Friederike Zwick

Exploring the limits of the *xylS/Pm* expression cassette for recombinant protein production in *Escherichia coli*

Thesis for the degree of Philosophiae Doctor

Trondheim, August 2013

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 978-82-471-4577-7 (printed ver.)
ISBN 978-82-471-4578-4 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2013:229

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 and the financial support was provided by the Faculty of Natural Sciences.

Professor Svein Valla has been my supervisor and I am very grateful for his constant help with all occurring challenges. He always contributed with new ideas and aspects and kept me busy by suggesting new promising experiments. At the same time he gave me the valuable freedom to realize my own ideas.
Equally I want to thank Rahmi Lale for good collaborations, many fruitful discussions, proofreading the thesis and especially for stepping in as co-supervisor in the very end. He always answered my questions and helped me to find good solutions.

Thanks are also given to my colleagues. Trond Erik Vee Aune assisted me through the starting period and offered me some interesting insight into the industrial site of our research by employing me for a short while in Vectron Biosolutions AS.
Veronika Kucharova often kept my company in the lab, especially in the early morning hours, and Simone Balzer and Laila Berg collaborated with me on smaller projects.

Finally I acknowledge the valuable help of the staff at the Department of Biotechnology, and the companionship of my office mates.

Abstract

Recombinant protein production is applied for a wide range of applications, however, so far no universal system exists which allows for production of sufficient amounts of any protein of interest. Thus improvement of existing systems and better understanding of their molecular function is still as important as the discovery and development of new expression tools. Particularly expression systems which allow a certain degree of flexibility and adaptation concerning growth conditions, expression levels, and choice of expression host appear promising.
The work described in this thesis concentrates on vectors with these characteristics. They combine the positively regulated XylS/*Pm* promoter system with the broad-host-range RK2 mini-replicon. They are well established and have been shown to be capable of industrial-level production of several medically important proteins. Intensive studies, especially with *bla* (coding for β-lactamase) as reporter gene, and mutagenesis of several of the control elements in the applied system have resulted in identification of variants that increase expression levels also for other reporter genes tested. In this thesis it was aimed to assess, whether there was potential for further improvements of the vectors, and indeed, a combination of several high-expression variants in one cassette resulted in higher expression levels than those achieved by single variants. With the reporter gene *bla* about 75 times as much protein as with the wild type system could be produced when variants for *Pm*, the DNA-region corresponding to the 5′-untranslated region of mRNA (5′-UTR) upstream of the reporter gene, and the gene of the positive regulator of the system, XylS, were combined. As a result of this combination corresponding transcript levels could also be increased up to 63-fold.
Testing of these new vectors with several other reporter genes demonstrated that expression of them in general could be increased, however, only within the limits of gene- and protein-specific boundaries. This was particularly investigated in the case of *celB* (coding for phosphoglucomutase), and a new role for the plasmid copy number was discovered: an increase in plasmid copy number does not only lead to higher expression levels due to increased production of transcript, but also facilitates better exploitation of the resources due to the larger distribution of mRNA across the cell.
The transcription factor XylS is of central importance for the here applied vec-

tors and a study of its expression revealed new insights into limitations and mechanisms by which it activates *Pm*. Its expression could not be increased by codon optimization, which was partly, but not only, caused by secondary structure formation in its translation initiation region. However, XylS expression could be improved either by exchange of its 5′-UTR and/or the promoter from which it was expressed. The choice of the positively regulated promoter system ChnR/*Pb* for transcription of *xylS* allowed variation of XylS expression levels with concurrent investigation of its mechanisms of activation of *Pm*. A synthetic operon construct permitted to detect XylS at physiological concentrations. This revealed the existence of a roof for induced expression from *Pm* depending on XylS expression levels, which probably is caused by oligomerization of the transcription factor at high levels inside the cell. This indicates that it is redundant to increase XylS expression levels beyond the saturating concentration, and the maximum induction ratio at *Pm* with the wild type system was found to be around 700-fold, with *bla* as reporter gene.

Regulation of XylS expression was also demonstrated to be useful in reduction of background expression from *Pm*, which often is increased, when induced expression is improved. This might provide benefits for expressing host-toxic proteins. Also background levels of the vectors which combine several high-expression variants, could be reduced by XylS regulation.

Another way of background reduction and enhancement of the induction ratio was demonstrated to be the usage of low-expression 5′-UTR variants, which turned out to be a promising approach for metabolic engineering purposes. This also revealed Rho-dependent termination of *celB*-transcription at low expression levels.

The combination of high-expression variants turned out to be useful in reduction of the metabolic load, which is imposed onto cells by recombinant protein production. The high expression levels which were achieved by this approach allowed expression from one single copy on the chromosome at levels higher than that of the wild type plasmid system.

The findings described here demonstrate that there is potential to further improve expression from vectors that combine the XylS/*Pm* expression system with the RK2 mini-replicon. In addition there already exist variants that cause varying phenotypes for several of the elements in these vectors, and it offers a large toolkit with many possibilities for combination of these features. Thus the system probably can be adapted to expression requirements of many proteins.

Some of the reported findings can also be seen in a broader context and potentially adapted to other systems, as for example the combination of variants with defined phenotype. Low-expression variants might be helpful to reduce background levels of strong promoters like in the largely applied T7-system. The discovery of the new role in plasmid copy number might be of importance for every expression system and the mechanisms which were revealed for XylS may potentially also exist for other and related transcription factors.

List of publications

Paper I

F. Zwick, R. Lale, and S. Valla. Strong stimulation of recombinant protein production in *Escherichia coli* by combining stimulatory control elements in an expression cassette. *Microb Cell Fact*, 11:133-140, 2012.

Paper II

F. Zwick, R. Lale, and S. Valla. Combinatorial engineering for heterologous gene expression. Addendum to Paper I. *Bioengineered, in press.*

Paper III

F. Zwick, R. Lale, and S. Valla. Increased plasmid copy number allows better utilization of cell resources. *Manuscript*

Paper IV

F. Zwick, R. Lale, and S. Valla, S. Regulation of the expression level of transcription factor XylS reveals new functional insight into its induction mechanism at the *Pm* promoter. *Submitted.*

Paper V

R. Lale, L. Berg, **F. Stüttgen**, R. Netzer, M. Stafsne, T. Brautaset, T. E. Vee Aune, and S. Valla. 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, 2011.

Abbreviations

σ	sigma factor
5′-UTR	5′ -untranslated region of mRNA
aa-tRNA	aminoacyl-tRNA complex
BCM	bicyclomycin
CAI	codon adaptation index
EF	elongation factor
GOI	gene of interest
IF	initiation factor
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
Rep	replication initiation protein
RF	release factor
RNAP	RNA polymerase
SD	Shine-Dalgarno sequence
TF	transcription factor
TIR	translation initiation region

Contents

1

Introduction

1.1 Recombinant protein production

1.1.1 *Escherichia coli* as host

Recombinant protein production can be defined as the expression of genes in foreign, sometimes not even related hosts, enabled by cloning and transformation techniques (275). Today recombinant protein production is used for a broad range of applications, as for example the production of medicines, vaccines, food-processing enzymes and biofuels (170). In many cases it offers an inexpensive production alternative to chemical synthesis or the possibility to overproduce compounds that naturally are found only in small amounts or are difficult to isolate (58, 116, 290). Recombinant protein production is of utmost importance for pharmaceutic biotechnology and over 150 different therapeutics produced by recombinant expression have been approved (116, 170). More than 30% of these are produced in the bacterium *Escherichia coli* (*E. coli*) (75, 170) and already in the first studies reporting recombinant protein production, *E. coli* was used as a host (45, 48).

E. coli is a Gram-negative, rod-shaped bacterium, which was first discovered in 1885 (108). In 1997 the whole genome sequence of *E. coli* K-12 was published (26), which opened new possibilities for the work with this bacterium. One of the main reasons for using this organism is that it is very well characterized, but there are also many other advantages connected to the use of *E. coli* as a host for recombinant protein production: It grows rapidly and can be cultivated at high cell densities, growth media are relatively inexpensive, and it is established

in almost every laboratory. In addition there are a large number of expression systems available for use in *E. coli* (260, 290).

Unfortunately not every protein can be produced in *E. coli*, in fact there are a range of limitations for the use of this bacterium as a host in recombinant protein production. Likewise other bacteria *E. coli* is not able to perform posttranslational modifications as for example glycolysation and disulfide-bond formation, that are necessary for the correct folding of many eukaryotic proteins. Folding problems can also occur due to the lack of appropriate chaperone molecules. Further challenges can be limitations in secretion, improper codon usage, degradation of produced proteins by host proteases, toxicity of the protein to the host and accumulation of endotoxins (153, 260, 290). Solutions have been suggested for many of these problems. Secretion can for example be improved by the use of signal sequences, coexpression of permeabilizing proteins or fusion partners (74, 153). A lack of chaperones can be circumvented by the coexpression of these molecules (51). For genes with a codon usage deviating from that in *E. coli*, it is either possible to exchange rarely used codons with frequently used ones or to coexpress rare tRNAs (153). Another possibility to avoid problems typically connected to the use of *E. coli* is to choose other hosts, which can be Gram-negative or Gram-positive bacteria or eukaryotic systems such as yeasts, fungi, mammalian, human or plant cells. Each host comes with its own advantages and limitations and there is no optimal expression platform that can serve as host for the production of all sorts of recombinant proteins (290).

Establishing a functioning expression system for a protein can be time-consuming. For most new applications a range of expression platforms has to be tested individually to find the right combination of host and expression system to achieve sufficient levels of protein production. Therefore it is still important to better understand and improve existing systems or develop new ones.

In the following sections elements that are essential for successful protein production, and their functions are described. As this study focuses on protein expression in *E. coli*, all parts of the introduction describe processes as observed in this bacterium.

1.1.2 Expression systems

Expression systems typically contain a set of certain elements, as shown in the upper panel of Figure 1.1: first of all the gene of interest (GOI), coding for the protein that is going to be produced, which should contain both a translational start and a stop codon. Second and third, there typically are a promoter and an untranslated DNA-stretch harboring a Shine-Dalgarno-sequence upstream of the gene, while transcriptional terminators often are placed downstream of it to avoid continuation of transcription. Plasmids, which often are used as expression vehicles, contain an origin of replication and typically selective markers (164). Last, the desire for tight control of gene expression leads to the presence of regulatory elements. Recombinant expression can be achieved either from

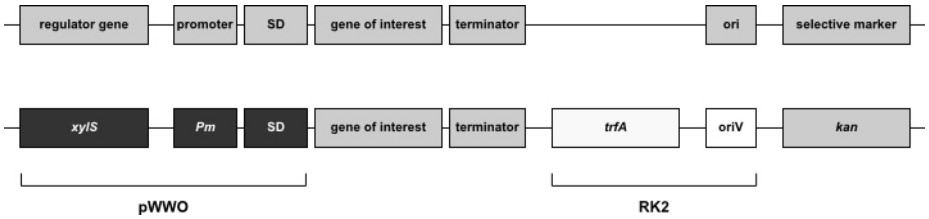

Figure 1.1: Typical expression system elements (for a description see this and the following sections); a general setup is displayed in the upper panel, the lower one shows corresponding elements for the expression vectors used in this study and their origin; SD: Shine-Dalgarno site; ori: origin of replication.

expression systems that are integrated into the host chromosome or that are, as named above, placed on plasmids, which are defined as extrachromosomal genetic elements (146). Plasmids found in nature normally provide the host cells with features that increase their fitness compared to cells without plasmid (60). The fact that they are not part of the chromosome makes plasmids popular tools in recombinant protein production. They can easily be isolated from cells, the copy number of plasmids is normally higher than that of the chromosome, and *in vitro* manipulation of plasmids is simpler than for chromosomal DNA due to their small size (197). In addition many plasmids can be replicated in a broad range of hosts, enabling relatively simple testing in different organisms.
Integration into the chromosome is typically achieved either by homologous recombination or by transposon- or bacteriophage-mediated integration at chromosomal attachment sites (13). Many systems that facilitate chromosomal in-

tegration have been constructed and one of them was used in this study. It is based on so-called conditional-replication, integration, and modular (CRIM) plasmids and mediates integration at several phage attachment sites. The system uses a combination of temperature-sensitive replication and expression of the essential integrase to ensure that cells are cured for the delivering plasmids. Successful integration can be confirmed by a simple PCR-procedure (109).

The RK2-XylS/*Pm* expression vectors

The main expression system used in this study combines the minimal replicon of plasmid RK2 with the strong *Pm* promoter from plasmid pWW0 (24). Typical elements of this system are depicted in Figure 1.1 and are described in the following sections.

The RK2 plasmid was first identified as antibiotic resistance mediating factor (151). RK2 is self-transmissible and appeared in different microorganisms at the Birmingham Accident Hospital. It has a size of 60 kb and is capable of replication in a broad range of hosts (133). The mini-replicon of RK2 consists only of the two regions that are essential for replication: the origin of vegetative DNA replication, *oriV*, and the gene coding for the replication initiation protein TrfA (*trans*-acting replication function) (232). Mini-RK2 is described in more detail in 1.2.3.

The *Pm* promoter derives from the *Pseudomonas putida* toluene-degradative plasmid pWWO, whose existence was discovered in 1974 (283). Genes coding for proteins which are required for degradation of toluene and xylenes are found on pWWO. They are clustered in two operons: the upper one is responsible for conversion of toluenes and xylenes into benzoates or alkylbenzoates, the lower or "*meta*-"operon is responsible for degradation of these compounds into Krebs-cycle intermediates (215). The *Pm* promoter controls the genes of the *meta*-operon, which consists of 13 different *xyl*-genes (110, 158). Expression from *Pm* is controlled by the transcriptional regulator protein XylS, which is encoded on pWW0 as well (215). Promoter and regulator protein are described more detailed in 1.3.3 (*Pm*) and 1.5.4 (XylS).

The first expression vectors that combined the XylS/*Pm*-system with the mini-RK2 replicon were shown to be capable of protein expression at high induction

ratios (24). They were further improved and demonstrated to function also in other bacteria than *E. coli* (25). Mini-RK2-XylS/*Pm* expression vectors have turned out to be useful both for fine-tuning of expression levels (12, 29, 30, 84, 285) and for industrial-level production of medically relevant, secreted proteins, one of them being host toxic (244, 245).

Other expression systems used in this study

Besides the mini-RK2-XylS/*Pm* system two other expression systems were used in this study: one that combines the strong promoter from bacteriophage T7 (see 1.3.4) and the replicon from plasmid pBR322 (see 1.2.4) in the pET-vector series from Novagen; and one, based on the *Pb* promoter from *Acinetobacter* and its positive regulator ChnR (described in 1.3.4) together with the pBBR1-replicon (see 1.2.4).

1.2 The replication machinery

Both chromosomal and plasmid DNA are duplicated preceding the formation of daughter cells, and in *E. coli* a range of different proteins are responsible for this process. Jacob and Brenner suggested in 1963 two elements as essential for replication of DNA in bacteria: a *trans*-acting initiator and a *cis*-acting replicator (120). Initiation proteins have been identified in many species and the replicator is today known as the origin of replication, *ori*. It typically contains regions for binding of the initiator protein(s), binding sites for helper proteins and regions with high AT-content (211).

In *E. coli* DnaA is the most important initiation protein for chromosomal replication, while the corresponding origin is called *oriC*. Upon binding of DnaA to *oriC*, the replication machinery, also called replisome, is responsible for unwinding of the double-stranded DNA, primer synthesis and polymerisation of nucleotides onto the new strands (19). Other proteins than DnaA that are involved in *E. coli* DNA replication are DnaB (helicase activity), DnaC (helper protein), DnaG (a primase), single-strand binding proteins, and the proteins building up the DNA polymerase holoenzyme (19, 78, 167). Two core enzymes are assembled together with other proteins in the replisome, with each one of

them replicating the leading and the lagging strand (205). This leads to a challenge in replication, because the replisome is moving unidirectionally and can synthezise DNA only in 5′-3′-direction, while leading and lagging strands of the DNA are, as originally proposed by Watson and Crick, oriented in different directions (277). This problem is solved by a relatively continuous replication of the leading strand and a discontinuous replication of the lagging strand as shown in Figure 1.2 (276). At the lagging strand DNA pieces with a size of

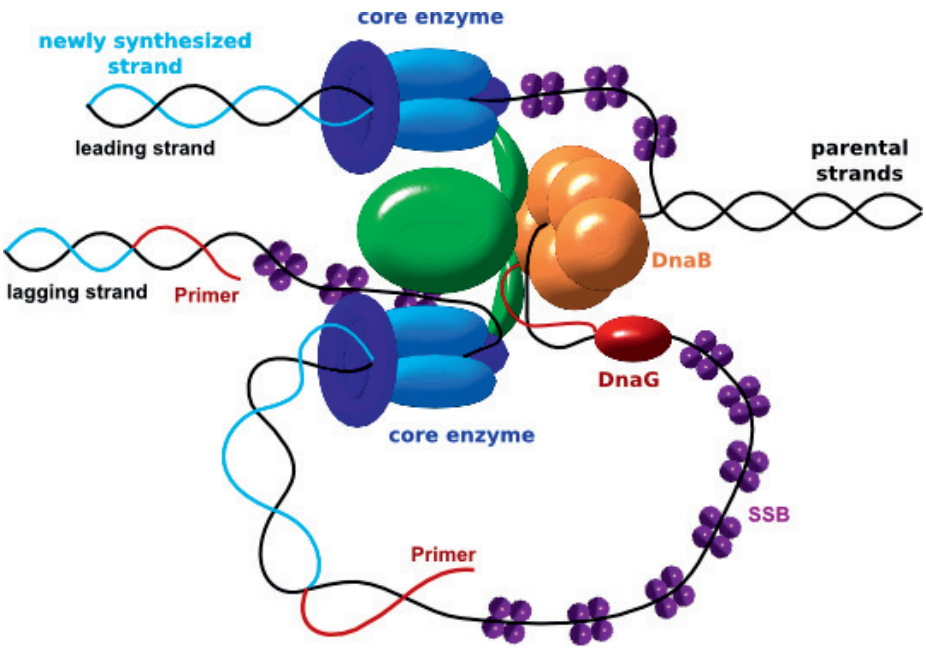

Figure 1.2: Simultaneous replication of leading and lagging strand by the replisome; the replisome consists of two DNA polymeraseIII enzymes (light blue) with sliding clamps attached (dark blue), a clamp loader (green), which assembles the sliding clamps on the DNA and connects the other parts, a DnaB helicase which unwinds the DNA, and DnaG, a primase; single-strand binding proteins (SSB) prevent secondary structure formation of single-stranded DNA. Adapted from (205).

1000-3000 base pairs (bp), also called Okazaki-fragments, are produced (243). They are later joined by ligase after replacement of the RNA primers with DNA by DNA polymerase I (205). The simultaneous replication of both strands leads to the formation of a replication fork, which has an Y-shaped structure (3).

1.2.1 Plasmid DNA replication

Typical for plasmids is a replication process which uses both plasmid-inherited determinants and parts of the host replication machinery. Plasmids that use a large proportion of the host enzymes can often replicate only in a small number of different hosts, while broad-host range typically is observed for plasmids that are relatively independent of host-encoded enzymes (141). Plasmid replication can be divided into three different types, which are roughly outlined in Figure 1.3: the theta-type mechanism, replication by strand displacement and the rolling-circle-type replication. Essential for replication is always an origin of replication, which is characteristic for each replicon. In many cases there are also genes encoding replication initiation proteins (Rep) and genes that are crucial for replication control found on the plasmids (57). Theta-type replication

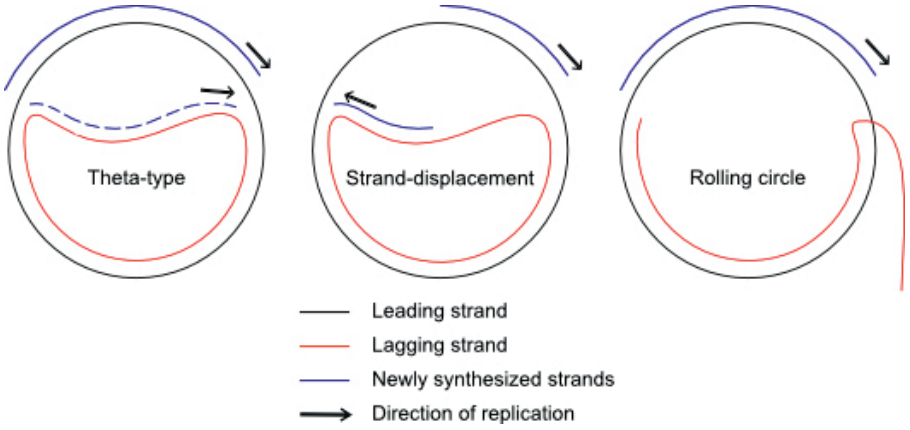

Figure 1.3: The three different mechanisms of plasmid replication; typical for theta-type mechanism is the coupled replication of both parental strands; replication by this mechanism can proceed uni- (as depicted) or bidirectional; the strand displacement mechanism is characterized by bidirectional replication; during rolling circle replication the lagging strand is cleaved and not copied before completion of the leading strand duplication.

typically starts with melting of the parental DNA strands, which occurs at an AT-rich region and is followed by synthesis of an RNA-primer and subsequent extension of this primer. The name derives from the Theta-shaped appearance of replication intermediates. Most plasmids replicating by this mechanism encode one or several Rep proteins, which bind to the origin of replication and form a complex with the host DnaA. This enables the complete assembly of the

replication machinery. Theta-type replication can be uni- or bidirectional and the continuous replication of the leading strand and the discontinuous replication of the lagging strand are coupled, leading to a process which is similar to chromosome replication in *E. coli* (57).
Replication by strand displacement is characterized by bidirectional and continuous replication of both leading and lagging strand. Many of the typical replication functions are provided by plasmid-encoded proteins and thus strand displacement is relatively independent of host enzymes (57).
In the rolling circle type of replication, which is typical for small plasmids, leading and lagging strand are replicated consecutively, both unidirectionally, and a plasmid encoded protein which is responsible for strand cleavage plays an important role (57, 130).
Different plasmids that share common replication features are often incompatible, which means that they cannot coexist in the same cell. This leads to the classification of plasmids in incompatibility groups (192).

1.2.2 Gene copy number

A general question to be addressed in recombinant protein production is that of the right choice of GOI copy number per cell. Chromosomal DNA typically exists at one copy per cell in *E. coli*. Excessive replication of DNA is prevented by autoregulation of DnaA expression (167). For plasmids the copy number per cell is characteristic for each replicon within a defined host (57). Plasmid-encoded control elements are responsible for maintenance of the plasmid at this number. This is mediated by negative regulation of the replication initiation step, which adjusts replication to one incident per plasmid copy and cell cycle (22, 56).
Besides the maintenance of the plasmid at a certain copy number, it is also necessary to ensure proliferation of plasmid copies to each daughter cell. Alternatively cells without plasmid can be subjected to cell death for example by selective media or toxin-antitoxin systems in order to avoid plasmid loss (83). Low copy number plasmids normally make use of partitioning systems to ensure segregation into daughter cells, while high copy number plasmids are thought to often rely on random segregation (70).

In theory higher copy numbers are desirable for recombinant protein expression, because more protein can be produced from a larger number of gene copies. In accordance with that most commercially available expression systems are based on plasmids with moderate to high copy numbers (122). However, even if expression rates in general increase when the copy number is elevated, this increase is not always proportional and there are some negative aspects arising from an increase in copy number (10, 164). One important term in this context is the metabolic load that production of foreign genes imposes on host cells. Metabolic load can be defined as the resources of the host cell that are required for maintenance and expression of the foreign DNA. This can be raw materials, as for example nucleotides or amino acids, as well as energy. Replication of foreign DNA can contribute to some extent to the metabolic load and large or high-copy number plasmids may imply a higher metabolic load to the host cell than small chromosomal inserts or low copy number plasmids (86). The presence of plasmids affects cell growth even if not to a great extent (114). In addition, expression of heterologous genes has an influence, and it has been shown that this affects cell growth significantly more than copy number and size (60). If one considers that protein expression requires energy for each addition of an amino acid (see 1.4.2), it is not surprising that a cell free of heterologous genes grows faster than one which is expressing recombinant proteins (86). In the former case cell growth is inversely proportional to plasmid copy number and plasmid size (60), while in the latter case growth rate is inversely proportional to the production rate of the recombinant protein (114). Cases have also been reported where cells that produce high levels of foreign proteins are not viable (10).

Plasmid loss in the absence of segregational control systems has been discussed as consequence of metabolic load caused by foreign protein expression. Another observed phenomenon is the mutation of foreign DNA with the result that the causer of metabolic load is removed, leading to non-producing cells (49). Typically antibiotic resistance genes are integrated on expression plasmids (or flanking the gene in the chromosome) to avoid plasmid or DNA loss. They provide the cell with an advantage over cells without plasmid during cultivation with antibiotics. However, antibiotic resistance genes cause an increase in the

metabolic load as well, as they have to be replicated and expressed (86). Other drawbacks derive from safety concerns connected to the use of antibiotic resistance markers and an alternative can be the complementation of auxotrophic strains by a plasmid-encoded gene (271).

There appears to be no perfect solution concerning the number of gene copies and often a balance between higher expression levels due to more gene copies and lower growth rates and metabolic changes due to the metabolic load has to be found. If the same expression level can be reached from either a high or a low copy number of the expression cassette, the lower copy number should be preferred, because this will reduce the costs for DNA maintenance.

In this context also promoters that can be regulated are of vital importance for recombinant gene expression. Initial growth of the cells without protein production reduces the metabolic load extensively. However, tight control is easier achievable if the foreign DNA is maintained at low copy numbers (164). If expression levels are high enough to integrate the expression system into the host chromosome this will give the double advantage of reduced replication resources due to the lower copy number and due to expandability of selective markers, as chromosomally integrated genes have been found to be in general more stable (86, 128). A promising approach is also the use of systems were both copy number and expression can be regulated so that the plasmid is maintained at one single copy per cell until protein production is induced (282).

1.2.3 The mini-RK2 replicon

The RK2 replicon belongs to the theta-type group of replicons and replication proceeds unidirectionally from *oriV* (57). Besides the plasmid-encoded initiator protein TrfA a range of host-encoded proteins is essential for RK2 replication (133). DnaA binds to four DnaA boxes in *oriV* and recruits DnaB and DnaC, but binding of TrfA is required in addition for replication initiation (see Figure 1.4). The gene *trfA* contains two alternative start codons leading to two forms of its protein product with different sizes: TrfA-33 and TrfA-44 (241). Both proteins are in *E. coli* expressed at relatively low levels and different negative regulators of their expression have been identified (67, 133). Overexpression of the two TrfA proteins leads only to a minor increase in plasmid copy number,

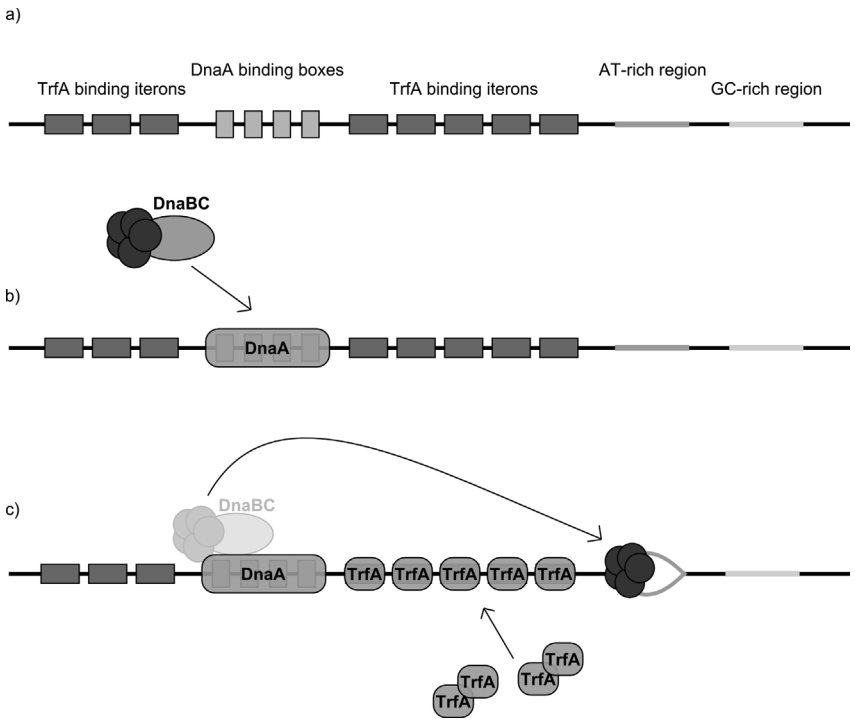

Figure 1.4: Replication mechanism of RK2; a) typical elements of *oriV*; b) host-encoded DnaA binds to four DnaA-boxes in *oriV*, helicase DnaB and helper protein DnaC can consecutively be recruited to the DNA; c) binding of TrfA in monomeric form to iterons of 17 bp is crucial for DNA-melting at the AT-rich region by DnaB. Adapted from (57, 133).

suggesting that copy number is controlled not only by regulation of *trfA* expression, but by other mechanisms in addition (67). In *E. coli* only TrfA-33 is essential for RK2 replication, while the larger TrfA protein is required in other hosts (242). TrfA forms dimers in solution, while it binds to the iterons in *oriV* in monomeric form (264). The handcuffing model, which is often referred to as regulation model for theta-replicating plasmids, ascribes an important role in copy number control to TrfA (57). It suggests that TrfA dimers couple two origins of replication together, resulting in a steric hindrance, which prevents replication, when the typical copy number is exceeded. This model is supported by the fact that *trfA* variants, that are defective in dimerization, increase the plasmid copy number (263). The typical copy number of RK2 in *E. coli* is four to eight copies per cell, but variants with copy numbers of up to 120 copies per cell have been identified (68, 133). Two such variants of *trfA* were used in this

study: cop271C and cop254D with a copy number of 17 to 19 and 77 to 89 copies per cell, respectively (68).

1.2.4 Other replicons used in this study

The pBR322-replicon

Besides plasmids based on the mini-RK2 replicon, pET-vectors (Novagen) were used in some parts of this study. These vectors contain a replicon which derives from plasmid pBR322 (222). Plasmid pBR322 was originally constructed from pMB1, a clinical isolate with a ColE1-like replicon (14). ColE1-like plasmids perform theta-type replication, which proceeds unidirectionally. It depends completely on host-encoded proteins and thus is not applicable to a broad range of hosts (141). Excess replication is prevented by a combination of a small RNA and a protein (41). The typical copy number of plasmids with this origin is 18 per cell (47).

The pBBR1-replicon

In contrast to pBR322 based vectors, plasmids that contain the replicon from pBBR1 stably replicate in a broad range of Gram-negative hosts. This is enabled by the use of a plasmid-encoded replication protein, named Rep (8, 142). However, details of the replication mechanism are still unknown (259). The replicon does not belong to any known imcompatibility group, which makes it attractive for use in expression vectors, as it can be combined with all other replicons (8, 142). Plasmids based on pBBR1 are typically maintained at a medium copy number of about 30 copies per cell (96) and variants with changed copy number behavior exist (259).

1.3 The promoter and its influence on transcription

Transcription is the first step of gene expression, where an RNA molecule is synthesized from a DNA template. The enzymes executing this process are the RNA polymerases (RNAP). The holo RNAP enzymes identify the promoter region at the 5′-end of a gene, which directs them to the transcriptional start site (33, 111). Promoter regions have to be unique in order to be recognized from the

much bigger amount of non-promoter DNA (292). In *E. coli* the most important promoter elements are two hexamers with conserved sequence, the -35 and the -10 region. The spacer in between these two hexamers has been found to have an optimal length for transcription initiation of 17 bp (148). Other regions with potential impact are: an extended -10 motif (18), an AT-rich region, located upstream of the other promoter elements in many promoters (the UP-element) (223), and recognition sites for regulator proteins (33). The degree of similarity of the different regions to the consensus sequences varies substantially among promoters and accounts for differences in promoter strength (111, 195). Not all elements are present in all promoters. For example in promoters with an extended -10 region the -35 is not strictly necessary for recognition (292). No natural promoters have been identified so far that match the consensus sequences in all nucleotides (33).

1.3.1 RNA Polymerase

In *E. coli* there is only one RNAP which is responsible for all transcriptional activity (111). The core enzyme consists of five subunits: two α-subunits that play a role in DNA binding, one β- and one β'- unit, forming the active site of the polymerase, and one ω-unit, which helps to assemble the core enzyme (33, 174). RNAP is described to have a crab-claw form, where the active site is found between the two pincers (292). RNA synthesis proceeds in 5′-3′-direction and is carried out by addition of nucleotide triphosphates (NTPs) to the 3′-OH end of the growing RNA chain, formation of phospho-diester bonds, and release of pyrophosphate (50). RNAP is, in contrast to DNA polymerase, able to synthesize nucleic acid chains *de novo* (292). Even if the core enzyme on its own is able to synthesize RNA, for successful transcription initiation it has to interact with a sigma factor (σ). σ is responsible for promoter recognition, positioning of the polymerase, and unwinding of the double stranded DNA, functions that it can perform only when bound to the core enzyme (33, 102). In *E. coli* seven different sigma factors have been identified (102). One of them, σ^{70}, is responsible for recognition of promoters in front of many housekeeping genes (33). For other sigma factors the composition of the promoter region deviates from the consensus for σ^{70}. Their action is controlled either by enrichment

of the factor in demand during specific stress situations, growth transitions or morphological changes or by interaction with anti-sigma factors, when their action is unrequested (33, 102).

1.3.2 Transcription

Initiation

A large proportion of non-transcribing RNAP molecules is attached to DNA prior to promoter recognition (124). Transcription is initiated when promoter DNA is recognized by a sigma factor, which recruits the core RNAP. There are several interactions between σ, core enzyme, and DNA (181). Both the -10 (consensus for σ^{70}: 5′-TTGACA-3′), the -35 (consensus for σ^{70}: 5′-TATAAT-3′), and the extended -10 region interact with different domains of σ, while the UP element, which is found in some promoters, interacts with the α-subunits of RNAP (33).

After identification of the promoter region RNAP catalyzes melting of the double-stranded DNA between positions -11 and +4 relative to the transcriptional start site to form a so called open complex (100, 292). Open complex formation enables the base in position +1 to pair with the first NTP, which preferably is an ATP (111, 292).

When RNA synthesis is started, RNAP has to change its activity from binding specifically to the promoter DNA to non-sequence-specific DNA binding in order to be able to move forward along the DNA and continue with RNA synthesis. This process is called promoter escape and is often accompanied by dissociation of σ from RNAP (292). However, cases have been observed, where σ does not dissociate from the core enzyme (17). Typically short abortive pieces of RNA are released in a process called abortive initiation, which has been observed both *in vitro* and *in vivo* (27, 88, 115). One suggested explanation for this is that a part of σ blocks the RNA exit channel of the RNAP. First upon displacement of σ and a conformational change of RNAP, transcription can continue (180, 181). In addition it has been shown recently that RNAP maintains interactions with the promoter region during early formation of the RNA chain, leading to a process called scrunching. The energy resulting from this scrunching leads either to breakage of the interactions between σ and the

promoter or to breakage of the complex with DNA and RNA, followed by the release of abortive transcripts (216). Scrunching can also be an explanation for the absence of promoters matching the consensus sequences in all bases in nature. This would lead to strong interactions with RNAP and consequently an excessive formation of abortive transcripts (9). Typical transcription initiation related mechanisms are outlined in Figure 1.5.

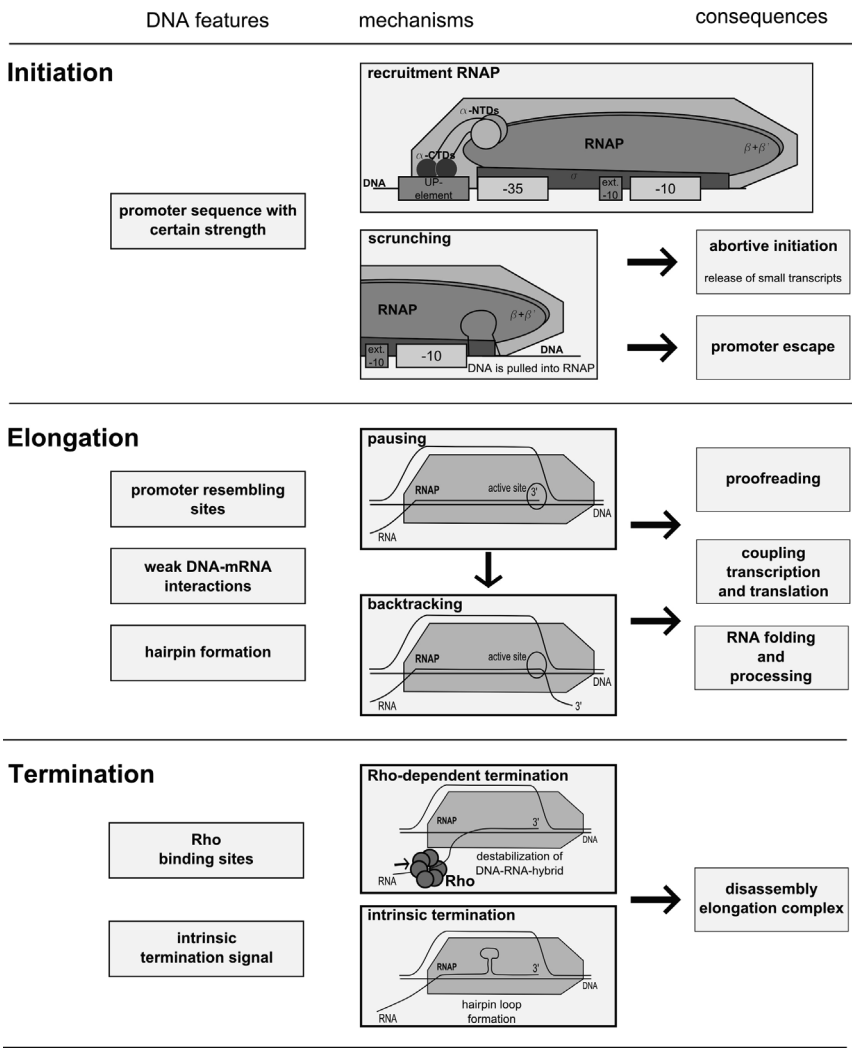

Figure 1.5: The transcriptional process in *E. coli*; typical DNA elements are described that are part of different mechanisms during the three phases of transcription. Adapted from (33, 98, 193, 216).

Elongation of transcription

Once promoter escape is accomplished, RNAP starts to elongate the RNA molecule (100). One template DNA can be transcribed to RNA by several RNAP molecules (27). It is separated into two strands near the active site of RNAP. These two strands take different paths through the enzyme, forming the so-called transcription bubble, before they reanneal. During the whole transcriptional process about 12 bp of the DNA template are held open (27, 292). If the RNA is prematurely released, RNAP is not able to restart transcription other than from the promoter sequence. RNAP, DNA template and RNA form an extremely stable complex during this phase, which prevents premature release, but pausing or arrest of the enzyme can be observed (137). Both can be caused either by DNA sequences, which interact with RNAP or form stable complexes with the synthesized RNA, or by DNA-binding proteins (98, 268). One such protein is NusA which decelerates transcription elongation (233), while NusG increases elongation rates by suppressing RNAP pausing at specific sites (35). It has been demonstrated that σ^{70} can stay associated to the core enzyme during transcription, which might lead to pausing at promoter resembling sites (178, 228).

Two studies from 1997 (136, 194) revealed sequence-dependent back- and forth-sliding of the RNAP in a process called backtracking. During this process the RNA molecule is removed from the active site of the enzyme, which becomes inactive, while the elongation complex remains stable. Backtracking has been observed for many pausing events. Proteins that stimulate RNAse activity of RNAP assist the enzyme in cleaving backtracked RNA, thus generating a new RNA $3'$ -OH end at the active site and enabling RNAP to continue with translation (160). Backtracking thus allows RNAP to carry out a certain proofreading activity: a mismatch in the DNA-RNA hybrid facilitates backtracking which then results in cleavage of the piece of RNA containing the wrong base (193). RNAP can also be reactivated by pushing of following RNAPs that transcribe the same template. This can serve as explanation for why transcripts from a strong promoter are elongated faster and why more pausing can be observed *in vitro* than *in vivo* (72).

Pausing of RNAP shortly after transcription initiation is regularly observed and

can be caused by both persistent contact of the RNAP with the promoter and easier backtracking due to a relatively short transcript (193). The backtracking process is visualized in Figure 1.5.

Termination of transcription

Termination of transcription is characterized by the disassembly of the stable elongation complex and can be achieved by two different mechanisms, which are shown in Figure 1.5. Common for both is that transcriptional pausing seems to be essential for successful termination (201).

At intrinsic terminators a stretch of U-residues probably causes backtracking and pausing of RNAP due to weak DNA-RNA-interactions. This is crucial for hairpin formation of the transcript at a typically GC-rich region (105, 193). The hairpin is formed in the RNA exit channel of RNAP which destabilizes the elongation complex (73).

The other mechanism of transcription termination is depending on the terminator protein Rho, a ring-shaped protein, that was discovered in 1969 (219). Rho binds to untranslated RNA, but there are no highly conserved binding sequences. However, Rho-terminators show two important features: the *rut*-site, which is rich in cytosine, about 80 nt long, and poor in secondary structures, and a second site, which contains the termination region (16). Rho binds to the *rut*-site and this stimulates ATPase-activity of Rho. The resulting energy enables Rho to translocate along the RNA in 5′-3′-direction (201). When the elongation complex is paused at the termination region, Rho can catch up, destabilize the DNA-RNA-hybrid, and dissociate the complex (217). Rho-dependent termination has been shown to be more effective during ineffective transcription, which can be explained by the requirement of Rho to approach the elongation complex to be able to terminate transcription (123, 201).

Rho-dependent termination can be studied by use of the antibiotic bicyclomycin (BCM). It interfers with the ATPase-activity of Rho and can thus directly inhibit termination of transcription (295). However, it has to be considered that an inhibition of Rho will lead to a broad spectrum of reactions, as transcription termination of all Rho-regulated genes will be changed (38).

1.3.3 The *Pm*-promoter

The *Pm* promoter originally controls expression of proteins of the *meta*-cleavage pathway of toluene degradation on plasmid pWW0, which enables *Pseudomonas putida* to grow on derivatives of benzoic acid (215). The -35 and -10 regions in *Pm* are not conserved, probably due to the fact that *Pm* is recognized by two different sigma factors, σ^{32} and σ^{38} (61, 159). σ^{32} is responsible for expression from *Pm* during early exponential phase, while σ^{38} gets active during late exponential and stationary phase (159). The putative -10 region of *Pm* has the sequence 5′-TAGGCT-3′ and is located at positions -7 to -12, relative to the transcriptional start site. Both sigma factors are able to establish binding to this region. The spacer between -10 and -35 has the optimal length of 17 bp (61). One important feature of *Pm* is its inducibility, which is mediated by the positive regulator XylS. XylS is expressed constitutively, but first upon interaction with an inducer (derivatives of benzoic acid) it activates transcription from *Pm* (79, 157). The binding site for XylS overlaps with the suggested -35 site by two base pairs (90). When bound to this site, XylS can interact both with the α-C-terminal domain of RNAP and the sigma factor (62, 224, 226).
Variants for *Pm* have been identified by two different approaches (11, 284). Expression levels for plasmids containing these variants were increased up to 14-fold and expression was still inducible.

1.3.4 Other promoters used in this study

The T7 promoter

Expression systems based on promoter sequences from bacteriophage T7 are extensively used. A specialty for these promoters is that they are recognized by the RNAP of the bacteriophage, T7-RNAP (253). Implemented in an expression system this has the advantage that transcription does not compete with transcription from other promoters with respect to the recruitment of RNAP (255). This leads to transcript elongation five times faster than for *E. coli* RNAP (260).
T7-RNAP is less complex than its bacterial analogues (42). It consists of only one subunit and is relatively small (98 kDa). Even if the polymerase domain of T7-RNAP is closely related to that of DNA polymerase I (249), it acts similar

to other RNAPs and is able to initiate oligonucleotide synthesis *de novo* (42). Binding to promoter sequences, melting of the double-stranded DNA as well as abortive initiation have been observed for T7-RNAP as for other RNAPs. T7 promoter sequences (17 bp long) are not recognized by host sigma factors, but by the promoter binding domain of T7-RNAP, which also is responsible for DNA melting (69).

Expression from T7 promoters requires the presence of T7-RNAP, and different host strains have been constructed, that harbor a genomic copy of the gene coding for it. The first expression system based on a T7 promoter was developed in 1986 and was demonstrated to be capable of accumulating target protein to more than 50% of the total cell protein. In this system T7-RNAP was either brought into the cell by lambda phage infection or by expression from a chromosomal copy behind the *lacUV5*-promoter (255). This second system is also used in the commercial pET-vectors from Novagen. In these vectors both expression of T7-RNAP and expression of the target gene are regulated by *lac*-repressor and can be induced by addition of Isopropyl-β-D-1-thiogalactopyranoside (IPTG). A disadvantage deriving from expression of T7-RNAP from the *lacUV5*-promoter is the leakiness of this promoter leading to high basal levels of expression. This can be partly improved by addition of T7-lysozyme to the system, typically via an extra plasmid, which acts as inhibitor of the polymerase (66, 254). The fact that T7-RNAP is much faster than other RNAPs explains the huge amounts of transcripts that can be produced by this RNAP (15). On the other hand genes expressed from T7 promoters are known to be ineffectively translated. The most probable explanation for this phenomenon is that T7-RNAP is moving faster than the host ribosomes, which causes a desynchronisation (150).

The *Pb* promoter

The *Pb* promoter derives from *Acinetobacter* sp. It controls expression of *chnB*, coding for cyclohexanone monooxygenase and is regulated by the AraC-XylS type regulator ChnR. Expression from *Pb* can be induced by cyclohexanone addition and up to 22-fold increase of expression levels could be observed upon induction (119). This promoter system has been demonstrated to be useful in

combination with a XylS/*Pm*-based expression system in alginate production (248).

The *P1* promoter

In plasmid pBR322 two promoters were identified in front of the β-lactamase gene, one of them being *P1* (32). This promoter is constitutive and there are about 150 bp in between its -10 region and the translational start site of the β-lactamase gene. Most probably *P1* is originally responsible for expression of a 14 kDa polypetide on plasmid pSC101, from which parts were taken to construct pBR322 (32).

1.4 The 5′-untranslated region of mRNA

1.4.1 Impact on translation

The product of transcription, RNA, can serve as template for translation, where the genetic code on the RNA defines the amino acid chain sequence. RNA molecules that are used as templates for translation are called messenger RNAs (mRNAs). An initiation codon (AUG coding for methionine in most cases, but other start codons are found as well) serves as starting point for translation and defines the reading frame (104, 173). Not every AUG can serve as start codon and the composition of the surrounding base pairs is decisive for turning an AUG-codon into a start codon.

Even though leaderless mRNAs exist and appear to be quite common (176), typically not the complete molecule is translated (145). Upstream of the start codon is the 5′-untranslated region of mRNA (5′-UTR), also called leader region, which contains the Shine-Dalgarno sequence (SD), with the consensus sequence 5′-GGAGGU-3′, which has been shown to play an important role in translation initiation (240). A spacer between SD and the start codon is of high importance (218). All three are part of the translation initiation region (TIR), which typically spans nucleotides -20 to +15, relative to the start codon (65). Interestingly, a high frequency of translated genes does not encode any SD-sequence on its mRNA, in *E. coli* SD-sites can be found only in 57.1% of all genes (152). Moreover, the combination of SD-site and AUG-codon alone is not sufficient for translation initiation (121). Several other mRNA regions that might have an

effect on translation initiation have been described. In Gram-negative bacteria a pyrimidine-rich region upstream of the SD has been found to interact with the translation machinery (135). Secondary structures in the TIR can act inhibitory on translation and accessibility of the SD is important for successful translation (154, 252).

1.4.2 The translational process

Translation is a highly energy-consuming process (2). One transcript can be translated several times and translation will start every few seconds (175). In bacteria translation is initiated before transcription is completed and both processes are tightly coupled, but different studies indicate that the main part of translation happens post-transcriptionally (145, 173). Anyhow, it has been shown that the velocities of transcription and translation are directly proportional, which can be explained by the ability of the translational machinery to push the RNAP, thus preventing backtracking (207). In addition the proteins NusG and NusE have been identified as linkers between the transcriptional and translational machineries, and the tight coupling of the two processes prevents excessive transcript formation and premature Rho-termination (34, 193).

The enzymatic complexes catalyzing translation are large macromolecular ribonucleoprotein complexes, called ribosomes (145). They are responsible both for reading the genetic code and catalyzing the polypeptide formation (161). Ribosomes are built up from two subunits: a 30S and a 50S subunit. Both contain RNA, which accounts for about two thirds, and a number of proteins (145, 250). The small subunit is composed of the 16S ribosomal RNA (rRNA) and 21 proteins, in the large subunit there are 34 proteins and 5S and 23S rRNAS (145). Decoding of mRNA takes place in the small subunit , while peptide bonds are formed in the large, and both processes are catalyzed by RNAs only (145, 190, 250).

The amino acids, which serve as precursors for polypeptide formation, are delivered to the ribosomes by transfer RNA molecules (tRNA) and the complexes of amino acid and tRNA are called aminoacyl-tRNAs (aa-tRNA). In *E. coli* 41 different tRNA species have been identified. They all contain different anticodons that interact with the codons on the mRNA in the decoding center of

the ribosomes and thus offer a way of translating the genetic code. There are three binding sites for transfer RNA inside the 30S subunit, named aminoacyl (A), peptidyl (P) and exit (E) site (145).

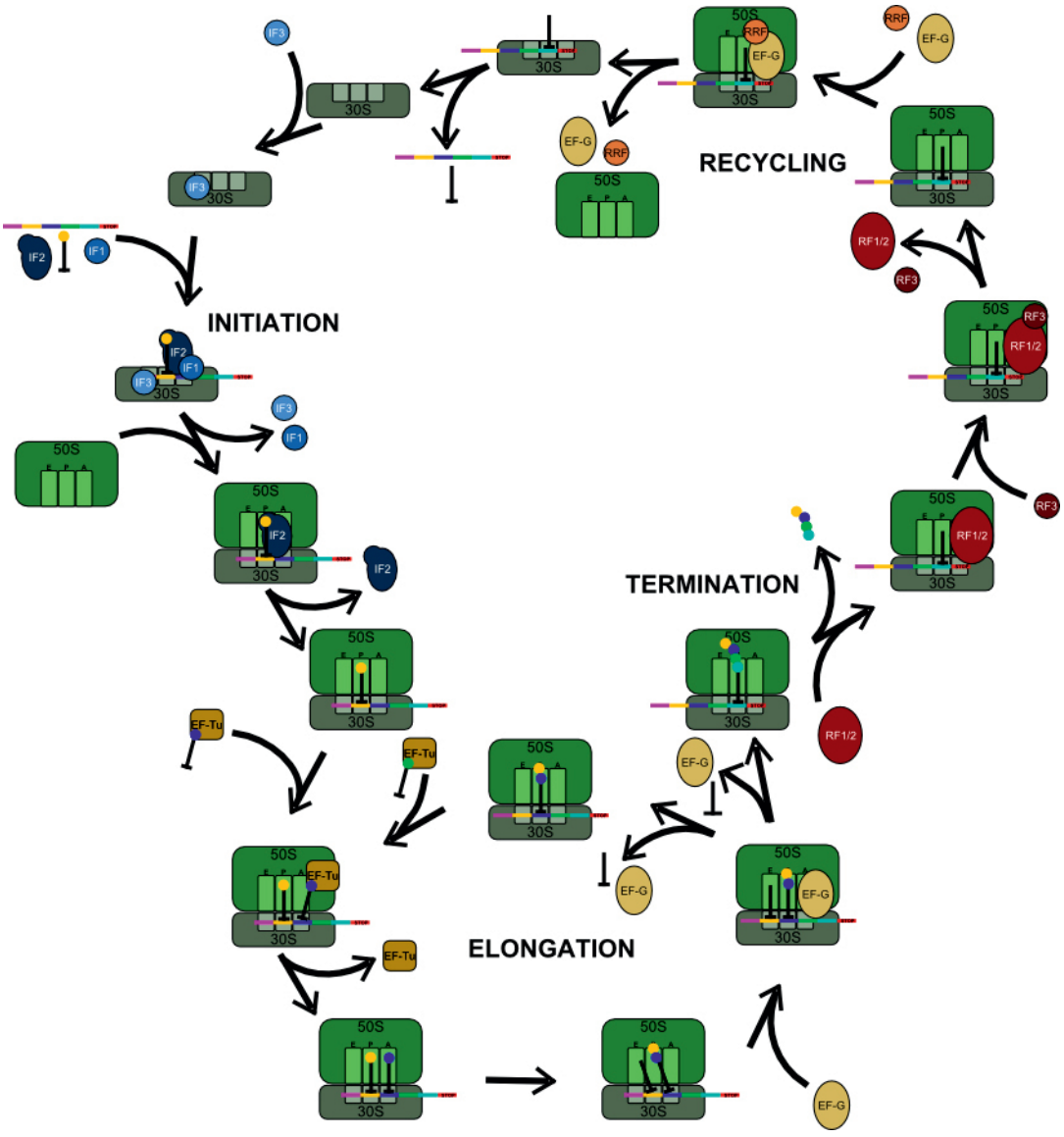

Figure 1.6: The phases of the translational process in *E. coli*; IF1, IF2 and IF3 help to assemble the ribosome during initiation; during elongation the ribosome is assisted by EF-G and EF-Tu; when a stop-codon (red) enters the A-site, RF1 or RF2 and RF3 help to terminate translation, before the parts of the ribosome are recycled assisted by EF-G, RRF and IF3. Adapted from (161).

Initiation

Translation can be divided into four main phases which are illustrated in Figure 1.6. In initiation phase, which normally happens in the order of seconds, the two subunits of the ribosome have to assemble on the TIR. Initiation is often considered as the rate-limiting step of translation, but this has been questioned lately (173). The first amino acid in the polypetide chain is normally N-formylated methionine and the cognate aa-tRNA is fMET-tRNA$_f^{Met}$ (145, 257). Three initiation factors (IF1, IF2, IF3) are essential during the initiation process. IF3 interacts with the 30S subunit and promotes dissociation of the 70S complex, starting a new round of translation. It also prevents premature association of the 50S subunit and destabilizes mismatched codon-anticodon interactions (2, 145, 202). IF1 and IF2 direct the first tRNA to the P-site (39, 145). A relativeley unstable preinitation complex consisting of the 30S subunit, the IFs, and the mRNA, which is bound codon-independently, is formed in a random order (104, 145). The more stable complex emerging from correct positioning of all components, which is achieved by interactions between SD (mRNA) and anti-SD (16S rRNA) as well as between codon (mRNA) and anticodon (tRNA), is called initiation complex (2, 154). In a last step the 50S subunit assembles to the 30S initiation complex resulting in the 70S initiation complex (173). The initiation factors are then released, which gives the A-site free for binding of the second aa-tRNA. Binding of cognate tRNA probably leads to a conformational change of the ribosome and the first peptide bond can then be formed (39, 104, 198).

Elongation

The elongation phase is characterized by sliding of the mRNA through the ribosome, where the genetic code is decoded and the polypetide is synthesized. New aa-tRNAs bind to the A-site, while the formed peptide chain is bound by the last recruited aa-tRNA to the P-site. During bond formation, the peptide chain is transfered to the new aa-tRNA and a deacylated tRNA is left in the P-site (2). This has to be transported into the E-site, while the tRNA in the P-site has to move into the A-site together with the adhering peptide chain to make space for a new aa-tRNA to enter (191). The mRNA has to be translocated

simultanously so that the next codon can be read (2). Observations from single molecule studies indicate that the 3′-end of the tRNAs, which are located in the 50S subunit, are translocated first, leading to different hybrid states (179). The protein elongation factors EF-Tu and EF-G assist the ribosomes during elongation (2). EF-Tu supports binding of new aa-tRNAs in the A-site, while EF-G promotes translocation (221).

The rate of elongation typically lies between 6 and 20 amino acids per second (161). A study that followed single ribosomes during translation showed that there typically are pauses between the single translocation steps that can take up to two minutes, one suggested explanation being interactions of the 16S rRNA with internal SD-sequences (280), another explanation might be codon-dependent variations in translocation rates (see 1.6.1).

Termination and recycling of the ribosomes

When the translating ribosomes reach a stop codon (UAA, UAG, UGA), the synthesized protein is released, followed by dissociation of the ribosome complexes. Upon binding of a stop codon in the A-site of the decoding center it is recognized by one of two release factors (RF1, RF2), which promote release of the peptide chain (2). A third release factor, RF3, causes dissociation of RF1 or RF2 from the ribosome and the resulting complex is termed posttermination ribosome complex (2, 77). It consists of the mRNA and the ribosome with deacylated tRNA in the P-site and an empty A-site. Dissociation of the ribosome is promoted by EF-G, ribosome recycling factor (RRF) and IF3 (200).

UAA has been identified as the favored stop codon for highly expressed genes and also the base directly following the stop codon seems to influence translation termination efficiency, with UAAU being the most effective translation termination signal (206, 236).

When ribosomes restart translation after termination from a nearby start codon, before they are completely released, this is called translational re-initiation (1). This phenomenon can be observed when translation is terminated prematurely or in operons consisting of several genes (291). In this process no SD-site is necessary for translation of the second gene, but translation can occur only when

ribosomes are translating the first gene. This is also called translational coupling (6). Posttermination ribosome complexes start translational re-initiation in cooperation with IF2 and IF3 (291). The process has been found to be most effective when the stop-codon of the first gene and the start-codon of the second overlap, resulting in the stop-start-codon ATGA or TGATG (138, 247).

1.4.3 The influence of the 5′-UTR on transcription

The 5′-UTR is known to be of high importance for the translational phase, however, it also influences transcription. It has for example been shown that 5′-UTR plays a role in promoter escape, probably by influencing the displacement of σ or the scrunching process (88, 115, 216). Due to the coupling of transcription and translation in bacteria, 5′-UTR can have an indirect effect on transcription, by accelerating translation which in turn increases transcription rates (207). Berg *et al.* found that also parts of the 5′-UTR which are located relatively far from the transcriptional start site can influence transcription and that this effect is promoter-independent, ascribing another important role to the 5′-UTR during transcription (20).

1.4.4 The *Pm*-5′-UTR

The 5′-UTR that naturally is found together with the *Pm* promoter is also the one which was mainly used in this study. It is 32 nucleotides long and contains the putative SD-site 5′-GGAG-3′. In former studies variants of this region that led to increased expression levels from *Pm* (up to 20-fold stimulation) have been identified. It is interesting to note that the SD-site was unchanged and that the variants showed similar behavior when combined with other promoters (20). A test with several reporter genes revealed a certain degree of gene-dependency for the stimulating influence of these variants on *Pm* (21).

1.5 Transcription factors

1.5.1 Regulation of gene expression

Not all proteins in a cell are expressed at similar levels and also expression of one single protein can vary over time and depending on growth conditions.

Regulation of protein expression can take place during all steps of expression. Transcription initiation is the step that is regulated most frequently (111). In this way overproduction of a protein can be prevented without excess formation of transcripts and unnecessary energy consumption.

Even if different sigma factors are employed to form the RNAP holo enzyme, it is always the same core enzyme molecules that take part (102). RNAP is constantly required for synthesis of householding genes and RNAs, in addition a not active part is bound to DNA. Thus there is a limitation of it leading to competition between the different sigma factors (33).

As mentioned in 1.2.2 a constitutive promoter is not always the best choice for expression of recombinant proteins. Promoters that can be regulated have a high impact on production of proteins that are toxic to the host cell, on expression at physiological levels, as typically requested in metabolic engineering approaches, but also for expression of proteins that require proper folding or secretion (31). The perfect promoter for recombinant protein expression in *E. coli* should be strong, but controllable, with minimal levels of background transcriptional activity, and it should be simply and cost-effectively inducible to varying degrees (122, 153). For induction a huge variety of mechanisms exist, as for example changed behavior due to temperature, pH or osmolarity changes, starvation or phage infection (153). Many of these mechanisms function via interacting regulator proteins, that are not part of the transcriptional enzymes, the transcription factors (TF) (293).

In *E. coli* about half of the promoters are regulated by TFs. TFs can be divided into two classes: transcriptional activators and transcriptional repressors. Some exhibit solely activating or repressing activity, while others can have both properties, depending on the target promoter (293). The activating or repressing function can also vary between different hosts (270). By a computational approach 314 TFs were identified for *E. coli* K12, whereof most acted as repressors (199). Some TFs act on a wide range of different promoters and in *E. coli* seven have been identified that control 50% of all regulated genes, while others only regulate expression from one single promoter (162).

The binding of a TF to its *cis*-acting DNA sequence, the TF-binding site, is often controlled by cofactor binding and a wide range of molecules can act as

cofactors. Some typical DNA-binding domains have been identified for bacterial TFs, as for example helix-turn-helix, winged helix and β-ribbon motifs, where the first one is the most common in *E. coli* (220, 293).
Many TFs bind as homodimers or homomultimeric protein complexes. The binding sites are typically of 16-20 nt length and show some form of symmetry like inverted repeats or direct repeats. The length of the spacer in between two repeats depends on the TF to bind, but typically it is a multiple of one DNA helix turn, ensuring that the binding sites face the same site of the DNA (220).

1.5.2 Repressors

Most commercial expression systems are based on promoters which are regulated by transcriptional repressors (31). They can function by different mechanisms as for example steric hindrance of RNAP binding, conformational DNA changes, caused by binding to different regions, or repression of the function of an activator (270). There are huge differences in repressor activity depending on where the TF-binding site is positioned relative to the promoter sequence (144). Most identified repressor sites are located between -60 and +60 (relative to the transcriptional start site), and from this it can be speculated that repression by sterical hindrance of RNAP binding is the most common mechanism of repression (220).

1.5.3 Activators

Also for the transcriptional activators different modes of action exist. Some TFs for example interact with the α-subunit of core RNAP, others promote σ-binding. In some cases conformational changes can cause activation or activators can neutralize repressors (270). Promoters which are activated by TFs often have defects in their promoter regions, whereby binding of RNAP is prevented in absence of TF (94). Most TFs bind to DNA prior to interaction with RNAP, but there are also examples for TFs that first bind to free RNAP (270).
Typical examples for positively regulated expression systems are those based on TFs of the AraC-XylS family (31). AraC was the first identified bacterial TF and

it regulates transcription of genes that are involved in L-arabinose catabolism (71).

1.5.4 XylS

Expression of XylS

XylS is the transcriptional activator that controls the TOL-plasmid *meta*-cleavage pathway for degradation of benzoates and toluates in *Pseudomonas putida*, where it derives from (286). It belongs to the AraC-XylS-family of transcriptional activators and is composed of 321 amino acids (118).

On the TOL-plasmid from *Pseudomonas putida xylS* is expressed from two different promoters. The first promoter, *Ps1*, is σ^{54}-dependent and controlled by the regulator of the upper pathway for toluene degradation, XylR. XylR induces expression from this promoter in presence of substrates of the upper pathway such as toluene, which leads to overproduction of XylS. The second promoter, *Ps2*, is a constitutive, but weak σ^{70}-promoter (79). XylS is produced at low levels from this promoter, it is found at about 200 molecules per cell (62). Not only more transcript is formed from *Ps1* than from *Ps2*, it has also been shown, that transcripts generated from the first promoter are translated more efficiently (91). Expression levels from *Ps2* are increased in the absence of expression from *Ps1* (158) and expression was found to be more effective at 30°C than at 37°C (214).

In the XylS/*Pm* expression systems used in this study XylS is only expressed from the *Ps2* promoter. The 5′-UTR of the mRNA obtained from *xylS* transcription from *Ps2* is only nine bp long and contains the putative SD-site 5′-GTGA-3′ (79).

XylS functions

XylS consists of two domains and for both models have been proposed, that are based on sequence alignments with other members of the AraC-XylS family of activators (62, 273). The N-terminal domain has been found to be involved in effector-binding, interaction with the α-C-terminal domain of RNAP and dimerization (127, 227). Single bases and regions have been identfied as essential for these processes (172, 224, 225, 226, 227). When a substrate of the *meta*-cleavage

pathway is added as effector, immediate expression from *Pm* can be observed, indicating that inducer binding renders XylS active (157). The inducer is probably bound to XylS in a nonsymmetrical way (212). Different molecules can act as inducers and m-toluate was found to be most effective (285). A big advantage of the XylS/*Pm* expression system is that the molecules acting as inducers enter the cell passively and thus no transport systems are required (143). Inducer molecules have been shown to influence transcription from *Pm* in several ways: First, XylS dimers are stabilized in the presence of inducer (227). Dimerized XylS can be considered as active, so effector binding shifts the equilibrium between inactive and active XylS towards dimerized and thus active XylS (62). Second, interaction between the two domains was suggested already in 1992 (171) and later it could be shown that the N-terminal domain represses action of the C-terminal domain. Inducer binding probably leads to a conformational change of XylS, which enables DNA-binding by the C-terminal domain (62). The same change in XylS conformation appears during dimerization in the absence of effector molecules and explains also why uninduced *Pm* still is activated when XylS is overexpressed (62, 79, 91, 166, 213): For overexpressed XylS the equilibrium is shifted towards more dimerized protein as well, and in the dimerized form the DNA binding C-terminal domain is accessible (62). As a consequence of inducer independency the ratio between induced and uninduced expression is seriously decreased for overexpressed XylS (129). A mutant that is deficient in dimerization binds to DNA in the presence of inducer, but is not able to activate transcription, demonstrating that the conformational change of XylS is not only caused by dimerization (62). Third, addition of inducer activates the heat shock response of the cell, which is necessary for an increase of the essential sigma factors σ^{32} and σ^{38}. This last effect probably also explains why the presence of inducer can stimulate transcription from *Pm* in the absence of XylS (159).

The proposed model for the N-terminal domain predicts a β-barrel and two α-helices and it was suggested that the β-barrel is involved in effector binding and the helices in dimerization (273).

The C-terminal domain is conserved among the members of the Arac-XylS family of activators. It consists of about 100 residues and is responsible for

DNA-binding and probably also involved in interaction with the sigma factors (64, 127). Seven α-helixes are proposed, that form two helix-turn-helix domains, flanked and linked by further α-helices (63). XylS binds as a dimer to two direct repeats with the sequence 5′-TGCAN$_6$GGNTA-3′ upstream of the -35 box of *Pm*, where each monomer binds to one of the two motifs. Residues in the third helix were found to bind to the first submotif in the XylS binding sequence (5′-TGCA-3′), while residues in the sixth helix bind to the second submotif (5′-GGNTA-3′) (63).The monomer binding proximal to *Pm* interacts mainly with the first submotif and if this one cannot bind, dimer formation is inhibited (80, 89). XylS binding reaches over four helical turns of the DNA and all the four stretches involved in binding lie on the same helical face (126). Also residues that are not located inside the helices are important for transcription activation and probably responsible for XylS structure (155). As mentioned in 1.3.3, XylS and RNAP binding sites overlap by two bases, and this overlap is crucial for successful transcription initiation from *Pm* (90). XylS recruits RNAP to the promoter, and probably the C-terminal domain of XylS, which is bound to the DNA, interacts with σ, which is bound to the promoter sequences (62). Interestingly, the C-terminal domain on its own can activate transcription from *Pm*, affirming the suggestion that its action is repressed by the N-terminal domain in the absence of inducer (127, 166).

Experiments with only the C-terminal domain of XylS indicate that the two XylS molecules bind sequentially, first the one proximal to the RNAP binding site, resulting in DNA bending, which enables the second to bind. This indicates that XylS dimerizes in solution before binding to DNA (64). Full activation of *Pm* by XylS cannot be obtained at typical XylS levels in the cell, only 30-40% of the promoter sequences are occupied by XylS and complete occupancy can probably be reached only by amounts of the activator that exceed its intracellular solubility. This was explained by a combination of weak XylS dimer bonds and weak interactions with the DNA (129).

The current model for activation of transcription from *Pm* by XylS is depicted in Figure 1.7.

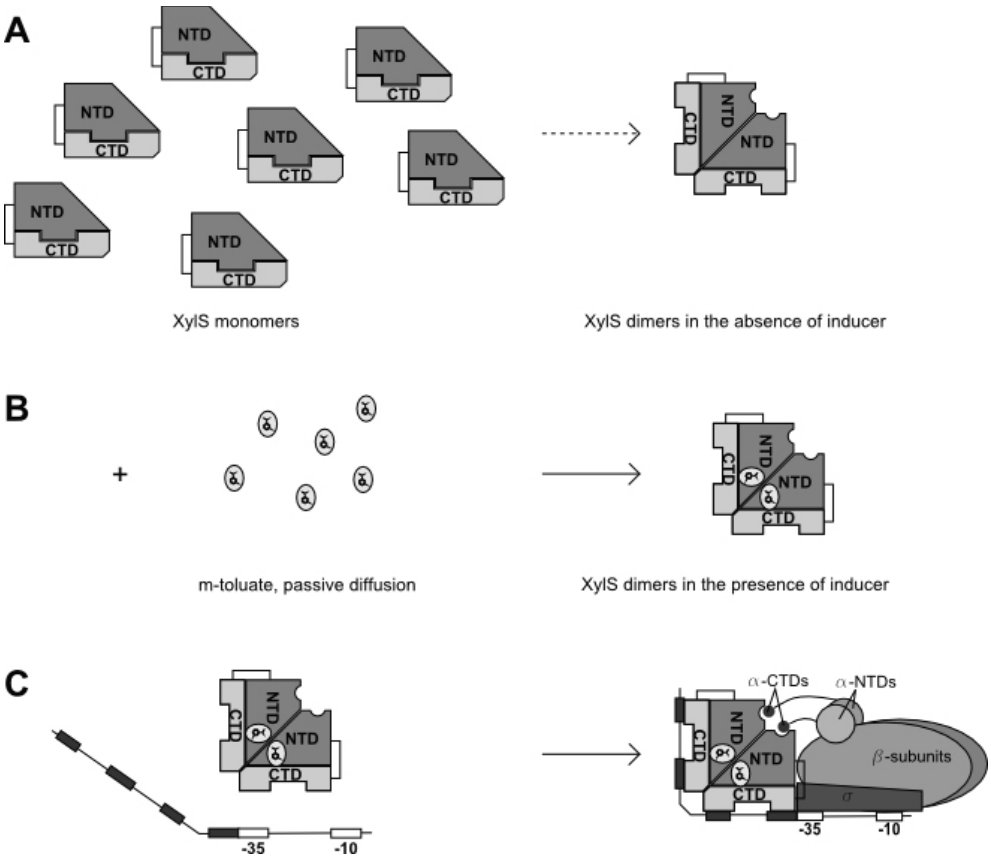

Figure 1.7: Transcription activation by XylS; a) uninduced conditions: the main part of XylS molecules is found as monomers; the N-terminal domain represses action of the C-terminal domain; some molecules dimerize, which leads to a conformational change; b) inducer binding: inducer binds to the N-terminal domain, which leads to the same conformational change as in a; dimerization, mediated by the N-terminal domain, is favoured; c) activation of *Pm*: one XylS monomer binds proximal to the -35 of *Pm*; this causes DNA-bending and the second monomer can bind; RNAP is recruited by both XylS domains. Adapted from (62, 64).

Engineering of XylS activity

Engineering of the activity of a transcription factor can be approached in two ways: either its function can be engineered or its expression levels modified. In case of XylS the first approach could for example include changes in dimerization ability, effector binding, DNA binding strength, and the ability to recruit RNAP. A range of different variants for *xylS* exist, as for example variants that activate *Pm* in the presence of untypical inducers, sometimes yielding higher expression

levels from *Pm* (212), variants that lead to an extended temperature range for XylS activity (214), and variants that mediate constitutive expression from *Pm* (294). Variants with several mutations in the N-terminal domain, that probably influence XylS activity at several levels have been identified by a library screening approach leading to about ten-fold increase of expression levels from *Pm* for the best variant (273).

There are also examples for modification of XylS levels inside the cell. XylS was found to be produced at high levels from the *tet* promoter, leading to expression from *Pm* in the absence of inducer, an effect that also was observed for expression from *Ps1* (166) . By fusing *xylS* to different promoters with a wide range of expression levels it could be observed that expression from *Pm* is highly dependent on cell-internal XylS levels (129). However, at high levels XylS looses its inducibility. This problem was solved by fusing *xylS* (in a mutant and wild type form) to a promoter which is activated in presence of inducers that also activate XylS. This leads to a double form of activation by inducer addition: XylS expression is increased and at the same time XylS is activated. By this approach the ratio between background and induced levels of expression from *Pm* could be increased to 180- to 240-fold (40).

1.6 The gene of interest

1.6.1 Features of the coding sequence

The suitability of genes for expression in foreign hosts varies greatly depending on features displayed by their coding sequence. There is a large variation in codon usage and GC-content among genes deriving from different organisms. In addition secondary structures or signal sequences can influence the ability of a gene to be overexpressed. Translational efficiency plays an important role for high expression levels. Both folding of the mRNA and codon usage have been shown to be factors determining translation initiation and elongation rates. In addition degradation of protein and mRNA influence the protein production yield (85).

The degeneracy of the genetic code allows a certain degree of freedom in choosing a codon, as most amino acids are encoded by several synonymous codons.

This implies that a GOI can be adapted to a host to a certain degree without changes in the amino acid sequence. This is less intricate than changing the primary structure of a protein, where the probability of changing secondary, tertiary or quaternary structure and thus the function of the protein is high. Also artificial gene design gets increasingly popular, connected to the decreasing prices for gene synthesis (281). However, adaptation and *de novo* design require either understanding of the underlying mechanisms or more laborious random approaches. Up to date there seems to be no perfect way of designing a DNA sequence for a certain protein, that will yield in high protein expression levels (279). A small change in the DNA sequence of a gene, even if silent, can influence a broad range of possibly interacting factors, as for example nucleotide frequency and transcription efficiency, mRNA structure, and local and global codon usage, it can create terminators, RNAse or recombination sites (107).
In the following sections some of the gene related features that influence expression levels are discussed. In general it is important to have in mind that a situation as observed for naturally expressed genes might differ from a situation, in which overexpression of a protein is desired. Overexpression comes along with high costs for the cell and is normally not favored by nature (85).

Codon bias

One of the main sequence encoded features discussed as determining translation efficiency is the codon usage. Different organisms show different patterns for the frequency of synonymous codons, termed codon usage bias, and related bacteria often show similar codon usage patterns (5, 203). There is a general tendency for favored codons to occur more often in highly expressed genes (92). One suggested explanation is that the identification of the cognate tRNA complex is the rate-limiting step in elongation (272). There is a correlation between the usage frequency of certain codons and the content of corresponding tRNAs in the cell (117). The competition between cognate and near-cognate tRNAs can explain the slower translation rate which can be observed for infrequently used codons (139, 246, 266). The lower abundance of cognate tRNAs and the associated slower translation rate are supposed to lead to ribosomal stalling, amino acid substitutions and frameshift events, and genes with a high codon bias have been

shown to be translated with lower error rates than other genes (234, 287). It has even been stated that codon bias is more important for successful translation than both the quality of the SD and the nature of the stop codon (149).

Based on these facts, one popular approach in recombinant gene expression is to adapt the codon usage of the GOI to the codon usage of the host. This approach has been successful in many cases. Expression levels in *E. coli* can typically be increased 5- to 15-fold by codon optimization, but also up to 140-fold improvement has been observed (106). However, codon optimization does not always lead to increased expression levels and can even reduce protein production (99). Especially at the 5′-end of the reading frame other factors seem to be more important than codon usage, the formation of secondary structure being one of them (see 1.6.1). Typically the codon adaptation index (CAI) or the tRNA adaptation index are used to measure the adaptation of a gene to the expression host. Both these indices do not consider local differences, but treat the whole gene equally, providing a possible explanation for why they not always correlate well with expression levels (7, 85). Recently it was proposed that codon bias does not correlate with gene expression levels, but only with local translation efficiency (140). However, this conclusion was drawn from studies with variants of one single recombinant gene.

Tuller *et al.* used a local tRNA adaptation index, that is calculated for single amino acids instead of for whole genes, and by comparison of different genomes they could identify a global accumulation of inefficiently translated codons in the beginning of open reading frames, with the exception of the second codon. This led to their ramp theory: they suggest that a ramp of unfavored codons close to the 5′-end reduces the ribosome density on translated mRNAs, by that reducing the probability for ribosomal traffic jams and abortion of translation. They also argued that under starvation conditions translation would be early aborted, because ribosomes would not be able to pass the ramp, which reduces the costs of redundant translation. They assigned a role in control of gene expression under different conditions to the ramp (265). Because of the reduction in ribosome density the ramp enables better usage of the pool of available ribosomes, which may increase protein yields for overexpressed genes (203). The ramp was found to have a typical length of 24 codons in *E. coli* and was higher

for mRNAs which exist at higher copy numbers and for genes which are expressed at high initiation rates (265).

Welch *et al.* tested the effect of synonymous codon changes in heterologous genes on expression levels in *E. coli* and came to the conclusion that codons interacting with the most highly charged tRNAs under amino acid starvation conditions were the most favored ones, not those that are most abundant in *E. coli* (278). This underlines the fact that rules can be different for heterologous, overexpressed genes than for endogenous genes.

Not only at the 5′-end of genes other factors can be more important for protein yields than codon bias. It has also been found that low-frequency codons can have a function in protein folding. Pausing of ribosomes caused by unfavored codons was found to enable proteins to fold in the correct way (46, 134). In general amino acid stretches that fold as α-helixes do more often contain favored codons than stretches that fold as β-sheets or loops (261). Another interesting observation was made by Cannarozzi *et al.*, who found for yeast that sequences that favor the reusage of tRNAs were expressed faster than sequences whose expression demands frequent tRNA-changes. They suggested that tRNAs stay associated with the ribosomes and can be recycled. A similar trend could also be observed for bacterial genes (37, 235). Yet another study revealed that pausing of ribosomes preferably occurs at SD-like sites and that this influenced translation efficiency more than codon bias (147).

There is a lot of discussion around the question if the codon usage patterns can be attributed to positive selection (203). The low codon bias observed for regulatory genes, that are present at small copy numbers in the cell, was suggested to be more a consequence of the absence of positive selection than of the presence of negative selection (237).

Secondary structure formation

Already in 1989 it was observed that bases in the coding region of a gene can shield the SD-site by basepairing and thus influence translation efficiency (231). In a more general study it was concluded that unfolding of the entire translation initiation region is crucial for successful translation initiation (52). Several

studies reporting decrease of expression levels or inhibition of translation by secondary structures support this finding (99, 252). By comparing expression levels of 154 variants of the same gene, Kudla *et al.* showed that mRNA folding near the translation initiation region plays an important role for expression levels, as this could explain half of the observed differences in the levels of expression (140). As named in the previous section, there did not exist any correlation between codon bias and expression levels. This was later explained by the fact that codon usage is important for expression levels only when no strong mRNA structures are present. For natural *E. coli* genes, which are expressed at lower levels and exhibit less secondary structure, the effect of codon bias seems to overrule the influence of secondary structure (258). This complies with the statement of Plotkin and Kudla that codon adaptation is critical for expression levels only when either translation initiation is not the rate limiting step or when proteins are overexpressed, which leads to a shortage of ribosomes (203). In addition there seems to be a general trend for high folding energy and thus few secondary structures at the 5′-end of mRNAs, which can be correlated to cell fitness but not to local translation efficiency, bringing back codon bias as an important player for translation efficiency (266). This trend is more distinct for highly expressed genes (103).

Not only mRNA structures at the 5′-end of the coding sequence influence translation, it could also be shown that local mRNA structures inside the whole gene sequence may have an influence, as they can retard translocation rates of the ribosomes (280). The mRNA has to be threaded through the ribosomes, which probably proceeds faster for less structured mRNA (85).

GC-content

The GC-content of genomic DNA varies from 25 to 75% or even more between different organisms (210), and in *E. coli* it is about 50% (183). In general it is similar for related organisms (256). Most synonymous codons vary in their third positions and this is reflected by GC-content variations in the third codon position, which can lie between 10 and 90% for some genes (238). Thus codon usage can be influenced by GC-preferences of an organism, and it was even stated that variations in codon usage are mainly driven by GC-content variations

(131). Interestingly, GC-content differs little in rRNA and tRNA, which are not translated into proteins and where RNA structure is of great importance for their function (183).

It has been widely discussed that there might be a correlation between the genomic GC-content and the optimal growth temperature of an organism, as seen for animals, because of the instability of AT-bonds at elevated temperatures (81, 182). Another factor that might influence the GC-content of DNA is the oxygen requirement of the host, and it was found that it often is higher for aerobic bacteria than for anaerobic ones (185). Recently, it has been revealed that there is a general bias in bacteria for mutations from G or C to A or T (112), while bacterial fitness correlates positively with the GC-content of a gene (210). As already discussed, factors other than codon bias in the 5′-region of mRNA seem to be of high importance for effective translation and besides low mRNA structures also high AU-content was found to play an important role (4).

Signal sequences

Even if protein production takes place in the cytoplasm, proteins can be secreted into the culture medium or in case of Gram-negative bacteria into the periplasm. This can be essential for proteins which need proper folding, especially for proteins containing disulfide-bonds, as the reducing environment of the cytoplasm does not allow disulfide-bond formation (260). Secretion also simplifies downstream processing, especially at the N-terminal end, reduces potential purification steps and problems with degradation of proteins by proteases (43). In addition the protein producing cells do not have to be disrupted for purificatio, when proteins are secreted (165), and in *E. coli* secretion of proteins into the culture medium is unusual, which leads to few other proteins in the culture medium (209). Different mechanisms have been identified for translocation of proteins over the inner membrane, and the presence of short N-terminal amino acid sequences (signal sequences) is necessary for secretion in any case. There is a broad range of signal sequences for use in *E. coli* and they show only some common features, as for example accumulation of hydrophobic amino acids (208).

Interestingly, also random sequences were shown to be capable of directing proteins to the periplasm (125). The signal peptides are typically cleaved off by a signal peptidase upon transport out of the cytoplasm (43). Secretion of proteins is widely used in recombinant gene expression due to the above mentioned advantages (82). Also expression of proteins that do not need secretion for proper folding can be improved by fusion of N-terminal signal sequences. This can be simply due to changes in the mRNA secondary structure formation and changes in codon usage at the 5′-end (279).

1.6.2 Reporter genes used in this study

bla, coding for β-lactamase, which confers resistance to ampicillin

The gene *bla* codes for the enzyme β-lactamase, which renders cells resistant to penicillin antibiotics like ampicillin. Penicillin and other β-lactam antibiotics act on the bacterial cell wall by binding to a membran-associated protein, which leads to cell lysis (262). β-lactamase is secreted into the periplasm of Gram-negative bacteria, where it can interact with the antibiotics. It inactivates penicillins by hydrolysis of their β-lactam ring (44). The resistance mediated by β-lactamase renders the *bla*-gene interesting for screening procedures, especially because a linear relation between amounts of β-lactamase and the resistance to ampicillin has been observed (267). Varying levels of β-lactamase expression can be easily monitored by plating of cells on agar medium containing different concentrations of ampicillin. This has been used to identify variants for different elements in the XylS/*Pm*-expression cassettes (11, 20, 273, 284).

The well expressed *celB* gene

celB is a gene which was identified in *Acetobacter xylinum* in 1991 (76). It was found to encode a protein with phosphoglucomutase activity, catalyzing the formation of D-glucose 6-phosphate from α-D-glucose 1-phosphate (28). A method for measurement of enzymatic activity has been described (76). The *celB* gene has previously been cloned into XylS/*Pm*-expression vectors, and the produced protein was clearly visible on a crude extract protein gel (25).

luc, coding for luciferase

A gene coding for luciferase from the North American firefly *Photinus pyralis* was used as reporter gene in this study. Firefly luciferase was purified already in 1956. In the presence of oxygen, luciferin, adenosine triphosphate and Mg^{2+} it catalyzes a reaction during which light is produced. This process leads to the bioluminescene observed for fireflies (97). The *luc*-gene was isolated and established in *E. coli* in 1985 (55). Assays for enzyme activity measurement exist and they are quite sensitive (196), making *luc* a promising reporter gene. When expressed from the strong T7 promoter high amounts of insoluble luciferase where formed (15).

gm-csf, coding for a medically important protein

A fourth gene used as reporter gene in this study was *gm-csf*, coding for one of four granulocyte-macrophage colony-stimulating factors, the glycoprotein GM-CSF (168). GM-CSF controls the formation of granulocytes and macrophages in white blood cells and thus it is clinically important (169). It has been shown to be producible from the XylS/*Pm*-expression system at industrially relevant levels (up to 1.7 g L^{-1}). Interestingly high production levels could first be obtained when a signal sequence was fused to the 5′-end of the gene. The highest production levels were observed with the *ompA* signal sequence (245).

The gene coding for host-toxic scFv-phOx

The last reporter gene used in this study was the gene coding for the single-chain antibody variable fragment scFv-phOx. In single-chain antibody fragments, the heavy and light antibody fragments are connected by a short linker and exist thus as a single chain (23). Shuffling of the two parts in this specific fragment improved affinity for hapten 2-phenyloxazol-5-one (phOx) (156). scFv-phOx has been demonstrated to be producable from the XylS/*Pm*- expression system at relatively high quantities (up to 2.3 g L^{-1}) (244, 245). Because of host toxicity problems in *E. coli*, expression conditions had to be carefully adapted in order to achieve high expression levels (244). Best results could be obtained, when the signal sequence *pelB* was fused to its 5′-end (245).

1.7 Subcellular organization in *Escherichia coli*

Bacterial cells do not contain any cell compartments, and for a long time it was taken for given that molecules freely diffuse in the cytoplasm. However, recently studies accumulate that contradict this assumption and indicate that most substances have certain addresses inside the cell. Important factors for positioning of molecules can be cytoskeletal elements, cell polarity, and membrane properties, securing effective processes in the cell (93). Cytoskeletal elements, for example, are responsible for effective chromosome segregation into the daughter cells (54, 95).

For plasmids it has been found that different replicons locate to different positions (113, 189, 204, 289). In case of the RK2 replicon two proteins, IncC and KorB, and binding sites on the DNA are responsible for partitioning and proper plasmid positioning (274). RK2 plasmids are mainly found at mid- and quartercell-positions, where they form clusters (204). Minimal RK2 replicons, which do not contain the genes encoding the two partitioning proteins, are mainly found at the cell poles (132, 274). They are generally more mobile inside the cell than complete RK2 replicons (59). Correct plasmid positioning inside the cell is depending on active transcription from a plasmid promoter (230, 288). Also high copy number plasmids form clusters, but only in 90% of the cells, while 30% contain also freely diffusing plasmids (204).

Visualisation of RNAP positioning during different conditions showed that most RNAP molecules are co-located with the nucleoid (36). This agrees with the fact that there is almost no free RNAP inside bacterial cells and that the enzyme probably is almost completely bound to DNA (239). For mRNAs different observations have been made. Transcripts could be identified which diffuse through the whole cell (87), while several researchers conclude from their results that mRNA in most cases is not spread throughout the cell during its lifetime, as could be expected from its diffusion coefficient. It can either stay close to the transcribed gene (177, 269) or migrate to the subcellular spot, where its protein product will be required (187, 188).

For ribosomes localisation depends on the presence or absence of transcription. While ribosomes that are depleted of mRNA diffuse through the cell, they are

located around the nucleoid and close to the cell poles when transcription is ongoing (163, 177).

2

Aims of the Study

The central goal of the work presented in this thesis was to improve the recombinant protein production from vectors that combine the XylS/*Pm* expression system with the mini-RK2 replicon, with a side goal to better understand the underlying mechanisms that cause changes in expression levels. Main focus was laid onto increasing expression under induced conditions. This was approached in two different ways; by combination of high-expression variants for different regulatory elements in one expression cassette, and by modification of expression of the transcription factor XylS.

For combination of variants the three regulatory elements promoter, *Pm*, 5′-UTR, and transcriptional activator gene, *xylS*, were chosen in the first place, with the option to include other variants, as for example for the replication protein gene, *trfA*, at a later stage. It was decided to test both direct combination of previously identified variants and screening procedures to reveal the best method for enhancement of expression levels.

XylS expression modification was chosen as a second approach because this did not only have the potential to increase expression levels from *Pm*. A secondary objective was to obtain new insight into the mechanisms by which the transcription factor activates *Pm*.

Increased expression under induced conditions typically is accompanied by raised background expression levels and often a reduction of cell fitness due to the metabolic load imposed on the cells by recombinant protein production. Thus ways to reduce the metabolic load and to increase induction ratios were explored as well. This was approached by integration of an optimized cassette

into the host chromosome, by the implementation of low-expression variants and by regulation of XylS expression. This last approach was later on used to reduce background expression of the expression cassettes which were optimized for high expression.

Even if the presented work mainly was focused on the XylS/*Pm* expression system and the host *E. coli*, general findings were thought to be transferable to other systems and possibly also to other hosts.

3

Results and Discussion

3.1 Maximization of protein production from a XylS/*Pm* expression cassette under induced conditions

3.1.1 Combination of stimulatory expression control elements

The XylS/*Pm* expression cassette has been intensively studied in our group and has been demonstrated to be useful tools for high-level production of recombinant proteins (see for example 244, 245). Variants of the different control elements in the XylS/*Pm* expression cassette, which lead to increased levels of protein expression, have previously been identified (11, 20, 273). One of the aims of this thesis was to investigate whether the XylS/*Pm* system could be further improved by means of combinatorial engineering.

As a first approach previously identified high-expression variants were directly combined in the same expression cassette. For this one variant each was picked for the elements promoter *Pm* (P), activator protein gene *xylS* (X) and 5′-UTR (U) (sequences in Paper I, Figure 1), while *bla* was chosen as reporter gene, which also had been used for the identification of the variants in the first place. Low inducer concentrations (0.1 mM 3-methylbenzoate) were used to avoid potential host-toxicity effects caused by high expression levels.

Prior to combinatorial testing the variants were individually characterized both at the transcriptional (qRT-PCR) and the translational level (β-lactamase en-

zyme activity) in the pTA16 backbone (Paper I, Figure 2). Previously it has been shown that *Pm*- and *xylS*-variants mainly act on the step of transcription (11, 273), while 5′-UTR-variants could be identified that mainly act on the level of translation (β-lactamase-activity) (20). In accordance with these findings expression for cells with either comP or comX was higher than for cells with the wild type plasmid and increased to similar extents both at the transcriptional and the translational level (Paper II, Figure 1a), which can be explained by improved transcription initiation. Raised transcript amounts are known to generally lead to an increase in translation at a similar degree, as long as the additional transcripts are not degraded or ineffectively translated (53). In contrast to comP and comX, the 5′-UTR-variant comU was found to primarily act on the level of translation. However, transcription was improved for comU as well, and this can be explained by the positive effect of improved translation on transcription because of the coupling of these two steps (see 1.4.3).

One would assume that combining variants that act on different steps of expression has a higher potential for additional improvement than combination of variants which individually influence the same step. This assumption was confirmed, when expression levels for plasmids that combined each two of the variants (comXP, comPU, comXU) were tested (Paper II, Figure 1b): for plasmid comXP expression levels were almost unchanged, compared to the plasmids with the single variants, while both for comPU and comXU expression levels were further increased. Interestingly, the ratio between transcript formation from *bla* and β-lactamase enzyme activity was constant (about 2.5-fold) between comU and comXU, while both were about four times increased from the latter plasmid, which correlates well with the transcription enhancement caused by the use of comX. This indicates that the variants for *xylS* and 5′-UTR, in contrast to the other two combinations, act in a directly additive manner. From these results it can be concluded that the effect on expression which can be achieved by combination of variants depends on the mechanisms by which they individually influence expression.

When all three stimulatory variants were combined (comXPU), expression levels could be further improved and an increase of about 75-fold at the protein level and about 63-fold at the transcript level, compared to the wild type system,

was obtained (Paper II, Figure 1c). On agar plates cells harboring comXPU tolerated 15 mg mL^{-1} ampicillin, when induced with 0.1 mM 3-methylbenzoate, which also accounts for a 75-fold increase compared to the wild type.

Direct combination versus screening procedures

Only combination of *xylS*- and 5′-UTR-variants resulted in a directly additive effect on expression, and the question arose if higher expression levels would be achievable by a screening procedure. Previously only screenings had been performed, in which one element was mutated, while all the others were kept in their wild type form. Here one element should be mutated, while another one was used in a high-expression variant form.
Three different library screenings were performed: in the first 5′-UTR-variants were screened for in comP; in the two other screenings comX was used as starting point to screen for *Pm*- or 5′-UTR-variants. In all three cases the improvement which could be achieved by this screening approach was comparable to or even lower than the improvement obtained by direct combination of previously identified variants (Paper II). Even if it cannot be excluded that better variant combinations could have been identified by a more extensive screening, based on the empirical findings the less laborious direct combination approach stands as a better approach.

Expression of different reporter genes from plasmids that combine stimulatory variants

To test the applicability of the combination expression cassettes for production of other recombinant proteins, the *bla* gene was substituted by four other reporter genes: *celB*, *luc*, *gm-csf*, and the gene coding for the human antibody fragment scFv-phOx (see 1.6.2).
For *celB* (coding for phosphoglucomutase) expression levels from the combination plasmids were measured by an enzyme assay and qRT-PCR (Paper I, Figure 4a) and expression was found to be significantly stimulated both at the transcriptional and at the translational level. However, the stimulation was not as strong as for *bla* and especially for comXPU the transcript formation was stimulated more than the phosphoglucomutase enzyme activity. A possible explanation would be inclusion body formation, leading to production of inactive

protein, however, as described in 3.1.2 it is possible to reach higher levels of active enzyme production from a comparable system with increased plasmid copy number. Based on this information inclusion body formation is not likely to be an issue. *celB* is well-expressed from the XylS/*Pm* wild type system and the protein is clearly visible on an SDS-PAGE gel (25). Thus it is more probable that, in contrast to expression of *bla*, the excessive *celB*-transcript amounts produced by comXPU cannot be handled by the translational machinery.

luc (coding for luciferase) has previously been used as reporter gene to test the effect of mutations in the promoter region on expression, and enzyme levels could be increased about four-fold by use of a *Pm*-variant instead of wild type *Pm* (11). No further increase in luciferase activity could be achieved by using plasmid comPU. A codon-optimized version of *luc* was ordered to exclude unfavorable codon usage due to its animal origin as limitation for efficient expression. However, also with this codon-optimized version expression levels could not be significantly increased. Tests in other combination plasmids were not undertaken, as it seemed that increased expression of *luc* from the XylS/*Pm* expression cassettes was limited and an explanation might be production of the protein in inactive form, as observed, when it was expressed from the strong T7-promoter (15).

For expression of the two last and medically important proteins, N-terminal signal sequences were fused to the genes (OmpA in case of *gm-csf*, PelB in case of *scFv-phOx*), as this had earlier been discovered to increase expression levels (245). Transcript amounts were measured by qRT-PCR (only for *gm-csf*) and protein amounts visualized by Western Blotting. From the wild type XylS/*Pm* casette not enough protein is produced to be detectable on a Western Blot (245), thus plasmids with elevated copy number (3-4x) were chosen as a second reference. This offered the additional advantage of direct comparability of new results to formerly published ones (245). Comparability of results was also the reason for using RV308 as a host, and not DH5α, which was used as a host in the expression studies with *bla*, *celB*, and *luc*.

In case of *gm-csf* (coding for a granulocyte-macrophage colony-stimulating factor) an increase in transcript amounts could be observed for the two tested combinations (comXU, comXPU, Paper I, Figure 5b) compared to the wild

type plasmid. Results from Western Blotting indicated that *gm-csf* expression was much more stimulated at the translational than at the transcriptional level: while transcript amounts for comXPU were lower than for the reference plasmid with elevated copy number, the Western Blot signal was stronger for comXPU than for the reference. To exclude host-specific differences the plasmids with *gm-csf* were tested in DH5α cells, as well, and similar results as with RV308 were obtained.

For the antibody fragment scFv-phOx less protein was produced from the combination plasmids than from the reference plasmid with increased copy number. However, the protein was detectable on a Western Blot, when produced from either comXU or comXPU, which was not the case for the wild type plasmid with normal copy number, indicating that a slight improvement could be achieved. Less protein was produced from comXPU than from comXU and it can be hypothesized that this is due to the host-toxicity of scFv-phOx in *E. coli*, which might cause difficulties in combination with comXPU because of high background expression levels of this construct (see also 3.2).

With all of the tested reporter genes improvement of expression levels compared to the wild type system could be achieved by the combination of stimulatory variants, albeit to varying extents. The best results were obtained for *bla* and the reason may be that the variants originally were identified with this gene and thus particularly well adapted to it. If one considers the importance of the composition of the 5′-end of genes for their successful expression (see 1.6), it seems obvious that different genes may respond differently to changes in the 5′-UTR. But also a stimulation of transcription initiation, as probably caused by comX and comP, might have varying effects on the expression of different genes depending on their gene-encoded features. In addition there might be some general limitations for improvement of expression levels, as observed for *celB* and *luc*. For GM-CSF and scFv-phOx high expression may be limited due to their eukaryotic origin by factors like inappropriate codon usage and proteolytic degradation, and for scFv-phOx it is known that expression conditions in *E. coli* have to be fine-tuned in order to achieve high-level production of the protein (244).

Based on the presented findings it is clear that combination of stimulatory variants is a powerful technique to increase gene expression levels also for other, unrelated genes than the one used for their identification within the limits of gene- and protein-specific boundaries.

3.1.2 The effects of gene dosage on expression

For *celB* expression was mainly stimulated at the transcriptional level, when expressed from combination constructs while enzyme activities stagnated at about 8.5 times wild type activity (Paper I, Figure 4a). This roof level could not be exceeded by variations in inducer concentration either (Paper III, Figure 1a), while transcript amounts could be increased up to over 25 times, compared to the wild type cassette, by this approach.

To test the influence of plasmid copy number on expression of *celB*, the wild type *trfA*-gene in comXPU was exchanged by the variants cop271 and cop254, respectively. Cop271 leads to three- to four-fold and cop254 to 24-fold increased plasmid copy number compared to the wild type system, for which it is four to eight plasmids per cell (68, 133). The resulting plasmids comXPU.271 and comXPU.254 with *celB* as reporter gene were tested with a wide range of different inducer concentrations and results can be seen in Paper III, Figure 1 b and c. Interestingly, the previously observed roof for translation could be exceeded with comXPU.271 (Paper III, Figure 1b). Thus this roof was not caused by a general incapability of the host cell to produce more active protein. However, also for comXPU.271 a new translational limit at around 22-fold wild type activity was observed, while transcript amounts could be increased almost 50-fold, compared to the wild type plasmid.

The highest enzyme activity was achieved by comXPU.254, when induced with 0.01 mM 3-methylbenzoate, and was about 38 times higher than that of the wild type, while about 180 times more transcript was produced (Paper III, Figure 1c).

Separation of total cell extracts by SDS-PAGE demonstrates that phosphoglucomutase is one of the dominating proteins in the cell, when produced from the wild type plasmid (Paper III, Figure 2). For the sample deriving from comXPU.254 a much stronger band, outnumbering those of all other proteins

in the cell by far, could be seen on the gel.

A standard curve established by qRT-PCR measurements with in *vitro* transcribed *celB* mRNA and 16S rRNA allowed for estimating the number of transcripts per cell. About 13.5 *celB*-mRNAs and 7280 16S-rRNAS per cell were calculated for the wild type system, grown in the presence of 0.1 mM 3-methylbenzoate. The calculated amount of 16S-rRNA molecules is significantly lower than what can be found in the literature (186), indicating that also the real amount of *celB*-mRNAs per cell might be higher. The main factor causing this discrepancy is probably loss during cell-lysis and RNA-isolation. Another possible source of error derives from the calculation of the cell mass via spectrophotometric OD-measurements, as this does not distinguish between viable and dead cells. It cannot be excluded either that the number of 16S rRNA is low in the used host. However, based on the calculated value for the reference case, mRNA molecules per cell were calculated for other settings as well, and considered as minimum amounts.

It is noteworthy that in the best case (comXPU.254, 0.1mM 3-methylbenzoate) at least 2460 molecules *celB*-mRNA are present as a cell average. Due to the stochastic nature of gene expression there might even exist some fluctuations between different cells, resulting in extremely high transcript amounts inside the cells, if one considers that there typically are just slightly less than 1000 mRNAs (accounting for about 600 different messengers) present in an average *E. coli* cell (186). A comparison of the mRNA amounts of samples, for which protein production levels off (comXPU, 10 μM 3-methylbenzoate and comXPU.271, 5 μM 3-methylbenzoate, Paper III, Figure 1) indicates that there is a certain amount of mRNAs per plasmid copy which can be handled by the translational machinery. If the estimated numbers are considered as correct, protein amounts reach the roof level, when slightly more than 20 transcripts per plasmid copy are produced in average.

A similar roof level as with comXPU could also be observed, when *Pm* was exchanged for the T7 promoter from pET16 (see 1.3.4), which is known to lead to production of large amounts of transcript, normally resulting in high levels of protein production (15). As it can be seen from Paper III, Figure 5, the transcript amounts were higher for plasmids with the T7 promoter than for plasmids

with *Pm* and could be increased when plasmids with higher copy number were used. Rising the concentration of inducer IPTG from 0.1 mM to 0.5 mM did not result in higher enzyme activities than from *Pm* either. From that the conclusion was drawn that the observed phenomenon is not promoter-specific.

Similar levels of transcript are not processed equally for varying copy numbers

In the experiments described above it could repeatedly be observed that constructs that led to varying amounts of transcript formation produced similar amounts of enzyme with *celB* as reporter gene. The question arose if it would be possible to produce the opposite situation, that is to find two plasmids with similar expression behavior at the transcript level, which show differences in expression at the translational level. This led to the construction of comX.254 (comX with the *trfA*-variant cop254). Based on the combinatorial study described above (3.1.1) less transcript is formed from comX than from comXPU at the same plasmid copy number, independent of the reporter gene used. Expression of *celB* from comX.254 led to transcript levels which covered a similar range as those from comXPU.271 (Paper III, Figure 3). However, stronger enzyme activity was generally observed for comX.254 than for comXPU.271 at similar transcript levels. In contrast to comX, comXPU contains promoter- and 5′-UTR-variants, but this cannot explain the difference in enzyme activities, as the variants in comXPU have been shown to result in more efficient expression of *celB* than the wild type elements in comX (11, unpublished results), while enzyme activities here were higher for the comX-version. For a direct comparison, enzyme activities were measured for comX.271 as well, which confirmed that expression per plasmid copy number is lower from comX than from comXPU (Paper III, Figure 3). Thus the difference in expression at the translational level in spite of the similarity of transcript amounts can only be explained by the change in copy number, strengthening the theory that the observed translational limitations do not reflect a general limit for the whole cell to produce more active protein.

If *celB*-mRNA was diffusing freely in the cell during its lifetime, one would expect to see no differences in the efficiency of its processing into protein for the

plasmids with different copy number. The observed results indicate rather that *celB*-mRNA stays associated with the transcribed genes and thus the plasmids, and this has been described as one possibility for mRNA localization in the cell (188). Transcription from a higher number of plasmids might then lead to a wider distribution of *celB*-mRNA, especially if one considers that high copy number plasmids are thought to display a higher mobility in the cell than low copy number plasmids (204). Accumulation of transcripts around the plasmids may either lead to a crowding effect, which prevents the translational machinery (or parts thereof) to access the mRNA, or it may result in local depletion of translational components due to the high concentration of mRNAs at one spot. In either case higher translation rates would be achievable, when a certain amount of transcript is produced at a higher number of locations, which can explain the observed results.

3.1.3 Modification of XylS expression

As alternative method to combination of stimulatory variants and copy number variation, modification of XylS expression should be examined as a tool to increase expression levels from XylS/*Pm* vectors. XylS is expressed from its natural promoter *Ps2* at low levels, which are sufficient for the performance of its role as transcription factor in activation of *Pm* (62). It is known that expression from *Pm* is strongly influenced by cell-internal XylS levels, while overexpression of XylS from a strong promoter causes a loss of induciblity of *Pm* (129, 166). This loss of inducibility seems to be avoidable by small changes in XylS expression levels and it should be tested to which degree that could be exploited.

Codon optimization of *xylS*

As described in 1.6.1 adaptation of the codon usage to the host can be an efficient tool to increase expression levels (106). XylS does not derive from *E. coli* and thus codon usage may be unfavorable for this host. The CAI for *xylS* in *E. coli* was calculated with help of the Java Codon Adaptation Tool (101) and turned out to be relatively low (0.222). A direct comparison of codon occurences in *xylS* with rare codons in *E. coli* (taken from reference 184) revealed that some

of these rare codons occur up to seven times in *xylS*.

Based on this a synthetic version of the gene (*synxylS*) was ordered from GenScript with an improved codon adaptation index of 0.438 in *E. coli*. Besides that unfavorable and uneven GC-content as well as unfavorable mRNA secondary structures were intended to be avoided. The synthetic gene was in the first place cloned into the XylS/*Pm*-based plasmid pTA13 (273), with *bla* as reporter gene. The goal was to study effects of this exchange on expression from *Pm*, with the possibility to express it from a stronger promoter if it turned out to be well expressed. Surprisingly, expression from *Pm* was heavily decreased, when *xylS* was exchanged by its synthetic version (pTA13.syn, for parts of the sequence and position of the mutations see Figure 3.1 a and b). This was not caused by changes in transcription of *xylS*, as transcript levels turned out to be similar for both versions.

In order to encircle the most problematic region for expression of the synthetic *xylS* gene, fragments of wild type *xylS* were exchanged for their counterparts from *synxylS* (leading to *xylS*-versions named syn[x-y]). The focus was laid onto the 5′-end of the gene, because it has been repeatedly described, that in this region factors other than codon usage can strongly influence expression efficiency (see 1.6.1). When tested on agar plates with varying ampicillin concentrations, syn[1-42] (with only eight mutations compared to the wild type gene) turned out to confer the same low level of expression as the whole synthetic gene. The effect of the eight mutations in syn[1-42] was studied individually and in combinations (results in Figure 3.1 b and c). With the knowledge that secondary structure formation at the 5′-end can slowdown translation initiation (99, 140), the different variants were tested by bioinformatic tools for potential secondary structure formation, even if this was supposed to be avoided in *synxylS*. A strong stem-loop structure was predicted for several of the variants with low-expression phenotypes. A possible explanation for the occurence of this structure is that the 5′-UTR of *xylS* was not considered when the synthetic gene was designed.

However, when a variant (synA, Figure 3.1 a) was tested, which was predicted to lead to expression levels between those for *synxylS* and wild type *xylS*, caused by a weakening of the stem-loop structure, only a slight improvement compared

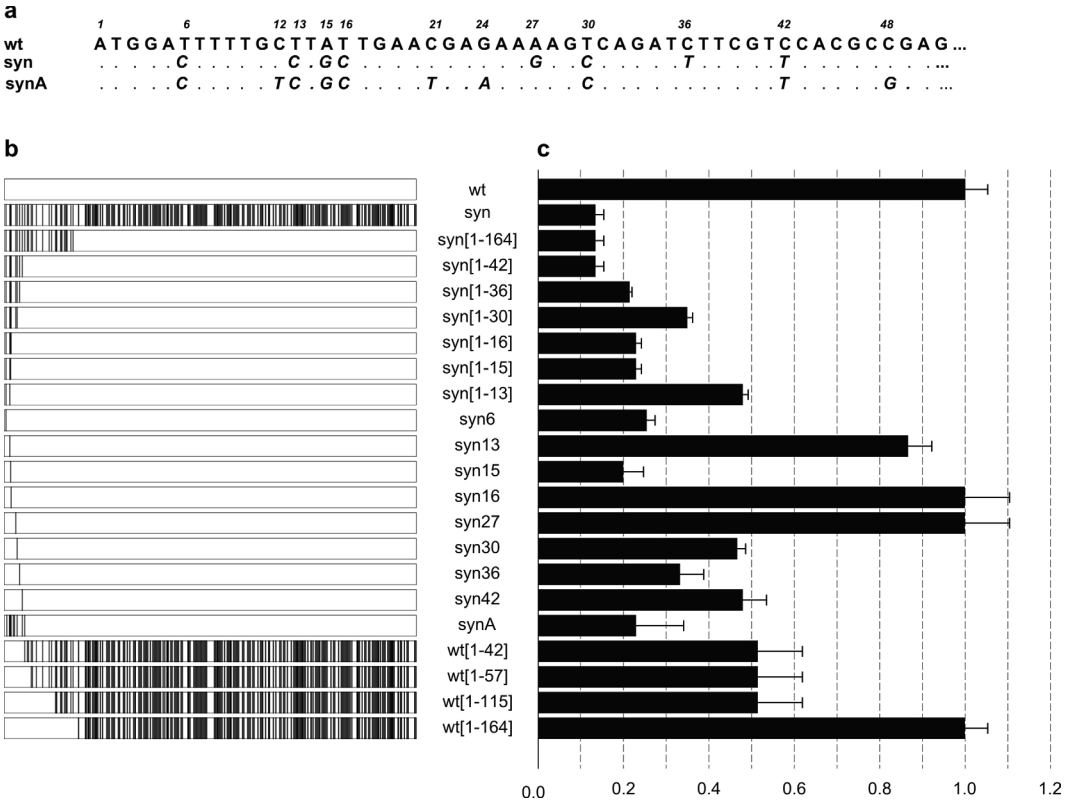

Figure 3.1: Effect of synonymous mutations in *xylS* on expression from *Pm*; a) synonymous mutations in two synthetic versions of *xylS*; b) placement of mutations inside *xylS* for different versions of the gene; c) expression levels from *Pm* for different versions of *xylS*, measured as resistance to ampicillin on plates; all values are relative to those for the wild type system, which is set to 1; error bars indicate ampicillin concentrations which inhibit growth; numbers in square brackets indicate the part of *synxylS* or wild type *xylS*, which was present, single numbers stand for single mutations in the corresponding position of *xylS*.

to *synxylS* could be observed (Figure 3.1 c). This indicated that problematic translation initiation is not the only causer of low expression in case of *synxylS*. *xylS*-variants were then constructed, in which the 5′-end of the synthetic gene was exchanged for their wild type counterparts (naming: wt[x-y], see Figure 3.1 b and c). In this way the formation of the stem-loop structure should be avoided, which should have the potential to increase expression levels due to the codon optimization. Other than expected, also these variants led to reduced expression levels from *Pm*, compared to wild type *xylS*, and first when a fragment of 164 base pairs was exchanged, levels similar to those for wild type (but not better) could be achieved (Figure 3.1 b and c). These results confirm that the

stem-loop structure is not the only reason for the decreased expression of the synthetic gene.

Even if no single mutations could be identified that were responsible for this phenotype, two of them (in positions 6 and 15) decreased expression levels almost as drastically as syn[1-42]. That led to the question if other single mutations with a similar, or even stronger impact, might exist. To test this a library was constructed, in which the 5′-end of *xylS* (the first 32 bp) was mutated such that only silent mutations (like in *synxylS*) were allowed. In total 576 out of 200 000 clones were plated on varying ampicillin concentrations. Sequencing of 25 of them, which did not grow on 100, 200 or 300 μg mL^{-1} ampicillin, revealed deletions or insertions in 20 of them, leaving five candidates, whereof two had one single mutation, which was the same as *synxylS* has in position 6 (Figure 3.1 a). Also the other three variants contained mutations which also were found in *synxylS*. Thus no new single mutations with similar effect could be identified. Transcript amounts were similar for wild type *xylS* and *synxylS* and the protein products can be expected to be the same (all mutations in *synxylS* are synonymous), indicating a reduction in cell-internal XylS-amounts as explanation for the decrease of expression from *Pm*. This assumption could not be proved by Western Blotting, as all efforts to detect XylS by this method failed.

To enable an alternative and precise detection of XylS expression levels, a plasmid was constructed, in which *luc* is fused 3′-terminally to *xylS* by an overlapping stop-start codon (TGATG at the DNA level). In this construct *luc* is supposed to be translated only by ribosomes which translate *xylS* first. When this construct was tested with different *xylS*-variants, luciferase activities were suspiciously similar for all variants. A potential SD-site at the 3′-end of *xylS* was discovered, which might allow *luc* to be translated without previous translation of *xylS*. Indeed, when a version of *xylS* with a His-tag at the 3′-end, functioning as spacer between the potential SD-site and the start codon of *luc*, was used (plasmid pFS7, Paper IV, Figure 1), variations in luciferase activities for different *xylS*-variants existed and correlated with the expression levels from *Pm* (Paper IV, Figure 2).

It could now be concluded that the low level of expression from *Pm*, which had been observed with *synxylS*, indeed was caused by translational bottlenecks.

The synthetic operon construct pFS7 had been proved useful as a tool for detection of variations in XylS amounts at small concentrations inside the cell.

A change of promoter and/or 5′-UTR leads to high expression levels both for wild type *xylS* and *synxylS*

The expected improvement of expression by use of a codon-optimized version of *xylS* could not be achieved. Nevertheless, both wild type *xylS* and *synxylS* in the synthetic operon setting described above were tested behind the strong T7-promoter in a pET16b-plasmid (Novagen). A host which coexpresses T7-RNAP had to be used, and ER2566 was chosen for this purpose. One would expect that the difference in expression between the two versions of *xylS* would be even stronger at higher expression levels. In contrast, no significant differences in luciferase activity could be observed between the two different strains. For both an 800- to 850-fold higher luciferase activity than for pFS7 was detected. To test the influence of the changed 5′-UTR-sequence in front of *xylS* in the pET16b-plasmid on expression levels, five alternative 5′-UTRs were cloned into plasmid pTA13 instead of the short natural 5′-UTR: two, which have been demonstrated to increase expression when fused to *Pm*: LII.10 (20) and H39 (= comU in 3.1.1); the T7-5′-UTR from plasmid pET16b; the same 5′-UTR including a His-tag, which is located between 5′-UTR and the gene of interest in pET16b; and the natural *Ps2*-5′-UTR together with the His-tag from pET16b (for sequences see Figure 3.2 a). For the last construct expression levels were slightly lower than for wild type 5′-UTR (Figure 3.2 b). For all the others, cells were growing on increased and interestingly similar ampicillin concentrations (3.5 mg mL^{-1}, wild-type: 0.6 mg mL^{-1}). Remarkably, for both constructs with a His-tag, the expression levels for wild type *xylS* and *synxylS* were leveled out. This indicated that it is not the high expression level which helps to overcome the translational problems of *synxylS*, when expressed from the T7 promoter, but rather the His-tag, probably by moving the stem-loop structure, which has been proposed for *synxylS*, away from the TIR. It was suggested that partial addition of the His-tag to *synxylS* should gradually raise expression levels to those of wild type *xylS*, but none of the constructs with only parts of the His-tag tolerated ampicillin at high concentrations. This was also the case, when the His-tag fragments were fused to wild type *xylS*, and might be due to production

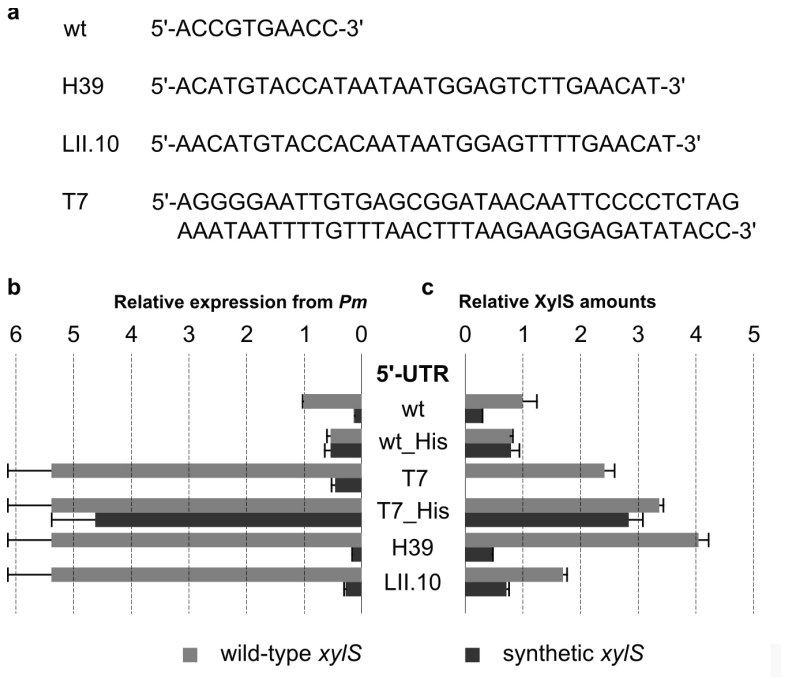

Figure 3.2: The effect of 5′-UTR-exchange on cell-internal XylS levels and expression from *Pm*; a) sequences for the applied 5′-UTRs; b) effect on expression from *Pm*; c) relative XylS expression levels; all values are given as relative values compared to those for pTA13 or pFS7; expression from *Pm* is measured as resistance to ampicillin on plates (1 mM 3-methylbenzoate) and error bars reflect ampicillin concentrations that led to growth inhibition; relative XylS expression levels were determined by luciferase activity measurement from constructs harboring the synthetic operon *xylS-luc* (data from at least two biological replicas); the missing measure point for *synxylS* (with T7-5′-UTR) was not taken.

of the protein in inactive form. Also the exchange of the His-tag by a fragment of the well-expressed *celB*-gene with similar length led to decreased expression from *Pm* and again the explanation might be that XylS was produced in inactive form.

It was striking that expression levels from *Pm* were similar for almost all constructs with changed UTR, and it was decided to clone the synthetic operon from pFS7 into these constructs, for determination of XylS amounts via luciferase activity measurements. Results can be seen in Figure 3.2 c and it is obvious, that XylS levels vary, while these variations are not reflected by the expression levels from *Pm*. This indicates some limitation for expression activation from *Pm* in response to changes in XylS expression.

Expression from *Pm* can be increased by an elevation in *xylS* copy number

A method of expression level improvement alternative to those described above is the elevation of the copy number of the gene to be expressed. To study the effect of an increase in *xylS* copy number on expression from *Pm*, the activator was expressed from a separate plasmid based on the pBBR1-replicon (see 1.2.4), which has about five-fold higher copy number than the RK2 replicon and is compatible with pTA13 / pFS7. The new plasmid was termed pFZ2A and on this plasmid XylS is expressed from its natural *Ps2* promoter. In plasmid pFS7 *xylS* and *luc* were deleted, giving pFS15, which was used as *Pm*-bearing plasmid.

Ampicillin resistance testing on agar plates with cells harboring both pFZ2A and pFS15 revealed that the five-fold increase in *xylS*-copy number led to about four times higher expression from *Pm*. The small discrepancy might be explained by the placement of *xylS* on a separate plasmid and the connected *in trans* activation of *Pm*. Background expression levels were lower for the two-plasmid-system than for pFS7, indicating a decreasing effect deriving from the placement on a separate plasmid. In the absence of inducer cells containing pFZ2A and pFS15 did not tolerate more than 5 μg mL^{-1} ampicillin on plates, which is comparable to the resistance of cells without plasmid. The resistance conferred by pFS7 (in the absence of inducer) is about four times higher, even if the copy number is lower in this plasmid. As described in 1.7 different replicons locate to different positions inside the cell and probably the activation capacity at *Pm* is influenced by the larger spatial distance between XylS expression and action. A possible explanation for this phenomenon might be that the protein will dimerize only occasionally in the absence of inducer and probably dissociate again before contact to *Pm* is established, while the chance of dimerized XylS to bind to *Pm* is much higher, when the protein is expressed from the same plasmid and thus closer to the promoter.

3.2 Drawbacks deriving from high-level production of recombinant proteins

Reduction of the bacterial fitness

As described in 1.2.2 recombinant protein production and the presence of foreign DNA, especially at high copy numbers, impose a metabolic load to the host cells. During the experiments with *celB* as the reporter both copy number and expression levels were heavily increased, compared to the wild type system, which is known to produce recombinant protein at high levels.

This was reflected by the growth behavior of the cells, as can be seen in Paper III, Figure 4. Plasmids with the highest copy number in the test (cop254) retarded cell growth, even if no reporter gene was present and thus no recombinant protein produced. This demonstrates the metabolic load, which the cells have to handle due to the presence and maintenance of the plasmid DNA. During production of phosphoglucomutase, the inhibiting effect on growth was even stronger, as with *trfA* variant cop254 only half the amount of cell mass compared to cells with the wild type plasmid was obtained after eight hours. This is not surprising if one considers the vast amounts of transcript and protein produced in these cells (see 3.1.2).

Also for variant cop271 the metabolic load, imposed on the cells, is reflected by reduced growth in the presence of *celB*, but to a lesser degree (Paper III, Figure 4). In case of the wild type copy number plasmids no significant difference in growth behaviour could be seen in the presence or absence of reporter gene.

During all experiments which were carried out for this thesis it could be observed that high protein production levels and high plasmid copy numbers led to a reduction in cell fitness. The cell mass which was reached, when samples were taken, was in general lower for these plasmids than for others. This indicates that plasmids that confer high production levels per cell do not necessarily have to be the best choice for high-level production of recombinant protein, because of the reduced growth behavior. Cells harboring a plasmid with slightly lower production levels, which in consequence grow better, might be the better choice.

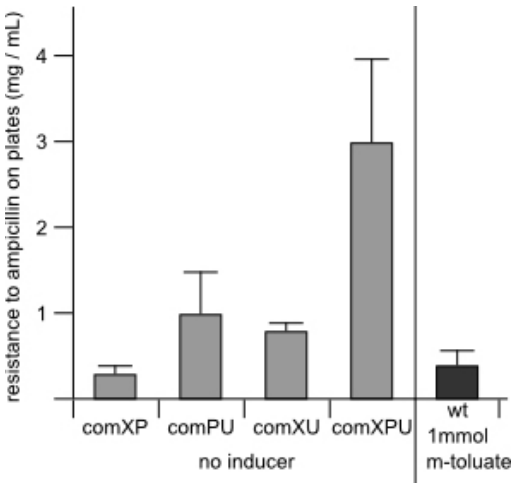

Figure 3.3: Background expression levels for the different combination constructs; values are given as resistance against ampicillin on plates in mg mL^{-1} and error bars indicate the ampicillin concentrations which led to growth inhibition; ampicillin resistance for the wild type plasmid pTA16 in the presence of 1mmol inducer is included as reference.

Increase in background expression levels

As a consequence of the improvements in the expression cassette which were described in the previous sections background expression levels were typically raised, as well. Background resistance levels for the combination plasmids with *bla* as reporter gene are depicted in Figure 3.3. As expected, the highest background expression could be observed for plasmid comXPU and cells containing this vector were able to grow on 3 mg mL^{-1} ampicillin in the absence of inducer. Thus they tolerated about 7.5 times more ampicillin than cells with the wild type plasmid in the presence of 1mmol 3-methylbenzoate (which is included as reference in Figure 3.3) and about 600 times more ampicillin than the wild type plasmid in the absence of inducer. This increased background expression is not problematic when the only goal is to produce as much protein as possible, but it can be challenging for the production of host-toxic proteins, which might explain why expression levels for scFv-phOx generally were low and for comXPU not as high as for comXU (see 3.1.1).

3.3 Reduction of the metabolic load and enhancement of the induction ratio

3.3.1 Chromosomal integration

As described in the previous section, the use of plasmids, especially at high copy numbers is accompanied by certain drawbacks. This led to the question, if the high expression levels, which were obtained by combination of stimulatory variants, would be sufficient to enable expression from a single copy of the expression cassette on the chromosome at levels comparable to the wild type plasmid system. This would cause a reduction of the metabolic load by removal of the plasmid.

To test this the comXPU expression cassette with different reporter genes was integrated into the *E. coli* chromosome by use of conditional-replication, integration and modular plasmids (see 1.1.2). After confirmation of successful integration, expression levels were characterized by qRT-PCR, enzyme activity measurements and/or Western Blotting and compared to levels from the wild type plasmids.

In case of *bla* about eight times more enzyme activity could be achieved from the integrated cassette than from the wild type plasmid (Paper I Figure 3b). It seems obvious that expression levels were not as high as for comXPU in plasmid state as the copy number was reduced from four to eight (plasmid) to one (chromosome). Thus the reduction (compared to comXPU in plasmid state) correlates almost to the reduction in copy number and it is reasonable to believe that the fold-changes would still be as high as in plasmid state if the chromosomally integrated comXPU was compared to a wild type cassette on the chromosome.

For *celB* the enhancement of expression levels in the plasmid system was only around eight times, measured in enzyme activity, and thus lower levels were also expected from the chromosomally integrated expression cassette. Enzyme activities were found to be slightly lower than for the wild type plasmid. As discussed in 3.1.2 the copy number may also here play an important role for the expression levels, which can be reached with *celB*.

For *gm-csf* protein produced from a chromosomally integrated gene copy could

not be detected by Western Blot. As mentioned before, protein produced from the wild type plasmid is not detectable, either. Transcript levels were about five-fold lower from the chromosomal cassette than from the wild type plasmid, and even if transcripts from comXPU are more effectively translated than transcripts from the wild type cassette, detectable protein amounts were unlikely to be expected.

Anyhow, the results described here indicate that relatively high expression levels can be reached by the engineered XylS/*Pm* expression system integrated into the chromosome. Some additional improvement potential may lie in the optimization of growth conditions as it has been done for the T7 system (251). Thus it can be concluded that combination of optimized elements and subsequent integration into the chromosome is a promising approach to reduce gene copy number and to achieve a higher stability of the expression cassette by avoiding the use of plasmids.

3.3.2 Reduction of background expression by 5′-UTR variants

Not only high-expression variants have been identified for the three regulatory elements *xylS*, *Pm* and 5′-UTR, also screenings for variants that lead to reduced expression have previously been conducted. Some of these variants were included into the combinatorial testing. Combination of low-expression variants with each other led generally to low expression levels. The function of down-regulating variants with an influence on transcription (*xylS* or *Pm*) could not be restored to wild type activity when combined with high-expression variants either, however, combination of the low-expression 5′-UTR- variant with either comP or comX led to enzyme activities similar to that of the wild type system and even higher transcript levels, demonstrating that ineffective translation can be reversed, when enough transcript is produced. This demonstrates the potential of the 5′-UTR-DNA region for improvement of the XylS/*Pm* expression cassette via background reduction with simultaneous retention of a functioning system.

In a more detailed study on low-expression 5′-UTR-variants, five such variants, identified by a screening (performed by Rahmi Lale), were chosen to be tested

with different reporter genes (Paper V). They acted similarly on expression of *bla* and *luc* (Paper V Figures 2a and 3) and also when fused to another promoter, the constitutive σ^{70}- promoter P1 (Paper V Figure 2c): they shifted both induced and uninduced expression towards lower levels. The induction ratio was either as high as in the wild type system or increased (up to 260-fold, Paper V Figure 5).

These variants might offer a possibility to reduce the high background levels, which were observed for the combination vectors, but this was not tested. Anyhow, their applicability for metabolic engineering purposes could be demonstrated: variants DI-3 and DI-8 were shown to reduce or completely remove background expression levels for the production of C50 carotenoid sarcinaxanthin from lycopene in a lycopene-producing *E. coli* host (Paper V Figure 7).

The effect of down-regulated translation on transcription and an important role for transcription terminator Rho

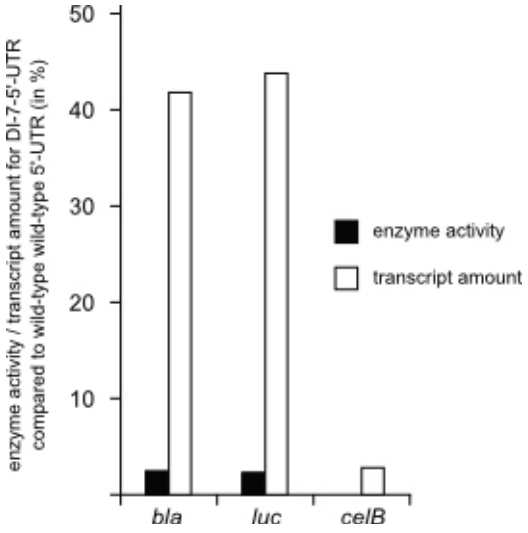

Figure 3.4: Comparison of expression with variant 5′-UTR (DI-7) at transcriptional and translational levels for three different reporter genes; values are given relative to those for expression with wild type 5′-UTR, which are set to 100.

The 5′-UTR is of great importance for the step of translation and thus variants in this region can be expected to mainly influence expression at the protein level (see 1.4). In agreement with that the 5′-UTR-variants described

above down-regulated expression mainly at the translational level, as demonstrated by the comparison of transcript amounts and enzyme activities for *bla* and *luc* in Figure 3.4. The choice of *celB* as reporter gene revealed that down-regulation of translation can also effect transcription. In contrast to the other two reporter genes, no enzyme activity could be detected for cells harboring plasmids with 5′-UTR-variants and transcript amounts were heavily decreased as well (Figure 3.4). This indicated that the reduced translation efficiency, caused by the variant, indirectly led to a reduction of transcript amounts, either by rapid degradation of non-translated transcripts or by early transcriptional termination. The first hypothesis could be disproved and evidence for the second provided by addition of the Rho-inhibitor BCM to cell cultures to a final concentration of 100 μg mL^{-1}. Transcript amounts for samples taken 30 minutes after induction and BCM addition, were determined with the three reporter genes *bla*, *luc*, and *celB*. Transcript formation of all three in the wild type plasmids was up-regulated about 10- to 15-fold in the presence of BCM, which can be explained by expression profile changes of all Rho-regulated proteins inside the cells. Among the Rho-terminated proteins are a wide range of transcrip-

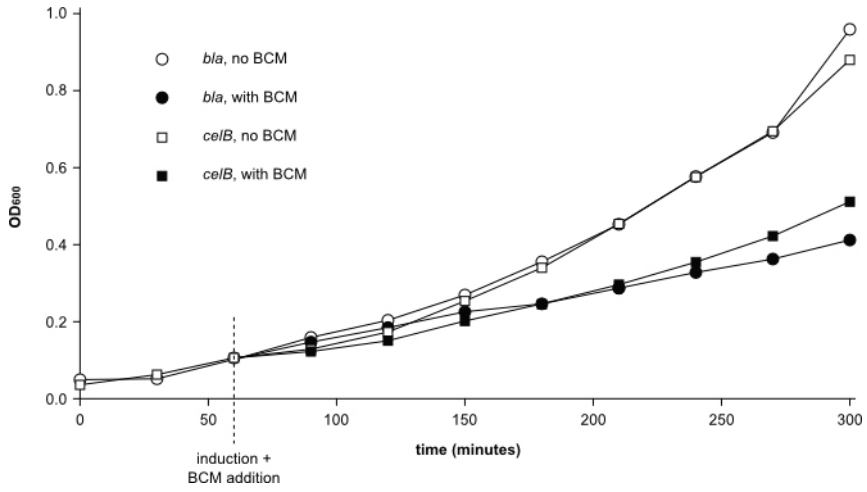

Figure 3.5: Growth curves for cells harboring plasmids with either *bla* or *celB* as reporter gene, with and without addition of BCM at the timepoint of induction.

tion factors (personal communication with Christopher Cardinale, Children's Hospital of Philadelphia), which in turn regulate expression of other proteins. Transcript measurements for *xylS* from cultures grown both in the presence

and absence of BCM revealed that also this transcription factor is regulated by Rho-termination. In addition the effect on expression of one of the two sigma factors, that are responsible for expression from *Pm* (σ^{38}) was studied and there was a negative correlation between expression levels and BCM addition. The complex specter of reactions to BCM addition is also reflected by the fact that cells grow slower in the presence of BCM, as tested for cells with plasmids that contain either *bla* or *celB* as reporter gene (Figure 3.5).

When *bla* was used as reporter gene transcript formation was similarly up-regulated for both wild type and variant 5′-UTR, in case of *celB* the effect of BCM addition on transcript production from the variant plasmid was much higher than for the wild type plasmid (Figure 3.6a). This supports the assump-

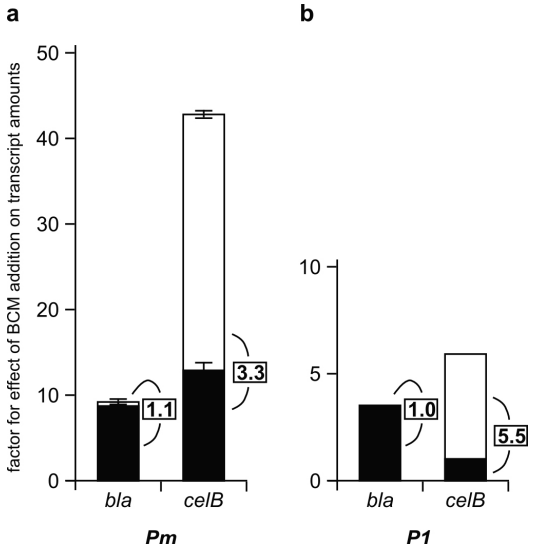

Figure 3.6: Transcription of *celB* is terminated at a gene-internal Rho-site, when expressed from a plasmid with a low-expression 5′-UTR-variant (DI-7); the bars represent the factor between expression levels for cultures grown in the presence and absence of BCM; black: factor for wild type 5′-UTR; white: factor for variant 5′-UTR; numbers in squares give the ratio between the two factors; a) transcription from the *Pm* promoter; b) transcription from the *P1* promoter.

tion that effective translation of *celB*-transcript (as in the case of the wild type 5′-UTR) prohibits transcriptional termination by Rho at an internal Rho-site. A similar effect was observed when the variants with the different reporter genes were fused to promoter P1, by which promoter-specific effects could be ruled out (Figure 3.6b).

As a conclusion it can be said that variation of the DNA-region corresponding to the 5′-UTR can reduce expression levels by down-regulating translation, which also might have an effect on transcription.

3.3.3 Maximizing induction ratio via regulation of XylS expression

As described in 3.1.3 placement of *xylS* on a separate plasmid increased expression from *Pm*, while reducing background expression, and by that an induction ratio of about 500 could be reached. It seemed obvious that there might be additional potential for increasing the induction ratio from *Pm* by regulating XylS expression. For that purpose the synthetic operon *xylS-luc* was placed behind the positively regulated *Pb* promoter on a plasmid similar to pFZ2A. The gene coding for the activator protein of *Pb* (*chnR*) was included as well and the resulting plasmid was named pFZ2B1. pFS15 was as before used as *Pm*-bearing target plasmid for XylS and cells with both plasmids were plated on different concentrations of ampicillin in combination with varying inducer concentrations (both 3-methylbenzoate, inducer for *Pm*, and cyclohexanone, inducer for *Pb*). A higher amount of inducer led in general to higher ampicillin resistance. To directly link expression levels from *Pm* (measured as resistance to ampicillin) to XylS amounts in the cell, luciferase activities were measured at different cyclohexanone concentrations (Paper IV, Figure 3a). Expression from *Pm* was found to be directly proportional to XylS expression, at least at relatively low concentrations of the regulator, but a maximum in ampicillin resistance was reached in the presence of 1 mM 3-methylbenzoate and 1 mM cyclohexanone. When the cyclohexanone concentration was increased to 2 mM no higher ampicillin resistance could be achieved and interestingly the maximum resistance was similar to that with 5′-UTR-variants (see 3.1.3). This suggested the existence of an upper limit for improvement of expression from *Pm* by increases in XylS expression, which could be confirmed by XylS expression from plasmids (pFZ2B2, pFZ2B3), in which expression of the activator for *Pb*, ChnR, is increased. For these constructs luciferase activities (and thus XylS expression levels) were further increased, up to 240-fold compared to pFZ2B1 in the absence of inducer.

In contrast, the cells did not tolerate higher ampicillin concentrations than observed before (Paper IV, Figure 3b).

When the variant StEP-13 (comX, see 3.1.1) was used instead of wild type *xylS*, the observed maximum was reached at lower concentrations of cyclohexanone (Paper IV, Figure 3a). This was not due to increased expression of XylS, as luciferase amounts turned out to be similar at similar conditions for wild type *xylS* and StEP-13. Also background expression was increased for plasmids with StEP-13 compared to the wild type system, so that the higher expression from *Pm* with StEP-13 at low XylS concentrations cannot be caused by better inducer binding only. A more probable explanation is that this variant acts via better dimerization and/or interaction with RNAP.

To test if the observed limit was caused by occupancy of all available *Pm* sites, the copy number of the target plasmid pFS15 was increased by exchange of the *trfA*-gene by its variant cop271. Even if ampicillin resistance of the corresponding cells was slightly increased, this did not correlate to the increase in copy number (Paper IV, Figure 3a). More importantly, a roof was reached at similar XylS expression levels as with the wild type copy number, both with wild type *xylS* and StEP-13, indicating that availibily of free *Pm*-sites was not expression-limiting, which is in agreement with previously published studies (129).

Interestingly, a stagnation in resistance against ampicillin could not be observed for background expression at XylS levels which led to saturation of the system under induced conditions (Paper IV, Figure 3b). Since cells react less to cyclohexanone-addition in the absence of 3-methylbenzoate than in its presence, induction ratios increase at low cyclohexanone concentrations. At XylS levels exceeding the saturating concentration for the induced system induction ratios decrease again, leading to a maximum ratio of 700 at the point, where the induced activity reaches its maximum. It has been described previously that *Pm* looses its inducibility at high XylS concentrations (129, 166), and indeed, when XylS was expressed from the T7 promoter in the presence of ITPG (in a pET16b-plasmid with a tetracycline resistance marker, which was combined with pFS15), no difference in ampicillin resistance could be observed between cells which were grown in the presence or absence of 3-methylbenzoate. This indicates that uninduced activity does not level off before the same roof level is

reached as under induced conditions. For XylS expressed from the T7 promoter a strong band was visible in the insoluble fraction on an SDS-PAGE gel (Paper IV, Figure 6).

A model was established by which most of the observed results can be explained (Paper IV, Figure 5): XylS monomers dominate in the absence of inducer (227). When inducer is added to the cells, this will lead to increased dimerization and dimers can be considered as active, resulting in activation of transcription from *Pm*. When a certain concentration of XylS is reached inside the cells, the protein will start to oligomerize and by that become inactive again. Thus increased production of XylS will lead to higher amounts of inactive protein, while the amount of active dimers will stay constant, explaining the observed maximum expression level. StEP-13 acts probably via better dimerization and will thus reach the dimer concentration at which further oligomerization starts at lower total XylS concentrations than the wild type protein. The fact that the variant reacts more to inducer addition than wild type *xylS* demonstrates that the better dimerization probably partly is caused by better inducer binding (Paper IV, Figure 4).

In the absence of inducer the amount of active XylS dimers will increase proportionally to total XylS amounts, while the percentage of dimers is lower than under induced conditions. Thus oligomerization will start at much higher total concentrations and not before the same amount of active dimers is reached as under induced conditions when expression levels out.

The described model fails to explain the high amount of insoluble protein, which is produced from the T7-system in the presence of IPTG (Paper IV, Figure 6). Explanations might be that the absence of a *Pm*-bearing plasmid during this experiment had an influence on the state of XylS or that there is no correlation between oligomerization and insolubility of the protein. It is also important to note that both the backbone of the *xylS*-expressing plasmid and the host were changed when *xylS* was expressed from the T7 promoter, which might complicate direct comparability of the results to those obtained with the other plasmids.

3.3.4 Reduction of background expression in high-level expression vectors

The combination vectors that are described in 3.1.1 are useful tools for the achievement of high expression levels in *E. coli*, but their high background levels disqualify them for expression of host-toxic proteins. XylS influences background expression to a high degree as can be seen from Paper I, Table 1. With the knowledge from 3.3.3 it was argued that regulation of XylS expression might allow a better control of background expression also in the combination vectors. Thus a plasmid (pFZ10) was constructed based on comPU, in which, similarly as in pFS15, *xylS* was deleted. pFZ10 was then combined with either pFZ2B1 (wild-type *xylS* behind the inducible promoter *Pb*) or pFZ2B1.StEP-13 (variant StEP-13 behind *Pb*). The first combination equates to comPU with regulated XylS expression, the second one to comXPU. Background resistance against ampicillin on plates was reduced in both cases, when the two-plasmid version, in which XylS expression is regulated, was compared to the version with only one plasmid (Figure 3.7). For comPU induced expression was slightly increased

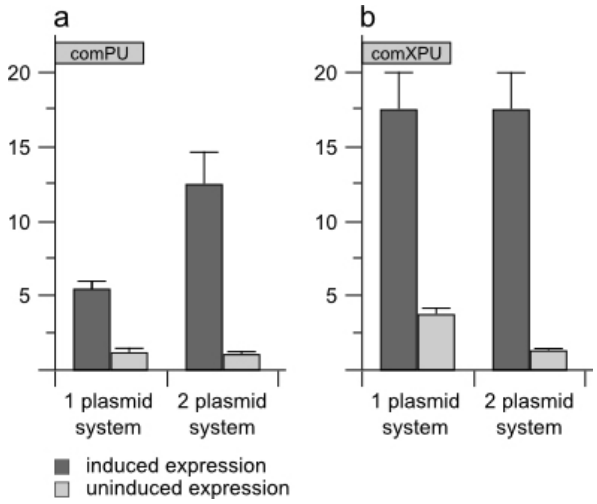

Figure 3.7: Reduction of background expression in combination vectors by XylS regulation; a) comPU; b) comXPU; expression levels under induced (0.1mmol 3-methylbenzoate, 1mmol cyclohexanone for two-plasmid systems) and uninduced conditions (no 3-methylbenzoate, no cyclohexanone) were measured as resistance against ampicillin on plates and are given in mg mL^{-1}; error bars indicate the ampicillin concentrations which led to growth inhibition.

for the two-plasmid system, which can be explained by the higher level of XylS

expression, while for comXPU induced expression was similarly high both with one or two plasmids. This seems resonable, as the roof level for expression enhancement from *Pm* is reached at lower levels of *xylS* expression with the variant comX. For the same reason, one would maybe expect to reach similar expression levels with the comPU two-plasmid system as with the comXPU two-plasmid system, however, in the experiments described here, only 0.1 mM 3-methylbenzoate were used to induce expression from *Pm*, to allow comparability to the results in 3.1.1. Thus with the used inducer concentrations the roof level was probably not reached with wild type *xylS* (in comPU), yet.

Comparison of resistance levels on plates can be inaccurate at high ampicillin concentrations. Therefore the comXPU-constructs were also compared by β-lactamase-assay, both under induced and uninduced conditions. Again the background level was reduced, when the two-plasmid system was used, while expression levels under induced conditions were similar for both constructs (Figure 3.8). To test if it would be possible to further increase induced expression

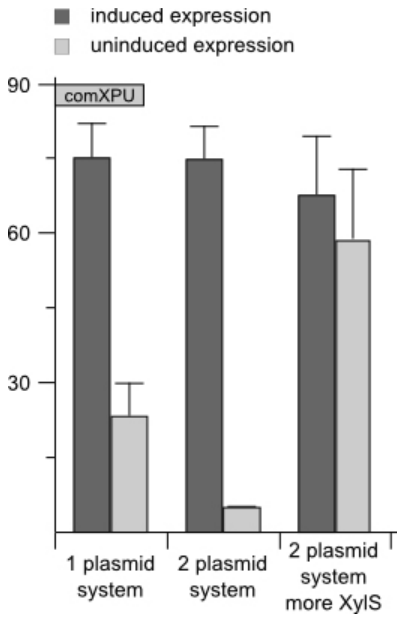

Figure 3.8: Reduction of background expression in combination vectors by XylS regulation at the enzyme level; expression levels under induced (0.1mmol 3-methylbenzoate, 1mmol cyclohexanone for two-plasmid systems) and uninduced conditions (no 3-methylbenzoate, no cyclohexanone) were measured as β-lactamase-activity from at least two biological replicas.

of comXPU, plasmid pFZ2B3.StEP-13, from which more XylS is produced than from pFZ2B1.StEP-13, was combined with pFZ10 and tested for β-lactamase-activity. Even if the background was heavily increased for this new construct, no higher induced expression than with the other two constructs could be achieved. Besides oligomerization of XylS at high concentrations, possible explanations for the observed limit include a cell-dependent limitation for active β-lactamase production at higher levels, translational limitations caused by high mRNA-amounts as described in 3.1.2 or a combination of these effects.

Based on the fact that the same maximum level of expression from *Pm* could be reached with wild type XylS as with StEP-13, it should be possible to reach expression levels similar to those of comXPU with the comPU two-plasmid system (wild type *xylS*), in the presence of high inducer concentrations. This would allow a further reduction of background levels for the combination system.

Even if background expression could be reduced by regulation of XylS expression, it was still high compared to the wild type system. Interestingly, cells containing pFZ10 only (no XylS inside the cells), are able to grow on more than 3 mg mL^{-1} ampicillin in the presence of 3-methylbenzoate, which demonstrates that activation of transcription has become highly independent of XylS by the combination of variant promoter and variant 5′-UTR.

4

Concluding remarks

One central conclusion from the results described in this thesis is that it is possible to get an additional effect from the combination of variants with known phenotype. This can be of high importance for the use of synthetic elements as 5′-UTR sequences suggested by the RBS-calculator (229) or the use of codon-optimized genes. The high expression levels, which could be reached by this approach, allow to reduce the metabolic load via copy number reduction. On the other hand a combination with variants that individually lead to decreased expression might help to reduce background expression.

However, combination of elements did not work equally good in all cases. The results obtained from combination of *xylS* and promoter variants indicate that an additional effect only can be obtained when the variants individually influence different steps of expression. Also the expression of other reporter genes from the combination plasmids did not lead to the same high levels as for *bla*, which probably was caused by the gene-dependency of at least some of the variants. It seems obvious that especially the 5′-UTR variants might act more gene-dependent than the other variants, and thus it might be a better strategy to screen for a gene-specific 5′-UTR variant, which then can be used in combination vectors.

Both the results from combination of high-expression variants and from mutation of the 5′-UTR demonstrate that there often is a correlation between expression levels under induced and uninduced conditions, making it a challenge to increase the induction ratio. For that regulation of XylS expression turned out to be a useful approach. At the same time the better understanding of the

activation mechanism at *Pm* by XylS allows to optimize expression conditions and as a consequence to minimize the metabolic load. For example it seems reasonable to express XylS at levels that do not exceed the saturating concentration, and when XylS expression is regulated, both use of the XylS-variant StEP-13 and copy number variation do not increase expression levels significantly. Nevertheless, increases in plasmid copy number can be meaningful as the expression studies with *celB* demonstrate.

The limitation of activation at *Pm* by the concentration of active XylS indicates that it might be a promising approach to search for variants of the activator, which can form higher concentrations of active dimers. These could also be promising candidates for XylS structure determination.

In general the experiments demonstrate that there still is not one optimal way to express any protein of interest, and that gene-specifities and the context have to be taken into account. Expression of host toxic proteins like the antibody fragment ScFv-phOx might require completely different conditions than expression of well expressed genes like *celB*.

However, the XylS/*Pm* expression system offers a large tool-kit and enables expression under completely different conditions, as the studies presented here demonstrate. A combination with other features as for example mutated signal sequences or optimized genes seems fully possible and might further increase expression levels. The better the mechanisms which limit expression of a certain protein are understood, the easier it will be to pick the right conditions. In any case the existence of a broad range of expression vectors with a similar design, but different characteristics (as existing for the XylS/*Pm* system) might enable quick identification of suitable settings.

During the work for this thesis, the broad-host range feature of the mini-RK2 replicon has not been considered, as all experiments were carried out in *E. coli*. Exploration of the findings described here for their transferability to other hosts might lead to even further possibilities to adapt expression conditions to the protein to be expressed.

References

1. M. R. Adhin and J. van Duin. Scanning model for translational reinitiation in eubacteria. *J Mol Biol*, 213:811–818, 1990.

2. C. E. Aitken, A. Petrov, and J. D. Puglisi. Single ribosome dynamics and the mechanism of translation. *Annu Rev Biophys*, 39:491–513, 2010.

3. B. Alberts. DNA replication and recombination. *Nature*, 421:431–435, 2003.

4. M. Allert, J. C. Cox, and H. W. Hellinga. Multifactorial determinants of protein expression in prokaryotic open reading frames. *J Mol Biol*, 402: 905–918, 2010.

5. S. G. Andersson and C. G. Kurland. Codon preferences in free-living microorganisms. *Microbiol Rev*, 54(2):198–210, 1990.

6. A. Andre, A. Puca, F. Sansone, A. Brandi, G. Antico, and R. A. Calogero. Reinitiation of protein synthesis in *Escherichia coli* can be induced by mRNA *cis*-elements unrelated to canonical translation initiation signals. *FEBS Lett*, 468:73–78, 2000.

7. E. Angov. Codon usage: Nature's roadmap to expression and folding of proteins. *Biotechnol J*, 6:650–659, 2011.

8. R. Antoine and C. Locht. Isolation and molecular characterization of a novel broad-host-range plasmid from *Bordetella bronchiseptica* with sequence similarities to plasmids from Gram-positive organisms. *Mol Microbiol*, 6:1785–1799, 1992.

9. I. Artsimovitch. Post-initiation control by the initiation factor sigma. *Mol Microbiol*, 68:1–3, 2008.

10. J. E. Bailey, N. A. Da Silva, S. W. Peretti, J. H. Seo, and F. Srienc. Studies of host-plasmid interactions in recombinant microorganisms. *Ann N Y Acad Sci*, 469:194–211, 1986.

11. I. Bakke, L. Berg, T. E. Aune, T. Brautaset, H. Sletta, A. Tondervik, and S. Valla. Random mutagenesis of the *Pm* promoter as a powerful strategy for improvement of recombinant-gene expression. *Appl Environ Microbiol*, 75:2002–2011, 2009.

12. K. Bakkevig, H. Sletta, M. Gimmestad, R. Aune, H. Ertesvag, K. Degnes, B. E. Christensen, T. E. Ellingsen, and S. Valla. Role of the *Pseudomonas fluorescens* alginate lyase (AlgL) in clearing the periplasm of alginates not exported to the extracellular environment. *J Bacteriol*, 187:8375–8384, 2005.

13. P. Balbas and G. Gosset. Chromosomal editing in *Escherichia coli*. Vectors for DNA integration and excision. *Mol Biotechnol*, 19:1–12, 2001.

14. P. Balbas, X. Soberon, E. Merino, M. Zurita, H. Lomeli, F. Valle, N. Flores, and F. Bolivar. Plasmid vector pBR322 and its special-purpose derivatives–a review. *Gene*, 50:3–40, 1986.

15. S. Balzer, V. Kucharova, J. Megerle, R. Lale, and S. Valla. A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in *Escherichia coli*. *Microb Cell Fact*, 12: 26–39, 2013.

16. S. Banerjee, J. Chalissery, I. Bandey, and R. Sen. Rho-dependent transcription termination: more questions than answers. *J Microbiol*, pages 11–22, 2006.

17. G. Bar-Nahum and E. Nudler. Isolation and characterization of σ^{70}-retaining transcription elongation complexes from *Escherichia coli*. *Cell*, 106:443–451, 2001.

18. K. A. Barne, J. A. Bown, S. J. Busby, and S. D. Minchin. 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, 1997.

19. S. J. Benkovic, A. M. Valentine, and F. Salinas. Replisome-mediated DNA replication. *Annu Rev Biochem*, 70:181–208, 2001.

20. L. Berg, R. Lale, I. Bakke, N. Burroughs, and S. Valla. 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, 2009.

21. L. Berg, V. Kucharova, I. Bakke, S. Valla, and T. Brautaset. 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, 2012.

22. L. E. Bingle and C. M. Thomas. Regulatory circuits for plasmid survival. *Curr Opin Microbiol*, 4:194–200, 2001.

23. R. E. Bird, K. D. Hardman, J. W. Jacobson, S. Johnson, B. M. Kaufman, S. M. Lee, T. Lee, S. H. Pope, G. S. Riordan, and M. Whitlow. Single-chain antigen-binding proteins. *Science*, 241:423–426, 1988.

24. J. M. Blatny, T. Brautaset, H. C. Winther-Larsen, K. Haugan, and S. Valla. 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, 1997.

25. J. M. Blatny, T. Brautaset, H. C. Winther-Larsen, P. Karunakaran, and S. Valla. Improved broad-host-range rk2 vectors useful for high and low regulated gene expression levels in gram-negative bacteria. *Plasmid*, 38(1): 35–51, 1997.

26. F. R. Blattner, G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. The complete genome sequence of *Escherichia coli* K-12. *Science*, 277:1453–1462, 1997.

27. S. Borukhov and E. Nudler. RNA polymerase: the vehicle of transcription. *Trends Microbiol*, 16:126–134, 2008.

28. T. Brautaset, R. Standal, E. Fjaervik, and S. Valla. Nucleotide sequence and expression analysis of the *Acetobacter xylinum* phosphoglucomutase gene. *Microbiology*, 140:1183–1188, 1994.

29. T. Brautaset, S. Petersen, and S. Valla. An experimental study on carbon flow in *Escherichia coli* as a function of kinetic properties and expression levels of the enzyme phosphoglucomutase. *Biotechnol Bioeng*, 58:299–302, 1998.

30. T. Brautaset, S. B. Petersen, and S. Valla. *In vitro* determined kinetic properties of mutant phosphoglucomutases and their effects on sugar catabolism in *Escherichia coli*. *Metab Eng*, 2:104–114, 2000.

31. T. Brautaset, R. Lale, and S. Valla. Positively regulated bacterial expression systems. *Microb Biotechnol*, 2:15–30, 2009.

32. J. Brosius, C. R. L., and A. P. Perlmutter. Precise location of two promoters for the β-lactamase gene of pBR322. S1 mapping of ribonucleic acid isolated from *Escherichia coli* or synthesized *in vitro*. *J Biol Chem*, 257: 9205–9210, 1982.

33. D. F. Browning and S. J. Busby. The regulation of bacterial transcription initiation. *Nat Rev Microbiol*, 2:57–65, 2004.

34. B. M. Burmann, K. Schweimer, X. Luo, M. C. Wahl, B. L. Stitt, M. E. Gottesman, and P. Rosch. A NusE:NusG complex links transcription and translation. *Science*, 328:501–504, 2010.

35. E. Burova, S. C. Hung, V. Sagitov, B. L. Stitt, and M. E. Gottesman. *Escherichia coli* NusG protein stimulates transcription elongation rates in vivo and in vitro. *J Bacteriol*, 177:1388–1392, 1995.

36. J. E. Cabrera and D. J. Jin. The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. *Mol Microbiol*, 50:1493–1505, 2003.

37. G. Cannarrozzi, N. N. Schraudolph, M. Faty, P. von Rohr, M. T. Friberg, A. C. Roth, P. Gonnet, G. Gonnet, and Y. Barral. A role for codon order in translation dynamics. *Cell*, 141:355–367, 2010.

38. C. J. Cardinale, R. S. Washburn, V. R. Tadigotla, L. M. Brown, M. E. Gottesman, and E. Nudler. Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science*, 320:935–938, 2008.

39. A. P. Carter, W. M. Clemons, D. E. Brodersen, R. J. Morgan-Warren, T. Hartsch, B. T. Wimberly, and V. Ramakrishnan. Crystal structure of an initiation factor bound to the 30s ribosomal subunit. *Science*, 291: 498–501, 2001.

40. A. Cebolla, C. Sousa, and V. de Lorenzo. Rational design of a bacterial transcriptional cascade for amplifying gene expression capacity. *Nucleic Acids Res*, 29:759–766, 2001.

41. G. Cesareni, M. Cornelissen, R. M. Lacatena, and L. Castagnoli. Control of pMB1 replication: inhibition of primer formation by Rop requires RNA1. *EMBO J*, 3:1365–1369, 1984.

42. G. M. Cheetham and T. A. Steitz. Structure of a transcribing T7 RNA polymerase initiation complex. *Science*, 286:2305–2309, 1999.

43. J. H. Choi and S. Y. Lee. Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl Microbiol Biotechnol*, 64: 625–635, 2004.

44. F. L. Cohen and D. Tartasky. Microbial resistance to drug therapy: a review. *Am J Infect Control*, 25:51–64, 1997.

45. S. N. Cohen and A. C. Chang. Recircularization and autonomous replication of a sheared R-factor DNA segment in *Escherichia coli* transformants. *Proc Natl Acad Sci U S A*, 70:1293–1297, 1973.

46. P. Cortazzo, C. Cerveansky, M. Martn, C. Reiss, R. Ehrlich, and A. Deana. Silent mutations affect in vivo protein folding in *Escherichia coli*. *Biochem Biophys Res Commun*, 293:537–541, 2002.

47. L. Covarrubias, L. Cervantes, A. Covarrubias, X. Soberon, I. Vichido, A. Blanco, Y. M. Kupersztoch-Portnoy, and F. Bolivar. Construction and characterization of new cloning vehicles. V. Mobilization and coding properties of pBR322 and several deletion derivatives including pBR327 and pBR328. *Gene*, 13:25–35, 1981.

48. R. Crea, A. Kraszewski, T. Hirose, and K. Itakura. Chemical synthesis of genes for human insulin. *Proc Natl Acad Sci U S A*, 75:5765–5769, 1978.

49. C. Dahlberg and L. Chao. Amelioration of the cost of conjugative plasmid carriage in *Eschericha coli* K12. *Genetics*, 165:1641–1649, 2003.

50. P. L. de Haseth, M. L. Zupancic, and J. Record, M. T. RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J Bacteriol*, 180:3019–3025, 1998.

51. A. de Marco, E. Deuerling, A. Mogk, T. Tomoyasu, and B. Bukau. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol*, 7:32–40, 2007.

52. M. H. de Smit and J. van Duin. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc Natl Acad Sci U S A*, 87:7668–7672, 1990.

53. R. de Sousa Abreu, L. O. Penalva, E. M. Marcotte, and C. Vogel. Global signatures of protein and mRNA expression levels. *Mol BioSyst*, 5:1512–1526, 2009.

54. W. de Souza. Prokaryotic cells: structural organisation of the cytoskeleton and organelles. *Mem Inst Oswaldo Cruz*, 107:283–293, 2012.

55. J. R. de Wet, K. V. Wood, D. R. Helinski, and M. DeLuca. Cloning of firefly luciferase cDNA and the expression of active luciferase in *Escherichia coli*. *Proc Natl Acad Sci U S A*, 82:7870–7873, 1985.

56. G. del Solar and M. Espinosa. Plasmid copy number control: an ever-growing story. *Mol Microbiol*, 37:492–500, 2000.

57. G. del Solar, R. Giraldo, M. J. Ruiz-Echevarria, M. Espinosa, and R. Diaz-Orejas. Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev*, 62:434–464, 1998.

58. A. L. Demain and P. Vaishnav. Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv*, 27:297–306, 2009.

59. A. I. Derman, G. Lim-Fong, and J. Pogliano. Intracellular mobility of plasmid DNA is limited by the ParA family of partitioning systems. *Mol Microbiol*, 67:935–946, 2008.

60. J. C. Diaz Ricci and M. E. Hernandez. Plasmid effects on *Escherichia coli* metabolism. *Crit Rev Biotechnol*, 20:79–108, 2000.

61. P. Domínguez-Cuevas, P. Marin, J. L. Ramos, and S. Marqués. RNA polymerase holoenzymes can share a single transcription start site for the Pm promoter. Critical nucleotides in the -7 to -18 region are needed to select between RNA polymerase with σ^{38} or σ^{32}. *J Biol Chem*, 280:41315–41323, 2005.

62. P. Domínguez-Cuevas, P. Marin, S. Busby, J. L. Ramos, and S. Marqués. Roles of effectors in XylS-dependent transcription activation: intramolecular domain derepression and DNA binding. *J Bacteriol*, 190:3118–3128, 2008.

63. P. Domínguez-Cuevas, P. Marin, S. Marqués, and J. L. Ramos. XylS-Pm promoter interactions through two helix-turn-helix motifs: identifying XylS residues important for DNA binding and activation. *J Mol Biol*, 375: 59–69, 2008.

64. P. Domínguez-Cuevas, J. L. Ramos, and S. Marqués. Sequential XylS-CTD binding to the Pm promoter induces DNA bending prior to activation. *J Bacteriol*, 192:2682–2690, 2010.

65. M. Dreyfus. What constitutes the signal for the initiation of protein synthesis on *Escherichia coli* mRNAs? *J Mol Biol*, 204:79–94, 1988.

66. J. W. Dubendorff and F. W. Studier. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with *lac* repressor. *J Mol Biol*, 219:45–59, 1991.

67. R. H. Durland and D. R. Helinski. Replication of the broad-host-range plasmid RK2: direct measurement of intracellular concentrations of the essential TrfA replication proteins and their effect on plasmid copy number. *J Bacteriol*, 172:3849–3858, 1990.

68. R. H. Durland, A. Toukdarian, F. Fang, and D. R. Helinski. Mutations in the *trfA* replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy numbers. *J Bacteriol*, 172:3859–3867, 1990.

69. K. J. Durniak, S. Bailey, and T. A. Steitz. The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science*, 322:553–557, 2008.

70. G. Ebersbach and K. Gerdes. Plasmid segregation mechanisms. *Annu Rev Genet*, 39:453–479, 2005.

71. E. Englesberg, J. Irr, J. Power, and N. Lee. Positive control of enzyme synthesis by gene C in the L-arabinose system. *J Bacteriol*, 90:946–957, 1965.

72. V. Epshtein and E. Nudler. Cooperation between RNA polymerase molecules in transcription elongation. *Science*, 300:801–805, 2003.

73. V. Epshtein, C. J. Cardinale, A. E. Ruckenstein, S. Borukhov, and E. Nudler. An allosteric path to transcription termination. *Mol Cell*, 28:991–1001, 2007.

74. D. Esposito and D. K. Chatterjee. Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotechnol*, 17:353–358, 2006.

75. N. Ferrer-Miralles, J. Domingo-Espín, J. Corchero, E. Vázquez, and A. Villaverde. Microbial factories for recombinant pharmaceuticals. *Microb Cell Fact*, 8:17–24, 2009.

76. E. Fjaervik, K. Frydenlund, S. Valla, Y. Huggirat, and M. Benziman. Complementation of cellulose-negative mutants of *Acetobacter xylinum* by the cloned structural gene for phosphoglucomutase. *FEMS Microbiol Lett*, 77:325–330, 1991.

77. D. V. Freistroffer, M. Y. Pavlov, J. MacDougall, R. H. Buckingham, and M. Ehrenberg. Release factor RF3 in *E.coli* accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *EMBO J*, 16:4126–4133, 1997.

78. D. N. Frick and C. C. Richardson. DNA primases. *Annu Rev Biochem*, 70:39–80, 2001.

79. M. T. Gallegos, S. Marqués, and J. L. Ramos. Expression of the TOL plasmid *xylS* gene in *Pseudomonas putida* occurs from a σ^{70}-dependent promoter or from σ^{70}- and σ^{54}-dependent tandem promoters according to the compound used for growth. *J Bacteriol*, 178:2356–2361, 1996.

80. M. T. Gallegos, S. Marqués, and J. L. Ramos. The $TACAN_4TGCA$ motif upstream from the -35 region in the σ^{70}-σ^{S}-dependent Pm promoter of the TOL plasmid is the minimum DNA segment required for transcription stimulation by XylS regulators. *J Bacteriol*, 178:6427–6434, 1996.

81. N. Galtier and J. R. Lobry. Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J Mol Evol*, 44:632–636, 1997.

82. G. Georgiou and L. Segatori. Preparative expression of secreted proteins in bacteria: status report and future prospects. *Curr Opin Biotechnol*, 16: 538–545, 2005.

83. K. Gerdes, P. B. Rasmussen, and S. Molin. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proc Natl Acad Sci U S A*, 83:3116–3120, 1986.

84. M. Gimmestad, H. Sletta, H. Ertesvåg, K. Bakkevig, S. Jain, S. J. Suh, G. Skjåk-Bræk, T. E. Ellingsen, D. E. Ohman, and S. Valla. The *Pseudomonas fluorescens* AlgG protein, but not its mannuronan C-5-epimerase activity, is needed for alginate polymer formation. *J Bacteriol*, 185:3515–3523, 2003.

85. H. Gingold and Y. Pilpel. Determinants of translation efficiency and accuracy. *Mol Syst Biol*, 7:481–493, 2011.

86. B. R. Glick. Metabolic load and heterologous gene expression. *Biotechnol Adv*, 13:247–261, 1995.

87. I. Golding and E. C. Cox. RNA dynamics in live *Escherichia coli* cells. *Proc Natl Acad Sci U S A*, 101:11310–11315, 2004.

88. S. R. Goldman, R. H. Ebright, and B. E. Nickels. Direct detection of abortive RNA transcripts in vivo. *Science*, 324:927–928, 2009.

89. M. M. González-Pérez, J. L. Ramos, M. T. Gallegos, and S. Marqués. Critical nucleotides in the upstream region of the XylS-dependent TOL meta-cleavage pathway operon promoter as deduced from analysis of mutants. *J Biol Chem*, 274:2286–2290, 1999.

90. M. M. González-Pérez, S. Marqués, P. Dominguez-Cuevas, and J. L. Ramos. XylS activator and RNA polymerase binding sites at the Pm promoter overlap. *FEBS Lett*, 519:117–122, 2002.

91. M. M. González-Pérez, J. L. Ramos, and S. Marqués. Cellular XylS levels are a function of transcription of *xylS* from two independent promoters and the differential efficiency of translation of the two mRNAs. *J Bacteriol*, 186:1898–1901, 2004.

92. M. Gouy and C. Gautier. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res*, 10:7055–7074, 1982.

93. S. Govindarajan, K. Nevo-Dinur, and O. Amster-Choder. Compartmentalization and spatio-temporal organization of macromolecules in bacteria. *FEMS Microbiol Rev*, 36:1005–1022, 2012.

94. J. D. Gralla. Activation and repression of *E. coli* promoters. *Curr Opin Genet Dev*, 6:526–530, 1996.

95. P. L. Graumann. Cytoskeletal elements in bacteria. *Annu Rev Microbiol*, 61:589–618, 2007.

96. S. Graupner and W. Wackernagel. A broad-host-range expression vector series including a P*tac* test plasmid and its application in the expression of the *dod* gene of *Serratia marcescens* (coding for ribulose-5-phosphate 3-epimerase) in *Pseudomonas stutzeri*. *Biomol Eng*, 17:11–16, 2000.

97. A. A. Green and W. D. McElroy. Crystalline firefly luciferase. *Biochim Biophys Acta*, 20:170–176, 1956.

98. S. J. Greive and P. H. von Hippel. Thinking quantitatively about transcriptional regulation. *Nat Rev Mol Cell Biol*, 6:221–232, 2005.

99. K. E. Griswold, N. A. Mahmood, B. L. Iverson, and G. Georgiou. Effects of codon usage versus putative 5′-mRNA structure on the expression of *Fusarium solani* cutinase in the *Escherichia coli* cytoplasm. *Protein Expression Purif*, 27:134–142, 2003.

100. C. A. Gross, C. Chan, A. Dombroski, T. Gruber, M. Sharp, J. Tupy, and B. Young. The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb Symp Quant Biol*, 63:141–155, 1998.

101. A. Grote, K. Hiller, M. Scheer, R. Munch, B. Nortemann, D. C. Hempel, and D. Jahn. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res*, 33:526–531, 2005.

102. T. M. Gruber and C. A. Gross. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol*, 57:441–466, 2003.

103. W. Gu, T. Zhou, and C. O. Wilke. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput Biol*, 6, 2010.

104. C. O. Gualerzi, A. Fabbretti, L. Brandi, P. Milon, and C. L. Pon. Role of the initiation factors in mRNA start site selection and fMet-tRNA recruitment by bacterial ribosomes. *Isr J Chem*, 50:80–94, 2010.

105. I. Gusarov and E. Nudler. The mechanism of intrinsic transcription termination. *Mol Cell*, 3:495–504, 1999.

106. C. Gustafsson, S. Govindarajan, and J. Minshull. Codon bias and heterologous protein expression. *Trends Biotechnol*, 22:346–353, 2004.

107. C. Gustafsson, J. Minshull, S. Govindarajan, J. Ness, A. Villalobos, and M. Welch. Engineering genes for predictable protein expression. *Protein Expr Purif*, 83:37–46, 2012.

108. J. Hacker and G. Blum-Oehler. In appreciation of Theodor Escherich. *Nat Rev Microbiol*, 5:902, 2007.

109. A. Haldimann and B. L. Wanner. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol*, 183:6384–6393, 2001.

110. S. Harayama and M. Rekik. The *meta* cleavage operon of TOL degradative plasmid pWW0 comprises 13 genes. *Mol Gen Genet*, 221:113–120, 1990.

111. S. P. Haugen, W. Ross, and R. L. Gourse. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat Rev Microbiol*, 6:507–519, 2008.

112. R. Hershberg and D. A. Petrov. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet*, 6, 2010.

113. T. Q. Ho, Z. Zhong, S. Aung, and J. Pogliano. Compatible bacterial plasmids are targeted to independent cellular locations in *Escherichia coli*. *EMBO J*, 21:1864–1872, 2002.

114. F. Hoffmann and U. Rinas. Stress induced by recombinant protein production in *Escherichia coli*. *Adv Biochem Eng/Biotechnol*, 89:73–92, 2004.

115. L. M. Hsu. Promoter clearance and escape in prokaryotes. *Biochim Biophys Acta*, 1577:191–207, 2002.

116. C. J. Huang, H. Lin, and X. Yang. Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *J Ind Microbiol Biotechnol*, 39:383–399, 2012.

117. T. Ikemura. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*, 2:13–34, 1985.

118. S. Inouye, A. Nakazawa, and T. Nakazawa. Nucleotide sequence of the regulatory gene *xylS* on the *Pseudomonas putida* TOL plasmid and identification of the protein product. *Gene*, 44:235–242, 1986.

119. H. Iwaki, Y. Hasegawa, M. Teraoka, T. Tokuyama, H. Bergeron, and P. C. Lau. Identification of a transcriptional activator (ChnR) and a 6-oxohexanoate dehydrogenase (ChnE) in the cyclohexanol catabolic pathway in *Acinetobacter sp.* Strain NCIMB 9871 and localization of the genes that encode them. *Appl Environ Microbiol*, 65:5158–5162, 1999.

120. F. Jacob and S. Brenner. [On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon]. *C R Hebd Seances Acad Sci*, 256:298–300, 1963.

121. N. Jacques and M. Dreyfus. Translation initiation in *Escherichia coli*: old and new questions. *Mol Microbiol*, 4:1063–1067, 1990.

122. S. Jana and J. K. Deb. Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Appl Microbiol Biotechnol*, 67:289–298, 2005.

123. D. J. Jin, R. R. Burgess, J. P. Richardson, and C. A. Gross. Termination efficiency at rho-dependent terminators depends on kinetic coupling between RNA polymerase and rho. *Proc Natl Acad Sci U S A*, 89:1453–1457, 1992.

124. H. Kabata, O. Kurosawa, I. Arai, M. Washizu, S. A. Margarson, R. E. Glass, and N. Shimamoto. Visualization of single molecules of RNA polymerase sliding along DNA. *Science*, 262:1561–1563, 1993.

125. C. A. Kaiser, D. Preuss, P. Grisafi, and D. Botstein. Many random sequences functionally replace the secretion signal sequence of yeast invertase. *Science*, 235:312–317, 1987.

126. N. Kaldalu, T. Mandel, and M. Ustav. TOL plasmid transcription factor XylS binds specifically to the *Pm* operator sequence. *Mol Microbiol*, 20: 569–579, 1996.

127. N. Kaldalu, U. Toots, V. de Lorenzo, and M. Ustav. Functional domains of the TOL plasmid transcription factor XylS. *J Bacteriol*, 182:1118–1126, 2000.

128. J. D. Keasling. Gene-expression tools for the metabolic engineering of bacteria. *Trends Biotechnol*, 17:452–460, 1999.

129. B. Kessler, M. Herrero, K. N. Timmis, and V. de Lorenzo. Genetic evidence that the XylS regulator of the *Pseudomonas* TOL *meta* operon controls the *Pm* promoter through weak DNA-protein interactions. *J Bacteriol*, 176:3171–3176, 1994.

130. S. A. Khan. Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid*, 53:126–136, 2005.

131. R. D. Knight, S. J. Freeland, and L. F. Landweber. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol*, 2:1–13, 2001.

132. K. Kolatka, M. Witosinska, M. Pierechod, and I. Konieczny. Bacterial partitioning proteins affect the subcellular location of broad-host-range plasmid RK2. *Microbiology*, 154:2847–2856, 2008.

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

134. A. A. Komar, T. Lesnik, and C. Reiss. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett*, 462:387–391, 1999.

135. A. V. Komarova, L. S. Tchufistova, M. Dreyfus, and I. V. Boni. AU-rich sequences within 5′-untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. *J Bacteriol*, 187:1344–1349, 2005.

136. N. Komissarova and M. Kashlev. RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. *J Biol Chem*, 272:15329–15338, 1997.

137. N. Korzheva, A. Mustaev, M. Kozlov, A. Malhotra, V. Nikiforov, A. Goldfarb, and S. A. Darst. A structural model of transcription elongation. *Science*, 289:619–625, 2000.

138. M. Kozak. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, 361:13–37, 2005.

139. E. B. Kramer and P. J. Farabaugh. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA*, 13:87–96, 2007.

140. G. Kudla, A. W. Murray, D. Tollervey, and J. B. Plotkin. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*, 324:255–258, 2009.

141. U. Kues and U. Stahl. Replication of plasmids in gram-negative bacteria. *Microbiol Rev*, 53:491–516, 1989.

142. R. Lale, T. Brautaset, and S. Valla. Broad-host-range plasmid vectors for gene expression in bacteria. *Methods Mol Biol*, 765:327–343, 2011.

143. R. J. Lambert and M. Stratford. Weak-acid preservatives: modelling microbial inhibition and response. *J Appl Microbiol*, 86:157–164, 1999.

144. M. Lanzer and H. Bujard. Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci U S A*, 85:8973–8977, 1988.

145. B. S. Laursen, H. P. Sørensen, K. K. Mortensen, and H. U. Sperling-Petersen. Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev*, 69:101–123, 2005.

146. J. Lederberg. Cell genetics and hereditary symbiosis. *Physiol Rev*, 32: 403–430, 1952.

147. G. W. Li, E. Oh, and J. S. Weissman. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, 484: 538–541, 2012.

148. S. Lisser and H. Margalit. Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Res*, 21:1507–1516, 1993.

149. G. Lithwick and H. Margalit. Hierarchy of sequence-dependent features associated with prokaryotic translation. *Genome Res*, 13:2665–2673, 2003.

150. P. J. Lopez, I. Iost, and M. Dreyfus. The use of a tRNA as a transcriptional reporter: the T7 late promoter is extremely efficient in *Escherichia coli* but its transcripts are poorly expressed. *Nucleic Acids Res*, 22:2434, 1994.

151. E. J. Lowbury, H. A. Lilly, A. Kidson, G. A. Ayliffe, and R. J. Jones. Sensitivity of *Pseudomonas aeruginosa* to antibiotics: emergence of strains highly resistant to carbenicillin. *Lancet*, 2:448–452, 1969.

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

153. S. C. Makrides. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev*, 60:512–538, 1996.

154. N. Malys and J. E. McCarthy. Translation initiation: variations in the mechanism can be anticipated. *Cell Mol Life Sci*, 68:991–1003, 2011.

155. M. Manzanera, S. Marqués, and J. L. Ramos. Mutational analysis of the highly conserved C-terminal residues of the XylS protein, a member of the AraC family of transcriptional regulators. *FEBS Lett*, 476:312–317, 2000.

156. J. D. Marks, A. D. Griffiths, M. Malmqvist, T. P. Clackson, J. M. Bye, and G. Winter. By-passing immunization: building high affinity human antibodies by chain shuffling. *Biotechnol*, 10:779–783, 1992.

157. S. Marqués, A. Holtel, K. N. Timmis, and J. L. Ramos. Transcriptional induction kinetics from the promoters of the catabolic pathways of TOL plasmid pWW0 of *Pseudomonas putida* for metabolism of aromatics. *J Bacteriol*, 176:2517–2524, 1994.

158. S. Marqués, M. T. Gallegos, M. Manzanera, A. Holtel, K. N. Timmis, and J. L. Ramos. Activation and repression of transcription at the double tandem divergent promoters for the *xylR* and *xylS* genes of the TOL plasmid of *Pseudomonas putida*. *J Bacteriol*, 180:2889–2894, 1998.

159. S. Marqués, M. Manzanera, M. M. Gonzalez-Perez, M. T. Gallegos, and J. L. Ramos. The XylS-dependent Pm promoter is transcribed *in vivo* by RNA polymerase with σ^{32} or σ^{38} depending on the growth phase. *Mol Microbiol*, 31:1105–1113, 1999.

160. M. T. Marr and J. W. Roberts. Function of transcription cleavage factors GreA and GreB at a regulatory pause site. *Mol Cell*, 6:1275–1285, 2000.

161. R. A. Marshall, C. E. Aitken, M. Dorywalska, and J. D. Puglisi. Translation at the single-molecule level. *Annu Rev Biochem*, 77:177–203, 2008.

162. A. Martínez-Antonio and J. Collado-Vides. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol*, 6:482–489, 2003.

163. J. Mascarenhas, M. H. Weber, and P. L. Graumann. Specific polar localization of ribosomes in *Bacillus subtilis* depends on active transcription. *EMBO Rep*, 2:685–689, 2001.

164. F. Mergulhão, G. Monteiro, J. Cabral, and M. Taipa. Design of bacterial vector systems for the production of recombinant proteins in *Escherichia coli*. *J Microbiol Biotechnol*, 14:1–14, 2004.

165. F. J. Mergulhão, D. K. Summers, and G. A. Monteiro. Recombinant protein secretion in *Escherichia coli*. *Biotechnol Adv*, 23:177–202, 2005.

166. N. Mermod, J. L. Ramos, A. Bairoch, and K. N. Timmis. The *xylS* gene positive regulator of TOL plasmid pWWO: identification, sequence analysis and overproduction leading to constitutive expression of meta cleavage operon. *Mol Gen Genet*, 207:349–354, 1987.

167. W. Messer. The bacterial replication initiator DnaA. DnaA and *oriC*, the bacterial mode to initiate DNA replication. *FEMS Microbiol Rev*, 26: 355–374, 2002.

168. D. Metcalf. The granulocyte-macrophage colony stimulating factors. *Cell*, 43:5–6, 1985.

169. D. Metcalf. Control of granulocytes and macrophages: molecular, cellular, and clinical aspects. *Science*, 254:529–533, 1991.

170. H.-P. Meyer and D. R. Schmidhalter. *Microbial Expression Systems and Manufacturing from a Market and Economic Perspective*. Innovations in Biotechnology. 2012.

171. C. Michan, B. Kessler, V. De Lorenzo, K. N. Timmis, and J. L. Ramos. XylS domain interactions can be deduced from intraallelic dominance in double mutants of *Pseudomonas putida*. *Mol Gen Genet*, 235:406–412, 1992.

172. C. Michan, L. Zhou, M. T. Gallegos, K. N. Timmis, and J. L. Ramos. Identification of critical amino-terminal regions of XylS. The positive regulator encoded by the TOL plasmid. *J Biol Chem*, 267:22897–22901, 1992.

173. P. Milon and M. V. Rodnina. Kinetic control of translation initiation in bacteria. *Crit Rev Biochem Mol Biol*, 47:334–348, 2012.

174. L. Minakhin, S. Bhagat, A. Brunning, E. A. Campbell, S. A. Darst, R. H. Ebright, and K. Severinov. Bacterial RNA polymerase subunit ω and eukaryotic polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly, journal = Proc Natl Acad Sci U S A, volume = 98, pages = 892-897, year = 2001.

175. N. Mitarai, K. Sneppen, and S. Pedersen. Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization. *J Mol Biol*, 382:236–245, 2008.

176. I. Moll, S. Grill, C. O. Gualerzi, and U. Blasi. Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Mol Microbiol*, 43:239–246, 2002.

177. P. Montero Llopis, A. F. Jackson, O. Sliusarenko, I. Surovtsev, J. Heinritz, T. Emonet, and C. Jacobs-Wagner. Spatial organization of the flow of genetic information in bacteria. *Nature*, 466:77–81, 2010.

178. J. Mukhopadhyay, A. N. Kapanidis, V. Mekler, E. Kortkhonjia, Y. W. Ebright, and R. H. Ebright. Translocation of σ^{70} with RNA polymerase during transcription: Fluorescence resonance energy transfer assay for movement relative to DNA. *Cell*, 106:453–463, 2001.

179. J. B. Munro, R. B. Altman, N. O'Connor, and S. C. Blanchard. Identification of two distinct hybrid state intermediates on the ribosome. *Mol Cell*, 25:505–517, 2007.

180. K. S. Murakami and S. A. Darst. Bacterial RNA polymerases: the wholo story. *Curr Opin Struct Biol*, 13:31–39, 2003.

181. K. S. Murakami, S. Masuda, and S. A. Darst. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science*, 296:1280–1284, 2002.

182. H. Musto, H. Naya, A. Zavala, H. Romero, F. Alvarez-Valín, and G. Bernardi. Correlations between genomic GC levels and optimal growth temperatures in prokaryotes. *FEBS Lett*, 573:73–77, 2004.

183. A. Muto and S. Osawa. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci U S A*, 84:166–169, 1987.

184. Y. Nakamura, T. Gojobori, and T. Ikemura. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res*, 28:292, 2000.

185. H. Naya, H. Romero, A. Zavala, B. Alvarez, and H. Musto. Aerobiosis increases the genomic guanine plus cytosine content (gc%) in prokaryotes. *J Mol Evol*, 55:260–264, 2002.

186. F. C. Neidhardt and R. Curtiss, editors. *Escherichia coli and Salmonella : cellular and molecular biology*. ASM Press, Washington, D.C. :, 2nd ed. edition, 1996.

187. K. Nevo-Dinur, A. Nussbaum-Shochat, S. Ben-Yehuda, and O. Amster-Choder. Translation-independent localization of mRNA in *E. coli*. *Science*, 331:1081–1084, 2011.

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

189. H. Niki and S. Hiraga. Subcellular distribution of actively partitioning F plasmid during the cell division cycle in *E. coli*. *Cell*, 90:951–957, 1997.

190. P. Nissen, J. Hansen, N. Ban, P. B. Moore, and T. A. Steitz. The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289:920–930, 2000.

191. H. F. Noller, M. M. Yusupov, G. Z. Yusupova, A. Baucom, and J. H. Cate. Translocation of tRNA during protein synthesis. *FEBS Lett*, 514:11–16, 2002.

192. R. P. Novick. Plasmid incompatibility. *Microbiol Rev*, 51:381–395, 1987.

193. E. Nudler. RNA polymerase backtracking in gene regulation and genome instability. *Cell*, 149(7):1438–45, 2012.

194. E. Nudler, A. Mustaev, E. Lukhtanov, and A. Goldfarb. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*, 89:33–41, 1997.

195. O. N. Ozoline, A. A. Deev, and M. V. Arkhipova. Non-canonical sequence elements in the promoter structure. Cluster analysis of promoters recognized by *Escherichia coli RNA polymerase*. *Nucleic Acids Res*, 25:4703–4709, 1997.

196. A. J. Palomares, M. A. De Luca, and D. R. Helinski. Firefly luciferase as a reporter enzyme for measuring gene expression in vegetative and symbiotic *Rhizobium meliloti* and other Gram-negative bacteria. *Gene*, 81:55–64, 1989.

197. L. A. Palomares, S. Estrada-Mondaca, and O. T. Ramrez. Production of recombinant proteins: challenges and solutions. *Methods Mol Biol*, 267: 15–52, 2004.

198. T. Pape, W. Wintermeyer, and M. Rodnina. Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J*, 18:3800–3807, 1999.

199. E. Pérez-Rueda and J. Collado-Vides. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res*, 28: 1838–1847, 2000.

200. F. Peske, M. V. Rodnina, and W. Wintermeyer. Sequence of steps in ribosome recycling as defined by kinetic analysis. *Mol Cell*, 18:403–412, 2005.

201. J. M. Peters, A. D. Vangeloff, and R. Landick. Bacterial transcription terminators: the RNA 3′-end chronicles. *J Mol Biol*, 412:793–813, 2011.

202. D. Petrelli, A. LaTeana, C. Garofalo, R. Spurio, C. L. Pon, and C. O. Gualerzi. Translation initiation factor IF3: two domains, five functions, one mechanism? *EMBO J*, 20:4560–4569, 2001.

203. J. B. Plotkin and G. Kudla. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*, 12:32–42, 2011.

204. J. Pogliano, T. Q. Ho, Z. Zhong, and D. R. Helinski. Multicopy plasmids are clustered and localized in *Escherichia coli*. *Proc Natl Acad Sci U S A*, 98:4486–4491, 2001.

205. R. T. Pomerantz and M. O'Donnell. Replisome mechanics: insights into a twin DNA polymerase machine. *Trends Microbiol*, 15:156–164, 2007.

206. E. S. Poole, C. M. Brown, and W. P. Tate. The identity of the base following the stop codon determines the efficiency of *in vivo* translational termination in *Escherichia coli*. *EMBO J*, 14:151–158, 1995.

207. S. Proshkin, A. R. Rahmouni, A. Mironov, and E. Nudler. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science*, 328:504–508, 2010.

208. A. P. Pugsley. The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev*, 57:50–108, 1993.

209. A. P. Pugsley, O. Francetic, O. M. Possot, N. Sauvonnet, and K. R. Hardie. Recent progress and future directions in studies of the main terminal branch of the general secretory pathway in Gram-negative bacteria–a review. *Gene*, 192:13–19, 1997.

210. R. Raghavan, Y. D. Kelkar, and H. Ochman. [a selective force favoring increased g+c content in bacterial genes.

211. M. Rajewska, K. Wegrzyn, and I. Konieczny. AT-rich region and repeated sequences – the essential elements of replication origins of bacterial replicons. *FEMS Microbiol Rev*, 36:408–434, 2012.

212. J. L. Ramos, A. Stolz, W. Reineke, and K. N. Timmis. Altered effector specificities in regulators of gene expression: TOL plasmid *xylS* mutants and their use to engineer expansion of the range of aromatics degraded by bacteria. *Proc Natl Acad Sci U S A*, 83:8467–8471, 1986.

213. J. L. Ramos, N. Mermod, and K. N. Timmis. Regulatory circuits controlling transcription of TOL plasmid operon encoding meta-cleavage pathway for degradation of alkylbenzoates by *Pseudomonas*. *Mol Microbiol*, 1:293–300, 1987.

214. J. L. Ramos, M. Gonzalez-Carrero, and K. N. Timmis. Broad-host range expression vectors containing manipulated *meta*-cleavage pathway regulatory elements of the tol plasmid. *FEBS Lett*, 226:241–246, 1988.

215. J. L. Ramos, S. Marqués, and K. N. Timmis. 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, 1997.

216. A. Revyakin, C. Liu, R. H. Ebright, and T. R. Strick. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*, 314:1139–1143, 2006.

217. J. P. Richardson. Rho-dependent termination and ATPases in transcript termination. *Biochim Biophys Acta*, 1577:251–260, 2002.

218. S. Ringquist, S. Shinedling, D. Barrick, L. Green, J. Binkley, G. D. Stormo, and L. Gold. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Mol Microbiol*, 6:1219–1229, 1992.

219. J. W. Roberts. Termination factor for RNA synthesis. *Nature*, 224:1168–1174, 1969.

220. D. A. Rodionov. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev*, 107:3467–3497, 2007.

221. M. V. Rodnina, H. Stark, A. Savelsbergh, H. J. Wieden, D. Mohr, N. B. Matassova, F. Peske, T. Daviter, C. O. Gualerzi, and W. Wintermeyer. GTPases mechanisms and functions of translation factors on the ribosome. *Biol Chem*, 381:377–387, 2000.

222. A. H. Rosenberg, B. N. Lade, D. S. Chui, S. W. Lin, J. J. Dunn, and F. W. Studier. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene*, 56:125–135, 1987.

223. W. Ross, K. K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov, and R. L. Gourse. A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase. *Science*, 262:1407–1413, 1993.

224. R. Ruiz and J. L. Ramos. Residues 137 and 153 of XylS influence contacts with the C-terminal domain of the RNA polymerase α subunit. *Biochem Biophys Res Commun*, 287:519–521, 2001.

225. R. Ruiz and J. L. Ramos. Residues 137 and 153 at the N terminus of the XylS protein influence the effector profile of this transcriptional regulator and the σ factor used by RNA polymerase to stimulate transcription from its cognate promoter. *J Biol Chem*, 277:7282–7286, 2002.

226. R. Ruiz, J. L. Ramos, and S. M. Egan. Interactions of the XylS regulators with the C-terminal domain of the RNA polymerase α subunit influence the expression level from the cognate Pm promoter. *FEBS Lett*, 491:207–211, 2001.

227. R. Ruíz, S. Marqués, and J. L. Ramos. 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, 2003.

228. R. M. Saecker, M. T. Record Jr, and P. L. Dehaseth. Mechanism of bacterial transcription initiation: RNA polymerase - Promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol*, 412:754–771, 2011.

229. H. M. Salis. The ribosome binding site calculator. *Methods Enzymol*, 498: 19–42, 2011.

230. M. A. Sánchez-Romero, D. J. Lee, E. Sánchez-Moran, and S. J. Busby. Location and dynamics of an active promoter in *Escherichia coli* K-12. *Biochem J*, 441:481–485, 2012.

231. B. Schauder and J. E. McCarthy. The role of bases upstream of the Shine-Dalgarno region and in the coding sequence in the control of gene expression in *Escherichia coli*: translation and stability of mRNAs in vivo. *Gene*, 78:59–72, 1989.

232. T. J. Schmidhauser, M. Filutowicz, and D. R. Helinski. Replication of derivatives of the broad host range plasmid RK2 in two distantly related bacteria. *Plasmid*, 9:325–330, 1983.

233. M. C. Schmidt and M. J. Chamberlin. Amplification and isolation of *Escherichia coli nusA* protein and studies of its effects on in vitro RNA chain elongation. *Biochemistry*, 23:197–203, 1984.

234. P. Shah and M. A. Gilchrist. Effect of correlated tRNA abundances on translation errors and evolution of codon usage bias. *PLoS Genet*, 6, 2010.

235. Z. Q. Shao, Y. M. Zhang, X. Y. Feng, B. Wang, and J. Q. Chen. Synonymous codon ordering: a subtle but prevalent strategy of bacteria to improve translational efficiency. *PLoS One*, 7, 2012.

236. P. M. Sharp and M. Bulmer. Selective differences among translation termination codons. *Gene*, 63:141–145, 1988.

237. P. M. Sharp and W. H. Li. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Res*, 14: 7737–7749, 1986.

238. P. M. Sharp, E. Bailes, R. J. Grocock, J. F. Peden, and R. E. Sockett. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res*, 33:1141–1153, 2005.

239. N. Shepherd, P. Dennis, and H. Bremer. Cytoplasmic RNA polymerase in *Escherichia coli*. *J Bacteriol*, 183:2527–2534, 2001.

240. J. Shine and L. Dalgarno. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A*, 71:1342–1346, 1974.

241. V. Shingler and C. M. Thomas. Analysis of the *trfA* region of broad host-range plasmid RK2 by transposon mutagenesis and identification of polypeptide products. *J Mol Biol*, 175:229–249, 1984.

242. V. Shingler and C. M. Thomas. Analysis of nonpolar insertion mutations in the *trfA* gene of IncP plasmid RK2 which affect its broad-host-range property. *Biochim Biophys Acta*, 1007:301–308, 1989.

243. N. K. Sinha, C. F. Morris, and B. M. Alberts. Efficient *in vitro* replication of double-stranded DNA templates by a purified T4 bacteriophage replication system. *J Biol Chem*, 255:4290–4293, 1980.

244. H. Sletta, A. Nedal, T. E. V. Aune, H. Hellebust, S. Hakvåg, R. Aune, T. E. Ellingsen, S. Valla, and T. Brautaset. 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, 2004.

245. H. Sletta, A. Tondervik, S. Hakvåg, T. E. Aune, A. Nedal, R. Aune, G. Evensen, S. Valla, T. E. Ellingsen, and T. Brautaset. 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, 2007.

246. M. A. Sorensen, C. G. Kurland, and S. Pedersen. Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol*, 207:365–377, 1989.

247. R. Sprengel, B. Reiss, and H. Schaller. Translationally coupled initiation of protein synthesis in *Bacillus subtilis*. *Nucleic Acids Res*, 13:893–909, 1985.

248. M. Steigedal and S. Valla. The *Acinetobacter* sp. *chnB* promoter together with its cognate positive regulator ChnR is an attractive new candidate for metabolic engineering applications in bacteria. *Metab Eng*, 10:121–129, 2008.

249. T. A. Steitz. The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Curr Opin Struct Biol*, 19:683–690, 2009.

250. T. A. Steitz and P. B. Moore. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci*, 28:411–418, 2003.

251. G. Striedner, I. Pfaffenzeller, M. Luchner, S. Nemecek, R. Grabherr, and K. Bayer. Plasmid-free T7-Based *Escherichia coli* expression systems. *Biotechnol Bioeng*, 105:786–794, 2010.

252. S. M. Studer and S. Joseph. Unfolding of mRNA secondary structure by the bacterial translation initiation complex. *Mol Cell*, 22:105–115, 2006.

253. F. W. Studier. Bacteriophage T7. *Science*, 176:367–376, 1972.

254. F. W. Studier. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol*, 219:37–44, 1991.

255. F. W. Studier and B. A. Moffatt. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*, 189: 113–130, 1986.

256. N. Sueoka. On the genetic basis of variation and heterogeneity of DNA base composition. *Proc Natl Acad Sci U S A*, 48:582–592, 1962.

257. R. M. Sundari, E. A. Stringer, L. H. Schulman, and U. Maitra. Interaction of bacterial initiation factor 2 with initiator tRNA. *J Biol Chem*, 251: 3338–3345, 1976.

258. F. Supek and T. Smuc. On relevance of codon usage to expression of synthetic and natural genes in *Escherichia coli*. *Genetics*, 185:1129–1134, 2010.

259. L. Tao, R. E. Jackson, and Q. Cheng. Directed evolution of copy number of a broad host range plasmid for metabolic engineering. *Metab Eng*, 7: 10–17, 2005.

260. K. Terpe. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol*, 72:211–222, 2006.

261. T. A. Thanaraj and P. Argos. Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci*, 5:1973–1983, 1996.

262. D. J. Tipper. Mode of action of beta-lactam antibiotics. *Pharmacol Ther*, 27:1–35, 1985.

263. A. E. Toukdarian and D. R. Helinski. TrfA dimers play a role in copy-number control of RK2 replication. *Gene*, 223:205–211, 1998.

264. A. E. Toukdarian, D. R. Helinski, and S. Perri. The plasmid RK2 initiation protein binds to the origin of replication as a monomer. *J Biol Chem*, 271: 7072–7078, 1996.

265. T. Tuller, A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman, and Y. Pilpel. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, 141: 344–354, 2010.

266. T. Tuller, Y. Y. Waldman, M. Kupiec, and E. Ruppin. Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci U S A*, 107:3645–3650, 2010.

267. B. E. Uhlin and K. Nordstrom. R plasmid gene dosage effects in Escherichia coli K-12: copy mutants of the R plasmic R1drd-19. *Plasmid*, 1:1–7, 1977.

268. S. M. Uptain, C. M. Kane, and M. J. Chamberlin. Basic mechanisms of transcript elongation and its regulation. *Annu Rev Biochem*, 66:117–172, 1997.

269. M. Valencia-Burton, A. Shah, J. Sutin, A. Borogovac, R. M. McCullough, C. R. Cantor, A. Meller, and N. E. Broude. Spatiotemporal patterns and transcription kinetics of induced RNA in single bacterial cells. *Proc Natl Acad Sci U S A*, 106:16399–16404, 2009.

270. S. A. van Hijum, M. H. Medema, and O. P. Kuipers. Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol Mol Biol Rev*, 73:481–509, 2009.

271. G. Vandermeulen, C. Marie, D. Scherman, and V. Preat. New generation of plasmid backbones devoid of antibiotic resistance marker for gene therapy trials. *Mol Ther*, 19:1942–9, 2011.

272. S. Varenne, J. Buc, R. Lloubes, and C. Lazdunski. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol*, 180:549–576, 1984.

273. T. E. Vee Aune, I. Bakke, F. Drabløs, R. Lale, T. Brautaset, and S. Valla. Directed evolution of the transcription factor XylS for development of improved expression systems. *Microb Biotechnol*, 3:38–47, 2010.

274. C. Verheust and D. R. Helinski. The *incC korB* region of RK2 repositions a mini-RK2 replicon in *Escherichia coli*. *Plasmid*, 58:195–204, 2007.

275. H. Waegeman and W. Soetaert. Increasing recombinant protein production in *Escherichia coli* through metabolic and genetic engineering. *J Ind Microbiol Biot*, 38:1891–1910, 2011.

276. T. C. Wang. Discontinuous or semi-discontinuous DNA replication in *Escherichia coli*? *Bioessays*, 27:633–636, 2005.

277. J. D. Watson and F. H. Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

278. M. Welch, S. Govindarajan, J. E. Ness, A. Villalobos, A. Gurney, J. Minshull, and C. Gustafsson. Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS One*, 4:e7002, 2009.

279. M. Welch, A. Villalobos, C. Gustafsson, and J. Minshull. Designing genes for successful protein expression. *Methods Enzymol*, 498:43–66, 2011.

280. J. D. Wen, L. Lancaster, C. Hodges, A. C. Zeri, S. H. Yoshimura, H. F. Noller, C. Bustamante, and I. Tinoco. Following translation by single ribosomes one codon at a time. *Nature*, 452:598–603, 2008.

281. K. Wetterstrand. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP), 2012. URL `http://www.genome.gov/sequencingcosts`.

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

283. P. A. Williams and K. Murray. Metabolism of benzoate and the methylbenzoates by *Pseudomonas putida (arvilla)* mt-2: evidence for the existence of a TOL plasmid. *J Bacteriol*, 120:416–423, 1974.

284. H. C. Winther-Larsen, J. M. Blatny, B. Valand, T. Brautaset, and S. Valla. *Pm* promoter expression mutants and their use in broad-host-range RK2 plasmid vectors. *Metab Eng*, 2:92–103, 2000.

285. H. C. Winther-Larsen, K. D. Josefsen, T. Brautaset, and S. Valla. Parameters affecting gene expression from the *Pm* promoter in gram-negative bacteria. *Metab Eng*, 2:79–91, 2000.

286. M. J. Worsey, F. C. Franklin, and P. A. Williams. Regulation of the degradative pathway enzymes coded for by the TOL plasmid (pWWO) from *Pseudomonas putida* mt-2. *J Bacteriol*, 134:757–764, 1978.

287. X. Wu, H. Jornvall, K. D. Berndt, and U. Oppermann. Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance. *Biochem Biophys Res Commun*, 313:89–96, 2004.

288. S. Yao, A. Toukdarian, and D. R. Helinski. Inhibition of protein and RNA synthesis in *Escherichia coli* results in declustering of plasmid RK2. *Plasmid*, 56:124–132, 2006.

289. S. Yao, D. R. Helinski, and A. Toukdarian. Localization of the naturally occurring plasmid ColE1 at the cell pole. *J Bacteriol*, 189:1946–1953, 2007.

290. J. Yin, G. Li, X. Ren, and G. Herrler. 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, 2007.

291. J. H. Yoo and U. L. RajBhandary. Requirements for translation reinitiation in *Escherichia coli*: roles of initiator tRNA and initiation factors IF2 and IF3. *Mol Microbiol*, 67:1012–1026, 2008.

292. B. A. Young, T. M. Gruber, and C. A. Gross. Views of transcription initiation. *Cell*, 109:417–420, 2002.

293. D. Zhou and R. Yang. Global analysis of gene transcription regulation in prokaryotes. *Cell Mol Life Sci*, 63:2260–2290, 2006.

294. L. M. Zhou, K. N. Timmis, and J. L. Ramos. Mutations leading to constitutive expression from the TOL plasmid meta-cleavage pathway operon are located at the C-terminal end of the positive regulator protein XylS. *J Bacteriol*, 172:3707–3710, 1990.

295. A. Zwiefka, H. Kohn, and W. R. Widger. Transcription termination factor rho: the site of bicyclomycin inhibition in *Escherichia coli*. *Biochemistry*, 32:3564–3570, 1993.

Paper I

Zwick et al. Microbial Cell Factories 2012, 11:133
http://www.microbialcellfactories.com/content/11/1/133

MICROBIAL CELL FACTORIES

RESEARCH Open Access

Strong stimulation of recombinant protein production in *Escherichia coli* by combining stimulatory control elements in an expression cassette

Friederike Zwick, Rahmi Lale and Svein Valla*

Abstract

Background: The XylS/*Pm* expression system has been used to produce recombinant proteins at industrial levels in *Escherichia coli*. Activation of transcription from the *Pm* promoter takes place in the presence of benzoic acid or derivatives of it. Previous mutagenesis studies resulted in identification of several variants of the expression control elements *xylS* (X), *Pm* (P) and the 5'-untranslated region (U) that individually gave rise to strongly stimulated expression. The goal of this study was to test if combination of such stimulatory mutations in the same expression vectors would lead to further increase of expression levels.

Results: We combined X, P and U variants that were originally identified due to their ability to strongly stimulate expression of the reporter gene *bla* (resistance to penicillin). Combination of optimized elements stimulated *bla* expression up to 75-fold (X, P and U combined) relative to the wild-type system, while accumulated transcript levels increased about 50-fold. This is much more than for the elements individually. We also tested combination of the variant elements on two other and unrelated genes, *celB* (encoding phosphoglucomutase) and the human growth factor gene *gm-csf*. Protein production from these genes is much more efficient than from *bla* in the wild-type system, but expression was still significantly stimulated by the combination of X, P and U variants, although not to the same extent as for *bla*.
We also integrated a single copy of the expression cassette with each gene into the *E. coli* chromosome and found that the expression level from this single copy was higher for *bla* than for the wild-type plasmid system, while it was lower for *celB* and *gm-csf*.

Conclusion: Our results show that combination of stimulatory expression control elements can be used to further increase production of different proteins in *E. coli*. For one reporter gene (*bla*) this allowed for more protein production from a single gene copy integrated on the chromosome, compared to the wild-type plasmid system. The approach described here should in principle be applicable for improvement of any expression cassette.

Keywords: Recombinant, XylS/Pm, UTR, Gene expression, Promoter, Mutant, Escherichia coli

* Correspondence: svein.valla@biotech.ntnu.no
Department of Biotechnology, Norwegian University of Science and Technology, Sem Sælands Vei 6/8, N-7491, Trondheim, Norway

Background

Plasmids are heavily utilized as expression tools for recombinant protein production in bacteria, mainly because they enable easy introduction of recombinant genes at high gene dosages, resulting in large amounts of transcripts from the gene of interest [1]. However, high plasmid copy numbers in itself represent a burden for the host [2-4], and this may lead to problems like plasmid instability, decreased growth rates, and plasmid DNA mutations [2,5-8]. Antibiotic selection markers included on plasmids for enhanced stability also impose an additional metabolic burden to the cell [9].

Ideally it would therefore be advantageous to avoid such autonomously replicating elements and chromosomal integration of expression cassettes stands as an attractive alternative for stable expression [10,11].

The main problem with single copy chromosomal integrations is that the strong reduction in gene dosage compared to high copy number plasmids leads to reduced transcript formation and will therefore require very efficient translation of each transcript. Increasing the number of copies of the gene of interest on the chromosome could potentially eliminate this problem, but this requires further modification of the host genome in order to ensure the stability of the multiple times integrated DNA [11]. It is therefore important to develop methods that aim for the highest possible levels of expression per gene copy, as this might allow reduction of the number of gene copies per cell. A recent study demonstrates an approach to ensure high levels of transcript formation, by placing the desired gene together with a tandem *tac* promoter cluster into the chromosome [12]. This method eliminates the problem of having multiple copies of the gene, however the promoter utilized is constitutive. For metabolic engineering type applications it is desirable to have time-dependent expression.

We have previously demonstrated that the strong and positively regulated *xylS/Pm* expression cassette in its wild-type form, combined with a mini-RK2 replicon, can serve as a tool to achieve industrial level production of recombinant proteins in *Escherichia coli* [13,14]. XylS belongs to the AraC-XylS transcriptional regulator family and in the presence of passively transported benzoic acid derivates it activates transcription from the *Pm* promoter.

More recently, we have shown that expression from *Pm* can be further strongly stimulated by introducing mutations in the promoter region [15], in the 5'-untranslated region of mRNA (5'-UTR) [16], or in the XylS regulator coding sequence [17]. In this study we show that combination of such previously identified variant DNA control elements leads to additive stimulatory effects on expression, and this approach should in principle be applicable to any expression system.

Results

The expression level from *Pm* can be strongly stimulated by combining mutated DNA elements previously shown to individually enhance expression

Initial studies involving combinations of previously isolated stimulatory control element variants (*Pm* promoter, its UTR and *xylS*) indicated that they at least to some extent acted additively, and based on these observations cells containing eight different plasmids were subjected to more detailed analyses. A plasmid with only wild-type elements was used as control, and in addition we constructed three plasmids containing one variant element only (comX, comP, comU), three constructs with each possible combination of two variant elements (comXP, comXU, comPU), and one construct (comXPU) in which all three elements were combined (Table 1). The three variants used were the *xylS* variant StEP-13 (X), which carries five amino acid substitutions and stimulates transcription [17]; a promoter variant (P) designated ML2-5 which carries five point mutations in *Pm*, also stimulating transcription [15]; and a 5'-UTR variant (U) designated H39 with two point mutations (Figure 1), which appears to predominantly stimulate translation.

To monitor the expression from the different variants and the combination of them we first used *bla* (encoding β-lactamase) as a reporter gene. This gene was also used as a reporter to originally identify the variant sequences,

Table 1 Expression profiles of *E. coli* DH5α cells containing the indicated constructs

Combinations	*xylS*	*Pm*	5'-UTR	Induced*	Uninduced*	Reference
wild-type	wild-type	wild-type	wild-type	0.2 - 0.4	0.005 - 0.01	This study
comP	wild-type	ML2-5	wild-type	1 – 2	0.005 - 0.01	[15]
comU	wild-type	wild-type	H39	2 – 3	0.06 - 0.08	This study
comX	StEP-13	wild-type	wild-type	1 – 2	0.02 - 0.03	[17]
comPU	wild-type	ML2-5	H39	5 – 6	1 - 1.5	This study
comXP	StEP-13	ML2-5	wild-type	3 – 4	0.3 - 0.4	This study
comXU	StEP-13	wild-type	H39	12 – 13	0.8 - 0.9	This study
comXPU	StEP-13	ML2-5	H39	15 – 20	3 - 4	This study

* Phenotypes are given as ampicillin resistance levels (mg mL^{-1}) under induced (0.1 mM m-toluate) and uninduced conditions.

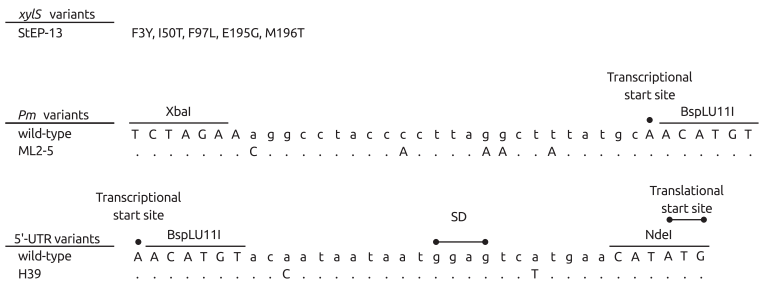

Figure 1 Sequence profiles of the variants that are used in this study. Restriction sites shown are unique. Nucleotides in lower case indicate randomly mutagenized bases, and introduced mutations are indicated for each construct (amino acid substitutions for XylS, and base changes for *Pm* and 5'-UTR variants). Identical nucleotides are indicated by dots, deletion mutations are indicated by short horizontal lines. Transcriptional and translational start sites are written in bold face and indicated with filled circles. SD is the putative Shine-Dalgarno sequence.

based on the observation that host tolerance against ampicillin correlates with the amounts of β-lactamase produced [15-17]. The *bla* gene was in all cases expressed from the relevant variant version of the plasmid pTA16 (Figure 2).

The expression level from *Pm* can be continuously adjusted by varying the inducer concentration [18-20], and in the tests of the different combination constructs we used a low concentration (0.1 mM) to avoid potential host toxicity effects caused by elevated expression levels of β-lactamase. Plating of the host strains containing the plasmid variants on agar medium supplied with varying concentrations of ampicillin demonstrated that under induced conditions cells containing the wild-type, comX, comP and comU plasmid constructs tolerated up to 0.2, 1, 1, and 2 mg mL^{-1} of the antibiotic, respectively (Table 1). Interestingly, all three combinations of the two variant elements, comXP, comXU and comPU, resulted in varying but further enhancement of target protein expression, observed as upper tolerance levels of 3, 12, and 5 mg mL^{-1} ampicillin, respectively. The effect was even more drastic for comXPU, which tolerated up to 15 mg mL^{-1} of ampicillin. This is 75-fold higher than for cells with the wild-type plasmid. The uninduced resistance levels also went up for all strains (up to 3 mg mL^{-1} for comXPU).

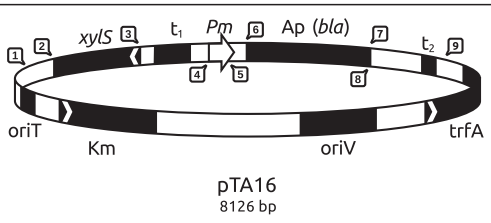

Figure 2 Map of plasmid pTA16. *Pm,* positively regulated promoter; *xylS,* gene encoding *Pm* activator; Ap^r *(bla),* ampicillin resistance gene encoding β-lactamase; Km^r, kanamycin resistance gene; *trfA,* gene encoding the essential replication protein; *oriV,* origin of vegetative replication; *oriT,* origin of conjugal transfer; *t1, rrnBT1T2* bidirectional transcriptional terminator; *t2,* bidirectional transcriptional terminator. Restriction sites shown are unique: 1-SfiI, 2-AgeI, 3-SacI, 4-XbaI, 5-BspLUIII, 6-NdeI, 7-NotI, 8-BamH1, 9-KpnI.

Ampicillin resistance is a good indicator of expression at the protein level, but cannot be used for accurate quantitative comparisons between clones. We therefore directly measured the corresponding β-lactamase enzyme activities. Combination of two elements in all cases stimulated expression at the protein level, about 6-fold for comXP, 25-fold for comPU, 50-fold for comXU and 75-fold for comXPU (Figure 3a). These data are in good agreement with the observed stimulation of ampicillin resistance.

The levels of accumulated transcripts of the target gene were also measured by relative quantification real-time RT-PCR (qRT-PCR), and as expected a strong stimulation was observed (up to about 63-fold for comXPU), although somewhat less than at the protein level (Figure 3a). This presumably reflects that some of the stimulation is resulting from improved translation of the target gene, most clearly demonstrated by comXU.

Combination of variant elements also leads to increased expression of two other tested reporter genes

As for any expression system individual proteins are expressed at quite varying levels from *Pm,* and β-lactamase is not among the highly expressed proteins. In the original identification of the X, P and U variants the *bla* gene was used as a reporter and it is therefore of interest to study if the variant combinations would also stimulate the expression of genes other than *bla*. We selected the bacterial *celB* gene (encoding phosphoglucomutase) and the human *gm-csf* gene (encoding cytokine granulocyte-macrophage colony-stimulating factor) as representative examples for such an analysis. Both of these genes were previously (in contrast to *bla*) shown to be efficiently

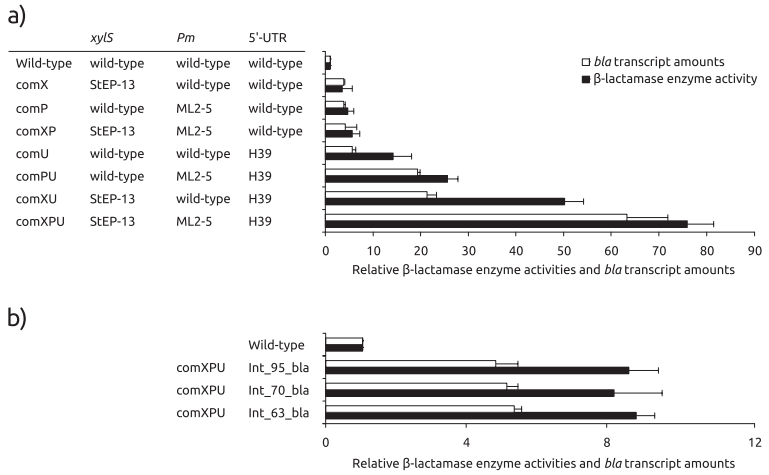

Figure 3 β-lactamase enzyme activities and *bla* transcript amounts for extra-chromosomally (a) and chromosomally (b) expressed combination constructs. Enzyme activities in white, transcript amounts in black, all values are relative to the wild-type (arbitrarily set to 1). The values are the average of at least two biological replicas. *E. coli* DH5α was used as host.

expressed from wild-type *xylS/Pm*. CelB was earlier found to be the clearly dominating protein on a crude gel when expressed from wild-type *xylS/Pm* [21] and also GM-CSF could be visualized on a protein gel when expressed from a plasmid with elevated copy number [14]. The *bla* gene in the constructs described above was therefore substituted with either the *celB* or the *gm-csf* gene. Phosphoglucomutase enzyme activities were then measured, while GM-CSF protein levels were visualized by Western Blotting (Figures 4a and 5a). In case of *gm-csf* another strain (RV308) was used as host to enable comparison of expression levels to previously published results [14]. Comparisons of ampicillin tolerances and analyses of *gm-csf* expression at the transcript and protein levels in DH5α andRV308 indicated that absolute values are slightly higher in RV308, while all relative values, compared to wild-type, were similar in both strains (data not shown).

Interestingly, expression of both genes was quite significantly stimulated at the protein level, particularly with the comXPU construct (about 8.5-fold for *celB*). For *gm-csf* protein production was difficult to detect by Western blotting, but we have shown before that it is easily detectable if a copy up (3–4 fold) version of the same plasmid is used [14]. Compared to this construct the comXPU variant (at wild-type plasmid copy number)

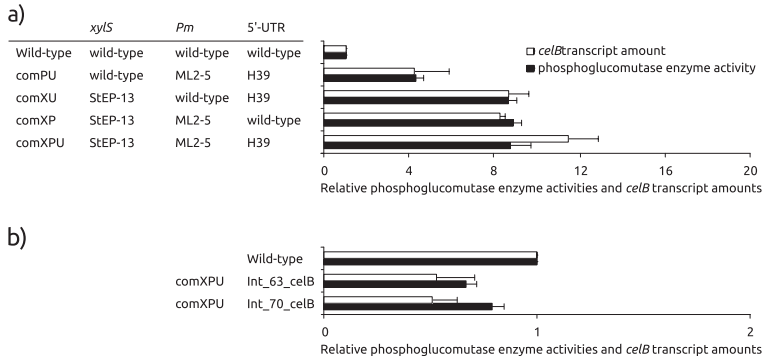

Figure 4 Phosphoglucomutase enzyme activities and *celB* transcript amounts for extra-chromosomally (a) and chromosomally (b) expressed combination constructs. Enzyme activities in white, transcript amounts in black, all values are relative to the wild-type (arbitrarily set to 1). The values are the average of at least two biological replicas. *E. coli* DH5α was used as host.

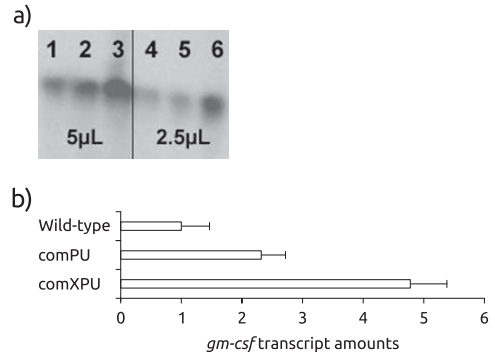

Figure 5 Western Blot (a) and *gm-csf* transcript amounts (b) for extra-chromosomally expressed combination constructs. (a) 1 and 4: wt plasmid with elevated copy number (3-4x); 2 and 5: comPU; 3 and 6 : comXPU; total cell protein was loaded on the gel (b) values are relative to the wild-type plasmid with elevated copy number (arbitrarily set to 1). The values are the average of at least two biological replicas. *E. coli* RV308 was used as host.

produced between 3 and 4-fold more GM-CSF. This means that the level of stimulation per gene copy is probably around ten-fold and the protein was now clearly visible on a crude gel, without the need for Western Blotting.

We also monitored the stimulation at the transcript level, and interestingly, for *celB* the comXPU construct stimulated accumulation of target gene transcripts more (about 15-fold) than at the protein level (Figure 4b). This might be interpreted to indicate inclusion body formation but we have found that this particular protein does not to any significant extent form inclusion bodies (data not shown). We therefore conclude that under the reported conditions (with comXPU) the excessive transcript amounts cannot be efficiently handled by the translational machinery. For *gm-csf* the transcriptional stimulation was also highest for comXPU (4.8-fold), compared to the wild-type plasmid with wild-type copy number (Figure 5b).

Based on all these studies we conclude that the variant expression control elements isolated in the context of the *bla* gene also are able to stimulate the expression of other and very unrelated genes, although not to the same extent. The reason for this difference may either be that the variant expression elements are particularly well adapted to *bla*, or that the more efficient expression of *celB* and *gm-csf* in the wild-type system gives less room for further improvements.

Chromosomal integration of genes may be a future option for optimized expression cassettes

Ideally chromosomal integrations can be seen as desirable in protein expression, due to greater genetic stability of the production strains. We have studied this for *bla*, *celB* and *gm-csf*, using the expression cassette with the comXPU variant, integrated into the *E. coli* chromosome (strain DH5α for *bla*, *celB*, and strain RV308 for *gm-csf*) by use of conditional-replication, integration and modular (CRIM) plasmids [22]. These replicons facilitate directed integration of a DNA-fragment at different locations in *E. coli* chromosome at bacterial attachment sites *(attB)*. The *xylS/Pm* expression cassette with the three target genes was PCR-amplified and cloned into the integration plasmids, and successful insertion of single copies was confirmed by a PCR procedure described by Haldimann and Wanner [22]. Expression levels of the resulting strains were tested both by enzyme assays or Western Blots and qRT-PCR. For each gene the results were similar independent of the integration sites, and for β-lactamase the final protein expression was about 8-fold higher relative to the wild-type plasmid system (Figure 3b). For *celB* the corresponding levels were slightly lower than that of the wild-type cassette in the plasmid (Figure 4b), and for GM-CSF the protein was not detectable on a Western Blot (data not shown).

At the transcriptional level stimulation of *bla* expression was somewhat lower than at the translational level, as for XPU in the plasmid state. In contrast, for *celB* the transcription was much more stimulated in the plasmid state. For *gm-csf* accumulated transcript levels were about 5-fold lower than for the corresponding wild-type plasmid with wild-type copy-number (not shown).

In conclusion the results show that combination of optimized variant sequences in the expression control system lead to additive effects that can be used to strongly stimulate expression, but not sufficiently for single copy chromosomal integrations to reach the maximum expression levels obtained by the use of optimized plasmids. It may be possible to overcome this bottleneck by other means (see Discussion).

Discussion

The wild-type *xylS/Pm* expression cassette has previously been found to express at least some proteins at industrial levels under high cell density conditions [13,14]. It is well known that no expression system can guarantee high levels of expression for any protein, and the reasons for this are obviously related to features of each specific gene or the protein encoded by it. It will in any case be important to use optimized inducible expression cassettes, particularly if the expression-stimulating features are acting on most genes. Most systems used for protein expression in *E. coli* are based on promoters that utilize the host RNA polymerase, with the exception of the bacteriophage T7 RNA polymerase based systems which appears to be prone to genetic instability [23].

From the results described here it is clear that the *xylS/Pm* cassette could be further improved relative to its wild-type version, and we believe that this potential has not yet been exhausted. It will clearly be important to reduce the enhanced uninduced expression level which was a consequence of the stimulated expression.

One important outcome of the studies reported here is that the improvement of the expression cassette seems to be valid at least to some extent for genes other than the one (*bla*) used to identify the variants. It appears likely that mutations stimulating transcription (X and P) might have a more universal effect than those stimulating translation (U), although we have found that sequences in the 5'-UTR region (or its corresponding DNA) may also affect transcription initiation rate [16]. The 5'-UTR may form secondary structures that potentially involve parts of the coding sequence of the gene to be expressed, but such problems can to some extent be predicted by bioinformatics analyses. It would be unrealistic to expect that all genes should respond quantitatively similar to changes in the sequences of the expression cassette, since the absolute expression levels vary over a wide range. This may explain most of the quantitative response differences observed among the genes tested in this study.

Chromosomal integration of genes to be expressed is an attractive approach in recombinant protein production, as it most likely will be associated with greater genetic stability. This approach appears to be possible for the T7 system if the growth conditions are modified and prolonged production times are considered acceptable [10], and also with a tandem used constitutive *tac* promoter system [12]. The study reported here indicates that under standard shake-flask conditions the integrated engineered XylS*/Pm* system can lead to enough transcript formation to allow expression levels comparable to those obtained by a plasmid-mediated system. Further improvement may be achieved by optimization of the growth conditions and we have also recently found (unpublished) that the level of expression is very sensitive to the intracellular concentration of the XylS regulator, a potential bottleneck that may be solved without the need for plasmids.

Conclusions

Our results clearly show that previously identified stimulatory expression control elements for the inducible *xyls/Pm* cassette can be combined in one expression vector to further improve expression levels. We found this valid also for genes other than the one used to identify the variants in the first place. For *bla* this resulted in higher expression from a single chromosomally integrated gene copy than from the corresponding wild-type plasmid system. The approach used here should be applicable also to other expression cassettes and such optimized systems should represent a good starting point for the goal of producing recombinant proteins without the need for extra-chromosomal replicons.

Methods

Strains and growth conditions

Escherichia coli DH5α (Bethesda Research Laboratories) was used as a host for cloning and expression studies unless otherwise stated. Cells were grown in Luria-Bertani (LB) broth (10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, and 5 g L^{-1} NaCl) or on LB agar (LB broth with 20 g L^{-1} agar) at 37°C, except for in expression studies, where 30°C was used. Kanamycin (50 μg μL^{-1}) was used for selection purposes.

For expression studies with *gm-csf Escherichia coli* RV308 (ATCC 31608) was used as host to facilitate comparison with previously published results [14]. Cells were grown at 30°C in HiYe medium (8.6 g L^{-1} Na_2H-$PO_4{\cdot}2H_2O$, 3 g L^{-1} KH_2PO_4, 1 g L^{-1} NH_4Cl, 0.5 g L^{-1} NaCl, 2 g L^{-1} glucose, 10 g L^{-1} glycerol, 10 g L^{-1} yeast extract, 2.5 mM $MgSO_4$, 250 μM Fe(III)-citrate, 49 μM H_3Bo_3, 79 μM $MnCl_2$, 23 μM EDTA, 9 μM $CuCl_2$, 10 μM Na_2MoO_4, 11 μM $CoCl_2$, 36 μM Zn-acetate) and induced by addition of 0.5 mM m-toluate.

Construction of plasmids and chromosomal integrations

The plasmid pTA16 was used as basis vector for construction of the combinations. pTA16 is a derivative of pIB11 [16], in which AgeI and SacI restriction sites were introduced at either end of the *xylS* gene. The *xylS* variants were obtained from the corresponding plasmids [17] upon digestion with AgeI and SacI. The *bla* gene was replaced by *celB* by the use of NdeI and BamHI restriction sites, and zby *gm-csf* by the use of NdeI and KpnI restriction sites. Agarose gel purifications were performed by QIAquick gel extraction kit (Qiagen). *Pm* and 5'-UTR variants were ordered synthetically (Eurofins MWG Operon, Germany) and were designed to carry overhangs suitable for cloning, XbaI/BspLUllI, and BspLUlll/NdeI, respectively. All constructs were confirmed by sequencing, performed by Eurofins MWG Operon (Germany).

The *xylS/Pm* expression cassette both with *bla, celB* and *gm-csf* as reporter gene was PCR-amplified with the primer set 5'-AAACACTAGTTCAGAGCTTGGAGAG-3' and 5'-CATAAAGCTGACTCTAGCTA-3' from plasmids pTA16_comXPU_bla, pTA16_comXPU_celBand pTA16_comXPU_gmcsf, respectively, and were cloned as BamHI/SpeI-fragments (*bla, celB*) or KpnI/SpeI-fragments (*gm-csf*) into the integration plasmids. Chromosomal integration of the resulting plasmids was performed as described by Haldimann and Wanner [22].

The plasmids pAH63, pAH70 and pAH95 with the corresponding helper plasmids pINT-ts (att_λ), pAH69

(att_{HK022}) and pAH121 (att_{P21}), respectively, were chosen for integration. Genomic location of the integration sites in the *E. coli* K-12 genome sequence are: for att_{λ} 806551, for att_{HK022} 1055412 and for att_{P21} 1210637 [24]. Successful integration could be confirmed for all three *bla*-constructs and for each two of the *celB*- and *gmcsf*-constructs. The integrants were named based on the pAH plasmids and the reporter genes used (Int_63_*bla*, Int_70_*bla*, Int_95_*bla*, Int_63_*celB*, Int_70_*celB*, Int_63_*gmcsf*, Int_70_*gmcsf*).

Standard DNA manipulations

Transformations of *E. coli* were performed with a modified RbCl protocol (Promega) in cloning experiments. Wizard Plus SV mini preps DNA purification kit (Promega) was used for plasmid DNA purifications. Enzymatic manipulations were performed as described by the manufacturers. PCR reactions were performed using the Expand high fidelity PCR system kit (Roche).

Enzyme assays

For enzyme assays growing strains were induced by addition of 0.1 mM m-toluate in exponential phase and samples were collected after four hours of continued growth at 30°C. Enzyme measurements (β-lactamase and phosphoglucomutase assay) were performed according to the procedures described previously [25,26]. All enzyme activity analyses were repeated at least twice, and measurements were carried out with minimum three technical recurrences.

Western blotting

Cell samples were harvested by centrifugation (8000 rpm, 5 min, 4°C). Cell lysis was performed by resuspension of the pellets in 50 mM Tris–HCl, pH 8.0, addition of the same volume of sucrose solution (40% sucrose containing 2 mM EDTA in 50 mM Tris–HCl, pH 8.0), 125 UmL^{-1} Benzonase Nuclease and 0.5 mg mL^{-1} lysozyme and incubation at room temperature with shaking for 1 h. 3x sample buffer (150 mM Tris, pH 6.8, 30% glycerol, 6% SDS, 0.3% bromophenolblue, 300 mM DTT) was added, samples were separated by SDS-PAGE (17%, ClearPage, CBS Scientific) and transferred to a nitro-cellulose membrane. The membrane was blocked with Blotto and GM-CSF was detected by His-Probe-HRP in combination with ECL-substrate (Thermo Scientific) in Kodak Image Station 2000R (Kodak).

A plasmid with the wild-type expression cassette but containing the *trfA* variant *cop271C* [27] was used as positive control. This plasmid is maintained at a copy number 3.5 times higher than for plasmids with wild-type *trfA*.

RNA isolation, cDNA synthesis and qRT-PCR

Strains were grown as described above (for enzyme assays). Cell cultures were stabilized with RNAprotect (Qiagen) prior to freezing and RNA was isolated from the frozen cell pellets using the RNAqeous kit (Ambion) as described by the manufacturers. The RNA preparations were treated with DNase (DNA-free, Ambion) and cDNA was produced from 3 mg total RNA as template using the First-Strand cDNA synthesis kit (Amersham Biosciences) with random pd(N)6 primers as described by the suppliers. Two-step qRT-PCR with the power SYBR green PCR master mix (Applied Biosystems) in a 7500 Real Time PCR System instrument (Applied Biosystems) was used for quantification of transcripts. PCR cycles were 95°C for 10 min, followed by 40 cycles of amplification (95°C for 15 s; 60°C for 1 min). Results were analysed using 7500 system software v1.3, and data were normalized by the $2^{-\Delta\Delta CT}$ method [28]. Primers were designed using the primer express software (Applied Biosystems). Primer pairs used for transcript quantification were 5'-ACGTTTTCCAATGATGAGCACTT-3' and 5'-TGCCCGGCGTCAACAC-3' for *bla*, 5'-GTCCT CTTAGTTAAATGG-3' and 5'-AGGAATCGAACCTG C-3' for *celB* and 5'-CCCTGGGAGCATGTGAATG-3' and 5'-CATCTCAGCAGCAGTGTCTCTACTC-3' for *gm-csf*. 16S rRNA gene (primer pair 5'-ATTGACGTT ACCCGCAGAAGAA-3' and 5'-GCTTGCACCCTCCG TATTACC-3') was used as a normalizer.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

FZ and RL were involved in all aspects of the experimental design and execution. All authors contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Acknowledgement

We would like to thank to Trond Erik Vee Aune for the construction of the plasmid pTA16.

Received: 21 June 2012 Accepted: 22 September 2012
Published: 2 October 2012

References

1. Kumar PK, Maschke HE, Friehs K, Schugerl K: **Strategies for improving plasmid stability in genetically modified bacteria in bioreactors.** *Trends Biotechnol* 1991, **9:**279–284.
2. Diaz Ricci JC, Hernandez ME: **Plasmid effects on *Escherichia coli* metabolism.** *Crit Rev Biotechnol* 2000, **20:**79–108.
3. Hoffmann F, Rinas U: **Stress induced by recombinant protein production in *Escherichia coli*.** *Adv Biochem Eng Biotechnol* 2004, **89:**73–92.
4. Bailey JE, Da Silva NA, Peretti SW, Seo JH, Srienc F: **Studies of host-plasmid interactions in recombinant microorganisms.** *Ann N Y Acad Sci* 1986, **469:**194–211.
5. Dahlberg C, Chao L: **Amelioration of the cost of conjugative plasmid carriage in *Eschericha coli* K12.** *Genetics* 2003, **165:**1641–1649.
6. Seo JH, Bailey JE: **Effects of recombinant plasmid content on growth properties and cloned gene product formation in *Escherichia coli*.** *Biotechnol Bioeng* 1985, **27:**1668–1674.
7. Summers DK: **The kinetics of plasmid loss.** *Trends Biotechnol* 1991, **9:**273–278.

8. Wang Z, Xiang L, Shao J, Wegrzyn A, Wegrzyn G: **Effects of the presence of ColE1 plasmid DNA in *Escherichia coli* on the host cell metabolism.** *Microb Cell Fact* 2006, **5:**34.
9. Baneyx F: **Recombinant protein expression in *Escherichia coli*.** *Curr Opin Biotechnol* 1999, **10:**411–421.
10. Striedner G, Pfaffenzeller I, Luchner M, Nemecek S, Grabherr R, Bayer K: **Plasmid-Free T7-Based *Escherichia coli* Expression Systems.** *Biotechnol Bioeng* 2010, **105:**786–794.
11. Tyo KE, Ajikumar PK, Stephanopoulos G: **Stabilized gene duplication enables long-term selection-free heterologous pathway expression.** *Nat Biotechnol* 2009, **27:**760–765.
12. Li M, Wang J, Geng Y, Li Y, Wang Q, Liang Q, Qi Q: **A strategy of gene overexpression based on tandem repetitive promoters in *Escherichia coli*.** *Microb Cell Fact* 2012, **11:**19.
13. Sletta H, Nedal A, Aune TE, Hellebust H, Hakvag 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:**7033–7039.
14. Sletta H, Tondervik A, Hakvag 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:**906–912.
15. Bakke I, Berg L, Aune TE, Brautaset T, Sletta H, Tondervik A, Valla S: **Random mutagenesis of the *Pm* promoter as a powerful strategy for improvement of recombinant-gene expression.** *Appl Environ Microbiol* 2009, **75:**2002–2011.
16. 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:**379–389.
17. Vee Aune TE, Bakke I, Drablos F, Lale R, Brautaset T, Valla S: **Directed evolution of the transcription factor XylS for development of improved expression systems.** *Microb Biotechnol* 2010, **3:**38–47.
18. Lale R, Berg L, Stuttgen 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:**2648–2655.
19. Ramos JL, Stolz A, Reineke W, Timmis KN: **Altered effector specificities in regulators of gene expression: TOL plasmid *xylS* mutants and their use to engineer expansion of the range of aromatics degraded by bacteria.** *Proc Natl Acad Sci USA* 1986, **83:**8467–8471.
20. 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:**79–91.
21. 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:**35–51.
22. Haldimann A, Wanner BL: **Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria.** *J Bacteriol* 2001, **183:**6384–6393.
23. Vethanayagam JG, Flower AM: **Decreased gene expression from T7 promoters may be due to impaired production of active T7 RNA polymerase.** *Microb Cell Fact* 2005, **4:**3.
24. Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, *et al*: **The complete genome sequence of *Escherichia coli* K-12.** *Science* 1997, **277:**1453–1462.
25. 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.
26. Fjaervik E, Frydenlund K, Valla S, Huggirat Y, Benziman M: **Complementation of cellulose-negative mutants of *Acetobacter xylinum* by the cloned structural gene for phosphoglucomutase.** *FEMS Microbiol Lett* 1991, **77:**325–330.
27. Durland RH, Toukdarian A, Fang F, Helinski DR: **Mutations in the *trfA* replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy numbers.** *J Bacteriol* 1990, **172:**3859–3867.
28. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) Method.** *Methods* 2001, **25:**402–408.

doi:10.1186/1475-2859-11-133
Cite this article as: Zwick *et al.*: Strong stimulation of recombinant protein production in *Escherichia coli* by combining stimulatory control elements in an expression cassette. *Microbial Cell Factories* 2012 **11**:133.

Paper II

Is not included due to copyright

Paper III

Is not included due to copyright

Paper IV

Is not included due to copyright

Paper V

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Apr. 2011, p. 2648–2655 Vol. 77, No. 8
0099-2240/11/$12.00 doi:10.1128/AEM.02091-10

Continuous Control of the Flow in Biochemical Pathways through 5′ Untranslated Region Sequence Modifications in mRNA Expressed from the Broad-Host-Range Promoter *Pm*▽†

Rahmi Lale,[1] Laila Berg,[1] Friederike Stüttgen,[1] Roman Netzer,[2] Marit Stafsnes,[1] Trygve Brautaset,[2] Trond Erik Vee Aune,[1]‡ and Svein Valla[1]*

Department of Biotechnology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway,[1] *and Department of Biotechnology, SINTEF Materials and Chemistry, SINTEF, 7465 Trondheim, Norway*[2]

Received 3 September 2010/Accepted 8 February 2011

The inducible *Pm* promoter integrated into broad-host-range plasmid RK2 replicons can be fine-tuned continuously between the uninduced and maximally induced levels by varying the inducer concentrations. To lower the uninduced background level while still maintaining the inducibility for applications in, for example, metabolic engineering and synthetic (systems) biology, we report here the use of mutations in the *Pm* DNA region corresponding to the 5′ untranslated region of mRNA (UTR). Five UTR variants obtained by doped oligonucleotide mutagenesis and selection, apparently reducing the efficiency of translation, were all found to display strongly reduced uninduced expression of three different reporter genes (encoding β-lactamase, luciferase, and phosphoglucomutase) in *Escherichia coli*. The ratio between induced and uninduced expression remained the same or higher compared to cells containing a corresponding plasmid with the wild-type UTR. Interestingly, the UTR variants also displayed similar effects on expression when substituted for the native UTR in another and constitutive promoter, *P1* ($P_{antitet}$), indicating a broad application potential of these UTR variants. Two of the selected variants were used to control the production of the C_{50} carotenoid sarcinaxanthin in an engineered strain of *E. coli* that produces the precursor lycopene. Sarcinaxanthin is produced in this particular strain by expressing three *Micrococcus luteus* derived genes from the promoter *Pm*. The results indicated that UTR variants can be used to eliminate sarcinaxanthin production under uninduced conditions, whereas cells containing the corresponding plasmid with a wild-type UTR produced ca. 25% of the level observed under induced conditions.

The initially used methods of deleting or overexpressing genes have been demonstrated to be inadequate for many applications in metabolic engineering (21, 31, 32). For example, when the goal is to optimize the expression level of a desired protein by engineering the relevant metabolic pathway, it might be necessary to change the expression of multiple enzymes simultaneously and to different levels (30, 38). Also, reducing the formation of particular by-products can increase the flux of the desired product (29). In addition, low basal expression is critical for applications such as the expression of toxic genes, metabolic engineering, and control analysis (2, 14, 27, 34). This has led to an increased focus on development of genetic tools to fine-tune gene expression to the desired levels. A commonly used strategy is to make so-called promoter libraries of constitutive promoters (1, 15, 17). Such promoters seem to be preferred over the corresponding inducible ones for industrial scale productions because of factors such as inducer costs, sensitivity to inducer concentration, and heterogeneity of expression caused by an all-or-none effect of induction (1, 19). However, the all-or-none induction effect may be eliminated if the inducer enters the cell interior by passive diffusion (20). Thus, regulatable promoter systems that eliminate as many problems as possible associated with their use are important tools, particularly considering the recent growing interest in systems and synthetic biology.

The broad-host-range XylS/*Pm*-positive regulator/promoter system (originating from the TOL plasmid [pWWO] of *Pseudomonas putida*) was previously in our laboratory inserted into a set of broad-host-range expression vectors (pJBn), together with the RK2 minimal replication elements *oriV* (origin of vegetative replication) and *trfA* (encodes replication initiation and copy number control protein) (6, 7). The expression level from the σ^{32}- and σ^{38}-dependent *Pm* promoter (26) can easily be varied by the specific benzoate derivative used as inducer (enters the cells by passive diffusion) and the concentration of it (33, 42). Further modifications of expression can be obtained by varying the vector copy number via mutations in *trfA*. This vector system has been used successfully in our group for metabolic engineering purposes in *Pseudomonas fluorescens* (4, 13) and also in combination with another newly developed expression system (40). Finally, one of the pJBn expression vectors has been shown to give rise to production of industrial levels of several recombinant proteins in *Escherichia coli* (36, 37). Recently, we have found that the *bla* gene (encoding β-lactamase) expression from the *Pm* promoter can be strongly stimulated in *E. coli* (up to 20-fold at the protein product level) by introducing mutations at the DNA level in its cognate 5′ untranslated region (UTR) (5). The inducible promoter phenotype was intact in all of the UTR variants.

* Corresponding author. Mailing address: Department of Biotechnology, Norwegian University of Science and Technology, Sem Sælands vei 6/8, N-7491 Trondheim, Norway. Phone: 47-73598694. Fax: 47-73591283. E-mail: svein.valla@biotech.ntnu.no.

‡ Present address: Vectron Biosolutions AS, P.O. Box 1262, Pirsenteret, 7462 Trondheim, Norway.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 18 February 2011.

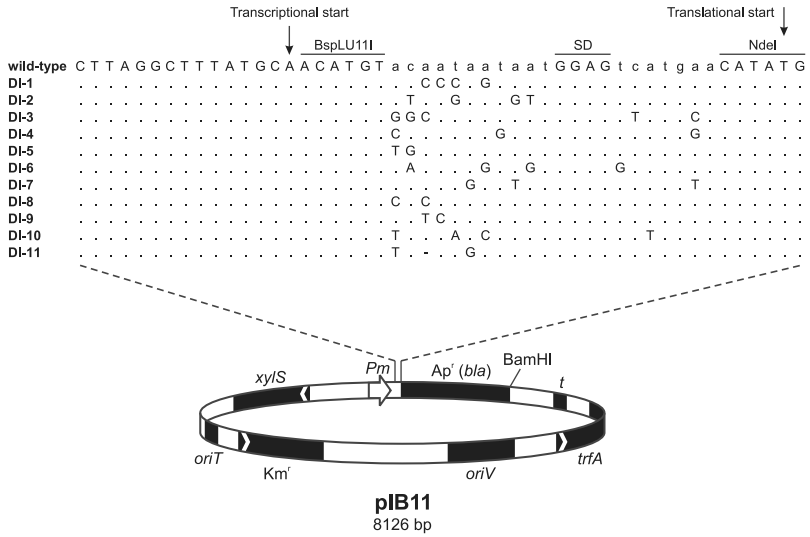

FIG. 1. Downregulating UTR sequences identified in DI library (pIB11), and map of the plasmid pIB11. Putative Shine-Dalgarno sequences are indicated by "SD". Identical bases are indicated by dots, and deletions are indicated by dashes (DI-11). Transcriptional and translational start sites are indicated with arrows, and bases subjected to mutagenesis are shown in lowercase. Features of the plasmid are shown in the map below the sequences. The indicated restriction enzyme sites are unique. *bla*, gene encoding β-lactamase; Km^r, kanamycin resistance gene; *trfA*, gene encoding the essential replication protein; *xylS*, gene encoding the transcriptional activator of the *Pm* promoter; *oriV*, origin of vegetative replication; *oriT*, origin of transfer; *t*, bidirectional transcriptional terminator.

The purpose of these earlier studies was to maximize protein expression, but we report here on a study of UTR variants that can be used to downregulate the uninduced gene expression level of different reporter genes both from *Pm* without losing inducibility and also from the constitutive *P1* promoter. The variants could also, in contrast to wild-type *Pm*, be used to tightly control the metabolic flow in the pathway, leading to synthesis of the carotenoid sarcinaxanthin.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. *E. coli* DH5α (Bethesda Research Laboratories) was used as a host, except for in the construction and screening of the DI UTR library and for heterologous production of carotenoids *E. coli* Gold (Stratagene) and *E. coli* XL1-Blue were used, respectively. All of the plasmid constructs used in the present study bear the same backbone, consisting of a RK2-based minimal replication system in which the *xylS*/*Pm* expression cassette has been integrated. The vectors used in the expression studies vary either with respect to the gene to be expressed (*bla* in plasmid pIB11 [5], *celB* in pLB11, and *luc* in pKT1) or to the promoter used (*P1*, also known as $P_{antitet}$ [9]) in pLB9. Details on the growth conditions and plasmid constructions can be found in the supplemental material.

Standard techniques. Plasmid isolation, enzymatic manipulations of DNA, and agarose gel electrophoresis were performed by the methods described elsewhere (35). The QIAquick gel extraction kit and QIAquick PCR purification kit (Qiagen) was used for DNA purifications from agarose gels and enzymatic reactions, respectively. A modified RbCl protocol (Promega) was used for transformations of *E. coli* in cloning experiments. PCR for cloning and amplified fragment length identification was performed by using the Expand High-Fidelity PCR-system (Roche Applied Science) and Dynazyme II (Finnzymes), respectively. Information on the primers used for PCR and DNA sequencing is given in the supplemental material.

Construction and screening of the UTR libraries. The DI library (consisting of about 22,500 transformants) was constructed in plasmid pIB11. The oligonucleotides used to construct the libraries consisted of the wild-type sequence and the complementary strand that was made by synthesis of randomly doped synthetic oligonucleotide mixture. The doping percentages of the nucleotide mixture were set to 76% of the wild-type nucleotide and 8% for each of the three others. Further details on the library construction can be found elsewhere (3, 5).

RNA isolation, cDNA synthesis, DNase treatment, and two-step qRT-PCR. Transcript amounts were determined by two-step quantitative real-time PCR (qRT-PCR). For further details, see the supplemental material.

Enzyme assays. Luciferase assays were performed by using the Luciferase Assay System (Promega) according to the manufacturer's protocol. Luciferase activity was measured with the GloMax 20/20 Luminometer (Promega). Phosphoglucomutase activities were measured as described previously (12). For all enzyme assays, measurements were carried out at least with two biological and three technical replicas.

Western blotting. SDS-PAGE was carried out according to the method of Laemmli (22), using 12% Tris-HCl Ready Gel (Bio-Rad). For further details, see the supplemental material.

Carotenoid analysis. Carotenoids produced by recombinant *E. coli* strains were extracted from liquid cell cultures and analyzed by liquid chromatography-mass spectrometry (LC-MS) analysis as described elsewhere (28).

RESULTS AND DISCUSSION

The expression level from *Pm* can be strongly reduced by mutating the UTR DNA region flanking the Shine-Dalgarno sequence. It is well established that mutations located near the transcriptional start site may affect expression at the transcriptional level and that the Shine-Dalgarno sequence is important for translational efficiency (8, 23, 24, 41). In our previous studies we found that nucleotide substitutions in the remaining parts of the UTR can also strongly affect expression (5), and for the study reported here we constructed a new UTR mutant library (designated DI, Fig. 1) in which the nucleotides close to the transcriptional start site and the Shine-Dalgarno sequence were not mutagenized. Plasmid pIB11 (Fig. 1) was used for the library construction, and β-lactamase was used as a reporter to

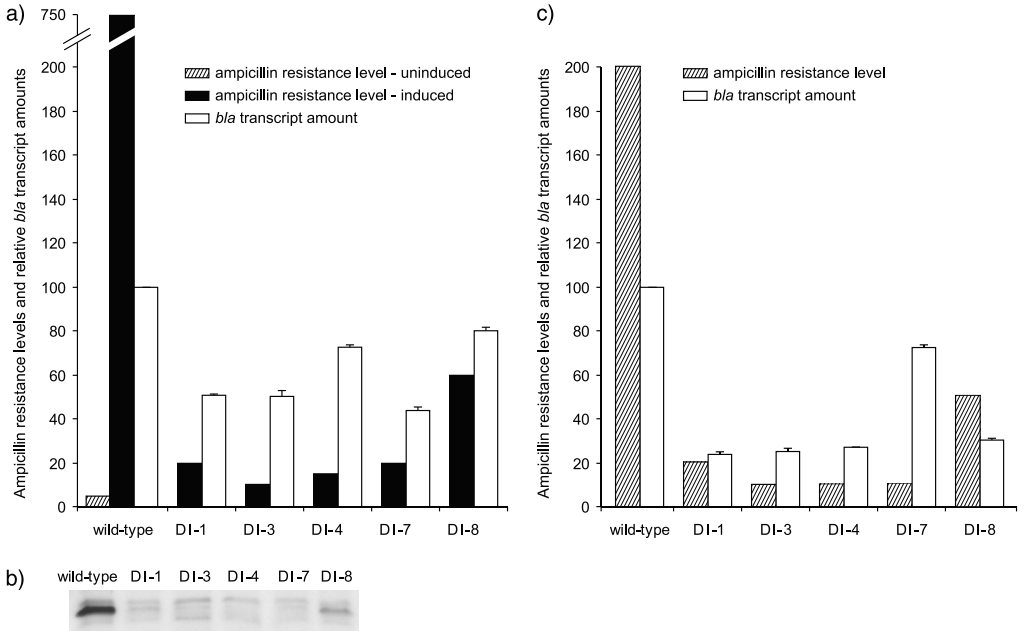

FIG. 2. Ampicillin resistance levels (under uninduced [▨] and induced [■] conditions) and *bla* transcript amounts (□) for the DI UTR variants in plasmid pIB11 (a) and pLB9 (c). (b) Western blot analysis of β-lactamase production from the strains used in panel a. The values in panels a and c are the average of at least two biological replicas, and the error bars indicate the standard deviations. Transcript amounts are normalized against the wild-type value, arbitrarily set to 100.

easily detect variation in expression levels. *E. coli* DH5α cells containing pIB11 (wild-type UTR) were found to grow on agar medium supplemented with maximally about 750 μg of ampicillin/ml under induced conditions. From a total of 4,773 randomly picked clones (colonies on agar medium supplied with kanamycin), 11 (DI-1 to -11, Fig. 1) were finally selected, based on their ampicillin resistance phenotypes and DNA sequences (see the supplemental material for further details concerning the selection). The corresponding UTR oligonucleotides were resynthesized, cloned into pIB11, and reconfirmed by DNA sequencing and phenotype analyses. Each variant carried two to five point mutations, and a total of 26 different substitutions were identified, involving 14 of the 18 bases in the wild-type mutagenized sequence.

UTR mutations seem to mostly act by reducing the translational efficiency. Five (DI-1, -3, -4, -7, and -8) of the 11 variants described above were subjected to further detailed studies. The maximum ampicillin tolerance levels of the *E. coli* DH5α host cells containing the corresponding plasmids were found to vary from 10 (pIB11-DI-3) to 60 μg/ml (pIB11-DI-6) under induced conditions, and under uninduced conditions none of the clones grew in the presence of 5 μg of ampicillin/ml on agar media (Fig. 2a), in contrast to cells containing pIB11 (wild-type UTR). Therefore, based on these results, it seemed clear that the variants have kept their inducible phenotypes, even though the exact uninduced maximal ampicillin resistance levels of them could not be determined because the background tolerance of the host cells themselves is too close to 5 μg/ml. Direct measurements of the β-lactamase activities were not feasible due to the relatively low sensitivity of the enzyme assay. Therefore, Western analysis was performed with β-lactamase antibody, and the results indicated that the ampicillin resistance results were consistent with the band intensities on the membrane (Fig. 2b).

If the mutations in the UTR are mainly leading to a reduction of the efficiency of translation, one would predict that the amounts of transcript might not be reduced to an extent comparable to that of the protein level. To investigate this, we directly measured the *bla* transcript levels for the selected five UTR variants by qRT-PCR and found that the amounts of accumulated *bla* mRNA varied between 43% (pIB11-DI-7) and 79% (pIB11-DI-8) of the wild-type quantity (Fig. 2a). These results are thus consistent with the hypothesis that the observed reduction of expression at the protein level is mainly the result of reduced translational efficiency.

UTR mutations also downregulate the expression level of two other tested reporter genes. The effects of the mutations in the UTR could be envisioned to display some gene dependency due to interactions with the coding sequence of the gene to be expressed. To monitor this, we substituted the *bla* gene with *luc* (encoding luciferase originating from *Photinus pyralis*, plasmid pKT1) and then measured the corresponding expression levels from plasmids containing each of the UTR variants DI-1, -3, -4, -7, and -8 in *E. coli* DH5α under induced conditions (Fig. 3). All of the DI variants resulted in a very strong reduction of expression at the protein levels ranging from 0.4%

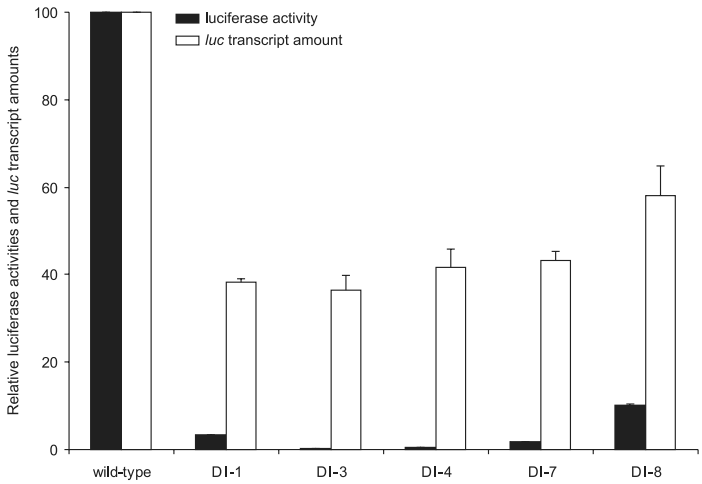

FIG. 3. Luciferase activities (■) and *luc* transcript amounts (□) for the DI UTR variants in plasmid pKT1 under induced conditions. The values are the average of at least two biological replicas, and the error bars are standard deviations. Both enzyme activities and transcript amounts are relative to the wild-type value, arbitrarily set to 100.

(pKT1-DI-3) to 11% (pKT1-DI-8) of the value obtained with pKT1 (wild-type UTR), while the corresponding accumulated transcript levels were found to range from 36% (pKT1-DI-3) to 58% (pKT1-DI-8) relative to pKT1 (wild-type UTR). It could therefore be concluded that the mutations in the UTR give rise to similar effects on the expression level of both *luc* and *bla*, further supporting the hypothesis that the UTR mutations primarily act by negatively affecting the translational efficiency in an apparently gene-independent manner.

Previously, we have found that the phosphoglucomutase gene *celB* (originating from *Acetobacter xylinum*, now *Gluconacetobacter hansenii*) can be expressed at unusually high levels from the wild-type XylS/*Pm* system in *E. coli* and other Gram-negative bacteria (7). *celB* was therefore used as a final reporter test gene to study the effect of the UTR variants on expression. The wild-type UTR DNA sequence in pLB11 was substituted by each of the UTR variants DI-1, -3, -4, -7, and -8, and the CelB protein activity levels of the corresponding *E. coli* DH5α host cells under induced *Pm* conditions were measured. Surprisingly, no CelB activity could be detected from any of the recombinant strains. The corresponding *celB* transcript levels were determined by qRT-PCR (Fig. 4a), and they were dra-

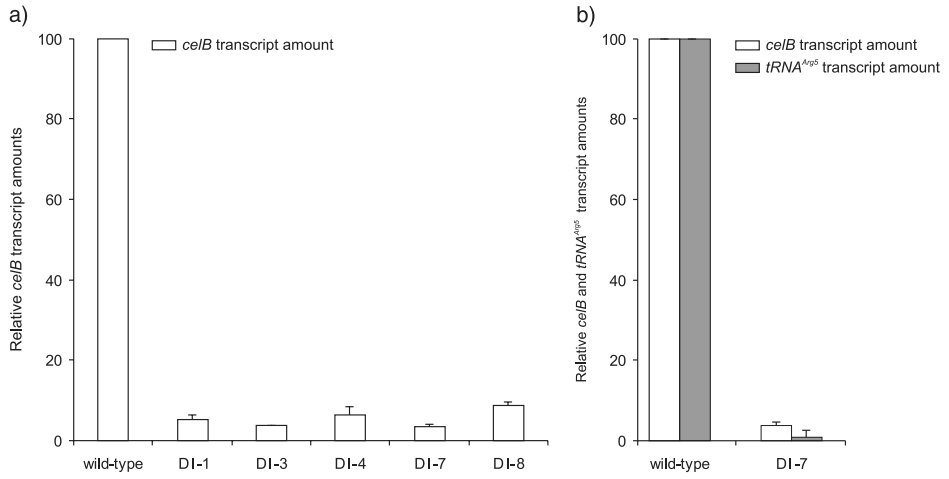

FIG. 4. *celB* transcript amounts (□) for the DI UTR variants in plasmid pLB11 (a) and *celB* (□) and $tRNA^{Arg5}$ (▤) transcript amounts for the plasmid pLB11Arg5 with the DI-7 sequence (b), relative to the wild type under induced conditions. The wild-type values are arbitrarily set to 100. The values are the average of at least two biological replicas, and the error bars indicate the standard deviations.

matically reduced for all five UTR variants tested, in contrast to what was observed with the *bla* and *luc* reporter genes (see above). *celB* therefore represents an example of a gene responding in a different way (compared to *bla* and *luc*) to the variant UTR sequences, although these gene-specific effects mainly appear to act at the transcriptional level. We also found it intriguing that the effects were similar for all of the UTR variants. If they were somehow caused by an interaction between the UTR and the *celB* gene coding sequence, one would not expect a similar response for all of the UTR variants.

The gene-specific reduction of *celB* transcript accumulation caused by mutations in the UTR is probably the result of enhanced Rho-dependent transcriptional termination. In the absence of translating ribosomes, transcripts are potentially more susceptible to degradation (11), and we therefore reasoned that the proposed drastic reduction of translational efficiency by the UTR mutations might somehow selectively lead to rapid degradation of *celB* mRNA. Since any gene-specific effect of the UTR variants could potentially become important in their future applications, we decided to study the somewhat unexpected specific effects on *celB* transcript accumulation in more detail. For this purpose, we used a method that involves fusion of the gene of interest (in this case *celB*) to a tRNA gene (encoding $tRNA^{Arg5}$) in such a way that both genes become expressed as one single transcript (generating plasmid pLB11Arg5). This experiment has been used previously by us for a related problem (5) and is based on the assumption that the $tRNA^{Arg5}$ part of the mRNA is essentially stable *in vivo* (25). Thus, even a drastic increase in the *celB* transcript turnover should not negatively affect the accumulation of the $tRNA^{Arg5}$ part of the transcript. The phosphoglucomutase activities and accumulated *celB* transcript levels under induced conditions of *E. coli* DH5α cells containing the plasmid pLB11Arg5 were found to be similar to those containing pLB11 (data not shown). The wild-type UTR DNA sequence of the plasmid pLB11Arg5 was then substituted with the corresponding DI-7 sequence, and the accumulated *celB* and $tRNA^{Arg5}$ transcript levels were determined by qRT-PCR and compared to the corresponding values from the pLB11Arg5 (Fig. 4b). These experiments clearly demonstrated that the accumulated $tRNA^{Arg5}$ transcript level was also strongly reduced in the presence of the DI-7 mutations (reduced to ca. 2% compared to the wild type), indicating that transcription of *celB* is inefficient or aborted prior to full-length synthesis. If so, one would also conclude that mRNA instability is probably not the primary cause of the selective reduction of transcript accumulation for the *celB* transcripts.

It is known that the absence of translating ribosomes on a transcript might activate mechanisms that prematurely terminate the transcription process (16). We therefore hypothesized that the *celB* transcript may contain a binding site for the transcription termination factor Rho in its coding sequence, thus potentially also explaining why all of the UTR variants act similarly and selectively on the *celB* gene (10). Bicyclomycin (BCM) is a natural inhibitor of Rho (43), and this compound was added to a final concentration of 100 μg/ml at the time of induction to cell cultures of *E. coli* DH5α containing pLB11-DI-7 or pLB11. As controls, the same two cultures were grown similarly but without the addition of BCM. The accumulated

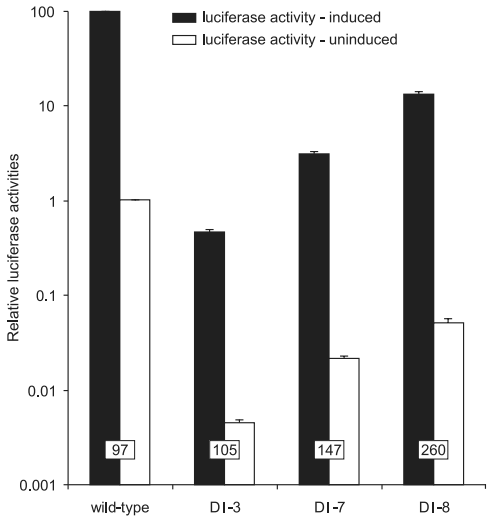

FIG. 5. Luciferase activities under induced (■) and uninduced (□) conditions for selected DI UTR variants in plasmid pKT1. The ratios between the induced and uninduced luciferase activity levels are displayed directly on the bars. Enzyme activities are shown in logarithmic scale. The values are the average of at least two biological replicas, and the error bars indicated the standard deviations. Activity values are relative to the wild type under induced conditions, which is arbitrarily set to 100 for the induced state.

celB transcript levels were then determined in the samples collected 30 min after the time of induction. Interestingly, in the strain with pLB11-DI-7 the accumulated transcript level under induced conditions was ~40-fold higher in the presence of BCM than in its absence (similar to cells containing pLB11 in the absence of BCM, data not shown). BCM also had a stimulatory effect on wild-type transcription, but to a lower extent (~10-fold). This indicates that Rho leads to premature transcription termination of the *celB* mRNA synthesis and that this effect is strongly stimulated when translation is inefficient (in the presence of the DI DNA sequences).

Based on this, it seemed likely that the primary effect of the UTR mutations is to cause reduced translational efficiency of any gene, but that this can lead to drastic reduction also of transcription for genes that are sensitive to early Rho-dependent termination of transcription. Importantly, such effects should not have any obvious negative impact for the generality of application of the UTR variants for control of biochemical pathways (see below).

The inducibility of *Pm* was maintained or improved by the UTR mutations. The luciferase enzyme assay is very sensitive compared to those of β-lactamase and CelB, and activities can be accurately monitored at both very low and high levels of expression. Therefore, we compared these enzyme activities in the host cells containing plasmids with three of the DI UTR variant sequences (pKT1-DI-3, -7, and -8) under induced (the same data are in Fig. 3) and uninduced (Fig. 5) conditions. The ratio between the induced and uninduced level of protein formation for the UTR variants ranged from 105 (pKT1-DI-3) to 260 (pKT1-DI-8), which is similar to or higher than for the wild

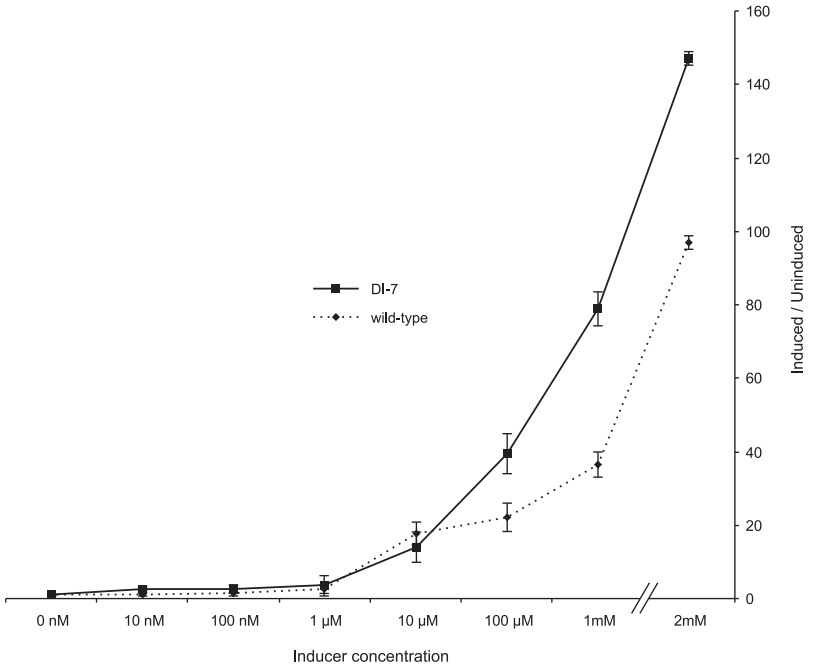

FIG. 6. Luciferase activities under different inducer concentrations for the wild type and UTR variant DI-7 in plasmid pKT1. The ratios between the induced and uninduced luciferase activity levels are plotted against the inducer concentrations. The values are the average of two biological replicas, and the error bars indicate the standard deviations. The activity values of the wild type and DI-7 are relative to their cognate uninduced levels.

type (ratio of 97). Furthermore, the induced expression levels of pKT1-DI-7 and pKT1-DI-8 are higher than the corresponding uninduced level from pKT1.

Previously, it has been demonstrated that the expression level from *Pm* can easily be controlled by varying the type and concentration of the inducer (42). We therefore also characterized the inducibility characteristics of pKT1 and pKT1-DI-7 at a range (0.01, 0.1, 1, 10, 100, 1,000, and 2,000 μM) of inducer concentrations (Fig. 6). This analysis also showed that the system can be turned on at levels as low as 1 μM, resulting in induction ratios of 1.9 and 2.6 for pKT1-DI-7 and pKT1, respectively. Furthermore, the induction levels increase gradually and according to a similar pattern for both wild type and variant as the inducer concentrations are increased. These results therefore clearly confirmed that UTR mutations can be applied to reduce the basal expression level without losing the inducibility feature. If we take into account the previously reported UTR variants leading to high expression levels (5), it also follows that expression can be continuously regulated over 5 orders of magnitude in this system.

UTR variants also lead to reduced gene expression when fused to the constitutive *P1* promoter. Although the properties of the UTR variants are clearly very useful in the XylS/*Pm* context they would be even more applicable if they would have a similar function in as many promoter systems as possible. To test this, we chose promoter *P1* (also known as $P_{antitet}$), which differs from *Pm* in that it is constitutive and σ^{70} dependent. It should therefore represent a good example for evaluation of the generality of the variant UTR functions. The native UTR in *P1* was thus substituted with the *Pm* wild-type UTR and the selected variants, using the same plasmid backbones as in the experiments described above and using *bla* as reporter. Determination of maximum ampicillin tolerance level of the corresponding host cells demonstrated a pattern that was nearly identical to that observed for the corresponding *Pm*-based constructs (Fig. 2c). This clearly indicates that the UTR variants have a much broader application potential than for the XylS/*Pm* system only. We also measured the accumulated transcript levels, and in all cases the amounts of transcript were also reduced, as in the XylS/*Pm* system. The relative patterns were in this case not as strikingly similar as at the protein level (see, for example, the outcomes for the DI-7 and -8 variants), but this is not necessarily surprising, since factors such as transcript formation kinetics from *Pm* and *P1* may not be similar. Note also that while host ampicillin resistance tolerances should allow for direct comparisons between Fig. 2a and c, the same is not the case for the transcript levels.

UTR variants DI-3 and DI-8 can be used to control heterologous production of the C_{50} carotenoid sarcinaxanthin in *E. coli*. Carotenoids represent a diverse class of natural molecules with numerous medical and industrial applications, and it is of interest to be able to produce such compounds in heterologous hosts, including *E. coli*. We recently demonstrated that the XylS/*Pm* promoter system can be used for efficient heterolo-

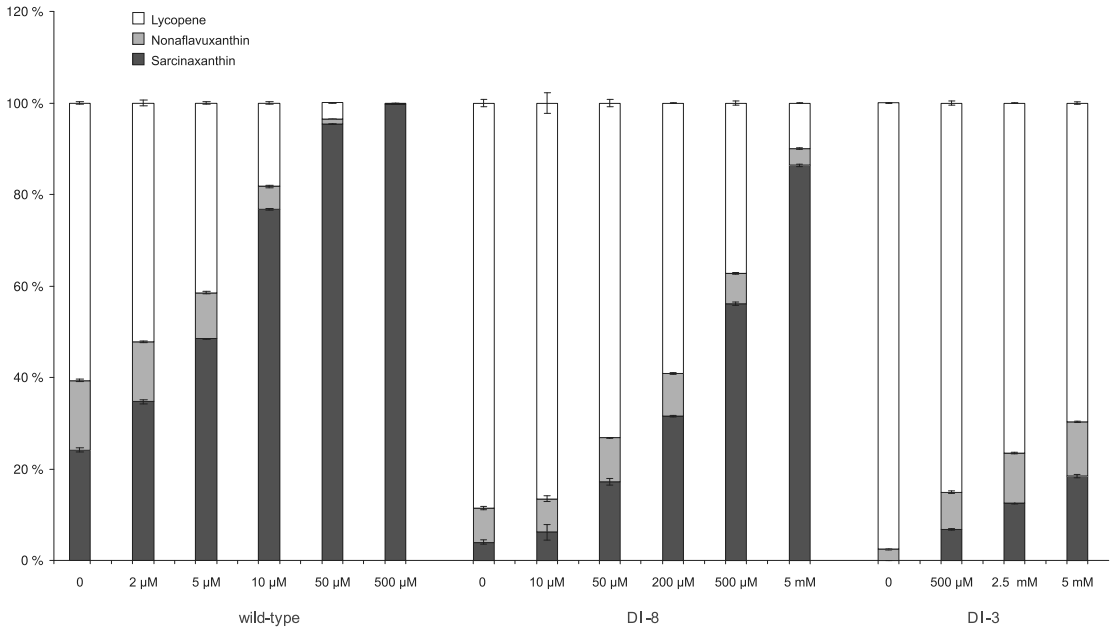

FIG. 7. Relative production levels of sarcinaxanthin in a lycopene-producing *E. coli* host expressing sarcinaxanthin biosynthetic genes from *Pm* and with different UTR variants (wild type, DI-8, and DI-3). Cells were cultivated in liquid cultures, and different concentrations of the *Pm* inducer *m*-toluate were added. Samples were harvested 24 h after induction and analyzed for carotenoid contents (for details see Materials and Methods).

gous production of the C_{50} carotenoid sarcinaxanthin in a constitutively lycopene (precursor for sarcinaxanthin)-producing strain of *E. coli* (28). Three *M. luteus* genes—*crtE2*, *crtYg*, and *crtYh*—were placed under the control of the *Pm* promoter and were introduced in a lycopene-producing *E. coli* host on a plasmid (28). It was observed that under induced conditions all of the lycopene in the cells was converted into sarcinaxanthin (2.3 mg per g of cell [dry weight]) (28), but substantial amounts of sarcinaxanthin were also produced in the absence of *Pm* induction. This system therefore represented a good test case for investigation of the applicability of the UTR variants to control a metabolic pathway, preferentially at all levels from zero production to the maximum level.

As tools for this purpose, we constructed plasmids pKT1-DI-8-CRT-E2YgYh-2665 and pKT1-DI-8-CRT-E2YgYh-2665 (see the supplemental material) that are both analogous to plasmid pCRT-E2YgYh-2665 (wild-type UTR) (28) but with the *Pm* UTR region substituted with the UTR variants DI-8 and DI-3, respectively. Both plasmids were transformed into the lycopene-producing strain *E. coli* XL1-Blue(pLYC), and the resulting recombinant strains were subjected to sarcinaxanthin production analyses in shake flasks essentially as described previously (28). Cultivations were performed in the presence of different concentrations of the *Pm* inducer, and carotenoid production was investigated by LC-MS analysis of cell extracts (Fig. 7). For the control strain (harboring the plasmid with the wild-type UTR), ca. 25% of the totally extracted carotenoid was identified as sarcinaxanthin in the absence of induction, and the remaining carotenoid fraction was identified as precursor carotenoids lycopene and nonaflavuxanthin. In the presence of a 500 μM concentration of inducer the entire extracted carotenoid fraction was sarcinaxanthin, and no lycopene or nonaflavuxanthin was detected. Note also that the total extracted carotenoid amount was similar for all analyzed samples (2.3 mg per g of cell [dry weight]).

For strain *E. coli* XL1-Blue(pLYC)(pKT1-DI-8-CRTE2YgYh-2665) grown under uninduced conditions, only 5% of the total carotenoid fraction was sarcinaxanthin, while 95% remained as C_{50} carotenoid precursors lycopene and nonaflavuxanthin. The sarcinaxanthin fraction increased gradually with higher inducer concentrations used during cultivation. In the presence of 500 μM inducer, the extracted carotenoid fraction consisted of 53% sarcinaxanthin. A further 10-fold increase in inducer concentration (5 mM) improved sarcinaxanthin production up to 90% of the totally extracted carotenoids. Furthermore, in strain *E. coli* XL1-Blue(pLYC)(pKT1-DI-3-CRTE2YgYh-2665) no sarcinaxanthin was detected under uninduced conditions, whereas the maximal production (5 mM inducer) was only 17% compared to that of the wild-type UTR control strain under the same conditions. Together, these data demonstrated that the UTR variants can be used to fine-tune the expression of a set of biosynthetic genes, resulting in controlled production of sarcinaxanthin at any level from zero and up to complete conversion of all precursors produced by the cells.

The importance of balanced gene expression has been shown for recombinant lycopene production in *E. coli*. Utilization of a low-copy-number plasmid instead of a high-copy-

number plasmid for arabinose-inducible overexpression of the rate-limiting enzyme DXS (1-deoxy-D-xylulose 5-phosphate synthase) enhanced lycopene production 2- to 3-fold and demonstrated that the overexpression of *dxs* at intracellular concentrations that exceed the availability of its glycolytic precursors pyruvate and G3P causes a significant metabolic burden for the cell and impaired growth significantly (18). This example demonstrates that genetic tools for fine-tuned regulation of gene expression can play an important role for the heterologous high-level production of complex compounds. The idea of using UTR variants to control metabolic pathways is not completely new (39), but the specific approach described here has the advantage that one can select specific UTR features that almost certainly cannot be rationally predicted.

ACKNOWLEDGMENTS

We thank Espen Fjærvik for his help in the robotic screening procedure. Bicyclomycin was obtained from Astellas Pharma, Inc. (Japan).

This study was supported by the Norwegian Research Council (grant 140533).

REFERENCES

1. **Alper, H., C. Fischer, E. Nevoigt, and G. Stephanopoulos.** 2005. Tuning genetic control through promoter engineering. Proc. Natl. Acad. Sci. U. S. A. **102:**12678–12683.
2. **Anthony, L. C., H. Suzuki, and M. Filutowicz.** 2004. Tightly regulated vectors for the cloning and expression of toxic genes. J. Microbiol. Methods **58:**243–250.
3. **Bakke, I., 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.
4. **Bakkevig, K., et al.** 2005. Role of the *Pseudomonas fluorescens* alginate lyase (AlgL) in clearing the periplasm of alginates not exported to the extracellular environment. J. Bacteriol. **187:**8375–8384.
5. **Berg, L., R. Lale, I. Bakke, N. Burroughs, and S. Valla.** 2009. The expression of recombinant genes in be strongly stimulated at the transcript production level by mutating the DNA-region corresponding to the 5′-untranslated part of mRNA. Microbial Biotechnol. **2:**379–389.
6. **Blatny, J. M., T. Brautaset, H. C. Winther-Larsen, K. Haugan, and S. Valla.** 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.
7. **Blatny, J. M., T. Brautaset, H. C. Winther-Larsen, P. Karunakaran, and S. Valla.** 1997. Improved broad-host-range RK2 vectors useful for high and low regulated gene expression levels in gram-negative bacteria. Plasmid **38:**35–51.
8. **Brautaset, T., S. B. Petersen, and S. Valla.** 2000. In vitro determined kinetic properties of mutant phosphoglucomutases and their effects on sugar catabolism in *Escherichia coli*. Metab. Eng. **2:**104–114.
9. **Brosius, J., R. L. Cate, and A. P. Perlmutter.** 1982. Precise location of two promoters for the beta-lactamase gene of pBR322. S1 mapping of ribonucleic acid isolated from *Escherichia coli* or synthesized in vitro. J. Biol. Chem. **257:**9205–9210.
10. **Ciampi, M. S.** 2006. Rho-dependent terminators and transcription termination. Microbiology **152:**2515–2528.
11. **Deana, A., and J. G. Belasco.** 2005. Lost in translation: the influence of ribosomes on bacterial mRNA decay. Genes Dev. **19:**2526–2533.
12. **Fjærvik, E., K. Frydenlund, S. Valla, Y. Huggirat, and M. Benziman.** 1991. Complementation of cellulose-negative mutants of *Acetobacter xylinum* by the cloned structural gene for phosphoglucomutase. FEMS Microbiol. Lett. **77:**325–330.
13. **Gimmestad, M., et al.** 2003. The protein, but not its mannuronan C-5-epimerase activity, is needed for alginate polymer formation. J. Bacteriol. **185:**3515–3523.
14. **Gruber, D. F., V. A. Pieribone, B. Porton, and H.-T. Kao.** 2008. Strict regulation of gene expression from a high-copy plasmid utilizing a dual vector system. Protein Expr. Purif. **60:**53–57.
15. **Hammer, K., I. Mijakovic, and P. R. Jensen.** 2006. Synthetic promoter libraries: tuning of gene expression. Trends Biotechnol. **24:**53–55.
16. **Henkin, T. M., and C. Yanofsky.** 2002. Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. Bioessays **24:**700–707.
17. **Jensen, P. R., and K. Hammer.** 1998. Artificial promoters for metabolic optimization. Biotechnol. Bioeng. **58:**191–195.
18. **Jones, K. L., S. W. Kim, and J. D. Keasling.** 2000. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. Metab. Eng. **2:**328–338.
19. **Keasling, J. D.** 2008. Synthetic biology for synthetic chemistry. ACS Chem. Biol. **3:**64–76.
20. **Khlebnikov, A., and J. D. Keasling.** 2002. Effect of *lacY* expression on homogeneity of induction from the P_{tac} and P_{trc} promoters by natural and synthetic inducers. Biotechnol. Prog. **18:**672–674.
21. **Koffas, M. A. G., G. Y. Jung, and G. Stephanopoulos.** 2003. Engineering metabolism and product formation in *Corynebacterium glutamicum* by coordinated gene overexpression. Metab. Eng. **5:**32–41.
22. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680–685.
23. **Laursen, B. S., H. P. Sørensen, K. K. Mortensen, and H. U. Sperling-Petersen.** 2005. Initiation of protein synthesis in bacteria. Microbiol. Mol. Biol. Rev. **69:**101–123.
24. **Lewis, D. E. A., and S. Adhya.** 2004. Axiom of determining transcription start points by RNA polymerase in *Escherichia coli*. Mol. Microbiol. **54:**692–701.
25. **Lopez, P. J., I. Iost, and M. Dreyfus.** 1994. The use of a tRNA as a transcriptional reporter: the T7 late promoter is extremely efficient in its transcripts are poorly expressed. Nucleic Acids Res. **22:**1186–1193.
26. **Marques, S., M. T. Gallegos, and J. L. Ramos.** 1995. Role of sigma S in transcription from the positively controlled *Pm* promoter of the TOL plasmid of *Pseudomonas putida*. Mol. Microbiol. **18:**851–857.
27. **Mijakovic, I., D. Petranovic, and P. R. Jensen.** 2005. Tunable promoters in systems biology. Curr. Opin. Biotechnol. **16:**329–335.
28. **Netzer, R., et al.** 2010. Biosynthetic pathway for gamma-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.
29. **Nielsen, J.** 2001. Metabolic engineering. Appl. Microbiol. Biotechnol. **55:**263–283.
30. **Pfleger, B. F., D. J. Pitera, C. D. Smolke, and J. D. Keasling.** 2006. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat. Biotechnol. **24:**1027–1032.
31. **Pharkya, P., A. P. Burgard, and C. D. Maranas.** 2003. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. Biotechnol. Bioeng. **84:**887–899.
32. **Pharkya, P., and C. D. Maranas.** 2006. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab. Eng. **8:**1–13.
33. **Ramos, J. L., S. Marqués and, and K. N. Timmis.** 1997. Transcriptional control of the *Pseudomonas* TOL plasmid catabolite operons is achieved through an interplay of host factors and plasmid-encoded regulators. Annu. Rev. Microbiol. **51:**341–373.
34. **Saïda, F., M. Uzan, B. Odaert, and F. Bontems.** 2006. Expression of highly toxic genes in *Escherichia coli*: special strategies and genetic tools. Curr. Protein Pept. Sci. **7:**47–56.
35. **Sambrook, J., and D. Russell.** 2000. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York, NY.
36. **Sletta, H., 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.
37. **Sletta, H., 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.
38. **Smolke, C. D., T. A. Carrier, and J. D. Keasling.** 2000. Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. Appl. Environ. Microbiol. **66:**5399–5405.
39. **Smolke, C. D., V. J. Martin, and J. D. Keasling.** 2001. Controlling the metabolic flux through the carotenoid pathway using directed mRNA processing and stabilization. Metab. Eng. **3:**313–321.
40. **Steigedal, M., and S. Valla.** 2008. The *Acinetobacter* sp. *chnB* promoter together with its cognate positive regulator ChnR is an attractive new candidate for metabolic engineering applications in bacteria. Metab. Eng. **10:**121–129.
41. **Walker, K. A., and R. Osuna.** 2002. Factors affecting start site selection at the promoter. J. Bacteriol. **184:**4783–4791.
42. **Winther-Larsen, H. C., K. D. Josefsen, T. Brautaset, and S. Valla.** 2000. Parameters affecting gene expression from the in gram-negative bacteria. Metab. Eng. **2:**79–91.
43. **Zwiefka, A., H. Kohn, and W. R. Widger.** 1993. Transcription termination factor rho: the site of bicyclomycin inhibition in *Escherichia coli*. Biochemistry **32:**3564–3570.