vector-Host Systems. Berkhamsted, Hertfordshire, United Kingdom: Labome- The world of laboratories.

13. Ferrer-Miralles N, Domingo-Espín J, Corchero JL, Vázquez E, Villaverde A (2009) Microbial factories for recombinant pharmaceuticals. Microb Cell Fact 8: 17.

14. Demain AL, Vaishnav P (2009) Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv 27: 297-306.

15. Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol 189: 113-130.

16. Wang YR, Zhang YHP (2009) Overexpression and simple purification of the *Thermotoga maritima* 6-phosphogluconate dehydrogenase in *Escherichia coli* and its application for NADPH regeneration. Microb Cell Fact 8.

17. Welch M, Villalobos A, Gustafsson C, Minshull J (2011) Designing genes for successful protein expression. Methods Enzymol 498: 43-66.

18. Terpe K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 72: 211-222.

19. Schumann W, Ferreira LCS (2004) Production of recombinant proteins in *Escherichia coli*. Genetics and Molecular Biology 27: 442-453.

20. Sletta H, Tøndervik A, Hakvåg S, Aune TE, Nedal A, et al. (2007) 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 73: 906-912.

21. Sørensen HP, Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J Biotechnol 115: 113-128.

22. Kapust RB, Waugh DS (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 8: 1668-1674.

23. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S (2005) Comparison of affinity tags for protein purification. Protein Expr Purif 41: 98-105.

24. Medina MG, Carbonell X, Villaverde A (2002) Connection between gene dosage and protein stability revealed by a high-yield production of recombinant proteins in an *E. coli* LexA1(Ind-) background. Biotechnol Bioeng 78: 722-730.

25. Berg L, Lale R, Bakke I, Burroughs N, Valla S (2009) 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 2: 379-389.

26. Zelcbuch L, Antonovsky N, Bar-Even A, Levin-Karp A, Barenholz U, et al. (2013) Spanning high-dimensional expression space using ribosome-binding site combinatorics. Nucleic Acids Res 41: e98.

27. Bakke I, Berg L, Aune TE, Brautaset T, Sletta H, et al. (2009) Random mutagenesis of the *Pm* promoter as a powerful strategy for improvement of recombinant-gene expression. Appl Environ Microbiol 75: 2002-2011.

28. De Mey M, Maertens J, Lequeux GJ, Soetaert WK, Vandamme EJ (2007) Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. BMC Biotechnol 7: 34.

29. Aune TEV, Bakke I, Drablos F, Lale R, Brautaset T, et al. (2010) Directed evolution of the transcription factor XylS for development of improved expression systems. Microbial Biotechnology 3: 38-47.

30. Lee SK, Chou HH, Pfleger BF, Newman JD, Yoshikuni Y, et al. (2007) Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters. Appl Environ Microbiol 73: 5711-5715.

31. Smolke CD, Carrier TA, Keasling JD (2000) Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. Appl Environ Microbiol 66: 5399-5405.

32. Nielsen J (1998) Metabolic engineering: techniques for analysis of targets for genetic manipulations. Biotechnol Bioeng 58: 125-132.

33. Keasling JD (2010) Manufacturing molecules through metabolic engineering. Science 330: 1355-1358.

34. Pfleger BF, Pitera DJ, Newman JD, Martin VJ, Keasling JD (2007) Microbial sensors for small molecules: development of a mevalonate biosensor. Metab Eng 9: 30-38.

35. Pfleger BF, Pitera DJ, Smolke CD, Keasling JD (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol 24: 1027-1032.

36. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, et al. (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol 27: 753-759.

37. Keasling JD (2012) Synthetic biology and the development of tools for metabolic engineering. Metab Eng 14: 189-195.

38. Fung E, Wong WW, Suen JK, Bulter T, Lee SG, et al. (2005) A synthetic gene-metabolic oscillator. Nature 435: 118-122.

39. Siuti P, Yazbek J, Lu TK (2013) Synthetic circuits integrating logic and memory in living cells. Nat Biotechnol 31: 448-452.

40. Zwick F, Lale R, Valla S (2012) Strong stimulation of recombinant protein production in *Escherichia coli* by combining stimulatory control elements in the expression cassette. Microb Cell Fact 11: 133.

41. Schwarz D, Klammt C, Koglin A, Löhr F, Schneider B, et al. (2007) Preparative scale cell-free expression systems: new tools for the large scale preparation of integral membrane proteins for functional and structural studies. Methods 41: 355-369.

42. Kues U, Stahl U (1989) Replication of Plasmids in Gram-Negative Bacteria. Microbiological Reviews 53: 491-516.

43. Suzuki H, Yano H, Brown CJ, Top EM (2010) Predicting plasmid promiscuity based on genomic signature. J Bacteriol 192: 6045-6055.

44. Kittleson JT, Cheung S, Anderson JC (2011) Rapid optimization of gene dosage in *E. coli* using DIAL strains. J Biol Eng 5: 10.

45. Couturier M, Bex F, Bergquist PL, Maas WK (1988) Identification and classification of bacterial plasmids. Microbiol Rev 52: 375-395.

46. Friehs K (2004) Plasmid copy number and plasmid stability. Adv Biochem Eng Biotechnol 86: 47-82.

47. Løbner-Olesen A (1999) Distribution of minichromosomes in individual *Escherichia coli* cells: implications for replication control. EMBO J 18: 1712-1721.

48. Jain A, Srivastava P (2013) Broad host range plasmids. FEMS Microbiol Lett 348: 87-96.

49. del Solar G, Giraldo R, Ruiz-Echevarria MJ, Espinosa M, Diaz-Orejas R (1998) Replication and control of circular bacterial plasmids. Microbiol Mol Biol Rev 62: 434-464.

50. Wild J, Szybalski W (2004) Copy-control tightly regulated expression vectors based on pBAC/*oriV*. Methods Mol Biol 267: 155-167.

51. Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T (1998) Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2 in *Escherichia coli*. Applied and Environmental Microbiology 64: 1694-1699.

52. Kirienko NV, Lepikhov KA, Zheleznaya LA, Matvienko NI (2004) Significance of codon usage and irregularities of rare codon distribution in genes for expression of BspLU11III methyltransferases. Biochemistry-Moscow 69: 527-535.

53. Tolia NH, Joshua-Tor L (2006) Strategies for protein coexpression in *Escherichia coli*. Nature Methods 3: 55-64.

54. Novick RP (1987) Plasmid Incompatibility. Microbiological Reviews 51: 381-395.

55. Silva F, Queiroz JA, Domingues FC (2012) Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*. Biotechnol Adv 30: 691-708.

56. Pedersen K, Gerdes K (1999) Multiple *hok* genes on the chromosome of *Escherichia coli*. Molecular Microbiology 32: 1090-1102.

57. Peubez I, Chaudet N, Mignon C, Hild G, Husson S, et al. (2010) Antibiotic-free selection in *E. coli*: new considerations for optimal design and improved production. Microb Cell Fact 9: 65.

58. Ingram LC, Richmond MH, Sykes RB (1973) Molecular Characterization of R-Factors Implicated in Carbenicillin Resistance of a Sequence of Pseudomonas-Aeruginosa Strains Isolated from Burns. Antimicrobial Agents and Chemotherapy 3: 279-288.

59. Greener A, Lehman SM, Helinski DR (1992) Promoters of the Broad Host Range Plasmid Rk2 - Analysis of Transcription (Initiation) in 5 Species of Gram-Negative Bacteria. Genetics 130: 27-36.

60. Blatny JM, Brautaset T, Winther-Larsen HC, Haugan K, Valla S (1997) Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. Appl Environ Microbiol 63: 370-379.

61. Toukdarian AE, Helinski DR (1998) TrfA dimers play a role in copy-number control of RK2 replication. Gene 223: 205-211.

62. Toukdarian A (2004) Plasmid strategies for broad-host-range replication in gram-negative bacteria. In: Funnell BE, Phillips GJ, editors. Plasmid Biology. Washington D.C.: ASM Press. pp. 1259-1270.

63. Pinkney M, Diaz R, Lanka E, Thomas CM (1988) Replication of mini RK2 plasmid in extracts of *Escherichia coli* requires plasmid-encoded protein TrfA and host-encoded proteins DnaA, DnaB, DnaG DNA gyrase and DNA polymerase III. Journal of Molecular Biology 203: 927-938.

64. Kline BC (1985) A review of mini-F plasmid maintenance. Plasmid 14: 1-16.

65. Kolatka K, Kubik S, Rajewska M, Konieczny I (2010) Replication and partitioning of the broad-host-range plasmid RK2. Plasmid 64: 119-134.

66. Thomas CM, Helinski DR (1989) Vegetative replication and stable inheritance of IncP plasmids. In: Thomas CM, editor. Promiscuous plasmids of Gram-negative bacteria. San Diego, CA: Academic Press. pp. 1-25.

67. Durland RH, Toukdarian A, Fang F, Helsinki DR (1990) Mutations in the *trfA* replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy numbers. J Bacteriol 172: 3859-3867.

68. Sletta H, Nedal A, Aune TE, Hellebust H, Hakvåg S, et al. (2004) 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 70: 7033-7039.

69. Zwick F, Lale R, Valla S (unpublished) Increased plasmid copy number allows better utilization of cell resources. Trondheim: Department of Biotechnology, Norwegian University of Science and Technology.

70. Verheust C, Helinski DR (2007) The *incC korB* region of RK2 repositions a mini-RK2 replicon in *Escherichia coli*. Plasmid 58: 195-204.

71. Easter CL, Schwab H, Helinski DR (1998) Role of the *parCBA* operon of the broad-host-range plasmid RK2 in stable plasmid maintenance. J Bacteriol 180: 6023-6030.

72. Niki H, Hiraga S (1999) Subcellular localization of plasmids containing the *oriC* region of the *Escherichia coli* chromosome, with or without the *sopABC* partitioning system. Mol Microbiol 34: 498-503.

73. Erdmann N, Petroff T, Funnell BE (1999) Intracellular localization of P1 ParB protein depends on ParA and *parS*. Proc Natl Acad Sci U S A 96: 14905-14910.

74. Berg L, Kucharova V, Bakke I, Valla S, Brautaset T (2011) Exploring the 5'-UTR DNA region as a target for optimizing recombinant gene expression from the strong and inducible *Pm* promoter in *Escherichia coli*. J Biotechnol 158: 224-230.

75. Sobecky PA, Easter CL, Bear PD, Helinski DR (1996) Characterization of the stable maintenance properties of the *par* region of broad-host-range plasmid RK2. J Bacteriol 178: 2086-2093.

76. Aakvik T, Degnes KF, Dahlsrud R, Schmidt F, Dam R, et al. (2009) A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. FEMS Microbiol Lett 296: 149-158.

77. Balbas P, Soberon X, Bolivar F, Rodriguez RL (1988) The plasmid, pBR322. Biotechnology 10: 5-41.

78. Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. J Bacteriol 177: 4121-4130.

79. Balbás P, Soberón X, Merino E, Zurita M, Lomeli H, et al. (1986) Plasmid vector pBR322 and its special-purpose derivatives - a review. Gene 50: 3-40.

80. Chiang CS, Bremer H (1988) Stability of pBR322-derived plasmids. Plasmid 20: 207-220.

81. Dong WR, Xiang LX, Shao JZ (2010) Novel antibiotic-free plasmid selection system based on complementation of host auxotrophy in the NAD *de novo* synthesis pathway. Appl Environ Microbiol 76: 2295-2303.

82. Ramos JL, Marqués S, Timmis KN (1997) Transcriptional control of the *Pseudomonas* TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators. Annu Rev Microbiol 51: 341-373.

83. Michán C, Zhou L, Gallegos MT, Timmis KN, Ramos JL (1992) Identification of critical amino-terminal regions of XylS. The positive regulator encoded by the TOL plasmid. J Biol Chem 267: 22897-22901.

84. Ruíz R, Marqués S, Ramos JL (2003) Leucines 193 and 194 at the N-terminal domain of the XylS protein, the positive transcriptional regulator of the TOL meta-cleavage pathway, are involved in dimerization. J Bacteriol 185: 3036-3041.

85. Ruíz R, Ramos JL (2002) Residues 137 and 153 at the N terminus of the XylS protein influence the effector profile of this transcriptional regulator and the sigma factor used by RNA polymerase to stimulate transcription from its cognate promoter. J Biol Chem 277: 7282-7286.

86. Ruíz R, Ramos JL (2001) Residues 137 and 153 of XylS influence contacts with the C-terminal domain of the RNA polymerase alpha subunit. Biochem Biophys Res Commun 287: 519-521.

87. González-Pérez MM, Marqués S, Dominguez-Cuevas P, Ramos JL (2002) XylS activator and RNA polymerase binding sites at the *Pm* promoter overlap. FEBS Lett 519: 117-122.

88. Marqués S, Manzanera M, González-Pérez MM, Gallegos MT, Ramos JL (1999) The XylS-dependent *Pm* promoter is transcribed *in vivo* by RNA polymerase with σ^{32} or sigma σ^{38} depending on the growth phase. Mol Microbiol 31: 1105-1113.

89. Lale R, Berg L, Stüttgen F, Netzer R, Stafsnes M, et al. (2011) Continuous control of the flow in biochemical pathways through 5' untranslated region sequence modifications in mRNA expressed from the broad-host-range promoter *Pm*. Appl Environ Microbiol 77: 2648-2655.

90. Heggeset TM, Kucharova V, Naerdal I, Valla S, Sletta H, et al. (2012) Combinatorial mutagenesis and selection of improved signal sequences and their application for high-level production of translocated heterologous proteins in *Escherichia coli*. Appl Environ Microbiol 79: 559-568.

91. Kucharova V, Skancke J, Brautaset T, Valla S (2013) Design and optimization of short DNA sequences that can be used as 5' fusion partners for high-level expression of heterologous genes in *Escherichia coli*. Appl Environ Microbiol.

92. Winther-Larsen HC, Blatny JM, Valand B, Brautaset T, Valla S (2000) *Pm* promoter expression mutants and their use in broad-host-range RK2 plasmid vectors. Metab Eng 2: 92-103.

93. Wilson BS, Kautzer CR, Antelman DE (1994) Increased protein expression through improved ribosome-binding sites obtained by library mutagenesis. Biotechniques 17: 944-953.

94. Zwick F, Lale R, Valla S (2013) Regulation of the expression level of transcription factor XylS reveals new functional insight into its induction mechanism at the *Pm* promoter. BMC Microbiol 13: 262.

95. Brautaset T, Lale R, Valla S (2009) Positively regulated bacterial expression systems. Microbial Biotechnology 2: 15-30.

96. Schleif R (2010) AraC protein, regulation of the $_{L}$-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. FEMS Microbiol Rev 34: 779-796.

97. Schleif R (2003) AraC protein: a love-hate relationship. Bioessays 25: 274-282.

98. Lobell RB, Schleif RF (1990) DNA looping and unlooping by AraC protein. Science 250: 528-532.

99. Martin K, Huo L, Schleif RF (1986) The DNA loop model for *ara* repression: AraC protein occupies the proposed loop sites in vivo and repression-negative mutations lie in these same sites. Proc Natl Acad Sci U S A 83: 3654-3658.

100. Lee NL, Gielow WO, Wallace RG (1981) Mechanism of *araC* autoregulation and the domains of two overlapping promoters, P_C and P_{BAD}, in the L-arabinose regulatory region of *Escherichia coli*. Proc Natl Acad Sci U S A 78: 752-756.

101. Carra JH, Schleif RF (1993) Variation of half-site organization and DNA looping by AraC protein. EMBO J 12: 35-44.

102. Khlebnikov A, Datsenko KA, Skaug T, Wanner BL, Keasling JD (2001) Homogeneous expression of the P(BAD) promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. Microbiology 147: 3241-3247.

103. Brown CE, Hogg RW (1972) A second transport system for L-arabinose in *Escherichia coli* B-r controlled by the *araC* gene. J Bacteriol 111: 606-613.

104. Miyada CG, Stoltzfus L, Wilcox G (1984) Regulation of the *araC* gene of *Escherichia coli*: catabolite repression, autoregulation, and effect on *araBAD* expression. Proc Natl Acad Sci U S A 81: 4120-4124.

105. Morgan-Kiss RM, Wadler C, Cronan JE, Jr. (2002) Long-term and homogeneous regulation of the *Escherichia coli araBAD* promoter by use of a lactose transporter of relaxed specificity. Proc Natl Acad Sci U S A 99: 7373-7377.

106. Nannenga BL, Baneyx F (2011) Enhanced expression of membrane proteins in *E. coli* with a P_{BAD} promoter mutant: synergies with chaperone pathway engineering strategies. Microb Cell Fact 10: 105.

107. Niland P, Hühne R, Müller-Hill B (1996) How AraC Interacts Specifically with its Target DNAs. J Mol Biol 264: 667-674.

108. iGEM_DTU (2011) Improving *araBAD*.

109. Steitz TA (2009) The structural changes of T7 RNA polymerase from transcription initiation to elongation. Curr Opin Struct Biol 19: 683-690.

110. Warren JW, Walker JR, Roth JR, Altman E (2000) Construction and characterization of a highly regulable expression vector, pLAC11, and its multipurpose derivatives, pLAC22 and pLAC33. Plasmid 44: 138-151.

111. Dubendorff JW, Studier FW (1991) Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with *lac* repressor. J Mol Biol 219: 45-59.

112. Novyan R, Morris B Use of glucose to control basal expression in the pET System. inNovations.

113. Li Z, Kessler W, van den Heuvel J, Rinas U (2011) Simple defined autoinduction medium for high-level recombinant protein production using T7-based *Escherichia coli* expression systems. Appl Microbiol Biotechnol 91: 1203-1213.

114. Iost I, Guillerez J, Dreyfus M (1992) Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. J Bacteriol 174: 619-622.

115. Vethanayagam JG, Flower AM (2005) Decreased gene expression from T7 promoters may be due to impaired production of active T7 RNA polymerase. Microb Cell Fact 4: 3.

116. Amann E, Ochs B, Abel KJ (1988) Tightly regulated *tac* promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene 69: 301-315.

117. Chen R (2012) Bacterial expression systems for recombinant protein production: *E. coli* and beyond. Biotechnol Adv 30: 1102-1107.

118. Schwartz M, Lo HS, Zhou JS, Yang P, Liao A (2011) Gene synthesis: a cost-effective alternative to traditional molecular cloning. Molecular Therapy 19: 1386-1386.

119. Hammarström M, Hellgren N, van Den Berg S, Berglund H, Härd T (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. Protein Sci 11: 313-321.

120. Westereng B, Ishida T, Vaaje-Kolstad G, Wu M, Eijsink VG, et al. (2011) The putative endoglucanase PcGH61D from *Phanerochaete chrysosporium* is a metal-dependent oxidative enzyme that cleaves cellulose. PLoS One 6: e27807.

121. Domínguez-Cuevas P, Ramos JL, Marqués S (2010) Sequential XylS-CTD binding to the *Pm* promoter induces DNA bending prior to activation. J Bacteriol 192: 2682-2690.

122. Knight T (2003) Idempotent Vector Design for Standard Assembly of Biobricks. MIT Artificial Intelligence Laboratory; MIT Synthetic Biology Working Group

123. Anderson JC, Dueber JE, Leguia M, Wu GC, Goler JA, et al. (2010) BglBricks: A flexible standard for biological part assembly. J Biol Eng 4: 1.

124. Silva-Rocha R, Martínez-García E, Calles B, Chavarría M, Arce-Rodríguez A, et al. (2013) The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. Nucleic Acids Res 41: D666-675.

125. Engler C, Gruetzner R, Kandzia R, Marillonnet S (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One 4: e5553.

126. Hunt I (2005) From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. Protein Expr Purif 40: 1-22.

127. Quan J, Tian J (2009) Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4: e6441.

128. Gibson DG, Smith HO, Hutchison CA, 3rd, Venter JC, Merryman C (2010) Chemical synthesis of the mouse mitochondrial genome. Nat Methods 7: 901-903.

129. Jeong JY, Yim HS, Ryu JY, Lee HS, Lee JH, et al. (2012) One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. Appl Environ Microbiol 78: 5440-5443.

130. Lorence A (2012) Recombinant Gene Expression; Walker JM, editor. New York, NY: Humana Press. 649 p.

131. Makino T, Skretas G, Georgiou G (2011) Strain engineering for improved expression of recombinant proteins in bacteria. Microb Cell Fact 10: 32.

132. Valderrama-Rincon JD, Fisher AC, Merritt JH, Fan YY, Reading CA, et al. (2012) An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. Nat Chem Biol 8: 434-436.

133. Fang H, Zhang X, Shen L, Si X, Ren Y, et al. (2009) RimJ is responsible for N^{α}-acetylation of thymosin α1 in *Escherichia coli*. Appl Microbiol Biotechnol 84: 99-104.

134. Young TS, Ahmad I, Yin JA, Schultz PG (2009) An enhanced system for unnatural amino acid mutagenesis in *E. coli*. J Mol Biol 395: 361-374.

135. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2: 2006 0008.

136. Dammeyer T, Steinwand M, Krüger SC, Dübel S, Hust M, et al. (2011) Efficient production of soluble recombinant single chain Fv fragments by a *Pseudomonas putida* strain KT2440 cell factory. Microb Cell Fact 10: 11.

137. Nogales J, Palsson BO, Thiele I (2008) A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. BMC Syst Biol 2: 79.

138. Víctor de Lorenzo's lab at Centro Nacional de Biotecnología - Research. Madrid.

139. Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol 25: 117-124.

140. Greive SJ, von Hippel PH (2005) Thinking quantitatively about transcriptional regulation. Nat Rev Mol Cell Biol 6: 221-232.

141. Browning DF, Busby SJ (2004) The regulation of bacterial transcription initiation. Nat Rev Microbiol 2: 57-65.

142. Gruber TM, Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol 57: 441-466.

143. Haugen SP, Ross W, Gourse RL (2008) Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat Rev Microbiol 6: 507-519.

144. Ishihama A (2000) Functional modulation of *Escherichia coli* RNA polymerase. Annu Rev Microbiol 54: 499-518.

145. Meinhardt S, Manley MW, Jr., Becker NA, Hessman JA, Maher LJ, 3rd, et al. (2012) Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression. Nucleic Acids Res 40: 11139-11154.

146. Stepanova EV, Shevelev AB, Borukhov SI, Severinov KV (2009) Mechanisms of action of RNA polymerase-binding transcription factors that do not bind to DNA. Biofizika 54: 773-790.

147. Mekler V, Kortkhonjia E, Mukhopadhyay J, Knight J, Revyakin A, et al. (2002) Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. Cell 108: 599-614.

148. Goldman SR, Ebright RH, Nickels BE (2009) Direct detection of abortive RNA transcripts *in vivo*. Science 324: 927-928.

149. Hsu LM (2002) Promoter clearance and escape in prokaryotes. Biochim Biophys Acta 1577: 191-207.

150. Brieba LG, Sousa R (2001) T7 promoter release mediated by DNA scrunching. EMBO J 20: 6826-6835.

151. Epshtein V, Cardinale CJ, Ruckenstein AE, Borukhov S, Nudler E (2007) An allosteric path to transcription termination. Mol Cell 28: 991-1001.

152. Banerjee S, Chalissery J, Bandey I, Sen R (2006) Rho-dependent transcription termination: more questions than answers. J Microbiol 44: 11-22.

153. Barne KA, Bown JA, Busby SJ, Minchin SD (1997) Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the 'extended-10' motif at promoters. EMBO J 16: 4034-4040.

154. Estrem ST, Ross W, Gaal T, Chen ZW, Niu W, et al. (1999) Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase α subunit. Genes Dev 13: 2134-2147.

155. Ross W, Ernst A, Gourse RL (2001) Fine structure of E. coli RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. Genes Dev 15: 491-506.

156. Durniak KJ, Bailey S, Steitz TA (2008) The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. Science 322: 553-557.

157. Mekler V, Severinov K (2013) Cooperativity and interaction energy threshold effects in recognition of the -10 promoter element by bacterial RNA polymerase. Nucleic Acids Res 41: 7276-7285.

158. Khandelwal G, Bhyravabhotla J (2010) A phenomenological model for predicting melting temperatures of DNA sequences. PLoS One 5: e12433.

159. Larson MH, Landick R, Block SM (2011) Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes. Mol Cell 41: 249-262.

160. Artsimovitch I, Landick R (1998) Interaction of a nascent RNA structure with RNA polymerase is required for hairpin-dependent transcriptional pausing but not for transcript release. Genes Dev 12: 3110-3122.

161. Artsimovitch I, Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc Natl Acad Sci U S A 97: 7090-7095.

162. Henkin TM, Yanofsky C (2002) Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. Bioessays 24: 700-707.

163. Pan T, Sosnick T (2006) RNA folding during transcription. Annu Rev Biophys Biomol Struct 35: 161-175.

164. Hatoum A, Roberts J (2008) Prevalence of RNA polymerase stalling at *Escherichia coli* promoters after open complex formation. Mol Microbiol 68: 17-28.

165. Perdue SA, Roberts JW (2010) A backtrack-inducing sequence is an essential component of *Escherichia coli* σ^{70}-dependent promoter-proximal pausing. Mol Microbiol 78: 636-650.

166. Zhilina EV, Miropolskaya NA, Bass IA, Brodolin KL, Kulbachinskiy AV (2011) Characteristics of σ-dependent pausing by RNA polymerases from *Escherichia col*i and *Thermus aquaticus*. Biochemistry (Mosc) 76: 1098-1106.

167. Epshtein V, Nudler E (2003) Cooperation between RNA polymerase molecules in transcription elongation. Science 300: 801-805.

168. Epshtein V, Toulme F, Rahmouni AR, Borukhov S, Nudler E (2003) Transcription through the roadblocks: the role of RNA polymerase cooperation. EMBO J 22: 4719-4727.

169. Kozak M (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 361: 13-37.

170. Laursen BS, Sørensen HP, Mortensen KK, Sperling-Petersen HU (2005) Initiation of protein synthesis in bacteria. Microbiol Mol Biol Rev 69: 101-123.

171. Simonetti A, Marzi S, Jenner L, Myasnikov A, Romby P, et al. (2009) A structural view of translation initiation in bacteria. Cell Mol Life Sci 66: 423-436.

172. Liu H, Pan D, Pech M, Cooperman BS (2010) Interrupted catalysis: the EF4 (LepA) effect on back-translocation. J Mol Biol 396: 1043-1052.

173. Peske F, Rodnina MV, Wintermeyer W (2005) Sequence of steps in ribosome recycling as defined by kinetic analysis. Mol Cell 18: 403-412.

174. Mata J, Marguerat S, Bähler J (2005) Post-transcriptional control of gene expression: a genome-wide perspective. Trends Biochem Sci 30: 506-514.

175. Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. Science 324: 255-258.

176. Osterman IA, Evfratov SA, Sergiev PV, Dontsova OA (2013) Comparison of mRNA features affecting translation initiation and reinitiation. Nucleic Acids Res 41: 474-486.

177. Ma J, Campbell A, Karlin S (2002) Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. J Bacteriol 184: 5733-5745.

178. Boni IV, Artamonova VS, Tzareva NV, Dreyfus M (2001) Non-canonical mechanism for translational control in bacteria: synthesis of ribosomal protein S1. EMBO J 20: 4222-4232.

179. Nudler E, Mironov AS (2004) The riboswitch control of bacterial metabolism. Trends Biochem Sci 29: 11-17.

180. Wen JD, Lancaster L, Hodges C, Zeri AC, Yoshimura SH, et al. (2008) Following translation by single ribosomes one codon at a time. Nature 452: 598-603.

181. Sørensen MA, Kurland CG, Pedersen S (1989) Codon usage determines translation rate in *Escherichia coli*. J Mol Biol 207: 365-377.

182. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, et al. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141: 344-354.

183. Belasco JG (2010) All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. Nat Rev Mol Cell Biol 11: 467-478.

184. Belasco JG, Higgins CF (1988) Mechanisms of mRNA decay in bacteria: a perspective. Gene 72: 15-23.

185. Xu F, Cohen SN (1995) RNA degradation in *Escherichia coli* regulated by 3' adenylation and 5' phosphorylation. Nature 374: 180-183.

186. Deana A, Celesnik H, Belasco JG (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. Nature 451: 355-358.

187. Pedersen S, Reeh S (1978) Functional mRNA half lives in *E. coli*. Mol Gen Genet 166: 329-336.

188. Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C (2003) Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. Genome Res 13: 216-223.

189. Miller OL, Jr., Hamkalo BA, Thomas CA, Jr. (1970) Visualization of bacterial genes in action. Science 169: 392-395.

190. Burmann BM, Rosch P (2011) The role of *E. coli* Nus-factors in transcription regulation and transcription:translation coupling: From structure to mechanism. Transcription 2: 130-134.

191. Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, et al. (2010) A NusE:NusG complex links transcription and translation. Science 328: 501-504.

192. Li J, Horwitz R, McCracken S, Greenblatt J (1992) NusG, a new *Escherichia coli* elongation factor involved in transcriptional antitermination by the N protein of phage lambda. J Biol Chem 267: 6012-6019.

193. Torres M, Balada JM, Zellars M, Squires C, Squires CL (2004) *In vivo* effect of NusB and NusG on rRNA transcription antitermination. J Bacteriol 186: 1304-1310.

194. Sullivan SL, Gottesman ME (1992) Requirement for *E. coli* NusG protein in factor-dependent transcription termination. Cell 68: 989-994.

195. Burmann BM, Scheckenhofer U, Schweimer K, Rösch P (2011) Domain interactions of the transcription-translation coupling factor *Escherichia coli* NusG are intermolecular and transient. Biochem J 435: 783-789.

196. Proshkin S, Rahmouni AR, Mironov A, Nudler E (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science 328: 504-508.

197. Roberts JW (2010) Molecular biology. Syntheses that stay together. Science 328: 436-437.

198. Gowrishankar J, Harinarayanan R (2004) Why is transcription coupled to translation in bacteria? Mol Microbiol 54: 598-603.

199. Iost I, Dreyfus M (1995) The stability of *Escherichia coli lacZ* mRNA depends upon the simultaneity of its synthesis and translation. EMBO J 14: 3252-3261.

200. Pato ML, Bennett PM, von Meyenburg K (1973) Messenger ribonucleic acid synthesis and degradation in *Escherichia coli* during inhibition of translation. J Bacteriol 116: 710-718.

201. Schneider E, Blundell M, Kennell D (1978) Translation and mRNA decay. Mol Gen Genet 160: 121-129.

202. Nilsson G, Belasco JG, Cohen SN, von Gabain A (1987) Effect of premature termination of translation on mRNA stability depends on the site of ribosome release. Proc Natl Acad Sci U S A 84: 4890-4894.

203. Arnold TE, Yu J, Belasco JG (1998) mRNA stabilization by the *ompA* 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation. RNA 4: 319-330.

204. Wagner LA, Gesteland RF, Dayhuff TJ, Weiss RB (1994) An efficient Shine-Dalgarno sequence but not translation is necessary for *lacZ* mRNA stability in *Escherichia coli*. J Bacteriol 176: 1683-1688.

205. Vytvytska O, Moll I, Kaberdin VR, von Gabain A, Bläsi U (2000) Hfq (HF1) stimulates *ompA* mRNA decay by interfering with ribosome binding. Genes Dev 14: 1109-1118.

206. Komarova AV, Tchufistova LS, Dreyfus M, Boni IV (2005) AU-rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. J Bacteriol 187: 1344-1349.

207. Vimberg V, Tats A, Remm M, Tenson T (2007) Translation initiation region sequence preferences in *Escherichia coli*. BMC Mol Biol 8: 100.

208. McDowall KJ, Lin-Chao S, Cohen SN (1994) A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. J Biol Chem 269: 10790-10796.

209. Kaberdin VR, Bläsi U (2006) Translation initiation and the fate of bacterial mRNAs. FEMS Microbiol Rev 30: 967-979.

210. Deana A, Belasco JG (2005) Lost in translation: the influence of ribosomes on bacterial mRNA decay. Genes Dev 19: 2526-2533.

211. Zhang G, Fedyunin I, Miekley O, Valleriani A, Moura A, et al. (2010) Global and local depletion of ternary complex limits translational elongation. Nucleic Acids Res 38: 4778-4787.

212. Roche ED, Sauer RT (1999) SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. EMBO J 18: 4579-4589.

213. Alper H, Stephanopoulos G (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype. Metab Eng 9: 258-267.

214. Klein-Marcuschamer D, Santos CN, Yu H, Stephanopoulos G (2009) Mutagenesis of the bacterial RNA polymerase α subunit for improvement of complex phenotypes. Appl Environ Microbiol 75: 2705-2711.

215. Voigt CA (2006) Genetic parts to program bacteria. Curr Opin Biotechnol 17: 548-557.

216. Henkin TM (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. Genes Dev 22: 3383-3390.

217. Osterman IA, Evfratov SA, Sergiev PV, Dontsova OA Comparison of mRNA features affecting translation initiation and reinitiation. Nucleic Acids Res 41: 474-486.

218. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 27: 946-950.

219. Rauhut R, Klug G (1999) mRNA degradation in bacteria. FEMS Microbiol Rev 23: 353-370.

220. Emory SA, Bouvet P, Belasco JG (1992) A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. Genes Dev 6: 135-148.

221. Cardinale S, Arkin AP (2012) Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems. Biotechnol J 7: 856-866.

222. Lou C, Stanton B, Chen YJ, Munsky B, Voigt CA (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. Nat Biotechnol 30: 1137-1142.

223. Care S, Bignon C, Pelissier MC, Blanc E, Canard B, et al. (2008) The translation of recombinant proteins in *E. coli* can be improved by *in silico* generating and screening random libraries of a -70/+96 mRNA region with respect to the translation initiation codon. Nucleic Acids Res 36: e6.

224. Qi L, Haurwitz RE, Shao W, Doudna JA, Arkin AP (2012) RNA processing enables predictable programming of gene expression. Nat Biotechnol 30: 1002-1006.

225. Salis HM (2011) The ribosome binding site calculator. Methods Enzymol 498: 19-42.

226. Na D, Lee S, Lee D (2010) Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. BMC Syst Biol 4: 71.

227. Seo SW, Yang JS, Kim I, Yang J, Min BE, et al. (2013) Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. Metab Eng 15: 67-74.

228. Kosuri S, Goodman DB, Cambray G, Mutalik VK, Gao Y, et al. (2013) Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. Proc Natl Acad Sci U S A 110: 14024-14029.

229. Tegel H, Ottosson J, Hober S (2011) Enhancing the protein production levels in *Escherichia coli* with a strong promoter. FEBS J 278: 729-739.

230. Walker KA, Osuna R (2002) Factors affecting start site selection at the *Escherichia coli fis* promoter. J Bacteriol 184: 4783-4791.

231. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, et al. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics 4: 1265-1272.

232. Schlegel S, Rujas E, Ytterberg AJ, Zubarev RA, Luirink J, et al. (2013) Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. Microbial Cell Factories 12.

233. Golomb M, Chamberlin M (1974) Characterization of T7-specific ribonucleic acid polymerase. IV. Resolution of the major in vitro transcripts by gel electrophoresis. J Biol Chem 249: 2858-2863.

234. Giacalone MJ, Gentile AM, Lovitt BT, Berkley NL, Gunderson CW, et al. (2006) Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system. Biotechniques 40: 355-364.

235. Wilms B, Hauck A, Reuss M, Syldatk C, Mattes R, et al. (2001) High-cell-density fermentation for production of L-*N*-carbamoylase using an expression system based on the *Escherichia coli rhaBAD* promoter. Biotechnol Bioeng 73: 95-103.

236. Lale R (2009) The *xylS*/*Pm* expression cassette: New functional insights and application potentials. Trondheim: NTNU. 75 p.

237. Berg L (2009) A mutational analysis of the biotechnologically useful broad-host-range *Pm* promoter in *Escherichia coli*. Trondheim: NTNU. 90 p.

238. Nevo-Dinur K, Govindarajan S, Amster-Choder O (2012) Subcellular localization of RNA and proteins in prokaryotes. Trends Genet 28: 314-322.

239. Esposito D, Chatterjee DK (2006) Enhancement of soluble protein expression through the use of fusion tags. Curr Opin Biotechnol 17: 353-358.

240. Leibly DJ, Nguyen TN, Kao LT, Hewitt SN, Barrett LK, et al. (2013) Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. PLoS One 7: e52482.

241. Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, et al. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma sp.* red fluorescent protein. Nat Biotechnol 22: 1567-1572.

242. Parrilli E, De Vizio D, Cirulli C, Tutino ML (2008) Development of an improved *Pseudoalteromonas haloplanktis* TAC125 strain for recombinant protein secretion at low temperature. Microb Cell Fact 7: 2.

243. Cornvik T, Dahlroth SL, Magnusdottir A, Herman MD, Knaust R, et al. (2005) Colony filtration blot: a new screening method for soluble protein expression in *Escherichia coli*. Nature Methods 2: 507-509.

244. Dahlroth SL, Nordlund P, Cornvik T (2006) Colony filtration blotting for screening soluble expression in *Escherichia coli*. Nat Protoc 1: 253-258.

245. Yumerefendi H, Desravines DC, Hart DJ (2011) Library-based methods for identification of soluble expression constructs. Methods 55: 38-43.

Paper I

Balzer et al. Microbial Cell Factories 2013, 12:26
http://www.microbialcellfactories.com/content/12/1/26

MICROBIAL CELL FACTORIES

RESEARCH Open Access

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

Simone Balzer[1], Veronika Kucharova[1], Judith Megerle[2], Rahmi Lale[1], Trygve Brautaset[3] and Svein Valla[1*]

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), LacI/P_{T7lac}, LacI/P_{trc} and AraC/P_{BAD} to control expression of different proteins with various origins. Generally and not unexpected high expression levels correlate with high replicon copy number and the LacI/P_{T7lac} 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 LacI/P_{T7lac} 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_{BAD} 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, LacI/P_{T7lac}, LacI/P_{trc}, AraC/P_{BAD}, Systematic comparison

* Correspondence: svein.valla@ntnu.no
[1]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_6 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 β-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 *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/*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 *Pm* promoter region [27] as well as the region corresponding to the *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/*Pm* (the native system; denoted in the figures as M); the high level expression variant *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

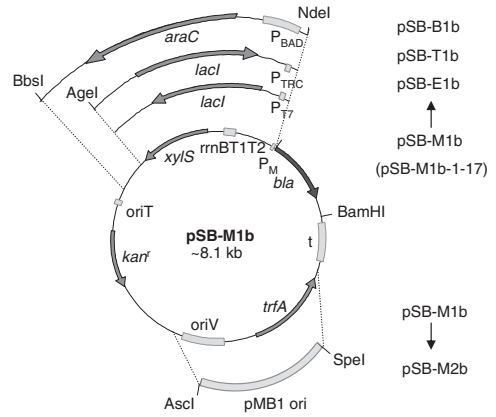

Figure 1 Illustration showing how the different constructs in the study were generated based on pSB-M1b. The upper part shows how the alternative regulator/promoter systems were incorporated. pSB-M1b-1-17 contains a variant of the *Pm* core promoter termed ML1-17 (see text). The lower part shows the *oriV/trfA* region in pSB-M1b that was replaced with the pMB1 *ori* described in Table 1.

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_{T7lac} with XylS/*Pm* because it carries a chromosomal copy of the T7 polymerase integrated into the *lac* operon (NEB). Since the LacI/P_{trc} 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_{trc} due to the specificity of this polymerase for its cognate promoter [40]. Expression from XylS/*Pm* compared to AraC/P_{BAD} was performed in *E. coli* DH10B which is unable to catabolize L-arabinose, the inducer of the AraC/P_{BAD} 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_{T7lac} 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/*Pm* ML1-17 and LacI/P_{T7lac} tended to produce most recombinant protein in the studies in strain ER2566 (Figure 2, Panels A-C). The mutations in the *Pm* core region were of vital importance, as XylS/*Pm* ML1-17 produced between 1.2- and 4.0- fold more active protein than the corresponding wild-type system. Note also that AraC/P_{BAD} generated similar amounts of active protein compared to XylS/*Pm* ML1-17 when the studies were done in an *ara* negative strain (DH10B; Figure 2, Panels D-F). LacI/P_{T7lac} 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/*Pm* 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/*Pm* ML1-17. We also noted in this and other related ongoing studies in our laboratory that to get stable expression from the LacI/P_{T7lac} 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[a]

Name	Key features		Source
pTA16/pSB-M1b[b]	*m*-toluate- inducible *Pm*, *xylS* activator gene, RK2 replicon, *bla* reporter, Kan^r		[31]
pET16b	IPTG-inducible P_{T7lac}, *lacI* repressor gene, Amp^r		Novagen
pBAD/gIII_calmodulin	L-arabinose- inducible P_{BAD}, *araC* activator gene, Amp^r		Invitrogen
pTrc99A	IPTG- inducible P_{trc} promoter, *lacI* repressor gene, Amp^r		Pharmacia
pIB11-luc_S	pIB11 [25] with luc_S under control of *xylS/Pm*, Kan^r		unpublished
pBAD24-GFP	pBAD24 with *gfpmut3* insert, Amp^r		[32]
pHOG-173-2-5-AP	pHOG plasmid with *scFv173-2-5-phoA* fusion gene insert, provided by Affitech AS, Oslo, Amp^r		unpublished
pMA-GH	pMA vector (GeneArt®, Invitrogen) with $GH1_S$ insert, provided by Vectron Biosolutions AS, Trondheim, Amp^r		unpublished
pMA-T-IL-1RA	pMA vector (GeneArt®, Invitrogen) with $IL1RN_S$ insert, provided by Vectron Biosolutions AS, Trondheim, Amp^r		unpublished
pSB-P0x	pSB-M1b variants with combinations of different features:		
	P... regulator/promoter system	M... *xylS/Pm*	
		M-1-17... *xylS/Pm* variant ML1-17	
		E... *lacI/P_{T7lac}* (from pET)	
		T... *$lacI^q$/P_{trc}* (from pTrc)	
		B... *araC/P_{BAD}* (pBAD)	
	0... origin of replication	1... RK2 replicon	
		2... pMB1 replicon	
	x... reporter gene	b... *bla*	
		l... luc_S	
		s... *scFv173-2-5-phoA*	
		g... *gfpmut3*	
		h... $GH1_S$	
		r... $IL1RN_S$	This study
e.g. pSB-M2l	*m*-toluate- inducible P_m, *xylS* activator gene, pMB1 *ori*, luc_S reporter, Kan^r		This study

[a] *bla:* β- lactamase gene; luc_S: 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; $GH1_S$: synthetic gene for human growth hormone, $IL1RN_S$: synthetic gene for human interleukin 1 receptor antagonist.
[b] pTA16 was named pSB-M1g in this study for consistency purposes.

The LacI/P_{T7lac} 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[a] 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 *E. coli* as soluble protein [36,37]
IL-1RA	industrially relevant protein, ~20.1 kDa, cytoplasmic localization, usually expressed in *E. coli* as soluble protein [38,39]

[a]Alkaline phosphatase.

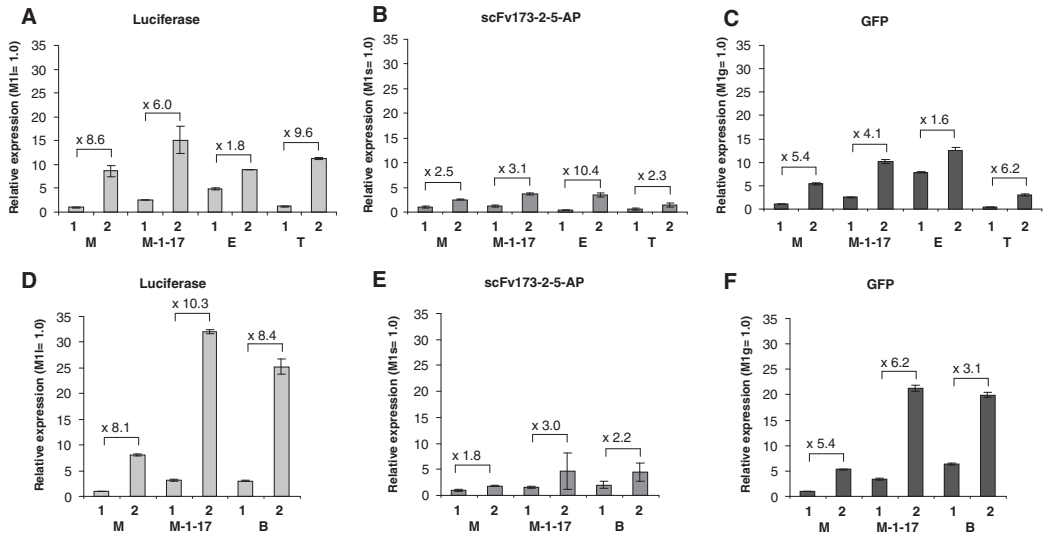

Figure 2 Maximum expression of three different genes placed under control of different regulator/promoter systems. Data represent relative expression levels under induced conditions where the activity of M1x (gene x under conrol of the *Pm* wildtype promoter, RK2 replicon) was set to 1.0. Expression was induced in a way that activity levels were maximized: 2 mM *m*-toluate for strains harboring XylS/*Pm*- based constructs, 1 mM IPTG for those with LacI/*PT7lac*, 0.2 mM IPTG for LacI/*Ptrc* and 0.015% L-arabinose for AraC/*PBAD*. The following *E. coli* strains were used as expression hosts. Panels **A**-**C**: ER2566. Panels **D**-**F**: DH10B. The naming code is the following: The capital letter represents the regulator/promoter system and the digit represents the origin of replication; for details see Table 1. The data presented are from independent biological replica.

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 LacI/P_{T7lac} system generated much more transcript than XylS/*Pm* (between 6.2 and 20 times more) and LacI/P_{trc} (between 3.9 and 206 times more) for all the five tested genes. XylS/*Pm* ML1-17 generated more transcript than LacI/P_{trc} as well (Figure 3), ranging from 3.3 times for luc_S mRNA and 88 times for $GH1_S$ mRNA, except for the special case with *scFv173-2-5-phoA*. Studies of AraC/P_{BAD} 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 LacI/P_{T7lac} 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 LacI/P_{T7lac}, and the final outcome at the protein level was more similar for these proteins than for luciferase. Generally, LacI/P_{T7lac} 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 LacI/P_{T7lac} 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 LacI/P_{T7lac} 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].

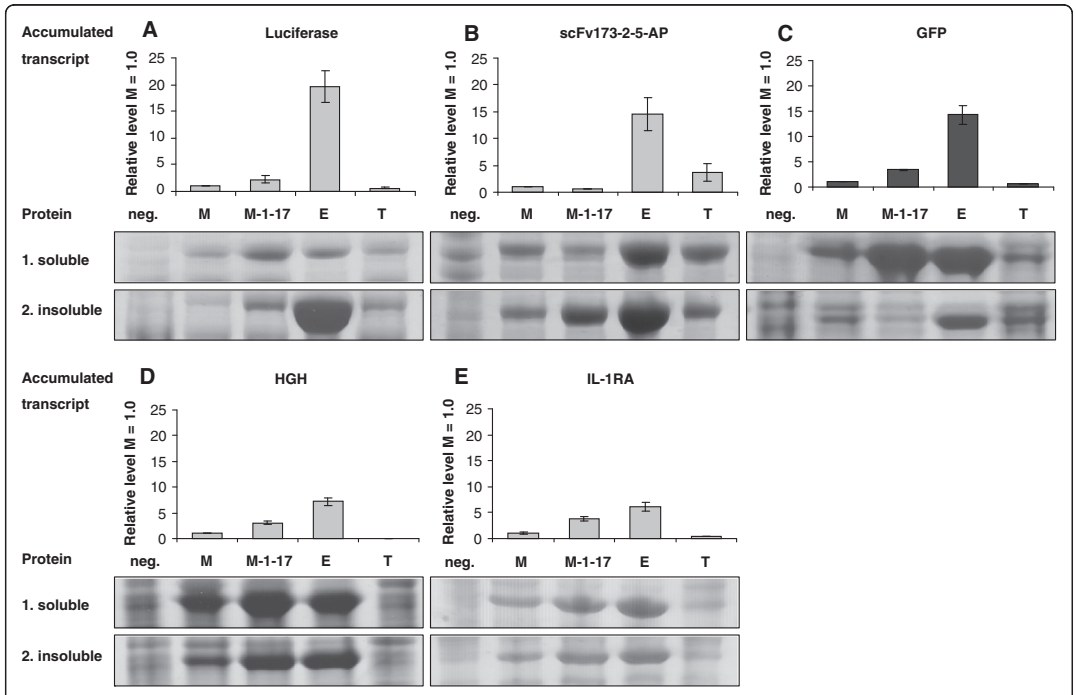

Figure 3 Correlation between the accumulated transcript and protein produced after induction. The five proteins (Panels **A**-**E**) were expressed in *E. coli* ER2566 harboring pMB1-based plasmids. Five hours after induction, samples were collected for relative quantification real-time RT-PCR (qRT-PCR) and SDS-PAGE. Accumulated transcript data were correlated to the XylS/*Pm* system (M2x; gene x under conrol of the *Pm* wildtype promoter, pMB1 replicon). The total protein fractions were separated into the soluble supernatant and the insoluble pellet fraction after sonication and separated through SDS-PAGE followed by staining with Coomassie Brilliant blue. Further information about the naming system can be found in Table 1. Neg: Negative control.

Uninduced expression levels are highest for LacI/P_{trc} and lowest for AraC/P_{BAD}

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, LacI/P_{trc} 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_{BAD} 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. LacI/P_{T7lac} 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 LacI/P_{T7lac} displayed the highest induction windows, while LacI/P_{trc} was by far the least inducible system (0.1-0.2 times compared to XylS/*Pm*). In DH10B, induction ratios for AraC/P_{BAD} 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_{BAD} 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_{BAD} system is that the inducer can be metabolized in most strains of *E. coli*.

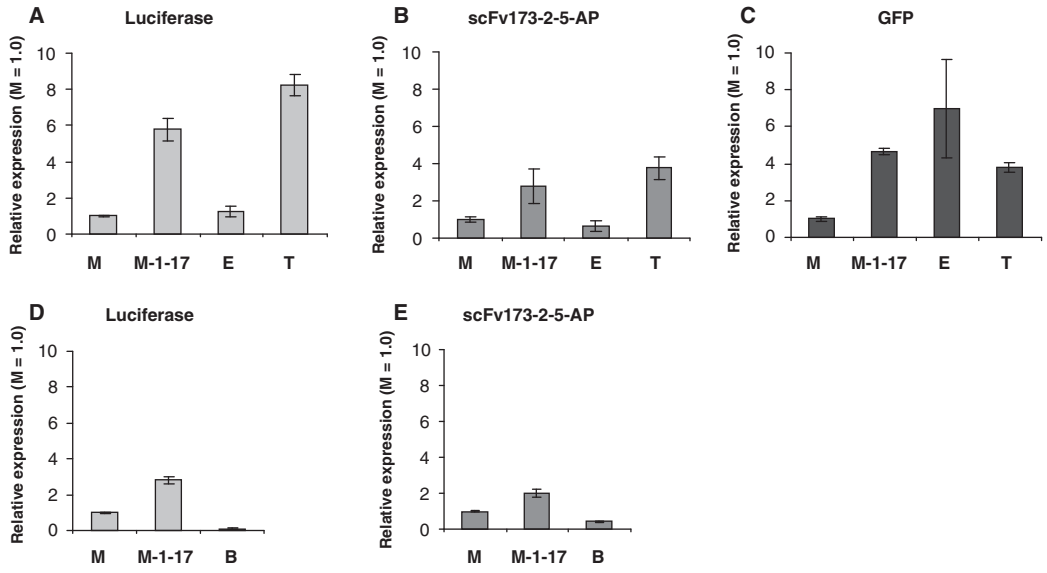

Figure 4 Investigation of the tightness of different regulator/promoter systems in combination with the pMB1 replicon. Protein activity was determined in parallel with induced cultures at the time point corresponding to five hours after induction. The data presented are from independet biological replica. The following *E. coli* strains were used as expression hosts. Panels **A**-**C**: ER2566. Panels **D**-**E**: DH10B. The capital letters represent the regulator/ promoter systems according to Table 1. Uninduced expression of GFP in DH10B was very close to the detection limit in LB medium and was left out.

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_{T7lac} for HGH and AraC/P_{BAD} for IL-1RA). Generally, the calculator predicted that the TIR values of the LacI/P_{T7lac} and the AraC/P_{BAD} RBSs were higher than those of XylS/*Pm* and LacI/P_{trc} RBSs, suggesting a more efficient translation. The relative differences between the TIRs of the LacI/P_{trc} and XylS/*Pm* 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_{T7lac} 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/*Pm* ML1-17 there appeared to be more protein per transcript compared to LacI/P_{T7lac} 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_{trc} 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

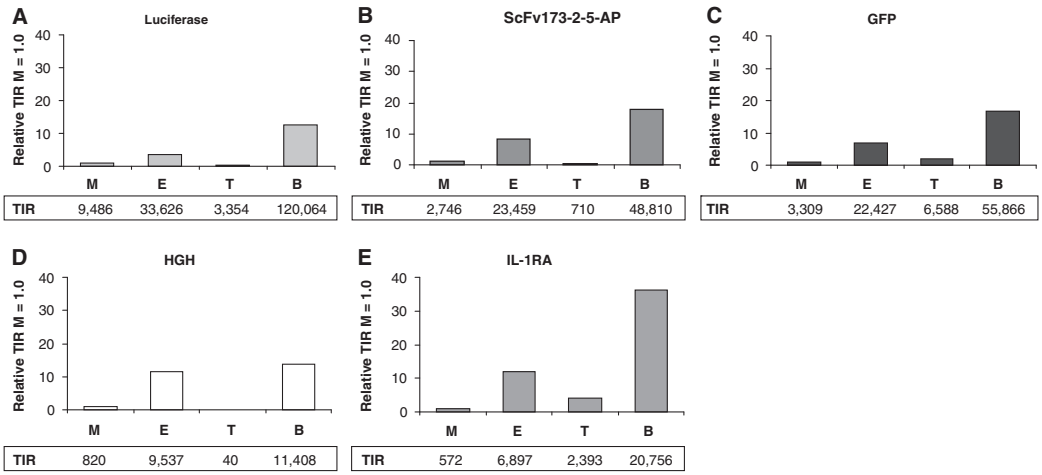

Figure 5 Theoretical analysis of the translational start site by calculating the translation initiation rate (TIR). The complete 5′-UTR sequences in combination with the first 50 nucleotides of the respective genes (luc_5 (Panel **A**), *scFv173-2-5-phoA* (**B**), *gfpmut3* (**C**), $GH1_5$ (**D**) and $IL1RN_5$ (**E**)) were used as input sequences for the RBS calculator [54].

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, $LacI/P_{T7lac}$ and AraC/P_{BAD} (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_{T7lac} (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

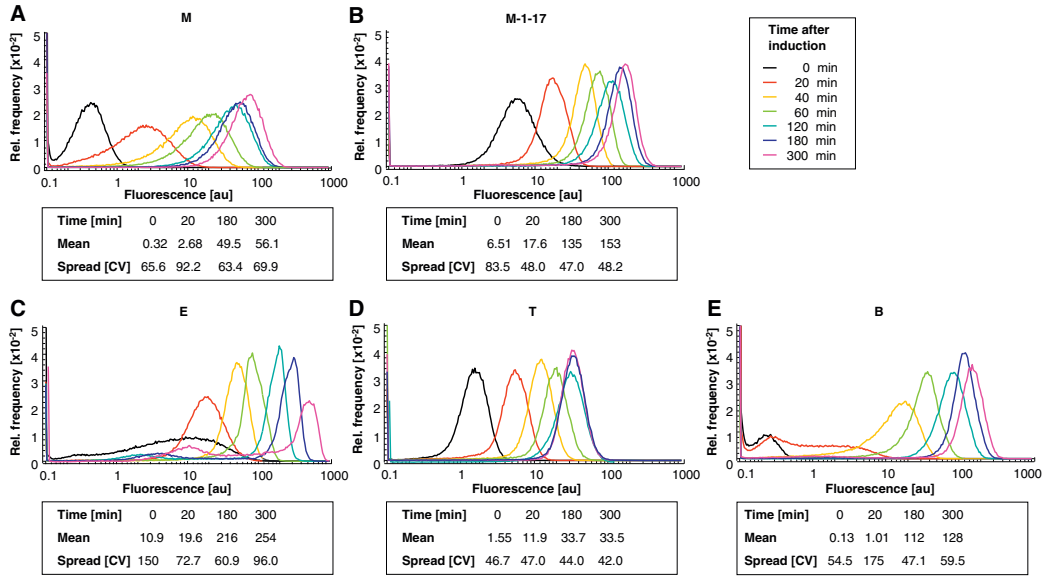

Figure 6 Analysis of the distribution of expression using flow cytometry. Strains were grown under standard conditions. At the time point of induction (t= 0 min) and at several points afterwards (t= 20–300 min), samples were collected, snap-frozen, and collectively analyzed with a flow cytometer. The spread is represented by the coefficient of variation (CV). Panels **A**-**D**: ER2566, Panel **E**: DH10B.

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

The LacI/P_{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 AraC/P_{BAD} system, displayed a similar behaviour as XylS/*Pm* 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 *Pm* 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 LacI/P_{T7lac} 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 XylS/*Pm* 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 XylS/*Pm* system can in some cases generate protein production at industrial levels it is clear that LacI/P_{T7lac} 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 XylS/*Pm* ML1-17). In contrast, for luciferase the amounts of transcript appears to be very important, but the potential in the LacI/P_{T7lac} system is in this case lost by the excessive production of inactive protein. LacI/P_{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. AraC/P_{BAD} seems to be best with respect to tight regulation of the uninduced state coupled with high expression when induced. The XylS/*Pm* system has a big

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

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

[a] Expression of lysozyme, the natural inhibitor of T7 RNA polymerase, reduces the basal transcription from P_{T7lac}.
[b] Supplementing glucose leads to catabolite repression which reduces basal transcription levels.
[c] 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/*Pm* ML1-17, LacI/P_{T7lac} and AraC/P_{BAD}. 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	ATCCACCGGAACTAGTCCCCTGCTC
Pwitw6_badF	AGACTAGTAAGCCCTCCCGTATCGTAGTTA
Pwitw6_badR	TGGCGCGCCAGATGCGTAAGGAGAAAA TACCG
ET_AgeI_fwd	GATGGCCCATATGATATCTCCTTCT
ET_NdeI_rev	GATCACCGGTCCAGTGATCGAA
BAD_BbsI_fwd	GGCCTTTCGTCTTCCCGGCATCCGCTTA CAGACA
BAD_NdeI_rev	GACGCCCATATGTAATTCCTCCTGTT AGCCCAAAAAACG
TRC_AgeI_fwd1	TGCATGTGTCACCGGTTTTCACCGTC
TRC_NdeI_rev1	GAGCTCGAATCATATGGTCTGTTTCCTG
pelB_fwd	AGCTACATATGAAATACCTATTGCCTACG
APhis_rev2	AGGATCCGAGCCTTTCGTTTTATTGATGC
b) qRT-PCR primers	
RT-synluc_fwd2	CCATGGCTTCGGCATGTT
RT-synluc_rev2	ACACGAAAGCCGCAAATCA
gfpmut3_fwd1	CATGGCCAACACTTGTCACT
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	CATTTGCAGATACAGCGTGTTCT
RT-16S-Fwd	ATTGACGTTACCCGCAGAAGAA
RT-16S-Rev	GCTTGCACCCTCCGTATTACC

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_{T7lac} 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 $lacI^q/P_{trc}$ region was amplified from pTrc99A using TRC_AgeI_fwd1and 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_{BAD} 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 luc_S gene from pIB11-luc_S 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 luc_S 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 $GH1_S$ and $IL1RN_S$ 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_{T7lac} and P_{trc} 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_{T7lac} and P_{BAD} 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_{600} of 0.05. Following incubation at 200 rpm and 30°C expression was induced at OD_{600}= 0.5-0.6 as follows: 2 mM *m*-toluate for strains harboring *Pm*- based constructs, 1 mM IPTG for those with P_{T7lac}, 0.2 mM IPTG for P_{trc} and 0.015% L-arabinose for P_{BAD}. 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_{600} of 0.5. 90 μL of this mixture was supplemented with 10 μL of K_2HPO_4, 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 $MgCl_2$). 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 decribed 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® 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™ 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 a very grateful to Laila Berg for the gift of pIB11-luc$_S$, 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
[1]Department of Biotechnology, NTNU, Sem Sælands vei 6, Trondheim 7491, Norway. [2]Faculty of Physics and CeNS, LMU Munich, Munich 80539, Germany. [3]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_{alkB}-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_{BAD} 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 Pm promoter in gram-negative bacteria.** *Metab Eng* 2000, **2**(2):79–91.
25. Lale R, Berg L, Stüttgen F, Netzer R, Stafsnes 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 *Pm*.** *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 *Pm* 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. Schlesinger 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-α 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, Bjerknes 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, Villanueba 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/.
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.

Paper II

Is not included due to copyright

Paper III

Is not included due to copyright