Veronika Kucharova

Expression of Recombinant Proteins in *Escherichia coli*: The Influence of the Nucleotide Sequences at the 5´ Ends of Target Genes

Thesis for the degree of Philosophiae Doctor

Trondheim, November 2012

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 978-82-471-3888-5 (printed ver.)
ISBN 978-82-471-3889-2 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2012:287

Printed by NTNU-trykk

For Naďa and Zdeněk

"…organisms of the most different sorts are constructed from the very same battery of genes. The diversity of life forms results from small changes in the regulatory systems that govern expression of these genes. "

-F. Jacob; *Of Flies, Mice and Men*

PREFACE

The work reported in this thesis was carried out at the Department of Biotechnology at the Norwegian University of Technology and Science (NTNU) and supported by a grant from Research Council of Norway.

It is difficult to list all the people who sincerely helped me during the past four years and I owe my gratitude to all of them who have made this dissertation possible. In the first place I would like to thank my main supervisor, Svein Valla, for the opportunity to be part of this exciting project and of his gene expression research group. Svein's patient guidance, encouragement and advice were among the main reasons for which I have truly enjoyed the work and successfully finished it. Secondly, I would like to thank Trygve Brautaset, my second supervisor from SINTEF, Department of Biotechnology. His constructive suggestions and support helped to overcome many challenges of the research work and showed me that even in the scientific world there exists a solution to any potential problem.

My appreciation also goes to all the former and current members in the Valla's research group. I would like to especially thank to my close collaborator and officemate Laila Berg who has always been there to listen and to provide very useful comments. During the work on this thesis I was very lucky to collaborate with several other excellent researchers; namely Tonje Marita Bjerkan Heggeset, Simone Balzer and Trine Aakvik Strand, and I want to thank them for their help and also patience with mine often too direct ways. For bringing valuable bioinformatic and mathematical element to my work, for pointing out alternative approaches of solving biological problems as well as for other scientific discussions and advice I would like to thank Jørgen Skancke and Eivind Almaas.

I would like to extend my appreciation to all of the staff and students at both NTNU and SINTEF Departments of Biotechnology, for welcoming me into their group and giving me encouragement throughout the course of this work. I would like to particularly thank Kathinka Q. Lystad, Kåre A. Kristiansen, Adrian E. Naas and Magnus Leithaug for their contribution to this thesis work.

Finally, I would like to express my heart-felt gratitude to my family. My parents are a constant source of inspiration to me and I thank them for providing me with the best condition for education. Last, Bernt Øyvind, thank you for your kindness, your faith in me and for pulling me back down on solid ground whenever the sanity seems to be departing.

ABSTRACT

The nucleotide sequence at the 5´ end of genes can be specified as the sequence of a promoter associated 5´ untranslated region (UTR) together with the initial coding sequence of a gene. Because this genetic region has been implicated in the control of translation, messenger RNA (mRNA) stability and even transcription, it can be looked at as one of the central control points in gene expression. Both the 5´-UTR and the coding sequence have often been included in optimization strategies targeted to simulate recombinant protein production in *E. coli* and numerous reports describe various sequence-dependent structural features that can positively influence the overall expression process. Nevertheless, the actual mechanisms by which the regulation of gene expression is exerted at the 5´ end remain obscure. The work reported in this thesis has involved various types of analyses of the functionality of the 5´ end, by using mutations as a major tool. The work can be seen as mainly a detailed empirical analysis of the relation between the specific nucleotide sequences at the 5' end of genes and the final outcome at the protein production level. The results also indicate that optimizations based on empirical laboratory protocols are currently unlikely to be exceeded by predictions based on bioinformatics software.

Sequence mutagenesis of elements in the XylS/*Pm* - positive regulator/promoter system coupled to high-throughput screening had been previously proven to be a powerful method for increasing the expression of recombinant genes from this expression cassette. At the beginning of this thesis work the effect of introducing random mutations in the DNA sequence of the *Pm* promoter associated 5´-UTR and two 5´ fusion partners, whose sequences correspond either to a consensus translocation signal peptide or the first 23 codons of a well-expressed *celB* gene (encoding a cytoplasmic phosphoglucomutase) was investigated. The core of the experimental work was construction of large combinatorial libraries of the different DNA sequences and subsequent selection for improved expression of a reporter gene (either ampicillin or apramycin resistance gene), that was indicated by an increase in antibiotic tolerance of the corresponding *E. coli* host cells. A shared result of the three

individual studies was the establishment of a collection of optimized sequences that generally improved protein production properties of both reporter and industrially relevant heterologous genes.

In addition to random mutagenesis, also synonymous mutations were introduced in the DNA sequence of the consensus signal peptide (CSP) and the consequent expression effects were evaluated. As a conclusion, the DNA changes that did not alter the amino acid sequence led to a lesser stimulation of expression of the *bla* reporter (ampicillin resistance) than when complete sequence randomization was applied. Moreover, similar results were obtained when synonymous codon usage of the first 9 codons of the medically important *ifn-α2b* gene was optimized by a bioinformatic method, followed by experimental determination of expression levels of several rationally selected *ifn-α2b* synonymous variants. These results indicated that optimization of the codon usage of the 5´ coding sequence has limited effects, probably due to the sequence intrinsic characteristics. However, the use of optimized 5´ fusion partners or 5´-UTR variants can often overcome such limitations.

Besides evaluating the expression at the protein level, the work also addressed how the changes of the 5´ end of a gene influence expression at the level of transcript accumulation and mRNA stability. For that purpose, a non-invasive method for accessing recombinant mRNA stability in bacteria was developed. The procedure was based on the removal of diffusible transcriptional inducers followed by qRT-PCR determination of mRNA levels at consecutive time-points. Among the principal findings was that a 5´ fusion partner (specifically: translocation signals *pelB* and *ompA*, together with the *celB*-based 5´ fusion) contributes to the stimulation of recombinant gene expression by enhancing the stability of the corresponding fusion mRNA. The stimulation of expression caused by specific mutations in the 5´-UTR and adjacent coding sequence (synonymous changes), on the other hand, surprisingly appeared to result from improved rate of mRNA synthesis. Three selected promoter systems (*Pm*, P_{tac} and the T7 based) were used in these studies, and part of the work also evaluated how fast each system responds to addition and removal of its inducer, respectively. The expression systems were found to affect both transcript accumulation and decay

in a specific way that correlated with the type of transcription regulation each system is subjected to.

Finally, a study comparing five bacterial expression systems (XylS/*Pm*, XylS/*Pm* ML1-17 (a *Pm* variant), the bacteriophage T7 RNA polymerase/promoter system, LacI/P_{trc} and AraC/P_{BAD}) with respect to their production capacity of five different recombinant proteins was carried out. The comparison revealed many expression system and model gene specific features and that none of the systems was superior in all evaluated aspects; which included system´s adaptability, maximum protein yield, basal expression in the absence of inducer, use of cellular resources and homogeneity of expression. However, particularly because of a large associated collection of optimized genetic elements (such as sequence variants of the *Pm* promoter, the XylS regulator, 5´-UTR and various translocation signals) and the possibility of simple genetic adjustments that can lead to both higher and lower expression levels, the XylS/*Pm* system appeared as a good starting point for optimization of various kinds of protein production processes.

TABLE OF CONTENTS

LIST OF PAPERS

Paper I

Berg, L., V. Kucharova, I. Bakke, S. Valla, and T. Brautaset. 2012. Exploring the 5´-UTR DNA region as a target for optimizing recombinant gene expression from the strong and inducible *Pm* promoter in *Escherichia coli*. Journal of Biotechnology 158:224-230.

Paper II

Heggeset, T. M. B., V. Kucharova, I. Nærdal, S. Valla, H. Sletta, and T. Brautaset. Combinatorial mutagenesis and selection of improved translocation signal sequences and their application for high level production of heterologous proteins in *Escherichia coli*. Manuscript under revision in Applied and Environmental Microbiology

Paper III

Kucharova, V., J. Skancke, T. Brautaset, and S. Valla. Design and optimization of short DNA sequences that can be used as novel fusion partners for high-level expression of heterologous genes in *Escherichia coli*. Manuscript

Paper IV

Kucharova, V., T. A. Strand, A. E. Naas, E. Almaas, T. Brautaset, and S. Valla. Non-invasive analysis of mRNA decay rates by a combination of inducer wash-out and qRT-PCR. Manuscript

Paper V

Balzer, S., V. Kucharova, J. Megerle, R. Lale, and S. Valla. The performance of the XylS/*Pm* system compared to other commercially available regulator/promoter systems used in bacterial recombinant expression. Manuscript

ABBREVIATIONS

5´-UTR	5´ untranslated region of mRNA
A	adenine
C	cytosine
CAI	codon adaptation index
CSP	consensus signal peptide
DNA	deoxyribonucleic acid
G	guanine
GRAS	generally recognized as safe
mRNA	messenger ribonucleic acid
RBS	ribosomal binding side
RNA	ribonucleic acid
RNAP	RNA polymerase
rRNA	ribosomal RNA
SD	Shine-Dalgarno
tRNA	transfer ribonucleic acid
U	uracil

1 TURNING BACTERIA INTO PROTEIN PRODUCTION FACTORIES

At present we are using biotechnology in countless applications; such as to help with the availability of food, to detect and control environmental pollution, to obtain alternative energy sources and to improve human health. Modern biotechnology can thus be regarded as one of the major contributors to the advancements towards more sustainable life-styles. Utilization of cell biosynthetic capacities is currently a well-established practice and most of the human population is profiting, knowingly or not, from this multidisciplinary field and its inventions in their everyday lives.

For centuries humans have used microorganisms to make useful products. The modern perception of biotechnology however developed in the 1960s and ´70s when a new field of study founded on emerging molecular and cellular technologies formed. By building up the knowledge of biosynthetic pathways and associated regulatory mechanisms, but also on account of the advancing progress in technology (particularly recombinant DNA), molecular biologists were able to devise methods to isolate and identify genes as well as to manipulate and insert them into other species (9, 68, 138). Shortly after that, the first genetically-engineered, synthetic human proteins were produced in bacterial cells by the artificial manipulation and modification of the corresponding DNA sequences (98, 137).

Over the years, the production of bioactive compounds by microorganisms from rather cheap media, compared to expensive and elaborate extraction processes, has built itself an unarguable position. Engineering of the cell biosynthetic machinery is being widely applied within university and industry based research centers for production of biological catalysts used in applications as diverse as cancer treatment and pulp bleaching (74, 205, 228, 261). Steadily increasing demand for active therapeutics together with a requirement of structural genomics and proteomics studies for large amounts of recombinant proteins (29, 83, 104, 155), can be regarded as important examples of motivation factors driving the field of recombinant protein expression and the production of bioactive compounds.

Among the many commonly used expression systems, derived from prokaryotic as well as eukaryotic cells, engineered bacterial expression systems are among the most attractive (228, 286). Nearly 30% of all of the recombinant pharmaceutical proteins is produced in *Escherichia coli*, the principal expression host and first-line production system for non-glycosylated proteins (114, 125, 244, 257). Since the work presented in this thesis has been carried out in *E. coli*, the general discussion on gene expression regulation and heterologous protein production in the following two introductory chapters will focus on this organism. Nevertheless, because significant progresses have been made in alternative bacterial expression systems during the past few years, few examples in this regard are going to be briefly mentioned.

1.1 Control of gene expression in E. coli

In all living cells, the genetic information encoded in DNA is first transcribed into mRNA which is then translated into one or more protein molecules. This hierarchical design made up from multiple consecutive steps, each of which is subjected to careful control, allows the cell to integrate and simultaneously coordinate different physiological processes (**Figure 1**). An important aspect that should be considered during genetic design of an expression system is to establish which control mechanisms are operative during expression of a particular recombinant gene. For that reason, a central theme of the three following sub-chapters is the description of our current understanding of gene expression regulation in prokaryotes.

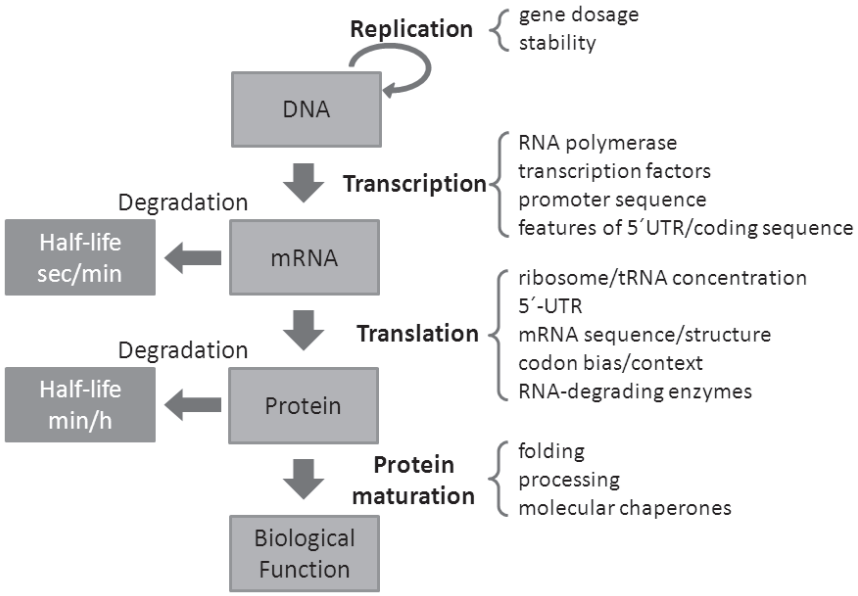


Figure 1 Factors influencing gene expression in E. coli*. Each step in the gene expression process may be modulated, including transcription, RNA turnover, translation, and post-translational processes. Several determinants that act along the path of expression are show, any of which could have both positive and negative regulatory functions.*

1.1.1 Transcriptional regulation

Gene expression in bacteria is primarily regulated at the first step; the transcription of the DNA template into its RNA copy, as it limits the energetic and materials cost of aborting expression at any subsequent stages (107). For accurate RNA polymerization the catalytic activity of DNA-dependent RNA polymerase (RNAP - defined as core enzyme with subunit composition $\alpha_2\beta\beta'\omega$) is obligatory (34). Additional activities of the polymerase accommodated by one of the discriminatory σ subunit species allow for specific promoter DNA sequence recognition, positioning of the RNAP holoenzyme (RNAP with associated sigma factor) at a target promoter and unwinding of the DNA duplex near the transcript start site (42). The initiation phase of transcription is controlled by two means; specific features of the promoter DNA sequence and management of the activity of the transcription initiation complex by regulatory proteins. Transcriptional regulators, such as DNA binding proteins, small ligands and

transcriptional factors that can bind directly to RNAP, are usually activated in response to environmental stimuli (26, 118, 135, 252).

Activation of gene expression in *E. coli* has classically been characterized either as σ^{70} dependent (including six of the sevens σ factors: σ^{70}, σ^{38}, σ^{32}, σ^{28}, σ^{24}, σ^{19}) or as σ^{54} dependent, a sigma factor which shares little sequence homology and promoter specificity with the σ^{70} class (26, 109). While the response of the RNAP enzyme to specialized σ subunits and transcription factors plays a crucial role in determining the cell´s transcription program under different growth conditions, the promoter sequences are major determinants of the wide range of gene expression levels observed (22, 221). Different sequence elements responsible for promoter recognition by RNAP holoenzyme have been identified. Hexamers –10 and –35 , which are located 10 and 35 base pairs (bp) upstream from the transcript start site (position +1), respectively, are the principal recognition elements and function as the primary determinants of the basal promoter strength (118, 119). In addition to these core promoter elements, other sequences throughout the extended promoter recognition region have been identified to have a large influence on the strength of investigated promoters. These sequence elements include a 3–4 bp motif located immediately upstream of the –10 and a ~ 20 bp sequence located upstream of the promoter –35 hexamer (24, 89, 220, 224).

The same promoter sequence can be recognized by more than one sigma factor as demonstrated for the *Pm* promoter, used in this work. In the early exponential growth phase the *Pm* promoter recognition is mediated by σ^{32}, but a switch to σ^{38} (also known as σ^{S}) takes place in the late exponential and early stationary phases (78, 212). The consensus sequences of the different promoter elements recognized by the RNAP complex with σ^{70}, σ^{32} or σ^{S} are illustrated in **Figure 2**, together with a schematic representation of interactions between RNAP holoenzyme and a promoter DNA sequence.

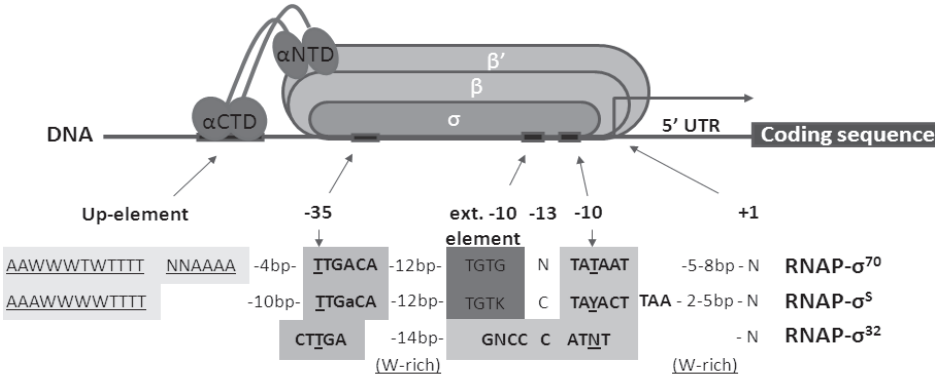


Figure 2 Promoter recognition by RNAP holoenzyme*. Illustration of known interactions between the promoter DNA core elements and RNAP holoenzyme during transcription initiation, together with a summary of all known consensus sequences for promoters recognized by the RNAP-σ^{70},-σ^{S} or –σ^{32}. N represents A, T, C or G; K represents T or G; W represents A or T; Y represents T or C (pyrimidine). UP: upstream. bp: base pairs. Adapted from (22, 42, 118, 263, 277).*

Once a dissociable σ factor has directed RNAP to a specific promoter site, the RNAP holoenzyme can either form a stable closed complex that is incompetent for transcription and which requires ATP hydrolysis by enhancer-binding proteins for additional transition (the case of σ^{54}) (118). Alternatively, it can spontaneously proceed to an open complex that is competent for transcription (the case of σ^{70}), in which the duplex DNA around the transcript start-point is unwound (16, 184). When the open complex has been formed, transcription is typically initiated through synthesis of short abortive transcripts that are repetitively released and re-synthesized while RNAP maintains its contacts with the promoter DNA sequence, as has been demonstrated both *in vitro* (124) and later *in vivo* (100). These studies furthermore showed that the about 25 bps 5´-part of the UTR DNA region (the initial transcribed region) can influence the promoter escape efficiency. To undergo promoter escape and proceed to the elongation phase, RNAP must relinquish its contacts to the core promoter elements (159). In order to do so, it has been proposed that excess template DNA must be transiently unwound and accommodated in the main channel of the RNAP in a process called DNA scrunching (38). This was supported experimentally by single-molecule DNA nanomanipulation (215), which detected template unwinding during

abortive initiation, and by single molecule fluorescence resonance energy transfer (FRET) experiments (147), which showed that RNAP remains fixed on promoter DNA sequence and pulls downstream DNA into itself.

Transient pauses during the transcription cycle

After RNAP has escaped into the productive elongation phase, the transcription complex may come across potential arrest sites at which the nascent RNA remains stably bound to the enzyme (12). Such pause sites are sequence-encoded and control the rate of mRNA synthesis by altering the fraction of time during which RNAP is actively elongating (163). Transcriptional pausing is believed to influence a variety of events. In addition to affecting the overall rate of RNA production, it enables the coordination of RNAP movement, interactions with external regulatory factors, folding of the nascent RNA and the synchronization of transcription with translation (13, 120, 206). Several models have been proposed to explain pausing during transcription elongation, including a model where sequences that resemble promoter elements are recognized by a σ factor that has failed to dissociate from RNAP (40, 199, 217). Although it was initially thought that the σ subunit dissociates from the transcription complex during the transition to elongation, the σ^{70} subunit can actually interact with the elongation complex (21, 193) and induce pauses during transcription (116, 209, 290). In summary, although transcriptional initiation has long been identified as a critical point of regulation, mechanisms for controlling expression levels during the elongation phase have added yet another layer of complexity to the transcriptional regulation.

1.1.2 Regulation of mRNA turnover

Undoubtedly, the concentration of mRNA which is available for translation directly determines the total protein production level. As a consequence, the control of mRNA stability is important for many expression strategies and in particular for design of expression cassette elements such as 5´-UTRs (11, 51-53). In contrast to rather stable DNA, bacterial mRNA molecules are intrinsically labile and continuously synthesized

(**Figure 3**). This provides bacteria with an effective way to regulate expression of their genes because those transcripts that are templates for proteins pertinent to a particular physiological state can be selectively maintained (27). Current knowledge of bacterial mRNA decay suggests that relative stability and how rapidly an mRNA is degraded is a combination of multiple sequence and structural features within the message itself, the efficiency of its translation and the cooperation of specific RNA processing enzymes (54, 143).

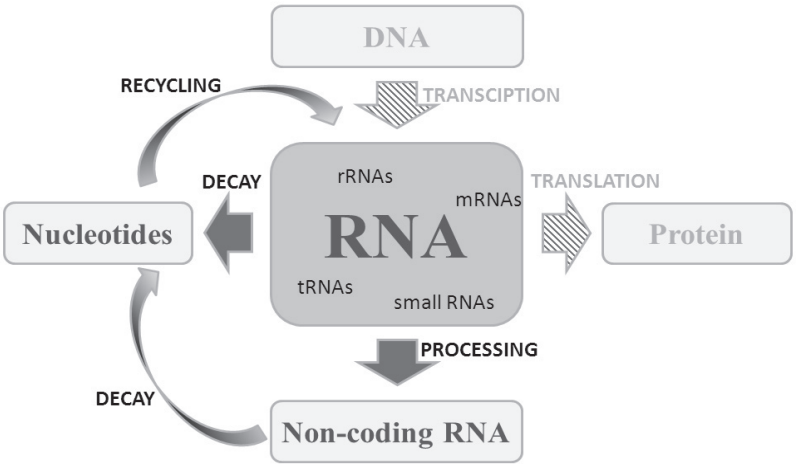


Figure 3 RNA metabolism in bacteria. *The RNA steady state level is a function of the molecule synthesis and decay rates. Different types of RNA are either directly involved in translation (mRNAs) or undergo further processing (pre-cursors of stable RNAs) or degradation (defective RNAs, untranslated or poorly translated mRNAs). The end products of RNA degradation, mononucleotides, are recycled in the next round of RNA synthesis. Adapted from (143).*

Enzymes of broad importance for cytoplasmic RNA decay

The intracellular RNA-degrading enzymes (ribonucleases or RNases) are generally classified by their mode of action into endonucleases which cut RNA internally and exonucleases that sequentially remove mononucleotides either from the 5´ or the 3´ end of RNA. In *E. coli*, several endonucleases and 3' exonucleases have been described (**Table 1**), but the organism appears to lack a 5' exonuclease (47, 123, 181, 216). The degrading enzymes often form a dynamic multiprotein complex (degradosome) primarily consisting of the essential endonuclease RNase E, polynucleotide

phosphorylase (PNPase), RNA helicase (RhlB), and enolase (48, 211, 267). Instead of a dedicated machinery for degradation of different RNA classes, the same enzymes have been described to participate in the degradation of both mRNAs and stable RNAs (ribosomal RNA (rRNA) and transfer RNA (tRNA)) (75, 76). Thus, the specificity of the RNase and the accessibility of the substrate seem to determine whether or not a particular RNA will be targeted for degradation (143). Although the detailed mechanism of prokaryotic RNA decay remains to be fully elucidated, both 5´ and 3´end-dependent degradation pathways have been described for *E. coli* and are presented in the following two sub-chapters.

Table 1 Important RNA-modifying enzymes that participate in mRNA turnover in *E. coli*.

Enzyme Class	Name	Specificity/ Function
Endonuclease	RNase E (RNase G homolog)	A/U rich single stranded regions, 5´-end dependent hydrolase
	RNase III	double-stranded RNA
	RNase G, RNase P	
3´ exonuclease	PNPase	single-stranded 3´ end
	RNase R, RNase II	single-stranded 3´ end
	Oligoribonuclease	RNA oligonucleotides
5´-end modification	RppH	single-stranded 5' terminus, pyrophospate removal
3´-end modification	Poly(A) polymerase	Polyadenylation
	PNPase	heteropolymetric tail addition

Internal and 3′-terminal degradative events in mRNA decay

Early evidence suggested that the mRNA lifetimes in *E. coli* are controlled primarily by internal events followed by 3´ exonuclease attack (6, 28). In support of this notion were observations that exonucleotic digestion of mRNA from the 3´ end is hindered by a characteristic stem-loop structure (**Figure 4**), and that the initial endonucleolytic cleavage (mainly mediated by RNase E) results in a pair of short-lived decay intermediates (7, 15, 203). Several reports of poly(A) tails on bacterial RNAs (197, 285) supported a hypothesis that addition of adenines to mRNA promotes 3′ exonucleolytic degradation of stem-loop structures in decay intermediates. The transient addition of poly(A) tails to bacterial RNAs is indeed a crucial step since it can present a single

stranded RNA segment as a start signal to the 3´ exonucleases such as PNPase and RNase R (63, 117, 284). Facilitation of 3' exonucleolytic degradation of mRNA decay intermediates by polyadenylation is illustrated in **Figure 5**.

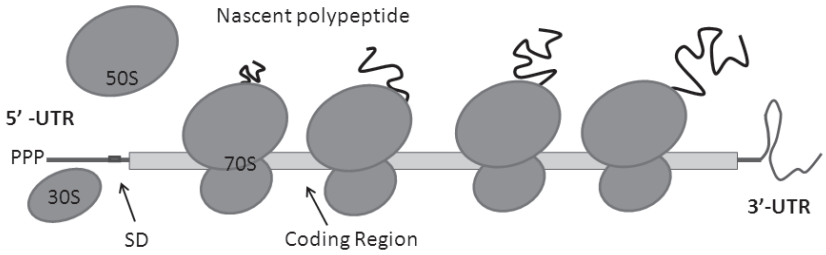


Figure 4 Illustration of translated prokaryotic mRNA. *Elements protecting the molecule from being degraded; a triphosphate at the 5´ end, a stem-loop at the 3´ end and translating ribosomes (spheres) are presented. The Shine-Dalgarno (SD) sequence is a part of ribosomal binding site. Description of translation is given in chapter 1.1.3. Adapted from (27).*

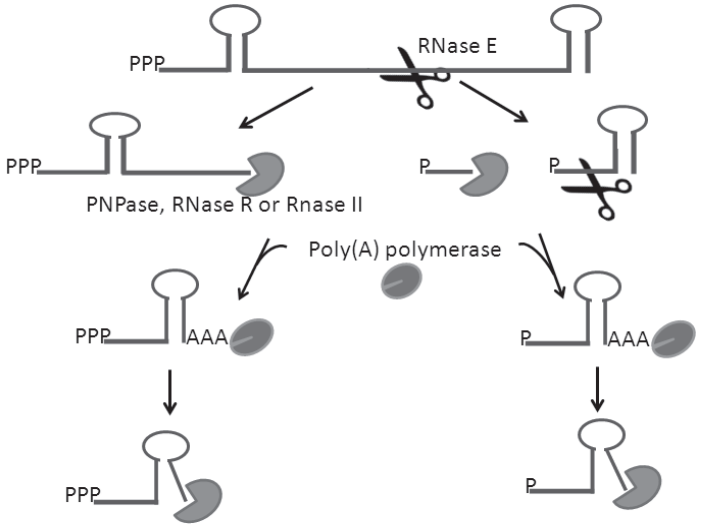


Figure 5 3' exonucleolytic degradation of mRNA. *Endonucleolytic cleavage of mRNA by RNase E initiates the decay in most cases. The decay intermediates, except the fragment with the original 3'-terminal stem-loop, undergo 3' exonucleolytic attack by PNPase, RNase R and/or RNase II. The fragments that contain stem-loops are polyadenylated by poly(A) polymerase, thereby enabling the exonucleases to reengage and degrade the stem loop. Adapted from (27).*

5′-terminal degradative events

A central finding for the characterization of the 5´-terminal degradation pathway was the 5′-UTR influence of the rate at which RNases degrade mRNA (172). Initially, a single phosphate at the 5´-terminus of mRNA was shown to affect the catalytic rate constant of RNase E by orders of magnitude compared to the triphosphorylated mRNA counterparts (171). At the same time, this 5´ end modification appeared not to accelerate the RNase E action by improving mRNA binding, but rather by enhancing the enzyme catalytic potency (141, 213). Determination of the crystal structure of RNase E later explained how the recognition of the 5´-terminus may trigger catalysis (45). The study suggested that the influence of 5′ phosphorylation on endonucleolytic cleavage by RNase E is a consequence of a discrete enzyme pocket in which monophosphorylated 5′ ends can bind and promote downstream cleavage.

Besides the 5´-phosphorylation status of mRNA, also the 5′-UTR propensity to form secondary structures has been pointed out as a determinant of mRNA stability. An increasing number of studies have documented that a stem-loop near the 5´ end helps to protect the message from degradation (10, 17, 37, 88). This finding could not be explained by the formerly proposed model of initial endonucleotic cleavage until a recent study re-examined the primary events of RNA decay and uncovered that mRNA degradation can be triggered by a 5′-terminal event (pyrophosphate removal) that marks transcripts for rapid turnover (58). In addition, the same authors showed that the protein RppH is the RNA pyrophosphohydrolase (initially designated NudH/YgdP) that initiates mRNA decay by this 5´ end dependent pathway and facilitate RNase E cleavage of primary transcripts by 5´ pyrophosphate removal (73). A proposed schematic pathway for 5′ end-dependent mRNA degradation in *E. coli* is presented in **Figure 6**.

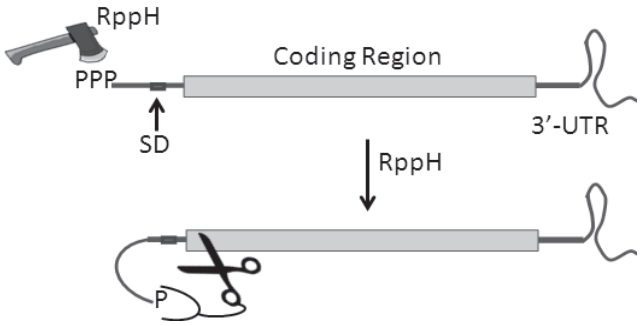


Figure 6 5´-end-dependent pathway of mRNA decay in E. coli. *Pyrophosphate removal by the enzyme RppH (axe) generates a 5'-terminal monophosphate that can bind to a discrete pocket of RNase E (scissors). This initial step facilitates mRNA cleavage at a downstream location. Shine-Dalgarno (SD) sequence is a part of ribosomal binding site (described in chapter 1.1.3). Adapted from (27).*

The influence of ribosomes on bacterial mRNA decay

One of the most perplexing factors influencing the lifetimes of individual mRNAs is their association with translating ribosomes. A couple of pioneering studies showed that antibiotic drugs that inhibit translation elongation (such as chloramphenicol, fusidic acid, and tetracycline) yield stabilization of cellular mRNAs, whereas drugs that inhibit initiation of translation (puromycin and kasugamycin) have the opposite effect (207, 229). The effects of these antibiotics were later shown to be more complex and the results judged as difficult to interpret due to cell-wide effects brought about by the drugs. A follow-up study showed that the inhibitors of protein synthesis can stabilize mRNAs by acting directly on the mRNA degradation machinery, irrespective of how they inhibit translation (169).

The effect of translational inhibitors can be mimicked by mutations that slow or block ribosome movement. Studies exploring early arrests of translation through the introduction of terminator codons and mutations in or near the ribosomal binding side (RBS) that result in reduced translation initiation frequency have suggested that some mRNAs can become less stable upon depletion of ribosomes or when ribosome density decreases (35, 200). Conversely, mutations that extended the complementarity between the Shine-Dalgarno (SD) sequence and 3´ end of 16S rRNA could increase translation and mRNA stability (10, 276). Furthermore, the demonstration that

untranslated *lacZ* mRNA is stabilized by RNase E inactivation suggested that the naked mRNA is more sensitive to the RNase attack (133).

The stabilizing effect of translation of many bacterial mRNAs has been explained by postulating that ribosomes can mask endoribonuclease sites which control the mRNA decay rate. The notion that ribosomes sterically block mRNA-degrading enzymes was further supported by i*n vitro* experiments with *ompA* mRNA that showed that 30S ribosomal subunits bound to the RBS can protect from RNase E cleavage in the 5´-UTR (272). Ribosome binding has been described to be stimulated by A/U rich regions flanking the start codon (110, 156, 270). Interestingly, such A/U-rich sequences frequently occur within RNase E cleavage sites (183). The location of coincident ribosome and RNase E binding sites could provide means for dual mRNA regulation of stability; either hindering the degradation by ribosome protection or, the other way around, providing a way to selectively eliminate mRNAs that are translationally inactive. Accordingly, this hypothesis implies that a rate-limiting step in mRNA decay depends on the binding affinity of either the 30S ribosomal subunit or RNase E for a given mRNA (144).

An alternative explanation (and to some extent complementary) is that by directing more efficient translation initiation, a closer spacing of translating ribosomes would improve steric protection of potential RNase E cleavage sites (72). Several studies addressed the question how the spacing between translating ribosomes along the length of the mRNA and the average degree of occupancy of the RBS influences mRNA degradation. In order to distinguish the two cases, the preferred initiation codon in *E. coli* (AUG) was mutated to a weaker variant (GUG, CUG, or UUG) or to a codon that cannot function in translation initiation (10, 182). Such genetic manipulation is assumed to reduce the frequency of translation initiation and thereby increase ribosomal spacing without significantly diminishing the ribosome-binding affinity of the RBS and its average occupancy by ribosomes. Although a high degree of ribosomal occupancy at the 5´ end appeared to be an important factor in determining mRNA longevity, on-going translation rather than ribosome binding per se (resulting in close ribosome spacing) was required to protect mRNA from degradation.

1.1.3 Regulation of translation

The cost of peptide bond formation in a cell is ~ 160 kJ/mol of free energy, which is gained from cleaving 4 high-energy phosphate bonds (113). However, for uncatalyzed peptide bond synthesis in dilute aqueous solution the required free energy change is only ~ 20 kJ/mol (77, 177). The extra energy consumed in the enzymatically catalyzed reactions is presumably needed for generating an accurate, mRNA-defined polypeptide sequence. Considering the high energy expense, bacteria must carefully regulate their protein production to be able to dynamically adjust to their usually ever-changing habitats. Control at the post-transcriptional level, through coordination of RNA turnover, the translation and post-translational processes, is thus an integral part of gene expression (180).

The translation process

The translation of genetic information from mRNA to protein is performed by macromolecular ribonucleoprotein complexes (ribosomes) and is commonly divided into four phases: initiation, elongation, termination and ribosome recycling (164). Likely due to the absence of nucleus in bacterial cells, ribosomes are able to initiate translation already on a nascent mRNA (transcription-translation coupling) (105). During the initial phase, which is just like in transcription the rate limiting step, three initiation factors (IF1-3) guide the ribosome in selection of mRNA and translational reading frame (189). In the subsequent elongation process ribosome utilizes three binding sites for the tRNA molecules that help to decode the genetic code in mRNA sequence into a protein: A site that binds aminoacyl-tRNAs (that function as the acceptors for the growing protein during peptide bond formation), a P site that interacts with peptidyl-tRNAs (the tRNA carrying the growing peptide chain), and an exit E site for deacylated tRNAs (192). The different mechanisms involved in protein synthesis are quite well understood at the molecular and atomic level, thanks to the crystallographic studies of the ribosome associated with different ligands and their high-resolution structures generated by X-ray and cryo-electron microscopy experiments (237).

A schematic view of the initiation step is shown in **Figure 7**. Translation commences with the formation of a 30S preinitiation complex (PIC) by the 30S ribosomal subunit, mRNA, initiation factors and the aminoacylated and formylated initiator tRNA (fMet-tRNA^{fMet}) (56, 188). The large ribosomal subunit 50S then joins the 30S PIC and forms a labile 70S initiation complex which is subsequently stabilized by IF2-dependent GTP hydrolysis (108, 259). Once the initiation factors have dissociated the fMet-tRNA^{fMet} is positioned in the P site of the resulting 70S complex. Following the binding of the first aminoacyl-tRNA to the A site and formation of the first peptide bond, the machinery then proceeds to the elongation phase. When the ribosome reaches the end of the mRNA coding sequence, indicated by a stop codon, it dissociates and releases a new protein into the cell (248).

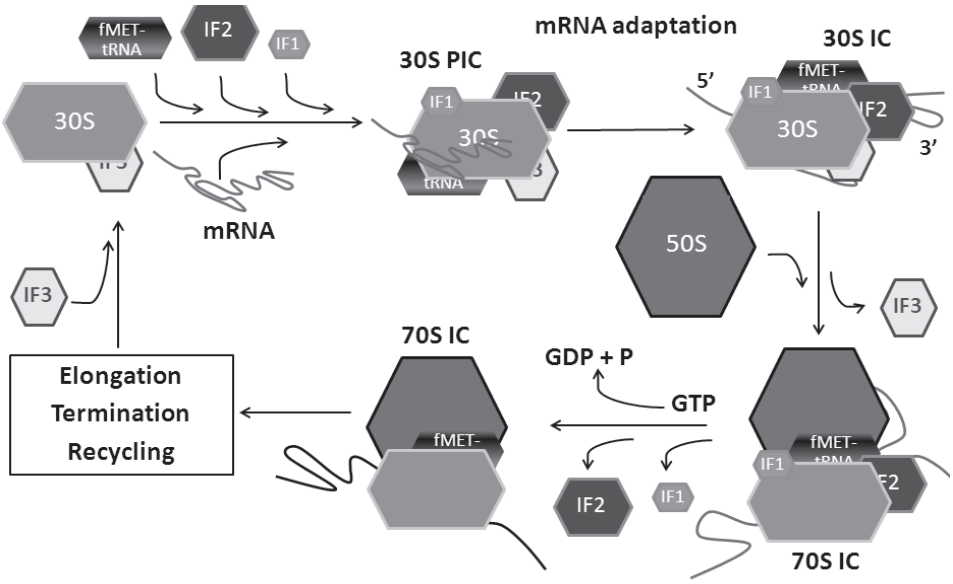


Figure 7 Illustration of the translation initiation. *Steps that include formation of 30S and 70S translation initiation complexes (IC), containing ribosomes, initiator fMet-tRNA^{fMet}, mRNA and initiation factors IF1, IF2 and IF3, are depicted. mRNA binds to 30S in two distinct steps. First the mRNA is anchored on the platform of the 30S subunit by the help of initiation factors (pre-initiation complex PIC). After that follows the accommodation of the mRNA in order to promote the codon-anticodon interaction in the P site. The resulting 30S IC engages the 50S subunit to form the 70S IC. Once initiation factors dissociate, the synthesis of the encoded protein can proceed through the elongation, termination and ribosome recycling phases. Adapted from (237).*

Sequence and structural elements that control initiation of translation

It is well established that the rate of ribosome binding to the 5´ end of mRNAs is control by the ribosomal binding site (RBS) located there and comprising ~ 25 nucleotides on each side of the translation initiation codon (129). Two principal sequence elements of the RBS are the SD sequence within the 5´-UTR (234), which facilitates 16S rRNA-specific binding and anchors the mRNA to the 30S subunit (287), and the start codon (most commonly ATG), which sets the reading frame for the rest of the translation process (158). In addition, the spacing region between the start codon and the SD sequence, the non-random distribution of nucleotides upstream of the SD and downstream of the initiation codon and the mRNA folding of the RBS have been shown to influence translation initiation (25, 61, 178, 218).

The complementary base-pairing between SD and a sequence motif at the 3´ end of the 16S rRNA (249), referred to as the anti-SD sequence, can occur under the condition that the SD sequence is not sequestered in a strong secondary structure. This was evidenced by quantitative analysis of the relationship between the stability of RBS secondary structure and translation efficiency (71). Furthermore, alterations of the SD or the anti-SD sequence that reduce their mutual complementarity were found to strongly inhibit protein synthesis (128, 139). At the same time, the SD sequence has been recognized as not essential for translation initiation (195, 196, 238). This is best documented by two alternative initiation mechanisms, which do not depend upon a consensus SD sequence. Leaderless mRNA that lacks its 5'-UTR can directly bind the ribosome and be efficiently translated (191, 264). Also, the ribosomal protein S1, which is a component of the 30S ribosomal subunit and binds to A/U-rich regions directly upstream of the SD sequence (144, 156, 230), can efficiently initiate translation regardless of the presence of the SD sequence (33, 157). Recent bioinformatics analyses, supported by experimental data, showed that mRNAs without a SD sequence are generally less structured in their translation initiation region (226). The authors therefore suggested that minimum mRNA folding around the start codon is necessary and sufficient to initiate SD–independent translation.

The notion that the structural features and the overall stability of the RBS quantitatively control the efficiency of translation, through modulating the accessibility of transcript to the ribosomes, is further supported by a number of investigations (25, 178, 227). Studies of various translation initiation regions generated by different rational design approaches, such as site-directed mutagenesis, revealed that A/U-rich sequences in the vicinity of the RBS stimulate translation (39, 85, 270, 288). Such sequences are believed to be less likely involved in strong secondary structures. The importance of mRNA folding at the beginning of a coding sequence has been directly or indirectly documented by a number of studies (57, 90, 160, 231) as well as the interplay between SD, initiation codon and the 5´ coding region in translation initiation (39, 103, 250, 251). Even though strong structures in the RBS are generally believed to act inhibitory (59, 289), initiation can occur when the structured element is positioned between the SD and the start codon or sufficiently downstream of the start codon (201, 208). In conclusion, minimal folding of mRNA and start codon accessibility appears to be of critical importance for efficient translation initiation.

Synonymous codon usage

The genetic code (on the mRNA level) is defined by triplets of four different nucleotides (A, C, G or U) and totally 61 sense codons are used to encode the 20 naturally occurring amino acids (three codons terminate translation). Synonymous codons are therefore used to code for the same amino acid. There exists a distinct codon bias; in most sequenced genomes synonymous codons are not used at equal frequencies (210). Codon usage can dramatically vary not only between different classes of genes but even within a single gene (67, 146). As one of the factors shaping codon usage patterns has been identified the availability of isoaccepting tRNAs (87), since in highly expressed genes codon preference was found to correlate with the composition of the tRNA pool (80, 131). Nevertheless, it is still not clear whether the pattern in codon usage is a cause or a consequence of the bias in tRNA abundance (194).

Indications that selective codon usage might be a translation control mechanism have come from studies that described a significant positive correlation between a gene's expression level and its codon bias (44, 91, 233), and several experimental studies showed that mRNA consisting of preferred codons is translated more efficiently than mRNA modified to contain rare codons (219, 255). This led to a view that high codon adaptation (defined by high codon adaptation index (232)) induces strong protein expression. However, the assumed stimulation of protein levels per mRNA by high codon adaptation does not agree well with the notion that initiation is generally the rate limiting step in translation, as illustrated in **Figure 8**. It has thus been hypothesized that when initiation is limiting it is unlikely that changes in the elongation rate can lead to an significant alteration in the rate of protein production (3, 44). In addition, specific synonymous codon usage has been described also for sequence motifs that are recognized by translational or by post-translational regulatory mechanisms; for example ribosomal pausing and protein folding (43, 225, 258). Thus, a deliberate location of well/ poorly adapted codons and consequent mRNA folding properties appear to be utilized for modulation of the performance of the translational apparatus.

Plotkin and Kudla recently suggested that a selection of synonymous codons against strong 5´ mRNA structure is an important evolutionary mechanism, in order to facilitate translation initiation (210). This was partly based on their own study of 154 genes coding for the same protein and the influence of synonymous codon usage on the protein expression level. Their results showed that mRNA folding near the RBS and associated rates of translation initiation play a central role in shaping expression levels of individual synonymous genes, while codon bias affects global translation efficiency and also cellular fitness. Similarly, a trend towards reduced codon adaptation and reduced mRNA stability near the translation initiation site in the 5´ region of *E. coli* genes has been documented (91, 110). Another systematic trend in the pattern of intragenic codon usage has been described by Tuller et al. (260). The study showed that poorly adapted (rare) codons at the beginning of genes might help prevent 'ribosomal traffic jams' by slowing elongation at the start of a gene. It is curious that

the 5´ region of poorly adapted codons identified by Tuller et al. overlaps to a great extent with the region in which synonymous codon choice systematically reduces mRNA stability (91, 110, 160). At present, it remains still unclear how the unusual and nearly universal pattern of 5´ codon usage modulates protein expression and which regulatory mechanisms it involves.

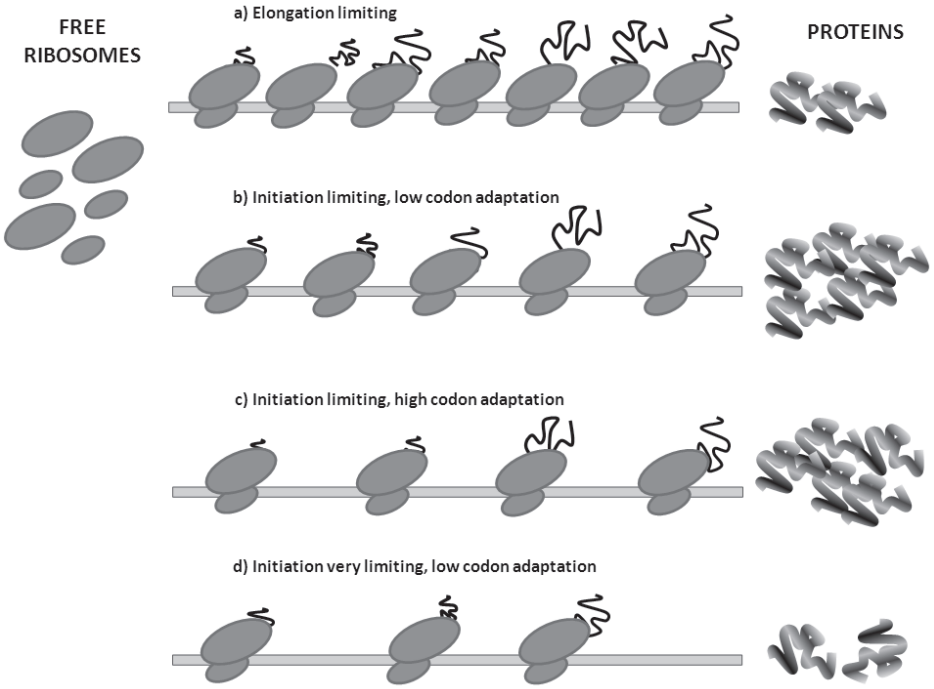


Figure 8 The rates of initiation and elongation determine the rate of protein synthesis. *When elongation is the rate-limiting step (a), mRNA will be covered densely by ribosomes and faster elongation will tend to increase the rate of protein synthesis. In case of limited initiation step (b, c and d) the transcript is not completely covered by ribosomes. For two genes with the same initiation rate(b and c), the mRNA with faster elongation will have a lower density of translating ribosomes (c versus b) but no greater rate of termination. In case of two genes with the same elongation rate, but different initiation rate (d versus c) the amount of protein that is produced will be lower for the mRNA that has the slower initiation rate (d). Adapted from (210).*

1.2 Heterologous protein overproduction

The main ambition of recombinant protein production is to simultaneously reach a high rate of protein synthesis, a high host cell density and most importantly a high product quality (235). Bearing this in mind, two main requirements from an expression system can be defined as: 1) efficient transcription and translation, and 2) acceptable stability of mRNA and protein molecules (18). Meeting the first prerequisite will result in a high rate of biomolecule synthesis, while stable macromolecules limit the degree of degradation and thus lead to better protein production. When considering the second requirement, central to any production strategy is the physiological status of the host cells, as a consequence of the intimate coupling between the heterologous gene expression and the overall cell fitness (122, 153, 154). Even though the basic goals of efficient gene expression have been pointed out, the inherent biodiversity of DNA sequences and their functional products makes the manufacture of a new protein (schematically depicted in **Figure 9**) an elaborate and complex task.

Many inventive and sophisticated strategies have been developed to meet the challenges in protein production (62, 205, 254, 257) as well as to address unfavorable attributes of the DNA coding sequence (210, 280). Accordingly, the following chapters aim to describe the recent advances in recombinant protein production together with various expression platforms currently adopted.

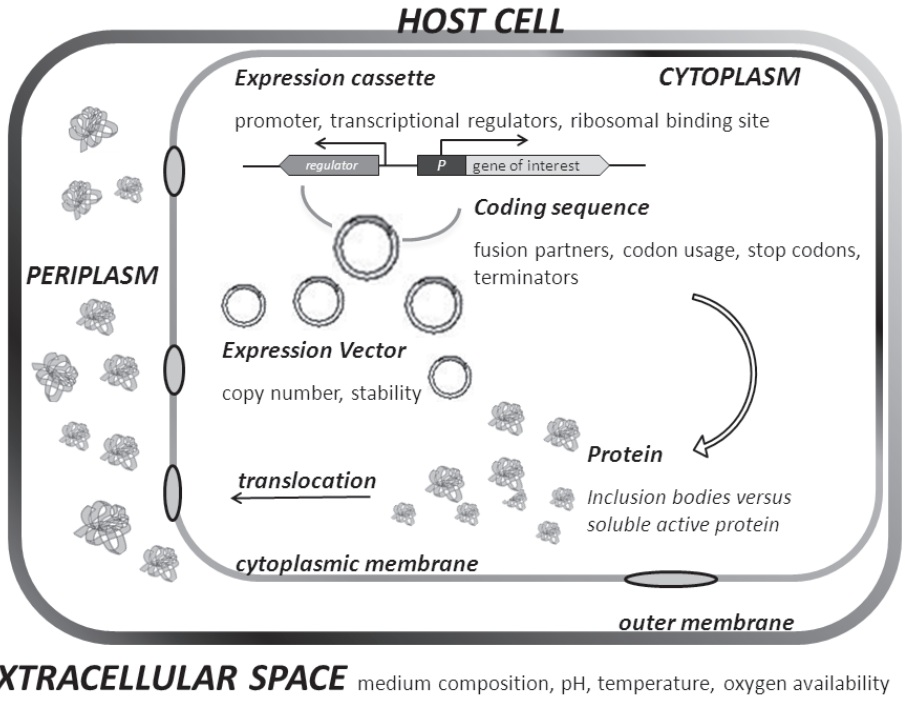


Figure 9 Schematic illustration of the main components of E. coli recombinant protein production system. *Traditional strategy involve transforming cells with DNA expression vector that contains the template and then culturing recombinant cells in defined media to allow the cells to produce the desired protein. Depending on the presence or absence of the corresponding targeting signals, the protein is usually expressed into the cytoplasm or periplasm. The efficiency of the protein production frequently depends on the unique characteristics of the targeted protein and its coding sequence. Several combinations of expression system, media and host are therefore often tested before achieving a viable production process.*

1.2.1 Expression host

By definition, an expression host provides a context for an expression vector to allow foreign gene function (that is, protein production). However, protein overproduction and especially synthesis of toxic proteins presents *E. coli* with growth conditions considerably different from its natural environment. This generally leads to deterioration in cell physiology and decreased protein yield. For these reasons, the use of genetically engineered host strains has been recognized as a fundamental step in effective recombinant protein production (65).

Genetic manipulations of the host cell have been either guided by the exact knowledge of specific biochemical pathways or approached from a more complex, genome-wide, perspective (**Figure 10**). When a specific cellular pathway is known

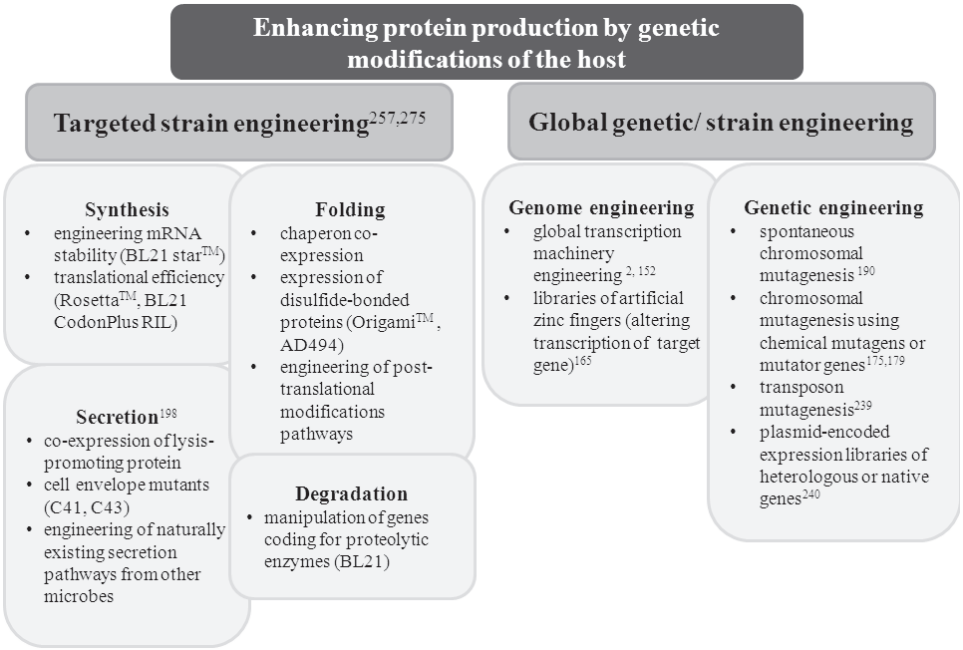


Figure 10 Engineering of host strain genetic make-up *(2, 152, 165, 175, 179, 190, 198, 239, 240, 257, 275). Both targeted and global genetic engineering (not specifically targeted) strategies aim to produce engineered E. coli strains with modifications in the expression level of stress-related sigma factors or with deletions of several proteases and mRNA endonucleases. The stress response related to heterologous protein overproduction can be also controlled by co-expression of key enzymes aiding protein biosynthesis. Examples of relevant commercial strains are given in the brackets.*

to affect protein biogenesis (including also transport, folding and stability) then targeted genetic modifications can be introduced (257, 275). Most of the efforts of targeted strain engineering can be classified into four major groups; 1) enhancing heterologous protein biosynthesis and stability, 2) controlling stress-related phenomena, 3) restoring host cell physiology and addressing problems with protein misfolding and solubility; 4) engineering of *E. coli* strains to allow extracellular protein production and/or post-translational modifications. If, on the other hand, the cause of

poor expression is not known one may manipulate the genetic background of the host by random mutagenesis of the entire genome or by co-expression of libraries in which the expression of originally chromosomally encoded genes can be up- or down-regulated (14, 49, 174). Several new approaches for the genome scale engineering of *E. coli* have been presented and their ultimate advantage is in simultaneous introduction and evaluation of multiple gene modifications (2, 278). These methodologies have facilitated the optimization of host strains in a manner analogous to metabolic engineering, where in order to produce a desired biochemical the cell's metabolism is altered by genetic manipulation (149). The next four subchapters aim to give an outline of current strategies to overcome limitations of *E. coli* as a recombinant protein producer.

Glycoprotein production in *E. coli*

Over the past few years significant progress has been made in diminishing several incapacities of *E. coli* when it comes to producing complex eukaryotic proteins. Due to evident biological dissimilarity between prokaryotic and eukaryotic cells the quality of recombinant proteins produced in bacteria is often not comparable to the grade of those produced from their natural sources (62, 125). One of the long-term issues in recombinant protein production was based on the belief that bacteria are incapable of supporting different posttranslational processes which yield essential modifications for correct cellular function of proteins. This notion was challenged by the discovery of an N-linked glycosylation system in the Gram-negative bacterium *Campylobacter jejuni* and successful transfer of the glycosylation pathway to *E. coli* (273, 274). As more than two thirds of eukaryotic proteins are predicted to be glycosylated (8), the ability to manipulate various glycosylation pathways and thereupon to produce engineered glycoproteins has further broadened the use of bacterial systems for production of high added value proteins (92, 127, 202).

Inclusion bodies

One of the important consequences of the inability of non-engineered *E. coli* cells to complete the biosynthetic process with specific posttranslational modification is the

formation of inclusion bodies (IBs); insoluble protein aggregates often regarded as biologically valueless material (20, 95). This outcome of protein overproduction is both due to missing posttranslational modifications and the limited capacity of the host cell to cope with non-physiological amounts of foreign proteins. It has been documented that the extent of cytoplasmic and periplasmic protein aggregation is determined by the combination of protein sequence specifics together with the process parameters such as; culture media composition, growth temperature, protein production rate and the availability of chaperones and proteases (205, 222, 257). All of these factors can be manipulated to enhance the protein solubility, which is considered the main macroscopic signal of successful protein conformation and functional quality at the molecular level (102).

Although IBs are generally regarded as a troublesome condition, the production of a protein in this form also offers several opportunities. The protein deposition in recombinant cells is fully reversible (55), and the separation from cell debris is simple because of the high portion of the recombinant protein deposited in IBs (more than 90%). That makes IBs a prospective source for *in vitro* refolding and protein recovery (266). More importantly, the polypeptide chains forming IBs retain a certain amount of native structure and a relevant fraction of IB proteins is actually functional (268). Recent insights into the physiology of IB formation and their molecular architecture have revealed a potential for applications of IBs as biocatalysts and as inert nanostructured material, and indicated future biotechnological directions in targeting protein production processes to obtained tailored IBs as the desired biomaterial (96, 97).

Targeting proteins into the periplasm or the culture media

For Gram-negative bacteria such as *E. coli*, translocating recombinant protein outside of the cell requires overcoming a complex cell envelope comprising of cytoplasmic membrane and cell wall plus an outer membrane (198). Therefore, overexpressed recombinant proteins typically accumulate either in the cytoplasm or periplasmic space (257). In basic research purposes and prior to any optimization action,

recombinant proteins are usually targeted to the cytoplasm, which has the advantage of higher product yield (74). On the other hand, in many applications such as high-throughput screening and production of toxic proteins, periplasmic production of recombinant product is desirable. Targeting recombinant proteins into the periplasm can be achieved by translocation signal sequences that direct the heterologous gene of interest through either sec-dependent or twin-arginine translocation path (243). The protein translocation presents several advantages such as authentic N-terminal, better folding environment with limited proteolytic degradation, simplified downstream processing and detection (186). The improved recombinant protein folding is aided by the oxidative environment of periplasm in which disulfide bonds, important structural features of many proteins, are preferentially formed (70). This is in contrast to the *E. coli* cytoplasm which is constantly maintained as a reducing environment.

Commonly used concepts to achieve extracellular protein secretion in *E. coli* are engineering of naturally existing secretion pathways from other microbes and fusing targeted protein to a carrier protein that is a native extracellular or outer membrane protein (5, 66, 142). Despite of strain engineering and exploitation of various secretion pathways, the lack of an efficient secretion mechanism is still considered one of the most significant barriers of using *E. coli* in advanced applications (64).

Growth conditions

Among factors influencing the expression of foreign proteins in *E. coli* is also the extracellular milieu determined by the media composition, pH and temperature. As continuous high level expression results in a drain of the cell's energy resources, the type and amount of nutrients available in the growth medium can greatly augment the fitness of bacterial cell (41, 222). One of many inconveniences in heterologous gene expression is the difference between the codon usage of *E. coli* and of the overexpressed protein natural producer. The tRNA pools of the host cell closely reflect the biased codon usage of the resident mRNA population (79) and the demand for one or more rare tRNAs can seriously impede protein synthesis (162). A couple of investigations also demonstrated that the rate of amino acid supply is restricting the

rate of protein elongation on ribosomes and hence protein synthesis (86, 87). As a solution to the negative effect of different codon biases can then be used an enrichment of the intracellular tRNA pool of the host either by over-expressing genes encoding the rare tRNAs (111) or by supplying the limiting amino acids in the culture medium (82). The medium composition and temperature also play an important role in controlling the relative level of soluble fraction and the total level of accumulation of a given protein (41). Although this is attainable mostly by trial and error, testing of several media types may lead to improved protein yields (245).

1.2.2 Expression vectors

The most favorite gene carriers with long history of use as cloning and expression vectors are plasmids; cytoplasmic autonomously replicating circular DNA molecules (125, 140, 205). They gained their popularity for simple means of manipulation, the flexibility to modulate gene dosage accessible for expression and for the option of employing alternative expression hosts when utilizing special shuttle vectors or broad-host-range replicons. The optimal plasmid copy number for recombinant gene expression depends on the host capacity to both maintain multiple plasmids and to cope with the general stress imposed by protein overproduction (19). In addition to plasmid copy number, other factors with an essential impact on system productivity are the plasmid structural and segregational stability (94, 150). Loss of stability can lead to unequal distribution of plasmids to daughter cells and eventually plasmid-free cells (segregational instability), to an altered DNA coding sequence (structural instability) and, in the worst case, to displacement of productive plasmids with non-productive ones and creation of cells that are resistant to selection pressure but not producing recombinant product.

The metabolic burden brought by maintaining the plasmid and the concomitant recombinant gene expression puts the cell at a selective disadvantage relative to a plasmid-free cell and leads to a gradual decline in its productivity (204). For basic research purposes this is not a serious problem, as plasmids can be maintained using antibiotic selection. Although widely used at laboratory scale, broader use of

antibiotics on an industrial scale is often not desirable due to the high associated cost and the potential risk of environmental pollution and spreading of antibiotics resistance genes (125, 140). Therefore alternative methods have been developed that could eliminate the use of antibiotics and administer stabile plasmid-bearing cultures (81, 99, 130, 269). The joint principle of the antibiotic-free expression is the culture dependence on the presence of a plasmid which encodes an essential gene for the recombinant cell survival.

Because of the intricate plasmid physiology and the build-in difficulties in controlling processes related to replication and segregation, biotechnological research has taken direction also towards expression strategies without plasmids (145). The most stable expression vector is the bacterial chromosome itself and several methods based on chromosomal integration have been proposed. A particular plasmid-free solution called chemically inducible chromosomal evolution (CIChE), supported by two- to fourfold increases in the yields of two useful biochemical products, was recently presented (262). Yet another method used integration of the target gene site specifically into the genome of the host (253). Central to any integration strategy is the number of copies and the location on the chromosome into which the gene of interest is inserted. This is because both can have a strong impact on the heterologous gene expression level and any native gene that is interrupted in the process (246).

Gene expression cassettes

As mentioned in **Chapter 1.1.1**, control of gene expression in prokaryotes is believed to be dominated by transcription and transcript processing (150). Because of that, the DNA region facilitating transcription (a promoter) plays an essential role in the design of an expression vector. An established fact is that the level of gene expression has direct effect on both the host cell fitness and the resulting plasmid stability, and constitutive expression is known to negatively affect both (121, 122, 241). Therefore, inducible systems are preferred and an optimal induction method should be simple, cost-effective, non-toxic and independent of the media components (18, 106). Systems commonly employed for heterologous protein expression require the addition of an

inducer molecule, the depletion or addition of a nutrient, or a shift in a physical or physicochemical factor. In addition to the regularly used expression system such as AraC/P_{BAD}, LacI repressed P_{lac}, $P_{tac,}$ and P_{trc}, and the bacteriophage T7 RNA polymerase/promoter system, there are also inducible promoters based on thermoinducible pR, pL promoters and nutritionally inducible promoters, such as P_{trp} and P_{phoA} (140, 185, 265). Characteristics of several commonly used expression systems are given in **Table 2**. The σ^{32}/σ^{38}-dependent *Pm* promoter together with its cognate positive regulator XylS, which is the primary expression system used in experimental studies presented later (chapter **Summary of the Results and Discussion**), is described in more detail below (sub-chapter XylS/*Pm* expression system).

When taking into consideration the preferred attributes of a promoter, strength is usually the first priority. A strong promoter can partly compensate for DNA coding sequence limitations, such as mRNA instability or poor translation efficiency, by producing large amounts of mRNA. On the other hand, overloading the cell machinery with recombinant mRNA can lead to host cell collapse. This is the case of the exceptionally active T7 expression system where the bacteriophage T7 RNA polymerase is directed exclusively to the transcription of the target gene (106, 125, 257). In addition, the T7 polymerase elongates transcripts much faster than the host RNA polymerase and this specific T7 polymerase property might eventually cause transcript instability due to decoupling of transcription and translation (134, 168). Use of special plasmids/ strains which are co-expressing the polymerase natural inhibitor T7 lysozyme or addition of glucose into the media is then needed to suppress T7 polymerase activity (247).

Table 2 Commonly used promoter systems for protein expression in *E. coli* (18, 106, 125, 140, 243, 247)

Promoter	Regulator's gene	Induction	Features
P_m	*xylS*	Benzoic acid derivatives	medium to high level expression, tight control, fine-tuning expression in a dose-dependent manner, inexpensive inducers
P_{BAD}	*araC*	L-arabinose	medium level expression, tight control, media limitations, specific host strain requirement, fine-tuning expression in a dose-dependent manner, inexpensive inducer
P_{lac}	*lacI, lacI*^{q}	IPTG	low level expression, high basal level, expensive induction
P_{trp}	*trpR*	Trp starvation	low level expression, high basal level, expensive induction
$P_{tac,}$ P_{trc}	*lacI, lacI*^{q}	IPTG	medium level expression, high basal level, specific host strain can reduce high background expression, expensive induction
T7	*λ cIts 857* (polymerase)	IPTG	very high level expression, genetic instability, low basal level depends on co-expressed vectors, specific host strain requirement, expensive induction

Besides the strength there are several other promoter attributes that are of importance. Firstly, regulation of the level of expression by varying inducer dosage is essential, for example; during toxic protein production as it plays an important role in balancing production yield and metabolic burden (223). Secondly, a promoter system should also have limited basal expression in the non-induced state. Leaky expression can cause metabolic stress by diverting the carbon and energy source before even reaching optimal cell density for recombinant protein production (18). This situation is particularly detrimental to the host cell when the expressed protein is toxic. Regulation of expression level by inducer titration goes hand in hand with an innate characteristic of each promoter system - its induction kinetics. Promoters with differing rate of induction (i.e. respond to the addition of inducer with respect to the kinetics of recombinant mRNA accumulation) can be used to regulate the rate of accumulation of the respective recombinant protein and eventually the fraction of soluble protein

formed. A slower rate, displayed for example by the P_{BAD} promoter (167), usually enables cells to grow continuously in spite of the metabolic burden. Yet another important aspect for the protein production process is homogeneous induction that guarantees that all cells in a culture are induced to the same level (151).

The XylS/*Pm* expression system

The inducible *Pm* promoter originally regulates the expression of the *meta*-cleavage pathway operon involved in the catabolism of aromatic hydrocarbons in the TOL plasmid (pWWO) of *Pseudomonas putida* (212). Its positive regulator XylS is activated by the binding of downstream products in the pathway (alkylbenzoates) that can function as effective low cost inducers (e.g. *m*-toluate). A significant advantage of using such low-molecular weight inducers is that they enter *E. coli* cells by passive diffusion which is believed to eliminate problems with all-or-none response of other promoters (148, 151). Furthermore, the expression levels from the *Pm* promoter can easily be controlled by varying the inducer amount and concentration (282). The XylS/*Pm* system was previously integrated into minimal replicon broad-host-range expression vectors (the pJBn vectors) based on the RK2 plasmid (31, 32). One of these vectors (pJB658) that has been used to produce industrial levels of several medically important proteins in *E. coli* under high-cell densities (242, 243) was also used as the basis for plasmids constructions in experimental studies reported in this work.

1.2.3 Alternative bacterial expression hosts

E. coli-based systems dominate the bacterial expression systems and currently represent about 34 % of all expression platforms used for the manufacture of biotherapeutics in the US and the EU (187). However, a great part of heterologous proteins cannot be expressed at all or only with difficulties (may also be non-functional) when *E. coli* is assigned to the task. Consequently, alternative bacterial hosts are explored and employed especially for production of extracellular proteins (62, 74). Unlike Gram-negative bacteria, the Gram-positives lack an outer membrane composed of lipopolysaccharides (236). This distinguishable feature offers two benefits; simplified protein secretion enabling effective downstream protein recovery

and absence of lipopolysaccharides that function as antigenic endotoxins (69). Because of these handicaps, biotechnologists have paid considerable attention to gram positive species such as *Bacillus*, *Lactoccocus, Streptomyces, Corynebacterium* and *Staphylococcus*. Furthermore, an increasing number of studies is documenting investigations and eventual genetically manipulations of both Gram-positive and Gram-negative bacteria in order to use them for both therapeutic and food applications.

The soil microorganism *B. subtilis* with generally recognized as safe (GRAS) status is known to naturally produce and secrete an abundance of various proteins into the growth medium (286). Besides *B. subtilis*, several other *Bacillus* expression systems (*B. megaterium*, *B. licheniformis* and *B. brevis*) have been established as major cell factories for homologous expression of secreted enzymes such as proteases and amylases and are also preferred for production of secreted heterologous proteins (74). The lactic acid bacterium *Lactococcus lactis* has emerged in recent years as an attractive alternative to the *Bacillus* model (84, 93, 161). Looking at alternative Gram-negative hosts, *Pseudomonas fluorescens* shares with *E. coli* the ability to overproduce a recombinant protein to yields more than 50% of total cell protein and to grow to a high cell density (214). In addition, *P. fluorescens* is less dependent on defined oxygen concentrations and does not accumulate acetate during fermentation (126). Yet another Gram negative-based expression system, presented as superior to *E. coli* with respect to inclusion body formation, has been developed in *Ralstonia eutropha* (23).

To summarize the second chapter on heterologous protein overproduction in bacteria; all the above considerations on properties and characteristics of various *E. coli* expression systems, together with the on-going exploration of alternative bacterial expression hosts, illustrate how to eventually tackle the empirical nature of protein expression. Increasing the number of unconventional bacterial expression systems, suitable for a particular category of difficult to express proteins, assists and complements the veteran systems such as *E. coli* whose possible applications are being constantly expanded.

2 AIMS OF THE STUDY

The presented PhD work was part of a comprehensive project entitled “A Combinatorial Mutagenesis Approach to Improve Microbial Expression Systems”. One of the principal objectives of this research program was to develop strategies for obtaining high-level production of industrially and medically important proteins in *E. coli* by using the XylS/*Pm* expression system. The more specific goal of this PhD work was to study the influence of the nucleotide sequences at the 5´ ends of target genes on their corresponding levels of protein expression. The DNA sequences primarily targeted for functional investigations were the *Pm* promoter associated 5´-UTR and adjacent 5´ coding regions of several genes of interest. In all cases, the experimental strategy involved sequence directed evolution (*in vitro* mutagenesis coupled to high throughput screening) which would eventually lead to sequence variants optimized for high-level expression at the protein level. Besides synonymously mutating the 5´ coding sequence also the effect of random mutations introduced into the DNA sequence of the 5´-UTR and of two 5´ fusion partners was subjected to thorough characterization.

As the work progressed it became evident that for an accurate description of the observed stimulatory effects caused by different DNA modifications it is essential to be able to directly evaluate and quantify gene expression at its different levels. Therefore the work also addressed how the introduced changes stimulate expression at the level of transcript accumulation and mRNA stability. Current methods for monitoring mRNA decay in *E. coli* were considered not to be satisfactory, and therefore a study aimed at developing a novel technique for evaluation of relative decay rates of recombinant mRNAs produced from the *Pm* promoter was initiated. Since the new methodology was based on a general principle of washing off diffusible transcriptional inducers of bacterial expression systems, the same methodology could be extended to study induction and decay kinetics of mRNAs produced from two other commonly used expression cassettes; containing the bacteriophage T7 and P_{tac} inducible promoters.

Finally, participation in a comparative study evaluating the relative performance of five bacterial promoter systems was undertaken. The main intention of that study was to describe how the different promoter systems respond to the generally empirical nature of gene expression, which is mainly a consequence of the target gene sequence.

3 SUMMARY OF RESULTS AND DISCUSSION

The results of this PhD work are described in Papers I-V and the following chapters are summarizing the major findings as well as discussing them in a broader context.

3.1 Functional analysis of DNA sequences corresponding to the 5´ end of recombinant mRNAs

The contribution of DNA sequences corresponding to the 5´ coding region and the adjacent UTR to the regulation of gene expression and to the determination of the resulting protein level is currently a well-established fact (see Introductory **Chapter 1**). Nevertheless, the understanding of the exact mechanisms by which these genetic elements can drive control of gene expression has been lagging behind the reasonably well characterized regulatory functions of the upstream promoter sequence. One possible explanation for this disparity is the many-sided role of the 5´ end that includes control of translation, mRNA decay and transcription. In addition, an accurate and reliable description of mRNA turnover and translation *in vivo* is often difficult to obtain with current experimental techniques. In view of the potential of DNA sequences corresponding to the 5´ end of mRNA for the control of heterologous gene expression, the work focused on: 1) development of empirical strategies for optimization of the 5´ ends for high-level protein production; 2) investigations of mechanisms behind the stimulation of recombinant gene expression.

3.1.1 Random mutagenesis of DNA region corresponding to the 5´ end of recombinant genes as an effective optimization strategy for achieving high-level protein production in *E. coli*

Previous investigations of the XylS/*Pm* system using the *bla* gene (coding for β-lactamase, enzyme capable of degrading ampicillin) as the expression reporter, showed that up to 20-fold stimulation at the protein level can be achieved by mutating DNA sequence corresponding to the 5´-UTR (30). This finding was followed up by more detailed analyses of the UTR region and its potential for improving expression of recombinant proteins from the XylS/*Pm* system (**Paper I**). The effects of the DNA

sequences of several UTRs (termed LII UTR variants) were studied also for two other genes; *celB* (encoding phosphoglucomutase) and *pelB-ifn-α2b$_5$* (encoding the pelB translocation signal peptide and human cytokine interferon α2b). The results showed that expression at the protein level of both genes could be further increased by at least one of the tested LII UTR variants (**Table 3 –** LII-10 and LII-12); despite the already high levels of expression of these genes from the *Pm* promoter and its associated wild type UTR (32, 243). The improvements (fold relative to the wild-type) of *bla*, *celB* and *pelB-ifn-α2b$_5$* expression at the protein level appeared highly gene-dependent and also a certain degree of context dependency between specific UTR sequence and the associated coding sequence was observed. The differences in fold improvements were not unexpected, since the absolute level of protein expression of *bla* is generally much lower than for the two other genes.

Because the LII UTR variants were selected on the basis of their efficiency to stimulate expression of *bla* also the possibility that there might exist specific UTR variants more optimal for expression of each individual gene was evaluated. For that purpose, a specific screening tool (termed synthetic operon) was designed (**Paper I**). The synthetic operon construct upon addition of inducer supported transcription from the *Pm* promoter of one single mRNA molecule consisting of two coding regions. The system functionality was based on the mechanism of translation reinitiation in which the translation initiation of the reporter gene relies on the translation efficiency of the upstream gene of interest (1, 4). With the choice of *celB* as the gene of interest and *bla* as the reporter, a new UTR mutant library (designated LV) was generated in the synthetic operon construct upstream of *celB*. Subsequently, LV library host cells were screened for mutants conferring improved ampicillin tolerance and thus indirectly for 5´-UTR variants that cause improved expression of *celB*. Analysis of DH5α strains expressing selected LV UTR variants in single gene context of *celB*, *bla* or *pelB-ifn-α2b* (**Table 3 – LV-3 and LV-4**) demonstrated that the overall protein production level of each tested gene with the most efficient UTR variant (either from the LII or LV library) is similar. Assuming that the screening procedure was efficient in itself; one possible interpretation of these results is that the attainable level of stimulation of

heterologous gene expression through optimization of the DNA sequence corresponding to the 5´-UTR is highly dependent on the gene itself.

Table 3 Relative protein production level of β-lactamase (Bla), phosphoglucomutase (CelB) and interferon α2b (IFN-α2b) when the respective genes were expressed from *Pm* in the presence of different UTR variants. For each protein, all values are compared to the original UTR (wt), arbitrarily set to 1.

UTR variant	Nucleotide sequence (5´ to 3´)	Bla protein/ activity[2]	CelB protein/ activity[2]	IFN-α2b protein level[3]
wt	aacatgtacaataataatggagtcatgaacat	1.0 ± 0.0	1.0 ± 0.1	1.0
LII-10[1]	.........c.c..........tt.....	16.2 ± 0.9	1.5 ± 0.1	~1.0
LII-12[1]	.......c.............-..-.....	16.0 ± 1.2	1.3 ± 0.1	~3.0-4.0
LV-3	..............ca...aca.....	13.4 ± 0.3	1.6 ± 0.1	~1.0
LV-4	............c...cca....a.....	13.7 ± 1.2	0.6 ± 0.2	~0.8

[1] The active protein production levels of Bla were collected from (30). [2] *E. coli* DH5α was used as expression host and was induced with 2 mM *m*-toluate. [3] *E. coli* RV308 was used as the expression host and was induced with 0.5 mM *m*-toluate.

Besides looking at gene expression at the protein level, the amount of transcript was determined both for the *pelB-ifn-α2b*$_S$ fusion and the *celB* gene when associated with the LII and LV UTR variants and compared to the effect of the same sequences on the *bla* transcript level (**Table 4**). As in the case of stimulation of expression at the protein level, context dependency between the DNA sequence of each specific UTR and the concomitant coding sequence of each model gene appeared to somehow influence the final transcript level. Moreover, the increase in transcript amount was not always correlated to the level of protein production (**Table 3**). Specifically, the LII-10 variant resulted in about 3-fold stimulation of transcript accumulation of both *celB* and *pelB-ifn-α2b*$_S$, but the corresponding protein levels did not change proportionally. The smaller fold improvement in transcript accumulation of both *celB* and *pelB-ifn-α2b*$_S$ (compared to more than 13-fold stimulation of the *bla* transcript amount) indicates again a certain degree of gene-specificity, but also possible coding sequence constraints when it comes to enhancing expression of already well-expressed genes.

Table 4 Relative transcript levels for *bla*, *celB* and *pelB-ifn-α2b*$_S$ when expressed from *Pm* using different UTR variants. For each transcript, all values are compared to the original UTR (wt), arbitrarily set to 1.

UTR variant	*bla* transcript level[2]	*celB* transcript level[2]	*pelB-ifn-α2b*$_S$ transcript level[2]
wt	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.2
LII-10[1]	5.6 ± 0.1	3.0 ± 0.3	3.6 ± 0.8
LII-12[1]	4.9 ± 0.4	1.6 ± 0.1	3.2 ± 0.3
LV-3	13.7 ± 1.1	2.9 ± 0.1	1.4 ± 0.1
LV-4	11.4 ± 0.4	1.6 ± 0.2	0.8 ± 0.1

[1] The accumulated *bla* transcript levels were collected from (30). [2] The experiments were performed as described in Table 3.

In addition to the 5´-UTR, in-frame additions of short DNA sequences corresponding to 5´ fusion partners to poorly expressed genes have been reported to positively affect the expression level (115, 243). Here, the same random mutagenesis strategy was used for two model 5´ fusion partners; specifically for the DNA sequence of a consensus translocation signal peptide (CSP) and a 23 codons long sequence originating from the *celB* gene (**Paper II and III,** respectively). The main difference between these 5´ fusion partners is related to their sequence function. Whereas CSP is involved in translocation of the mature protein into the periplasm, the *celB* gene encodes a cytoplasmic protein. In the second case the translocation process was therefore excluded as a potential parameter affecting the protein production level.

To identify optimized sequence variants, antibiotic resistance genes were used as the reporters of expression from the XylS/*Pm* system. Screening of CSP random library was carried out by using *bla* and in that manner the procedure was also coupled to the protein translocation into periplasm. For screening of random library of *celB*$_{23}$ sequence the *aac(3)-IV* gene was used, coding for cytoplasmic aminoglycoside-(3)-acetyltransferase IV that can inactivate apramycin. In **Paper II**, a collection of optimized CSP-based signal sequences was reported and found to increase the expression of two additional model genes: up to 8-fold at the active protein level for *phoA* (coding for alkaline phosphatase) and up to 3-fold at the total protein level for *ifn-α2b*$_S$ (during high-cell density cultivation), as compared to the CSP wild type. In

Paper III, an optimized DNA sequence of *celB*$_{23}$ (termed *celB*$_{D11}$) was reported to lead up to 4-fold stimulation of the protein production level of *ifn-α2b*$_S$ and three other human genes, coding for the medically important granulocyte-colony stimulating factor (G-CSF), granulocyte macrophage-colony stimulating factor (GM-CSF) and tumor necrosis factor (TNF-α1a). In contrast to CSP, *celB*-based fusion partner was stimulating the corresponding protein production levels independent of translocation. In summary, the studies reported in **Papers I-III** showed that targeting the DNA sequence of the 5´ end with random mutagenesis approach can be a powerful optimization step in the process of improving recombinant protein production.

The same model gene, *ifn-α2b*$_S$, and its expression from the XylS/*Pm* system was investigated in all three studies (**Paper I-III**) together with a reference for high-level expression at the protein level, the fusion gene *pelB-ifn-α2b*$_S$ (243). It should be noted that the protein production of *ifn-α2b*$_S$ was in all three studies found below detection level of a Western blot unless a 5´ fusion partner (*pelB*, CSP- or *celB*-based) was used. Comparison of the protein production levels of *ifn-α2b*$_S$, when different DNA sequences (corresponding to specific variants of 5´-UTR/pelB, CSP or *celB*$_{23}$) were used, showed that the fold improvements relative to *pelB-ifn-α2b*$_S$ are in all cases very similar (**Table 5**). It therefore appears that optimization of DNA sequences corresponding to the 5´-terminal of *ifn-α2b*$_S$ mRNA has certain limitations, when it comes to maximum attainable protein production level, presumably due to other limiting factors characteristic of this specific gene.

Table 5 Relative protein production levels of *ifn-α2b*$_S$, when expressed from *Pm* (*E. coli* RV308, 0.5 mM *m*-toluate) containing combinations of either LII-12 or the wild type UTR and different 5´ fusion partners.

gene	5´-UTR	5´ fusion partner	IFN-α2b protein level
ifn-α2b$_S$	wt	None	not detectable
pelB-ifn-α2b$_S$	wt	*pelB*	1
pelB-ifn-α2b$_S$	LII-12	*pelB*	3- 4
S2-ifn-α2b$_S$[1]	wt	*S2 (CSP variant)*	2-3
celB$_{D11}$*-ifn-α2b*$_S$	wt	*celB*$_{D11}$	3-4

[1] Results obtained under high-cell density fermentation conditions.

3.1.2 The outcome of codon usage optimization of the 5´ coding sequence appears to be strongly gene-dependent

In addition to the random mutagenesis, a codon usage optimization strategy was undertaken to evaluate the role of codon bias of the 5´ coding sequence in shaping expression levels of selected genes (**Paper II-IV**). Two different experimental designs for constructing and screening of synonymous codon libraries were used and both implied that the protein amino acid sequence remains the same. In the first case, the advantage of the near linear relationship between the expression level of the *bla* reporter and the host ampicillin tolerance was again exploited. The DNA sequence of either CSP (**Paper II**) or the native *bla* translocation signal (**Paper IV**) was cloned in frame with the rest of the *bla* gene encoding the mature part of the periplasmic β-lactamase (**Figure 11**). Even though the same gene was used as the reporter in both libraries, the "wild type" expression level (measured as ampicillin tolerance of the DH5α host) differed more than 5-fold (0.15±0.02 g/L for native *bla* and 1.00±0.20 g/L for CSP, under induced conditions of 20 μM *m*-toluate). This difference indicates that the CSP sequence confers more effective *bla* expression, translocation of the corresponding protein or a combination of both, compared to that of the native *bla* sequence.

Comparison of the best performing mutants obtained from screening of either the CSP or the native *bla* signal synonymous library (designated Si and SII, respectively) indicated that the codon usage of the CSP sequence is close to optimal. The Si2 variant of the CSP sequence (with changes in the 3rd codon (Lysin: AAA to AAG), 8th codon (Leucin: TTA to TTG) and 13th codon (Leucin: TTA to TTC)) and C19 variant of the *bla* wild type signal (with changes in the 2nd codon (Serine: AGT to TCT) and 8th codon (Valin: GTC to GTT)) resulted in similar ampicillin tolerance levels of the corresponding DH5α host cells (1.90±0.20 g/L and 1.50±0.20 g/L, respectively, induction of *Pm* with 20 μM *m*-toluate). The assumption that CSP has an already optimized DNA sequence for high-level expression of downstream fusion genes is based on the relatively smaller improvement of ampicillin tolerance level of the corresponding host cells caused by CSP variants (up to 90% of the original CSP). On the other hand, variants of the *bla*

native signal caused up to 10-fold increase in the ampicillin tolerance levels of the corresponding host cells, as compared to the *bla* wild type level.

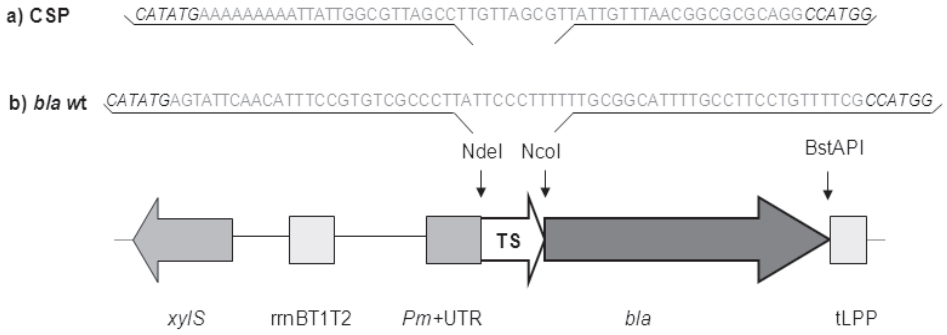


Figure 11 *Schematic representation of the expression related features of two vectors carrying either the CSP Si library (**a**) or the SII library of the native bla translocation signal sequence (**b**). TS – translocation signal; Pm+UTR, positively regulated promoter with associated UTR; xylS, gene encoding the Pm activator; bla, ampicillin resistance gene encoding mature part of β-lactamase; tLPP, transcriptional terminator; rrnBT1T2, bidirectional transcriptional terminator. Details of nucleotide sequence corresponding to the CSP and bla native translocation signal are displayed above the plasmid maps with NdeI and NcoI unique restriction sites in italics.*

Interestingly, during screening of both synonymous libraries a high occurrence of mutants with non-synonymous mutations was observed, even though these were not intended (**unpublished data**). This could be explained to be the result of efficient selection of infrequently occurring unintended mutations introduced during oligonucleotide synthesis or cloning. Furthermore, results from screening of two different libraries of CSP; one with completely randomized sequence (described in the previous chapter) and the Si library containing only synonymous codon changes suggested that there is likely a roof when it comes to improving gene expression through synonymous mutations. This could be attributed to unknown intrinsic features of the sequence itself, perhaps also at the amino acid level (**Paper II**).

In a second investigation of the effect of synonymous codon changes, *ifn-α2b*$_S$ was chosen as the model gene (**Paper III**). Because no effective screening system for improved *ifn-α2b*$_S$ expression was available, a bioinformatic approach was used for both synonymous library generation and its subsequent characterization. In such *in*

silico library constructions the number of included synonymous codons exponentially increases with each additional codon (281). For *ifn-α2b*$_S$, mutating the first 12 codons would yield 3 686 400 different synonymous variants, while 15 codons would yield 235 929 600 variants. Not only does the generation of large combinatorial libraries *in silico* represents a computational challenge, competent screening parameters that would result in identification of sequence variants conferring high-level expression are currently still elusive (112, 281). Nevertheless, an unstructured mRNA sequence near RBS (corresponding to high folding energy) and infrequent synonymous codon usage has been suggested to positively influence translation (260, 279). Accordingly, a combinatorial synonymous library of the first 9 codons of *ifn-α2b*$_S$ was first generated *in silico* and then several *ifn-α2b*$_S$ variants with either the highest folding free energy of the mRNA region comprising nucleotides -32 to +30 (+1 is the adenine of the translation start codon) or most rare codons in the modified region, were selected.

Subsequent analysis of DH5α strains expressing the selected *5´* coding sequence variants of *ifn-α2b*$_S$ (induction of *Pm* with 0.5 mM *m*-toluate) showed no detectable increase in protein production. Interestingly, however; a significant improvement (up to 8-fold) in accumulated transcript level could be observed, as compared to the amount generated from the original *ifn-α2b*$_S$ gene. One possible interpretation of this is that the amounts of protein produced from this particular gene (under the conditions used) are not limited by the amounts of transcript, but by some sequence features negatively influencing processes downstream of transcription. In agreement with this suggestion is the observation that the use of a 5´ fusion partner (*pelB*, *celB*-based or CSP-based) was many-fold more effective in stimulating *ifn-α2b*$_S$ expression at the protein level than the *in silico* codon usage optimization, but at the same time displayed similar transcript levels. In summary, codon optimization strategies that included two model genes (*bla* and *ifn-α2b*$_S$) and were carried out in two different ways, could not outperform approaches by which the 5´ coding sequence (CSP or *celB*$_{23}$) was randomized also at the amino acid level, when considering the protein production level.

3.1.3 Mutating DNA sequences corresponding to the 5´ coding region can result in stimulation of the rate of transcription

The total amount of each mRNA in a cell is determined by the rates of synthesis and decay (27, 144), and therefore any potential stimulation of transcript level might stem either from more stable mRNA, faster rate of transcription or a combination of both. By using the *bla* gene as reporter for the XylS/*Pm* expression system, Berg et al., 2009, examined the contribution of transcription rate and mRNA stability for a case of increased transcript production caused by specific mutations in the 5´-UTR (LV-2 variant). Based on several indirect experiments, the authors suggested the stimulation of the rate of transcript accumulation as the underlying reason for increased *bla* transcript amounts. More direct evidence for these interesting findings was provided by a novel qRT-PCR based method for monitoring recombinant mRNA decay kinetics (the inducer wash out method) introduced in **Paper IV**.

The determination of the decay of two *bla* mRNAs (containing either the wild type UTR or LV-2 variant) confirmed the increase in the rate of transcript accumulation as the dominating mechanism leading to the enhanced *bla* gene expression. In addition to LV-2, another variant (LII-11) that stimulated the *bla* gene expression both at the transcript and protein levels (30) (8-fold and 20-fold, respectively) was tested. The corresponding *bla* gene mRNA stability could not be distinguished from the stability of *bla* mRNA containing the wild-type (or LV-2) UTR (**Table 6)**. These results illustrated that despite the variation in *bla* expression at the transcript and protein level, caused by mutating the DNA sequence of associated 5´-UTR, mRNA stability can stay unchanged. Similar UTR-based stimulation of transcript accumulation kinetics cannot be ruled out also in case of *pelB-ifn-a2b*$_S$ fusion and *celB*, discussed in **chapter 3.1.1**.

Table 6 Relative mRNA decay rates for three pairs of *bla* transcripts, as determined by the inducer wash out method.

Transcripts compared in parallel	Decay rates[1]	Transcript level[2]
bla[3] (wt UTR) vs *bla* (LV-2 UTR)	0.14 (0.13-0.17) vs 0.13 (0.12-0.16)	1.00 vs 9.78
bla[3] (wt UTR) vs *bla* (LII-11 UTR)	0.16 (0.16-0.17) vs 0.17 (0.16-0.21)	1.00 vs 4.88
bla[4] vs *bla* (C19)	0.16 (0.14-0.18) vs 0.19 (0.15-0.23)	1.00 vs 4.66

[1] Decay rates were computationally estimated by using methodology described in Paper IV. 95% confidence intervals are given in brackets.
[2] Accumulated transcript level are given compared to the wild type (arbitrary set to 1), at time point 0 minutes after inducer wash out.
[3] pIB11 expression construct (described in (30))
[4] pBSP1bla expression construct (described in Paper II)

The mRNA stability was further studied in the context of synonymous codon changes of the 5´ coding sequence of *bla*. The above described C19 variant of the *bla* gene (**Chapter 3.1.2**), was found as a suitable candidate for comparison of its mRNA decay kinetics with the wild type *bla* mRNA. Similarly to the *bla* associated 5´-UTR variants, mRNA stability was not significantly affected (as jugged from the mathematically determined relative decay rates listed in **Table 6**) despite more than 4-fold increase in transcript accumulated levels. The overall findings from **Paper IV** are suggesting that DNA sequence corresponding to the 5´-terminal of mRNA can have a rather big impact on transcription itself. Couple of studies described that the DNA sequence of the *lac* operator (binding site for the Lac repressor that is located in the 5´-UTR) resembling a promoter –10 consensus (TATAAT) can induced σ^{70}-dependent transcriptional pausing (40, 199). In case of the tryptophanase operon, RNA polymerase pause sites have been described in the 220-nucleotides long spacer region (separating the coding regions of TnaC leader peptide and TnaA tryptophanase) and implicated in coupling of translation with transcription presumably through allowing time for a ribosome to bind to the transcript and begin translation (101). One possibility is therefore that mutations within either the 5´-UTR or the *bla* coding region affect pause sites in the wild-type *bla* DNA sequence in a way that is stimulatory for transcription.

3.1.4 5´-terminal fusions can confer more stable mRNA for poorly expressed heterologous genes

Analogous to the analysis of stimulation of gene expression caused by mutating the 5´ UTR or the *bla* 5´ coding sequence, examination of the stimulatory effect of several 5´ fusion partners was carried out (**Paper III, Paper IV**). Sletta et al. (2007) described stimulation of the *gm-csf* gene expression at the transcript and protein levels by 5´ fusion of the *ompA* translocation signal sequence. To determine whether the increase in transcript levels originates from stimulation of transcript synthesis or more stable mRNA, the inducer wash-out method was again employed **(Paper IV)**. The addition of the *ompA* fusion partner was found to lead to a 3-fold decrease in the corresponding mRNA decay rate. Furthermore, a similar effect (about 2-fold reduction in the mRNA decay rate) was observed for the *pelB* translocation signal sequence, when used as 5´ fusion to *ifn-α2b*$_S$. Thus, in these two specific cases the use of translocation signal sequence as 5´ fusion partner led to enhanced transcript stability of the corresponding recombinant genes.

The importance of the translocation function (via *ompA* and *pelB*) for efficient protein production was indirectly examined on a model example of *ifn-α2b*$_S$ (**Paper III**). A major finding of that study was that a 5´-terminal coding region of the well-expressed *celB* gene can replace *pelB* as an effectively 5´ fusion partner, stimulating the protein production of *ifn-α2b*$_S$ to a similar level. The stimulation of *ifn-α2b*$_S$ expression was highest when using *celB*-based fusions of 20-25 codons, resulting in more than 7-fold and at least 60-fold stimulation in transcript and protein levels, respectively. In order to have characterization also at the level of mRNA stability, one of the best performing *celB*-based 5´ fusion partners was used for determination of relative mRNA decay rate by the inducer wash-out method. Both the experimental data and their subsequent mathematical fitting supported up to 2-fold reduction in mRNA decay rate for the construct expressing *ifn-α2b*$_S$ with in-frame 5´ *celB* fusion.

Addition of a 5´ fusion partner may result in changes of, for example; mRNA secondary structure, RNaseE-mediated mRNA decay and also protein folding (115, 243). The observed increase in the rate of transcript accumulation and prolonged

mRNA half-lives may predict a possible protection of the transcripts by translating ribosomes due to more efficient translation initiation, as was previously observed in other studies (132, 272). In accordance with those reports, 5´ fusion partners were found to stimulate both mRNA stability and expression at the protein level. However, considering the observed multifold expression differences at the protein level, it appears likely that improvement in mRNA stability is a secondary effect of better translation, and that the primary reason for the very big increase in protein production is strongly improved translation.

3.2 Comparison of the performance of commonly used protein expression systems for E. coli

Recombinant protein production is to a great extent an empirical process. It is mostly because of various features of each specific protein coding sequence, which limit the upper level of expression by control mechanisms that are still not fully understood (112). However, appropriate selection of expression host, vector and culturing conditions can greatly aid in accomplishing effective expression (125, 223, 245). It was therefore of interest to compare how commonly used expression cassettes (comprised of specific inducible promoters and their accessory regulatory proteins) together with their recommended hosts, can influence the expression outcome. It is generally believed that different expression systems respond in a difficult to predict way towards each individual heterologous gene and its protein product. However, comparative studies characterizing such systems with respect to their relative performance towards selected model genes are still limited in numbers (256, 283). The prediction value of such studies is also often restricted unless all or most of the potential parameters influencing protein expression (e. g. vector design, host genetic make-up) are kept the same.

3.2.1 Evaluation of four expression cassettes based on the host RNA polymerase and one based on a bacteriophage T7 RNA polymerase

The last study included in this thesis (**Paper V**) was based on a set of identical vector backbones containing the regulator/promoter regions of XylS/*Pm*, XylS/*Pm* ML1-17 (a

Pm variant), LacI/P_{trc}, AraC/P_{BAD} and the bacteriophage-based LacI/P_{T7lac} expression system (also known as the T7 RNA polymerase/promoter system, hereafter referred to as T7). For the performance evaluation, five genes coding for luciferase, an antibody fragment fused in frame to alkaline phosphatase (scFv173-2-5-AP), green fluorescent protein (GFP), human growth hormone (HGH) and interleukin-1RA (IL-1RA), respectively, were selected as models for their distinct properties.

The results revealed many expression system and model gene specific features and, most importantly, that none of the expression systems was in all evaluated aspects superior to the others. While the T7 system had an apparent advantage by its capacity for producing large amounts of transcript, this potential was often not reflected at the translational level as judging from determination of the protein levels by enzymatic assays and SDS-PAGE analysis. This disparity was likely due to limitations in the protein production and/or folding capacity of the host cell. The advantage of XylS/*Pm* and the AraC/P_{BAD} systems was on the other hand low basal expression level and at the same time high inducibility. Basal expression can be also reduced for the T7 system, but then a special expression host harboring the *lacI*^{q} gene, whose mutated promoter increases LacI repressor expression 10-fold (46), is required. It should be noted that the ratio between the induced and the uninduced expression levels was protein dependent, with a relatively small induction window for toxic svFv173-2-5-AP (1.2-25) and large for luciferase (60-3,000). The XylS/*Pm* system was, as expected, found to be host-independent, in contrast to the T7 system functional dependency on a strain with a chromosomal copy of the T7 polymerase gene and the requirement of the AraC/P_{BAD} system for a strain unable to metabolize the expression inducer, arabinose. Although LacI/P_{trc} produced the least amount of protein in four out of five investigated cases, results from flow cytometry analysis indicated that this system might give rise to fast production kinetics. Such feature may be of some value for example for production of proteins with low stability.

The comparison of the most commonly used expression systems revealed that an expression system needs to be evaluated for each specific case and that the observed differences in performance between the various systems are likely a direct

consequence of the properties of the expression cassettes. By taking into account the relevant features characteristic of each expression cassette together with specifics of a desired protein to be produced, it might eventually become easier to make the best choice with respect to the goal of achieving high protein production either in functional, soluble form or non-functional, but high yield form such as inclusion bodies.

3.2.2 Accumulation and decay kinetics of recombinant mRNAs are influenced by the expression cassette used

The response to the inducer addition of two IPTG-inducible expression systems; LacI/P_{tac} and T7, was studied by previously published qRT-PCR based methodology (30), and compared to the accumulation kinetics of transcripts originating from the XylS/*Pm* system (induced by *m*-toluate) (**Paper IV and unpublished results**). The results showed that each of the systems affects transcript accumulation in a way that correlates with the type of transcription regulation the systems are subjected to.

Two *E. coli* ER2566 strains, harbouring expression vectors containing the *IL-1RA*$_S$ gene (coding for interleukin-1RA) that was governed either by the T7 or *Pm* promoter, were used to inspect the kinetics of the recombinant transcript accumulation. The positive regulation of the *Pm* promoter by the XylS transcriptional factor, which involves regulator dimerization before activation of transcription (36), was characterized by a gradual increase in accumulated transcript level for the first 20 minutes after induction, consistent with previous findings (30). In contrast, the T7 system was characterized by a 15 minutes lag period before the onset of a rapid accumulation phase, presumably reflecting the time needed to first synthesize the T7 polymerase (whose gene is encoded by the host genome and its transcription is activated by IPTG) (**Paper IV**). Yet another type of accumulation kinetics was observed with the negatively regulated (by the LacI repressor) P_{tac} promoter. Even though different *E. coli strains* were used for the LacI/P_{tac} and XylS/*Pm* systems (NEB Express Iq and DH5α, respectively), the same recombinant transcripts were measured (*gm-csf* and *ompA-gm-csf*) and therefore an indirect comparison could be made (**Figure 12**). The NEB Express Iq strain was employed because of significantly reducing the

background expression level from the P_{tac} promoter, one of the prerequisites for the functionality of the inducer wash-out method. The transcript accumulation from the LacI/P_{tac} system displayed rapid accumulation kinetics and reached saturation within a few minutes. One way of interpreting this is to assume that the DNA binding property of the LacI repressor is immediately inactivated by the addition of the IPTG inducer, providing fast access for the host RNA polymerase. Correspondingly, the establishment of the XylS/RNA polymerase complex may be formed by a slower kinetics.

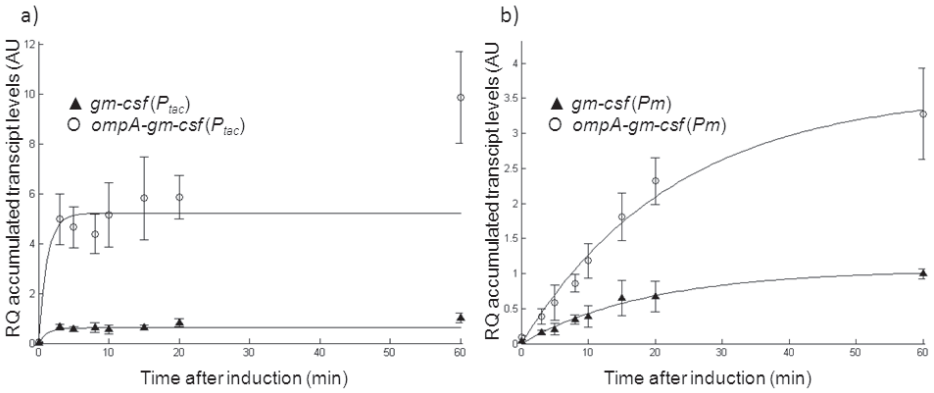


Figure 12 *Accumulation kinetics for gm-csf and ompA-gm-csf transcripts produced from the* ***a)*** P_{tac} *or* ***b)*** *Pm promoter. All transcript amounts are represented relative to the gm-csf accumulated transcript level at time 60 min, arbitrarily set to one. Solid lines represent the best fit to the data calculated according to the mathematical model described in Paper IV. Expression of the gm-csf and ompA-gm-csf genes was carried out in NEB Express Iq and DH5α strains harboring LacI/Ptac (0.5 mM IPTG) and XylS/Pm (0.5 mM m-toluate), respectively.*

The decay kinetics of the recombinant mRNAs produced from the LacI/P_{tac} and T7 systems was also analyzed by the inducer wash-out method (**Paper IV**). In the T7 system an apparent and expected slow decay rate was observed. This is almost certainly due to the continued presence of the T7 polymerase after inducer wash-out, so that transcript production to a significant extent continues throughout the time-frame of the experiments. However, indirect comparison of decay kinetics of *gm-csf* transcripts (with and without *ompA* 5´ fusion partner) produced either from P_{tac} or *Pm* (expression strains DH5α and NEB Express Iq, respectively) suggested that almost identical mRNAs can have different decay rates when produced in different expression

systems. In contrast to the XylS/*Pm* system, the transcripts produced by the LacI/P_{tac} system seemed to decay more rapidly, with about 3-fold increase in relative decay rates of both *gm-csf* and *ompA-gm-csf* (**Paper IV**). Even though the ratio of transcript decay rates (*ompA-gm-csf* versus *gm-csf*) was found similar when using XylS/*Pm* or LacI/P_{tac} (about 3 fold), the P_{tac} promoter apparently led to 2-fold higher transcript production of *ompA-gm-csf* relative to *gm-csf* (3.4±0.3 and 7.4±0.5 for *Pm* and P_{tac}, respectively). Direct comparison of *gm-csf* and *ompA-gm-csf* transcripts production from the two expression systems could determine whether different amounts are produce from *Pm* and P_{tac} promoters. One of the plausible reasons of different decay kinetics is the presence of each promoter cognate 5´-UTR and its individual influence on recombinant mRNA stability (10, 37, 88). In conclusion, these investigations of accumulation and decay showed that the use of different expression systems results in characteristic kinetic properties of recombinant transcript and likely also in slightly different mRNA stability.

3.3 Detailed analysis of the transcript and protein levels as a first step in an efficient optimization of recombinant gene expression

The comparative analysis of five bacterial expression systems (**Paper V**) showed that in many cases the final protein level is positively correlated with the level of transcripts. This is presumably also the underlying reason for the prevalent observation that protein production often positively correlates with the copy number of an expression vector (32, 50, 242). However, maximizing transcript production does not automatically lead to efficient protein production. The work in this thesis particularly illustrated this point on the inefficient protein production from several human genes (such as *ifn-α2b*$_S$ and *gm-csf*) unless a fusion partner was added (**Papers III and IV**) and on the exceptionally transcriptionally active T7 system (168), which often produces large amounts of insoluble and inactive (unfolded) protein (the case of luciferase described in **Paper V**). That the mRNA concentration is not always a good proxy of protein levels was further documented by the 5´ coding sequence variants of *ifn-α2b*$_S$ (**Paper III**). Thus, among all proteins there is a high probability that efficient expression

will face major problems due to coding sequence-specific features, and these problems can often not be solved in a satisfactory way by modifying features of the expression cassette only.

Several recent studies described genome-wide expression at the mRNA and corresponding protein levels and for the most part only limited correlations have been reported (173, 271). Specifically for *E. coli*, correlation of protein levels with mRNA abundance data obtained using SAGE and DNA microarrays, suggested that about half of protein levels are determined by transcriptional regulation (170). Although these studies are not directly related to recombinant protein production, the fact that processes downstream of transcription could explain as much variation in protein levels as mRNA levels themselves did, should put especially expression-related features of coding sequences into the forefront of heterologous gene expression optimization strategies.

Unfortunately, only a handful of studies investigated more systematically coding sequence variables and also their findings do not completely agree with each other. Among the first was a systematic analysis of effects of codon variation on heterologous expression of green fluorescent protein, eGFP (160). A library of 154 synonymous genes was expressed in *E. coli* and the corresponding eGFP protein levels showed 250-fold variation across the library (measured as fluorescence intensity). Multivariate analysis of the results suggested that the majority of the differences in eGFP levels are caused by mRNA folding near the translation start side. Although the experimental design during the library construction aimed to produce variations in GC content and codon adaptation index (CAI), no significant correlation was found between eGFP production and these variables.

Another quantitative description of the relationship between protein production levels and the coding sequence was given by Allert et al. (2010). The authors created 285 genes encoding three different proteins (120, 39 and 126 gene variants) with alteration in GC% and CAI (based on observed codon biases in natural *E. coli* genes) and determined the corresponding expression levels *in vitro* using *E. coli* extracts. The study revealed that expression is strongly dependent on the presence of high A/T

content and low secondary structure in the 5´ coding region. In a third approach, in which the gene engineering was based on systematic sampling and machine learning (279), no correlation between predicted mRNA structure near the 5´ end and heterologous gene expression was observed. In this study, two sets of 40 gene variants (encoding two commercially valuable proteins) were expressed in *E. coli* and the corresponding protein levels varied from undetectable to 30% of cellular protein. Multivariate analysis of parameters reported to affect expression identified that the amount of protein produced was strongly dependent on the codons used to encode a subset of amino acids. However, this subset did not contain the most abundant codons in highly expressed *E. coli* proteins (often described by high CAI of the gene), but rather codons that are predominantly read by tRNAs most charged during amino acid starvation.

The examples above, together with the results of this thesis, illustrate that there are indeed many ways a coding sequence can influence its expression. **Figure 13** summarizes all sequence-based determinants that have been identified to influence gene expression to date. Besides the notoriously mentioned 5´ end of a gene that influences efficiency of all basic processes of expression (transcription, mRNA turnover, translation) also the determinants of elongation speed of both transcription and translation has been shown to influence expression levels (166, 290).

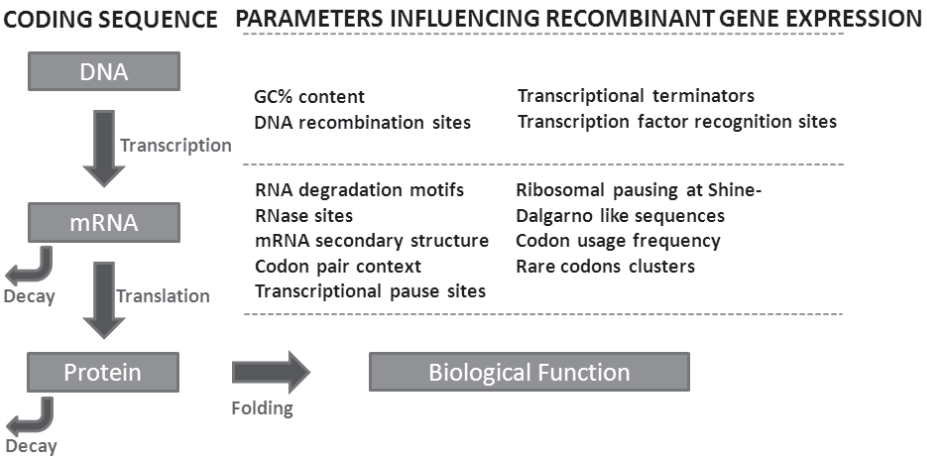


Figure 13 Summary of coding sequence features at the DNA and mRNA level that can influence recombinant protein expression*. Protein abundances are determined by a balance of regulation of both mRNA and protein production and turnover. Each protein's coding sequence significantly contributes to this regulation through the sequence features that can either enhance or reduce rates of transcription, mRNA degradation or translation. Based on (112, 210, 281).*

The gene coding sequence appears to influence its own expression at multiple levels and to varying extent. Moreover, it often appears to be one of the most critical determinants of the protein production level. At the same time, the determination of gene expression at the transcript level in **Paper I** and **III-V** suggests that efficient transcription is an essential requirement for high level heterologous gene expression. Vogel and Marcotte (2012) recently proposed a model explaining why the mRNA concentration in a cell may or may not be a good indicator of detectable protein expression. According to this model the probability of observing a given protein can be expressed as a function of mRNA abundance that has symmetrical sigmoid shape; if mRNA levels are below a certain threshold the protein level remains below detection limit, but then the likelihood of detectable protein rises sharply at higher mRNA levels (271). The results of another study reporting a large dynamic range of protein concentrations in a cell (136) further supported this idea of a stochastic switch between 'on' and 'off' states that is dependent on the level of mRNA.

The choice of expression cassette that supports production of large amounts of recombinant transcript can therefore be seen as one of the first steps on the way to efficient heterologous gene expression. Nonetheless, to meet the numerous

challenges presented by specific coding sequences can still be a difficult task. While the number of potentially useful regulator/promoter systems is expanding, more work is needed to characterize particular systems and optimize their elements with respect to different classes of heterologous proteins, especially those difficult to express such as insoluble and toxic to the host. A detailed description of an expression outcome at the level of transcript production/degradation and protein production/degradation could be then used to establish what level represents the main bottleneck and which should then be the target for optimization.

In this thesis work the *ifn-α2b*$_5$ gene represented a model of a heterologous gene which has been codon optimized (243) but still cannot not be efficiently expressed without a 5´ fusion partner (**Paper III**).There exists a vast number of different nucleic acid sequences that can be translated into the same amino acid sequence, and among such sequences few might confer high-level expression (281). However, without an effective screening system and intensive testing of potential candidates, finding the desired gene variant is almost unachievable. The rapid development of gene synthesis technologies with combination of experimental testing of expression levels and description of the observed correlations promises certain advancement. Eventual building up of a collection of credible coding sequence parameters (defined in a certain context and under specific expression conditions) may in the future serve as the basis for gene design principles as well as provide direction for gene improvement actions (280). In the end this may lead to rational design of gene sequences that are efficiently expressed.

In a more short-term perspective an often promising alternative way to improve the yield of protein is to join it to a fusion partner, as demonstrated in **Paper III**. Today´s abundance of convenient fusion tags allows for multiple options when assessing expression-related issues (60, 176). In addition, testing in parallel of several host strains with engineered properties for enhanced recombinant protein production as well as testing diverse culturing conditions has a good chance of resulting in a better expression outcome.

4 CONCLUDING REMARKS AND PERSPECTIVES

Despite intensive research efforts over many decades and by numerous groups in both academia and industry, recombinant protein production is still often dependent on trial and error approaches. The outcomes frequently depend on the unique characteristics of the targeted protein and its gene coding sequence, the expression system used, growth media and host selection. For this reason, most production processes go through a series of optimization steps (*i.e.* each protein needs individual optimization) before obtaining satisfactory levels. The work presented in this thesis focused on the 5´-terminal of recombinant genes, one of potential targets for optimization, since this genetic region is a main contributor determining the level of expression at the transcript and protein level.

The generation and characterization of optimized DNA sequences of the *Pm* promoter associated 5´-UTR and of the 5´ fusion partners, emphasized that random mutagenesis combined with competent selection system can be used for successful optimization of recombinant protein production. Moreover, the effects of introducing changes into the 5´-terminal of several recombinant genes appear to be gene specific and also to some degree dependent on the sequence context. The overall findings are suggesting that a collection of improved genetic elements could serve as an effective strategy for increasing protein production of potentially any gene of interest. *In vitro* sequence mutagenesis that utilizes mutated oligonucleotides might also be adapted for larger regions (currently up to 120 nucleotides long) comprising of several genetic elements. This could limit often unpredictable effects of context dependency between DNA sequences, as for example described in this work for the 5´-UTR and the neighboring coding sequence.

The analysis of several recombinant genes expression presented in this work was facilitated by the development of new methodology for monitoring recombinant mRNA decay, based on the wash out of transcriptional inducer combined with qRT-PCR. This non-invasive method is likely to give more reliable results than the already established approaches based on global blocking of transcription. The cases used to

establish the method demonstrated several different means of expression regulation associated with the mRNA 5´-terminal. The case of *ifn-α2b*$_5$ illustrated that despite the capacity of the *Pm* promoter to generate large amounts of transcripts, inefficient translation initiation of bacterial ribosomes on the mRNA is most likely the reason for the lack of detectable protein production. The basis for this conclusion came from studying the effects of using *pelB* and *celB*-based 5´ fusion partners for *ifn-α2b*$_5$ expression on the respective mRNA amount and stability. Although a slight increase in mRNA stability of the respective fusion genes was identified, the use of 5´ fusion partners appeared to directly modulate translation, while observed increase in transcript amounts and improved stability is presumably a secondary effect of the improved translation.

In contrast to the mRNA stabilization effect caused by inserting a 5´ fusion, specific synonymous mutations in the coding sequence of *bla* and *ifn-α2b*$_5$ seemed to act by stimulating the rate of transcription. Transcriptional regulation of gene expression is believed to be mainly exerted through the DNA sequence of a promoter, but as it appear, also the sequence of the promoter associated 5´-UTR and within a gene itself can influence, presumably both positively and negatively, transcription and consequently protein production.

From the results of this work, it is evident that the 5´ end of recombinant genes affects expression in a complex way and many different techniques should be used for monitoring all levels of the process. A detailed analysis of the functionality of all steps involved in the determination of the level of recombinant protein production would allow better formulated approaches for expression optimization, adapted to each specific case. Eventually, once the number of such detailed analyses have been carried out for a sufficient number of cases, methods for optimization of recombinant gene expression based on rational design strategies will probably become possible.

REFERENCES

1. **Adhin MR, van Duin J.** 1990. Scanning model for translational reinitiation in eubacteria. Journal of Molecular Biology **213:**811-818.
2. **Alper H, Stephanopoulos G.** 2007. Global transcription machinery engineering: A new approach for improving cellular phenotype. Metabolic Engineering **9:**258-267.
3. **Andersson SG, Kurland CG.** 1990. Codon preferences in free-living microorganisms. Microbiological Reviews **54:**198-210.
4. **Andrè A, Puca A, Sansone F, Brandi A, Antico G, Calogero RA.** 2000. Reinitiation of protein synthesis in *Escherichia* coli can be induced by mRNA cis-elements unrelated to canonical translation initiation signals. FEBS Letters **468:**73-78.
5. **Angkawidjaja C, Kuwahara K, Omori K, Koga Y, Takano K, Kanaya S.** 2006. Extracellular secretion of *Escherichia coli* alkaline phosphatase with a C-terminal tag by type I secretion system: purification and biochemical characterization. Protein Engineering Design and Selection **19:**337-343.
6. **Apirion D.** 1973. Degradation of RNA in *Escherichia coli*. Molecular and General Genetics MGG **122:**313-322.
7. **Apirion D.** 1978. Isolation, genetic mapping and some characterization of a mutation in *Escherichia coli* that affects the processing of ribonucleic acid. Genetics **90:**659-671.
8. **Apweiler R, Hermjakob H, Sharon N.** 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochimica et Biophysica Acta (BBA) - General Subjects **1473:**4-8.
9. **Arber W, Linn S.** 1969. DNA modification and restriction. Annual Review of Biochemistry **38:**467-500.
10. **Arnold TE, Yu J, Belasco JG.** 1998. mRNA stabilization by the *ompA* 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation. RNA **4:**319-330.
11. **Arraiano CM, Andrade JM, Domingues S, Guinote IB, Malecki M, Matos RG, Moreira RN, Pobre V, Reis FP, Saramago M, Silva IJ, Viegas SC.** 2010. The critical role of RNA processing and degradation in the control of gene expression. FEMS Microbiology Reviews **34:**883-923.
12. **Artsimovitch I, Landick R.** 1998. Interaction of a nascent RNA structure with RNA polymerase is required for hairpin-dependent transcriptional pausing but not for transcript release. Genes & Development **12:**3110-3122.
13. **Artsimovitch I, Landick R.** 2000. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proceedings of the National Academy of Sciences **97:**7090-7095.
14. **Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H.** 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol **2**.
15. **Babitzke P, Kushner SR.** 1991. The Ams (altered mRNA stability) protein and ribonuclease E are encoded by the same structural gene of *Escherichia coli*. Proceedings of the National Academy of Sciences **88:**1-5.
16. **Bai L, Santangelo TJ, Wang MD.** 2006. Single-molecule analysis of RNA polymerase transcription, p. 343-360, Annual Review of Biophysics and Biomolecular Structure, vol. 35. Annual Reviews, Palo Alto.
17. **Baker KE, Mackie GA.** 2003. Ectopic RNase E sites promote bypass of 5'-end-dependent mRNA decay in *Escherichia coli*. Molecular Microbiology **47:**75-88.

18. **Balbás P.** 2001. Understanding the art of producing protein and nonprotein molecules in *Escherichia coli*. Molecular Biotechnology **19:**251-267.
19. **Baneyx F.** 1999. Recombinant protein expression in *Escherichia coli*. Current Opinion in Biotechnology **10:**411-421.
20. **Baneyx F, Mujacic M.** 2004. Recombinant protein folding and misfolding in *Escherichia coli*. Nature Biotechnology **22:**1399-1408.
21. **Bar-Nahum G, Nudler E.** 2001. Isolation and characterization of σ70-Retaining transcription elongation complexes from *Escherichia coli*. Cell **106:**443-451.
22. **Barnard A, Wolfe A, Busby S.** 2004. Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. Current Opinion in Microbiology **7:**102-108.
23. **Barnard GC, Henderson GE, Srinivasan S, Gerngross TU.** 2004. High level recombinant protein expression in *Ralstonia eutropha* using T7 RNA polymerase based amplification. Protein Expression and Purification **38:**264-271.
24. **Barne KA, Bown JA, Busby SJW, Minchin SD.** 1997. Region 2.5 of the *Escherichia coli* RNA polymerase σ70 subunit is responsible for the recognition of the 'extended -10' motif at promoters. EMBO J **16:**4034-4040.
25. **Barrick D, Villanueba K, Childs J, Kalil R, Schneider TD, Lawrence CE, Gold L, Stormo GD.** 1994. Quantitative analysis of ribosome binding sites in *E.coli*. Nucleic Acids Research **22:**1287-1295.
26. **Beck LL, Smith TG, Hoover TR.** 2007. Look, no hands! Unconventional transcriptional activators in bacteria. Trends in microbiology **15:**530-537.
27. **Belasco JG.** 2010. All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. Nature Reviews Molecular Cell Biology **11:**467-478.
28. **Belasco JG, Biggins CF.** 1988. Mechanisms of mRNA decay in bacteria: a perspective. Gene **72:**15-23.
29. **Benita Y, Wise MJ, Lok MC, Humphery-Smith I, Oosting RS.** 2006. Analysis of high throughput protein expression in *Escherichia coli*. Molecular and Cellular Proteomics **5:**1567-1580.
30. **Berg L, Lale R, Bakke I, Burroughs N, Valla S.** 2009. The expression of recombinant genes in *Escherichia coli* can be strongly stimulated at the transcript production level by mutating the DNA-region corresponding to the 5'-untranslated part of mRNA. Microbial Biotechnology **2:**379-389.
31. **Blatny JM, Brautaset T, Winther-Larsen HC, Haugan K, Valla S.** 1997. Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. Applied and Environmental Microbiology **63:**370-379.
32. **Blatny JM, Brautaset T, Winther-Larsen HC, Karunakaran P, Valla S.** 1997. Improved broad-host-range RK2 vectors useful for high and low regulated gene expression levels in gram-negative bacteria. Plasmid **38:**35-51.
33. **Boni IV, Isaeva DM, Musychenko ML, Tzareva NV.** 1991. Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. Nucleic Acids Research **19:**155-162.
34. **Borukhov S, Nudler E.** 2003. RNA polymerase holoenzyme: structure, function and biological implications. Current Opinion in Microbiology **6:**93-100.
35. **Braun F, Le Derout J, Regnier P.** 1998. Ribosomes inhibit an RNase E cleavage which induces the decay of the *rpsO* mRNA of *Escherichia coli*. EMBO J **17:**4790-4797.
36. **Brautaset T, Lale R, Valla S.** 2009. Positively regulated bacterial expression systems. Microbial Biotechnology **2:**15-30.

37. **Bricker AL, Belasco JG.** 1999. Importance of a 5' Stem-Loop for Longevity of *papA* mRNA in *Escherichia coli*. Journal of Bacteriology **181:**3587-3590.
38. **Brieba LG, Sousa R.** 2001. T7 promoter release mediated by DNA scrunching. EMBO J **20:**6826-6835.
39. **Brock JE, Paz RL, Cottle P, Janssen GR.** 2007. Naturally occurring adenines within mRNA Coding Sequences affect ribosome binding and expression in *Escherichia coli*. Journal of Bacteriology **189:**501-510.
40. **Brodolin K, Zenkin N, Mustaev A, Mamaeva D, Heumann H.** 2004. The sigma 70 subunit of RNA polymerase induces l*acUV5* promoter-proximal pausing of transcription. Nature Structural & Molecular Biology **11:**551-557.
41. **Broedel SEJ, Papciak SM, Jones WR.** 2001. The selection of optimum media formulations for improved expression of recombinant proteins in *E. coli*. Technical Bulletin - Athena Enzyme SystemsTM Group **2**.
42. **Browning DF, Busby SJW.** 2004. The regulation of bacterial transcription initiation. Nature Reviews Microbiology **2:**57-65.
43. **Buchan JR, Stansfield I.** 2007. Halting a cellular production line: responses to ribosomal pausing during translation. Biology of the Cell **99:**475-487.
44. **Bulmer M.** 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics **129:**897-907.
45. **Callaghan AJ, Marcaida MJ, Stead JA, McDowall KJ, Scott WG, Luisi BF.** 2005. Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. Nature **437:**1187-1191.
46. **Calos MP.** 1978. DNA sequence for a low-level promoter of the *lac* repressor gene and an 'up' promoter mutation. Nature **274:**762-765.
47. **Carpousis AJ.** 2007. The RNA degradosome of *Escherichia coli*: An mRNA-degrading machine assembled on RNase E. Annual Review of Microbiology **61:**71-87.
48. **Carpousis AJ, Van Houwe G, Ehretsmann C, Krisch HM.** 1994. Copurification of *E. coli* RNAase E and PNPase: Evidence for a specific association between two enzymes important in RNA processing and degradation. Cell **76:**889-900.
49. **Carr PA, Church GM.** 2009. Genome engineering. Nature Biotechnology **27:**1151-1162.
50. **Carrier T, Jones KL, Keasling JD.** 1998. mRNA stability and plasmid copy number effects on gene expression from an inducible promoter system. Biotechnology and Bioengineering **59:**666-672.
51. **Carrier TA, Keasling JD.** 1997. Controlling messenger RNA stability in bacteria: Strategies for engineering gene expression. Biotechnology Progress **13:**699-708.
52. **Carrier TA, Keasling JD.** 1997. Engineering mRNA stability in *E. coli* by the addition of synthetic hairpins using a 5′ cassette system. Biotechnology and Bioengineering **55:**577-580.
53. **Carrier TA, Keasling JD.** 1999. Library of Synthetic 5′ Secondary Structures To Manipulate mRNA Stability in Escherichia coli. Biotechnology Progress **15:**58-64.
54. **Carrier TA, Keasling JD.** 1997. Mechanistic modeling of prokaryotic mRNA decay. Journal of Theoretical Biology **189:**195-209.
55. **Carrió MM, Villaverde A.** 2001. Protein aggregation as bacterial inclusion bodies is reversible. FEBS Letters **489:**29-33.
56. **Caserta E, Tomšic J, Spurio R, La Teana A, Pon CL, Gualerzi CO.** 2006. Translation initiation factor IF2 interacts with the 30 S ribosomal subunit via two separate binding sites. Journal of Molecular Biology **362:**787-799.

57. **Cèbe R, Geiser M.** 2006. Rapid and easy thermodynamic optimization of the 5'-end of mRNA dramatically increases the level of wild type protein expression in *Escherichia coli*. Protein Expression and Purification **45:**374-380.
58. **Celesnik H, Deana A, Belasco JG.** 2007. Initiation of RNA decay in *Escherichia coli* by 5' pyrophosphate removal. Molecular Cell **27:**79-90.
59. **Chang JT, Green CB, Wolf RE.** 1995. Inhibition of translation initiation on *Escherichia coli gnd* mRNA by formation of a long-range secondary structure involving the ribosome binding site and the internal complementary sequence. Journal of Bacteriology **177:**6560-6567.
60. **Chatterjee DK, Esposito D.** 2006. Enhanced soluble protein expression using two new fusion tags. Protein Expression and Purification **46:**122-129.
61. **Chen H, Bjerknes M, Kumar R, Jay E.** 1994. Determination of the optimal aligned spacing between the Shine–Dalgarno sequence and the translation initiation codon of *Escherichia coli* m RNAs. Nucleic Acids Research **22:**4953-4957.
62. **Chen R.** 2011. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. Biotechnology Advances.
63. **Cheng Z-F, Deutscher MP.** 2005. An important role for RNase R in mRNA decay. Molecular Cell **17:**313-318.
64. **Choi JH, Lee SY.** 2004. Secretory and extracellular production of recombinant proteins using *Escherichia coli*. Applied Microbiology and Biotechnology **64:**625-635.
65. **Chou C.** 2007. Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*. Applied Microbiology and Biotechnology **76:**521-532.
66. **Chung C, You J, Kim K, Moon Y, Kim H, Ahn J.** 2009. Export of recombinant proteins in *Escherichia coli* using ABC transporter with an attached lipase ABC transporter recognition domain (LARD). Microbial Cell Factories **8:**11.
67. **Clarke TFIV, Clark PL.** 2008. Rare codons cluster. PLoS ONE **3:**e3412.
68. **Cohen SN, Chang ACY, Boyer HW, Helling RB.** 1973. Construction of biologically functional bacterial plasmids *in vitro*. Proceedings of the National Academy of Sciences **70:**3240-3244.
69. **De Castro C, Parrilli M, Holst O, Molinaro A.** 2010. Chapter Five - Microbe-associated molecular patterns in innate immunity: Extraction and chemical analysis of Gram-negative bacterial lipopolysaccharides, p. 89-115. *In* F. Minoru (ed.), Methods in Enzymology, vol. Volume 480. Academic Press.
70. **de Marco A.** 2009. Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. Microbial Cell Factories **8:**26.
71. **de Smit MH, van Duin J.** 1990. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proceedings of the National Academy of Sciences **87:**7668-7672.
72. **Deana A, Belasco JG.** 2005. Lost in translation: the influence of ribosomes on bacterial mRNA decay. Genes & Development **19:**2526-2533.
73. **Deana A, Celesnik H, Belasco JG.** 2008. The bacterial enzyme RppH triggers messenger RNA degradation by 5´ pyrophosphate removal. Nature **451:**355-358.
74. **Demain AL, Vaishnav P.** 2009. Production of recombinant proteins by microbes and higher organisms. Biotechnology Advances **27:**297-306.
75. **Deutscher MP.** 2006. Degradation of RNA in bacteria: comparison of mRNA and stable RNA. Nucleic Acids Research **34:**659-666.
76. **Deutscher MP.** 2003. Degradation of stable RNA in bacteria. J. Biol. Chem. **278:**45041-45044.

77. **Doig AJ, Williams DH.** 1992. Binding energy of an amide-amide hydrogen bond in aqueous and nonpolar solvents. Journal of the American Chemical Society **114:**338-343.
78. **Domínguez-Cuevas P, Marín P, Ramos JL, Marqués S.** 2005. RNA polymerase holoenzymes can share a single transcription start site for the *Pm* promoter. J. Biol. Chem. **280:**41315-41323.
79. **Dong H, Nilsson L, Kurland CG.** 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. Journal of Molecular Biology **260:**649-663.
80. **Dong H, Nilsson L, Kurland CG.** 1995. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. Journal of Bacteriology **177:**1497-1504.
81. **Dong W-R, Xiang L-X, Shao J-Z.** 2010. Novel antibiotic-free plasmid selection system based on complementation of host auxotrophy in the NAD *de novo* synthesis pathway. Applied and Environmental Microbiology **76:**2295-2303.
82. **Donovan RS, Robinson CW, Glick BR.** 1996. Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of the *lac* promoter. Journal of Industrial Microbiology & Biotechnology **16:**145-154.
83. **Dortay H, Akula UM, Westphal C, Sittig M, Mueller-Roeber B.** 2011. High-Throughput Protein Expression Using a Combination of Ligation-Independent Cloning (LIC) and Infrared Fluorescent Protein (IFP) Detection. PLoS ONE **6:**e18900.
84. **Drew D, Slotboom D-J, Friso G, Reda T, Genevaux P, Rapp M, Meindl-Beinker NM, Lambert W, Lerch M, Daley DO, Van Wijk K-J, Hirst J, Kunji E, De Gier J-W.** 2005. A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. Protein Science **14:**2011-2017.
85. **Dreyfus M.** 1988. What constitutes the signal for the initiation of protein synthesis on *Escherichia coli* mRNAs? Journal of Molecular Biology **204:**79-94.
86. **Elf J, Ehrenberg M.** 2005. Near-critical behavior of aminoacyl-tRNA pools in *E. coli* at rate-limiting supply of amino acids. Biophysical Journal **88:**132-146.
87. **Elf J, Nilsson D, Tenson T, Ehrenberg M.** 2003. Selective charging of tRNA isoacceptors explains patterns of codon usage. Science **300:**1718-1722.
88. **Emory SA, Bouvet P, Belasco JG.** 1992. A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. Genes & Development **6:**135-148.
89. **Estrem ST, Ross W, Gaal T, Chen ZWS, Niu W, Ebright RH, Gourse RL.** 1999. Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase α subunit. Genes & Development **13:**2134-2147.
90. **Etchegaray J-P, Inouye M.** 1999. Translational enhancement by an element downstream of the initiation codon in *Escherichia coli*. J. Biol. Chem. **274:**10079-10085.
91. **Eyre-Walker A, Bulmer M.** 1993. Reduced synonymous substitution rate at the start of enterobacterial genes. Nucleic Acids Research **21:**4599-4603.
92. **Fisher AC, Haitjema CH, Guarino C, Çelik E, Endicott CE, Reading CA, Merritt JH, Ptak AC, Zhang S, DeLisa MP.** 2011. Production of secretory and extracellular N-linked glycoproteins in *Escherichia coli*. Applied and Environmental Microbiology **77:**871-881.
93. **Frelet-Barrand A, Boutigny S, Moyet L, Deniaud A, Seigneurin-Berny D, Salvi D, Bernaudat F, Richaud P, Pebay-Peyroula E, Joyard J, Rolland N.** 2010. *Lactococcus lactis* an alternative system for functional expression of peripheral and intrinsic *Arabidopsis* membrane proteins. PLoS ONE **5:**e8746.

94. **Friehs K.** 2004. Plasmid copy number and plasmid stability, p. 47-82. *In* T. H. Scheper (ed.), New Trends and Developments in Biochemical Engineering, vol. 86. Springer Berlin, Heildelberg.
95. **Garcìa-Fruitòs E, Gonzàlez-Montalbàn N, Martìnez-Alonso M, Rinas U, Villaverde A.** 2009a. Systems-level analysis of protein quality in inclusion body-forming *Escherichia coli* cells, p. 295-326. *In* S. Y. Lee (ed.), Systems Biology and Biotechnology of *Escherichia coli*. Springer Netherlands.
96. **Garcìa-Fruitòs E, Rodríguez-Carmona E, Díez-Gil C, Ferraz RM, Vázquez E, Corchero JL, Cano-Sarabia M, Ratera I, Ventosa N, Veciana J, Villaverde A.** 2009b. Surface cell growth engineering assisted by a novel bacterial nanomaterial. Advanced Materials **21:**4249-4253.
97. **García-Fruitós E, Vázquez E, Díez-Gil C, Corchero JL, Seras-Franzoso J, Ratera I, Veciana J, Villaverde A.** 2012. Bacterial inclusion bodies: making gold from waste. Trends in Biotechnology **30:**65-70.
98. **Goeddel DV, Kleid DG, Bolivar F, Heyneker HL, Yansura DG, Crea R, Hirose T, Kraszewski A, Itakura K, Riggs AD.** 1979. Expression in *Escherichia coli* of chemically synthesized genes for human insulin. Proceedings of the National Academy of Sciences **76:**106-110.
99. **Goh S, Good L.** 2008. Plasmid selection in *Escherichia coli* using an endogenous essential gene marker. BMC Biotechnology **8:**61.
100. **Goldman SR, Ebright RH, Nickels BE.** 2009. Direct detection of abortive RNA transcripts *in vivo*. Science **324:**927-928.
101. **Gong F, Yanofsky C.** 2003. A transcriptional pause synchronizes translation with transcription in the tryptophanase operon leader region. Journal of Bacteriology **185:**6472–6476.
102. **Gonzalez-Montalban N, Garcia-Fruitos E, Villaverde A.** 2007. Recombinant protein solubility - does more mean better? Nature Biotechnology **25:**718-720.
103. **Gonzalez de Valdivia EI, Isaksson LA.** 2004. A codon window in mRNA downstream of the initiation codon where NGG codons give strongly reduced gene expression in *Escherichia coli*. Nucleic Acids Research **32:**5198-5205.
104. **Gordon E, Horsefield R, Swarts HGP, de Pont JJHHM, Neutze R, Snijder A.** 2008. Effective high-throughput overproduction of membrane proteins in *Escherichia coli*. Protein Expression and Purification **62:**1-8.
105. **Gowrishankar J, Harinarayanan R.** 2004. Why is transcription coupled to translation in bacteria? Molecular Microbiology **54:**598-603.
106. **Graumann K, Premstaller A.** 2006. Manufacturing of recombinant therapeutic proteins in microbial systems. Biotechnology Journal **1:**164-186.
107. **Greive SJ, von Hippel PH.** 2005. Thinking quantitatively about transcriptional regulation. Nature Reviews Molecular Cell Biology **6:**221-232.
108. **Grigoriadou C, Marzi S, Kirillov S, Gualerzi CO, Cooperman BS.** 2007. A Quantitative kinetic scheme for 70S translation initiation complex formation. Journal of Molecular Biology **373:**562-572.
109. **Gruber TM, Gross CA.** 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annual Review of Microbiology **57:**441-466.
110. **Gu W, Zhou T, Wilke CO.** 2010. A universal trend of reduced mRNA stability near the translation-initiation site in Prokaryotes and Eukaryotes. PLoS Computational Biology **6:**e1000664.

111. **Gustafsson C, Govindarajan S, Minshull J.** 2004. Codon bias and heterologous protein expression. Trends in Biotechnology **22:**346-353.
112. **Gustafsson C, Minshull J, Govindarajan S, Ness J, Villalobos A, Welch M.** 2012. Engineering genes for predictable protein expression. Protein Expression and Purification **83:**37-46.
113. **Guynn RW, Veech RL.** 1973. The equilibrium constants of the adenosine triphosphate hydrolysis and the adenosine triphosphate-citrate lyase reactions. J. Biol. Chem. **248:**6966-6972.
114. **Hacker J, Blum-Oehler G.** 2007. In appreciation of Theodor Escherich. Nature Reviews Microbiology **5:**902-902.
115. **Hansted JG, Pietikäinen L, Hög F, Sperling-Petersen HU, Mortensen KK.** 2011. Expressivity tag: A novel tool for increased expression in *Escherichia coli*. Journal of Biotechnology **155:**275-283.
116. **Hatoum A, Roberts J.** 2008. Prevalence of RNA polymerase stalling at *Escherichia coli* promoters after open complex formation. Molecular Microbiology **68:**17-28.
117. **Haugel-Nielsen J, Hajnsdorf E, Regnier P.** 1996. The *rpsO* mRNA of *Escherichia coli* is polyadenylated at multiple sites resulting from endonucleolytic processing and exonucleolytic degradation. EMBO **15:**3144-3152.
118. **Haugen SP, Ross W, Gourse RL.** 2008. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nature Reviews Microbiology **6:**507-519.
119. **Helmann JD, deHaseth PL.** 1999. Protein−nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners. Biochemistry **38:**5959-5967.
120. **Henkin TM, Yanofsky C.** 2002. Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. BioEssays **24:**700-707.
121. **Hoffmann F, Rinas U.** 2001. On-line estimation of the metabolic burden resulting from the synthesis of plasmid-encoded and heat-shock proteins by monitoring respiratory energy generation. Biotechnology and Bioengineering **76:**333-340.
122. **Hoffmann F, Rinas U.** 2004. Stress onduced by recombinant protein production in *Escherichia coli* p. 73-92, Physiological Stress Responses in Bioprocesses, vol. 89. Springer Berlin / Heidelberg.
123. **Houseley J, Tollervey D.** 2009. The Many Pathways of RNA Degradation. Cell **136:**763-776.
124. **Hsu LM.** 2002. Promoter clearance and escape in prokaryotes. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression **1577:**191-207.
125. **Huang C, Jr., Lin H, Yang X.** 2012. Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. Journal of Industrial Microbiology & Biotechnology **39:**383-399.
126. **Huang K-x, Badger M, Haney K, Evans SL.** 2007. Large scale production of *Bacillus thuringiensis* PS149B1 insecticidal proteins Cry34Ab1 and Cry35Ab1 from *Pseudomonas fluorescens*. Protein Expression and Purification **53:**325-330.
127. **Hug I, Zheng B, Reiz B, Whittal RM, Fentabil MA, Klassen JS, Feldman MF.** 2011. Exploiting bacterial glycosylation machineries for the synthesis of a Lewis antigen-containing glycoprotein. J. Biol. Chem. **286:**37887-37894.
128. **Hui A, de Boer HA.** 1987. Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. Proceedings of the National Academy of Sciences **84:**4762-4766.

129. **Hüttenhofer A, Noller H.** 1994. Footprinting mRNA-ribosome complexes with chemical probes. EMBO J. **13:**3892-3901.
130. **Hägg P, de Pohl JW, Abdulkarim F, Isaksson LA.** 2004. A host/plasmid system that is not dependent on antibiotics and antibiotic resistance genes for stable plasmid maintenance in *Escherichia coli*. Journal of Biotechnology **111:**17-30.
131. **Ikemura T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. Journal of Molecular Biology **151:**389-409.
132. **Iost I, Dreyfus M.** 1995. The stability of *Escherichia coli lacZ* mRNA depends upon the simultaneity of its synthesis and translation. EMBO Journal **14**.
133. **Iost I, Dreyfus M.** 1995. The stability of *Escherichia coli lacZ* mRNA depends upon the simultaneity of its synthesis and translation. EMBO J **14**.
134. **Iost I, Guillerez J, Dreyfus M.** 1992. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes *in vivo*. Journal of Bacteriology **174:**619-622.
135. **Ishihama A.** 2000. Functional modulation of *Escherichia coli* RNA polymerase. Annual Review of Microbiology **54:**499-518.
136. **Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl FU, Kerner M, Frishman D.** 2008. Protein abundance profiling of the *Escherichia coli* cytosol. BMC Genomics **9:**102.
137. **Itakura K, Hirose T, Crea R, Riggs A, Heyneker H, Bolivar F, Boyer H.** 1977. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science **198:**1056-1063.
138. **Jackson DA, Symons RH, Berg P.** 1972. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: Circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. Proceedings of the National Academy of Sciences **69:**2904-2909.
139. **Jacob WF, Santer M, Dahlberg AE.** 1987. A single base change in the Shine-Dalgarno region of 16S rRNA of *Escherichia coli* affects translation of many proteins. Proceedings of the National Academy of Sciences **84:**4757-4761.
140. **Jana S, Deb JK.** 2005. Strategies for efficient production of heterologous proteins in *Escherichia coli*. Applied Microbiology and Biotechnology **67:**289-298.
141. **Jiang X, Belasco JG.** 2004. Catalytic activation of multimeric RNase E and RNase G by 5'-monophosphorylated RNA. Proceedings of the National Academy of Sciences **101:**9211-9216.
142. **Jong WSP, Saurí A, Luirink J.** 2010. Extracellular production of recombinant proteins using bacterial autotransporters. Current Opinion in Biotechnology **21:**646-652.
143. **Kaberdin V, Singh D, Lin-Chao S.** 2011. Composition and conservation of the mRNA-degrading machinery in bacteria. Journal of Biomedical Science **18:**23.
144. **Kaberdin VR, Bläsi U.** 2006. Translation initiation and the fate of bacterial mRNAs. FEMS Microbiology Reviews **30:**967-979.
145. **Kachroo AH, Jayaram M, Rowley PA.** 2009. Metabolic engineering without plasmids. Nature Biotechnology **27:**729-731.
146. **Kane JF.** 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coil*. Current Opinion in Biotechnology**:**494-500.
147. **Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebright RH.** 2006. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science **314:**1144-1147.

148. **Keasling JD.** 1999. Gene-expression tools for the metabolic engineering of bacteria. Trends in Biotechnology **17:**452-460.
149. **Keasling JD.** 2010. Manufacturing molecules through metabolic engineering. Science **330:**1355-1358.
150. **Keasling JD.** 2008. Synthetic Biology for Synthetic Chemistry. ACS Chemical Biology **3:**64-76.
151. **Khlebnikov A, Datsenko KA, Skaug T, Wanner BL, Keasling JD.** 2001. Homogeneous expression of the *PBAD* promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. Microbiology **147:**3241-3247.
152. **Klein-Marcuschamer D, Santos CNS, Yu H, Stephanopoulos G.** 2009. Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes. Applied and Environmental Microbiology **75:**2705-2711.
153. **Klumpp S.** 2011. Growth-rate dependence reveals design principles of plasmid copy number control. PLoS ONE **6:**e20403.
154. **Klumpp S, Zhang Z, Hwa T.** 2009. Growth rate-dependent global effects on gene expression in bacteria. Cell **139:**1366-1375.
155. **Koehn J, Hunt I.** 2009. High-Throughput Protein Production (HTPP): a review of enabling technologies to expedite protein production. Methods in Molecular Medicine™ **498:**1-18.
156. **Komarova AV, Tchufistova LS, Dreyfus M, Boni IV.** 2005. AU-Rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. Journal of Bacteriology **187:**1344-1349.
157. **Komarova AV, Tchufistova LS, Supina EV, Boni IV.** 2002. Protein S1 counteracts the inhibitory effect of the extended Shine-Dalgarno sequence on translation. RNA **8:**1137-1147.
158. **Kozak M.** 2005. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene **361:**13-37.
159. **Krummel B, Chamberlin MJ.** 1989. RNA chain initiation by *Escherichia coli* RNA polymerase. Structural transitions of the enzyme in early ternary complexes. Biochemistry **28:**7829-7842.
160. **Kudla G, Murray AW, Tollervey D, Plotkin JB.** 2009. Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. Science **324:**255-258.
161. **Kunji ERS, Slotboom D-J, Poolman B.** 2003. Lactococcus lactis as host for overproduction of functional membrane proteins. Biochimica et Biophysica Acta (BBA) - Biomembranes **1610:**97-108.
162. **Kurland C, Gallant J.** 1996. Errors of heterologous protein expression. Current Opinion in Biotechnology **7:**489-493.
163. **Larson MH, Landick R, Block SM.** 2011. Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes. Molecular Cell **41:**249-262.
164. **Laursen BS, Sørensen HP, Mortensen KK, Sperling-Petersen HU.** 2005. Initiation of protein synthesis in bacteria. Microbiology and Molecular Biology Reviews **69:**101-123.
165. **Lee JY, Sung BH, Yu BJ, Lee JH, Lee SH, Kim MS, Koob MD, Kim SC.** 2008. Phenotypic engineering by reprogramming gene transcription using novel artificial transcription factors in *Escherichia coli*. Nucleic Acids Research **36:**e102.
166. **Li G-W, Oh E, Weissman JS.** 2012. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature **484:**538-541.

167. **Lim HK, Jung KH, Park DH, Chung SI.** 2000. Production characteristics of interferon-α using an L-arabinose promoter system in a high-cell-density culture. Applied Microbiology and Biotechnology **53:**201-208.
168. **Lopez PJ, lost I, Dreyfus M.** 1994. 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 Research **22:**1186-1193.
169. **Lopez PJ, Marchand I, Yarchuk O, Dreyfus M.** 1998. Translation inhibitors stabilize Escherichia coli mRNAs independently of ribosome protection. Proceedings of the National Academy of Sciences **95:**6067-6072.
170. **Lu P, Vogel C, Wang R, Yao X, Marcotte EM.** 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nature Biotechnology **25:**117-124.
171. **Mackie GA.** 1998. Ribonuclease E is a 5'-end-dependent endonuclease. Nature **395:**720-724.
172. **Mackie GA.** 2000. Stabilization of circular *rpsT* mRNA demonstrates the 5′-End dependence of RNase E action *in vivo*. J. Biol. Chem. **275:**25069-25072.
173. **Maier T, Güell M, Serrano L.** 2009. Correlation of mRNA and protein in complex biological samples. FEBS Letters **583:**3966-3973.
174. **Makino T, Skretas G, Georgiou G.** 2011. Strain engineering for improved expression of recombinant proteins in bacteria. Microbial Cell Factories **10:**32.
175. **Makino T, Skretas G, Kang T-H, Georgiou G.** 2011. Comprehensive engineering of *Escherichia coli* for enhanced expression of IgG antibodies. Metabolic Engineering **13:**241-251.
176. **Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR.** 2006. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. Protein Science **15:**182-189.
177. **Martin RB.** 1998. Free energies and equilibria of peptide bond hydrolysis and formation. Biopolymers **45:**351-353.
178. **Marzi S, Myasnikov AG, Serganov A, Ehresmann C, Romby P, Yusupov M, Klaholz BP.** 2007. Structured mRNAs regulate rranslation initiation by binding to the platform of the ribosome. Cell **130:**1019-1031.
179. **Massey-Gendel E, Zhao A, Boulting G, Kim H-Y, Balamotis MA, Seligman LM, Nakamoto RK, Bowie JU.** 2009. Genetic selection system for improving recombinant membrane protein expression in *E. coli*. Protein Science **18:**372-383.
180. **Mata J, Marguerat S, Bähler J.** 2005. Post-transcriptional control of gene expression: a genome-wide perspective. Trends in Biochemical Sciences **30:**506-514.
181. **Matos RG, Pobre V, Reis FP, Malecki M, Andrade JM, Arraiano CM.** 2011. Structure and degradation mechanisms of 3′ to 5′ exoribonucleases, p. 193-222. *In* A. W. Nicholson (ed.), Ribonucleases. Springer Berlin Heidelberg.
182. **Matsunaga J, Simons EL, Simons RW.** 1997. *Escherichia coli* RNase III (*rnc*) autoregulation occurs independently of *rnc* gene translation. Molecular Microbiology **26:**1125-1135.
183. **McDowall KJ, Lin-Chao S, Cohen SN.** 1994. A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. J. Biol. Chem. **269:**10790-10796.
184. **Mekler V, Kortkhonjia E, Mukhopadhyay J, Knight J, Revyakin A, Kapanidis AN, Niu W, Ebright YW, Levy R, Ebright RH.** 2002. Structural organization of bacterial RNA

polymerase holoenzyme and the RNA polymerase-promoter open complex. Cell **108:**599-614.

185. **Menart V, Jevševar S, Vilar M, Trobiš A, Pavko A.** 2003. Constitutive versus thermoinducible expression of heterologous proteins in *Escherichia coli* based on strong pR,pL promoters from phage lambda. Biotechnology and Bioengineering **83:**181-190.
186. **Mergulhão FJM, Summers DK, Monteiro GA.** 2005. Recombinant protein secretion in *Escherichia coli*. Biotechnology Advances **23:**177-202.
187. **Meyer H-P, Schmidhalter DR.** 2012. Microbial expression systems and manufacturing from a market and economic perspective. *In* E. C. Agbo (ed.), vol. Innovations in Biotechnology. InTech.
188. **Milon P, Carotti M, Konevega AL, Wintermeyer W, Rodnina MV, Gualerzi CO.** 2010. The ribosome-bound initiation factor 2 recruits initiator tRNA to the 30S initiation complex. EMBO Rep **11:**312-316.
189. **Milón P, Maracci C, Filonava L, Gualerzi CO, Rodnina MV.** 2012. Real-time assembly landscape of bacterial 30S translation initiation complex. Nature Structural & Molecular Biology **advance online publication**.
190. **Miroux B, Walker JE.** 1996. Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of somemembrane proteins and globular proteins at high levels. Journal of Molecular Biology **260:**289-298.
191. **Moll I, Huber M, Grill S, Sairafi P, Mueller F, Brimacombe R, Londei P, Bläsi U.** 2001. Evidence against an Interaction between the mRNA Downstream Box and 16S rRNA in Translation Initiation. Journal of Bacteriology **183:**3499-3505.
192. **Moore PB, Steitz TA.** 2011. The roles of RNA in the synthesis of protein. Cold Spring Harbor Perspectives in Biology **3**.
193. **Mukhopadhyay J, Kapanidis AN, Mekler V, Kortkhonjia E, Ebright YW, Ebright RH.** 2001. Translocation of σ70 with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. Cell **106:**453-463.
194. **Najafabadi HS, Goodarzi H, Salavati R.** 2009. Universal function-specificity of codon usage. Nucleic Acids Research **37:**7014-7023.
195. **Nakagawa S, Niimura Y, Miura K-i, Gojobori T.** 2010. Dynamic evolution of translation initiation mechanisms in prokaryotes. Proceedings of the National Academy of Sciences **107:**6382-6387.
196. **Nakamoto T.** 2006. A unified view of the initiation of protein synthesis. Biochemical and Biophysical Research Communications **341:**675-678.
197. **Nakazato H, Venkatesan S, Edmonds M.** 1975. Polyadenylic acid sequences in E. coli messenger RNA. Nature **256:**144-146.
198. **Ni Y, Chen R.** 2009. Extracellular recombinant protein production from *Escherichia coli*. Biotechnology Letters **31:**1661-1670.
199. **Nickels BE, Mukhopadhyay J, Garrity SJ, Ebright RH, Hochschild A.** 2004. The sigma 70 subunit of RNA polymerase mediates a promoter-proximal pause at the *lac* promoter. Nature Structural & Molecular Biology **11:**544-550.
200. **Nilsson G, Belasco JG, Cohen SN, von Gabain A.** 1987. Effect of premature termination of translation on mRNA stability depends on the site of ribosome release. Proceedings of the National Academy of Sciences **84:**4890-4894.
201. **Nivinskas R, Malys N, Klausa V, Vaiškunaite R, Gineikiene E.** 1999. Post-transcriptional control of bacteriophage T4 gene 25 expression: mRNA secondary structure that enhances translational initiation. Journal of Molecular Biology **288:**291-304.

202. **Nothaft H, Szymanski CM.** 2010. Protein glycosylation in bacteria: sweeter than ever. Nature Reviews Microbiology **8:**765-778.
203. **Ono M, Kuwano M.** 1979. A conditional lethal mutation in an *Escherichia coli* strain with a longer chemical lifetime of messenger RNA. Journal of Molecular Biology **129:**343-357.
204. **Ow DS-W, Nissom PM, Philp R, Oh SK-W, Yap MG-S.** 2006. Global transcriptional analysis of metabolic burden due to plasmid maintenance in *Escherichia coli* DH5α during batch fermentation. Enzyme and Microbial Technology **39:**391-398.
205. **Palomares LA, Estrada-Moncada S, Ramírez OT.** 2004. Production of recombinant proteins, p. 15-51, vol. 267.
206. **Pan T, Sosnick T.** 2006. RNA folding during transciption. Annual Review of Biophysics and Biomolecular Structure **35:**161-175.
207. **Pato ML, Bennett PM, Von Meyenburg K.** 1973. Messenger ribonucleic acid synthesis and aegradation in *Escherichia coli* during inhibition of translation. Journal of Bacteriology **116:**710-718.
208. **Paulus M, Haslbeck M, Watzele M.** 2004. RNA stem–loop enhanced expression of previously non-expressible genes. Nucleic Acids Research **32:**e78.
209. **Perdue SA, Roberts JW.** 2010. A backtrack-inducing sequence is an essential component of *Escherichia coli* σ(70)-dependent promoter-proximal pausing. Molecular Microbiology **78:**636-650.
210. **Plotkin JB, Kudla G.** 2011. Synonymous but not the same: the causes and consequences of codon bias. Nature Reviews Genetics **12:**32-42.
211. **Py B, Higgins CF, Krisch HM, Carpousis AJ.** 1996. A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. Nature **381:**169-172.
212. **Ramos JL, Marqués S, Timmis KN.** 1997. Transcriptional control of the *Pseudomonas* TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators. Annual Review of Microbiology **51:**341-373.
213. **Redko Y, Tock MR, Adams CJ, Kaberdin VR, Grasby JA, McDowall KJ.** 2003. Determination of the catalytic parameters of the N-terminal half of *Escherichia coli* ribonuclease E and the identification of critical functional groups in RNA substrates. J. Biol. Chem. **278:**44001-44008.
214. **Retallack DM, Jin H, Chew L.** 2012. Reliable protein production in a *Pseudomonas fluorescens* expression system. Protein Expression and Purification **81:**157-165.
215. **Revyakin A, Liu C, Ebright RH, Strick TR.** 2006. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science **314:**1139-1143.
216. **Richards J, Sundermeier T, Svetlanov A, Karzai AW.** 2008. Quality control of bacterial mRNA decoding and decay. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms **1779:**574-582.
217. **Ring BZ, Yarnell WS, Roberts JW.** 1996. Function of *E. coli* RNA polymerase σ factor-σ70 in promoter-proximal pausing. Cell **86:**485-493.
218. **Ringquist S, MacDonald M, Gibson T, Gold L.** 1993. Nature of the ribosomal mRNA track: Analysis of ribosome-binding sites containing different sequences and secondary structures. Biochemistry **32:**10254-10262.
219. **Robinson M, Lilley R, Little S, Emtage JS, Yarranton G, Stephens P, Millican A, Eaton M, Humphreys G.** 1984. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. Nucleic Acids Research **12:**6663-6671.

220. **Ross W, Ernst A, Gourse RL.** 2001. Fine structure of *E. coli* RNA polymerase-promoter interactions: α subunit binding to the UP element minor groove. Genes & Development **15:**491-506.
221. **Ross W, Gourse RL.** 2009. Analysis of RNA polymerase-promoter complex formation. Methods **47:**13-24.
222. **Sahdev S, Khattar S, Saini K.** 2008. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Molecular and Cellular Biochemistry **307:**249-264.
223. **Saida F, Uzan M, Odaert B, Bontems F.** 2006. Expression of highly toxic genes in *E. coli*: special strategies and genetic tools. Current Protein and Peptide Science **7:**47-56.
224. **Sanderson A, Mitchell JE, Minchin SD, Busby SJW.** 2003. Substitutions in the *Escherichia coli* RNA polymerase σ70 factor that affect recognition of extended −10 elements at promoters. FEBS Letters **544:**199-205.
225. **Saunders R, Deane CM.** 2010. Synonymous codon usage influences the local protein structure observed. Nucleic Acids Research.
226. **Scharff LB, Childs L, Walther D, Bock R.** 2011. Local absence of secondary structure permits translation of mRNAs that lack ribosome-binding sites. PLoS Genetics **7:**e1002155.
227. **Schauder B, McCarthy JEG.** 1989. 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.
228. **Schmidt F-R.** 2006. From gene to product: the advantage of integrative biotechnology, p. 1-52, Handbook of Pharmaceutical Biotechnology. John Wiley & Sons, Inc.
229. **Schneider E, Blundell M, Kennell D.** 1978. Translation and mRNA decay. Molecular and General Genetics MGG **160:**121-129.
230. **Sengupta J, Agrawal RK, Frank J.** 2001. Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA. Proceedings of the National Academy of Sciences **98:**11991-11996.
231. **Seo SW, Yang J, Jung GY.** 2009. Quantitative correlation between mRNA secondary structure around the region downstream of the initiation codon and translational efficiency in Escherichia coli. Biotechnology and Bioengineering **104:**611-616.
232. **Sharp PM, Li W-H.** 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Research **15:**1281-1295.
233. **Sharp PM, Li WH.** 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Molecular Biology and Evolution **4:**222-230.
234. **Shine J, Dalgarno L.** 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proceedings of the National Academy of Sciences **71:**1342-1346.
235. **Shokri AS, Sandén AS, Larsson GL.** 2003. Cell and process design for targeting of recombinant protein into the culture medium of *Escherichia coli*. Applied Microbiology and Biotechnology **60:**654-664.
236. **Silhavy TJ, Kahne D, Walker S.** 2010. The bacterial cell envelope. Cold Spring Harbor Perspectives in Biology **2**.
237. **Simonetti A, Marzi S, Jenner L, Myasnikov A, Romby P, Yusupova G, Klaholz B, Yusupov M.** 2009. A structural view of translation initiation in bacteria. Cellular and Molecular Life Sciences **66:**423-436.

238. **Skorski P, Leroy P, Fayet O, Dreyfus M, Hermann-Le Denmat S.** 2006. The highly efficient translation initiation region from the *Escherichia coli rpsA* gene lacks a Shine-Dalgarno element. Journal of Bacteriology **188:**6277-6285.
239. **Skretas G, Georgiou G.** 2009. Genetic analysis of G protein-coupled receptor expression in *Escherichia coli*: Inhibitory role of DnaJ on the membrane integration of the human central cannabinoid receptor. Biotechnology and Bioengineering **102:**357-367.
240. **Skretas G, Georgiou G.** 2010. Simple genetic selection protocol for isolation of overexpressed genes that enhance accumulation of membrane-integrated human G protein-coupled receptors in *Escherichia coli*. Applied and Environmental Microbiology **76:**5852-5859.
241. **Sleight S, Bartley B, Lieviant J, Sauro H.** 2010. Designing and engineering evolutionary robust genetic circuits. Journal of Biological Engineering **4:**12.
242. **Sletta H, Nedal A, Aune TE, Hellebust H, Hakvag S, Aune R, Ellingsen TE, Valla S, Brautaset T.** 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*. Applied and Environmental Microbiology **70:**7033-7039.
243. **Sletta H, Tøndervik A, Hakvåg S, Aune TEV, Nedal A, Aune R, Evensen G, Valla S, Ellingsen TE, Brautaset T.** 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*. Applied and Environmental Microbiology **73:**906-912.
244. **Sorensen H.** 2010. Towards universal systems for recombinant gene expression. Microbial Cell Factories **9:**27.
245. **Sorensen H, Mortensen K.** 2005. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microbial Cell Factories **4:**1.
246. **Sousa C, de Lorenzo V, Cebolla A.** 1997. Modulation of gene expression through chromosomal positioning in *Escherichia coli*. Microbiology **143:**2071-2078.
247. **Spehr V, Frahm D, Meyer TF.** 2000. Improvement of the T7 expression system by the use of T7 lysozyme. Gene **257:**259-267.
248. **Spirin AS.** 2002. Termination of translation p. 261-270, Ribosomes. Springer US.
249. **Steitz JA, Jakes K.** 1975. How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. Proceedings of the National Academy of Sciences **72:**4734-4738.
250. **Stenström CM, Isaksson LA.** 2002. Influences on translation initiation and early elongation by the messenger RNA region flanking the initiation codon at the 3' side. Gene **288:**1-8.
251. **Stenström CM, Jin H, Major LL, Tate WP, Isaksson LA.** 2001. Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in Escherichia coli. Gene **263:**273-284.
252. **Stepanova E, Shevelev A, Borukhov S, Severinov K.** 2009. Mechanisms of action of RNA polymerase-binding transcription factors that do not bind to DNA. Biophysics **54:**555-568.
253. **Striedner G, Pfaffenzeller I, Markus L, Nemecek S, Grabherr R, Bayer K.** 2010. Plasmid-free T7-based *Escherichia coli* expression systems. Biotechnology and Bioengineering **105:**786-794.

254. **Sørensen HP, Mortensen KK.** 2005. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. Journal of Biotechnology **115:**113-128.
255. **Sørensen MA, Kurland CG, Pedersen S.** 1989. Codon usage determines translation rate in *Escherichia coli*. Journal of Molecular Biology **207:**365-377.
256. **Tegel H, Ottosson J, Hober S.** 2011. Enhancing the protein production levels in *Escherichia coli* with a strong promoter. FEBS Journal **278:**729-739.
257. **Terpe K.** 2006. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Applied Microbiology and Biotechnology **72:**211-222.
258. **Thanaraj TA, Argos P.** 1996. Ribosome-mediated translational pause and protein domain organization. Protein Science **5:**1594-1612.
259. **Tomsic J, Vitali LA, Daviter T, Savelsbergh A, Spurio R, Striebeck P, Wintermeyer W, Rodnina MV, Gualerzi CO.** 2000. Late events of translation initiation in bacteria: a kinetic analysis. EMBO J **19:**2127-2136.
260. **Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y.** 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell **141:**344-354.
261. **Turner P, Mamo G, Karlsson E.** 2007. Potential and utilization of thermophiles and thermostable enzymes in biorefining. Microbial Cell Factories **6:**9.
262. **Tyo KEJ, Ajikumar PK, Stephanopoulos G.** 2009. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nature Biotechnology **27:**760-765.
263. **Typas A, Becker G, Hengge R.** 2007. The molecular basis of selective promoter activation by the σ^S subunit of RNA polymerase. Molecular Microbiology **63:**1296-1306.
264. **Udagawa T, Shimizu Y, Ueda T.** 2004. Evidence for the translation initiation of leaderless mRNAs by the intact 70 S Ribosome without Its dissociation into subunits in Eubacteria. J. Biol. Chem. **279:**8539-8546.
265. **Valdez-Cruz N, Caspeta L, Perez N, Ramirez O, Trujillo-Roldan M.** 2010. Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda pL and/or pR promoters. Microbial Cell Factories **9:**18.
266. **Vallejo L, Rinas U.** 2004. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. Microbial Cell Factories **3:**11.
267. **Vanzo NF, Li YS, Py B, Blum E, Higgins CF, Raynal LC, Krisch HM, Carpousis AJ.** 1998. Ribonuclease E organizes the protein interactions in the *Escherichia coli* RNA degradosome. Genes & Development **12:**2770-2781.
268. **Ventura S, Villaverde A.** 2006. Protein quality in bacterial inclusion bodies. Trends in Biotechnology **24:**179-185.
269. **Vidal L, Lopez-Santin J, Caminal G, Ferrer P.** 2006. Development of an antibiotic-free plasmid selection system based on glycine auxotrophy for recombinant protein overproduction in *Escherichia coli*. Microbial Cell Factories **5:**P85.
270. **Vimberg V, Tats A, Remm M, Tenson T.** 2007. Translation initiation region sequence preferences in *Escherichia coli*. BMC Molecular Biology **8:**100.
271. **Vogel C, Marcotte EM.** 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nature Reviews Genetics **13:**227-232.
272. **Vytvytska O, Moll I, Kaberdin VR, von Gabain A, Bläsi U.** 2000. Hfq (HF1) stimulates *ompA* mRNA decay by interfering with ribosome binding. Genes & Development **14:**1109-1118.

273. **Wacker M, Feldman MF, Callewaert N, Kowarik M, Clarke BR, Pohl NL, Hernandez M, Vines ED, Valvano MA, Whitfield C, Aebi M.** 2006. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. Proceedings of the National Academy of Sciences **103:**7088-7093.
274. **Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, Aebi M.** 2002. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science **298:**1790-1793.
275. **Waegeman H, Soetaert W.** 2011. Increasing recombinant protein production in *Escherichia coli* through metabolic and genetic engineering. Journal of Industrial Microbiology & Biotechnology **38:**1891-1910.
276. **Wagner LA, Gesteland RF, Dayhuff TJ, Weiss RB.** 1994. An efficient Shine-Dalgarno sequence but not translation is necessary for *lacZ* mRNA stability in *Escherichia coli*. Journal of Bacteriology **176:**1683-1688.
277. **Wang Y, deHaseth PL.** 2003. Sigma 32-Dependent Promoter Activity In Vivo: Sequence Determinants of the groE Promoter. Journal of Bacteriology **185:**5800-5806.
278. **Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LBA, Gill RT.** 2010. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nature Biotechnology **28:**856-862.
279. **Welch M, Govindarajan S, Ness JE, Villalobos A, Gurney A, Minshull J, Gustafsson C.** 2009. Design parameters to control synthetic gene expression in *Escherichia coli*. PLoS ONE **4:**e7002.
280. **Welch M, Villalobos A, Gustafsson C, Minshull J.** 2011. Designing genes for successful protein expression. Methods Enzymol **498:**43-66.
281. **Welch M, Villalobos A, Gustafsson C, Minshull J.** 2009. You're one in a googol: optimizing genes for protein expression. Journal of The Royal Society Interface **6:**S467-S476.
282. **Winther-Larsen HC, Josefsen KD, Brautaset T, Valla S.** 2000. Parameters affecting gene expression from the *Pm* promoter in gram-negative bacteria. Metab Eng **2:**79-91.
283. **Woestenenk EA, Hammarstrom M, van den Berg S, Hard T, Berglund H.** 2004. His tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors. J Struct Funct Genomics **5:**217-229.
284. **Xu F, Cohen SN.** 1995. RNA degradation in Escherichia coli regulated by 3' adenylation and 5' phosphorylation. Nature **374:**180-183.
285. **Xu F, Lin-Chao S, Cohen SN.** 1993. The Escherichia coli pcnB gene promotes adenylylation of antisense RNAI of ColE1-type plasmids in vivo and degradation of RNAI decay intermediates. Proceedings of the National Academy of Sciences **90:**6756-6760.
286. **Yin J, Li G, Ren X, Herrler G.** 2007. Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. Journal of Biotechnology **127:**335-347.
287. **Yusupova GZ, Yusupov MM, Cate JHD, Noller HF.** 2001. The path of messenger RNA through the ribosome. Cell **106:**233-241.
288. **Zhang J, Deutscher MP.** 1992. A uridine-rich sequence required for translation of prokaryotic mRNA. Proceedings of the National Academy of Sciences **89:**2605-2609.
289. **Zhang W, Xiao W, Wei H, Zhang J, Tian Z.** 2006. mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in Escherichia coli. Biochemical and Biophysical Research Communications **349:**69-78.

290. **Zhilina E, Miropolskaya N, Bass I, Brodolin K, Kulbachinskiy A.** 2011. Characteristics of σ-dependent pausing by RNA polymerases from *Escherichia coli* and *Thermus aquaticus*. Biochemistry (Moscow) **76:**1098-1106.

Paper I

Journal of Biotechnology 158 (2012) 224–230

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Exploring the 5′-UTR DNA region as a target for optimizing recombinant gene expression from the strong and inducible *Pm* promoter in *Escherichia coli*

Laila Berg [a], Veronika Kucharova [a], Ingrid Bakke [a], Svein Valla [a], Trygve Brautaset [b,*]

[a] Department of Biotechnology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway
[b] Department of Biotechnology, SINTEF Materials and Chemistry, Sem Sælands vei 2A, N-7465 Trondheim, Norway

ARTICLE INFO

Article history:
Received 10 February 2011
Received in revised form 27 June 2011
Accepted 4 July 2011
Available online 22 July 2011

Keywords:
Combinatorial mutagenesis
Synthetic operon
Recombinant gene expression
Pm/xylS
Interferon-α2b

ABSTRACT

By using the strong and inducible *Pm* promoter as a model, we recently reported that the β-lactamase production (encoded by *bla*) can be stimulated up to 20-fold in *Escherichia coli* by mutating the DNA region corresponding to the 5′-untranslated region of mRNA (UTR). One striking observation was the unexpected large stimulatory effect some of these UTR variants had on the *bla* transcript production level. We here demonstrate that such UTR variants can also be used to improve the expression level of the alternative genes *celB* (encoding phosphoglucomutase) and *inf-α2b* (encoding human cytokine interferon α2b), which both can be expressed to high levels even with the wild-type *Pm* UTR DNA sequence. Our data indicated some degree of context dependency between the UTR DNA and concomitant recombinant gene sequences. By constructing and using a synthetic operon, we demonstrated that UTR variants optimized for high-level expression of probably any recombinant gene can be efficiently selected from large UTR mutant libraries. The stimulation affected both the transcript production and translational level, and such modified UTR sequences therefore clearly have a significant applied potential for improvement of recombinant gene expression processes.

1. Introduction

The 5′-untranslated region (the UTR) of bacterial mRNAs is well-known to be critical for efficient translation, and the features generally believed to be important are the SD sequence (Shine and Dalgarno, 1974) and A/U richness, which serves to minimize the formation of strong secondary structures in the ribosome binding site (Kozak, 2005; Simonetti et al., 2009). Ribosomes seem to preferably bind to unstructured regions of the mRNA (Gu et al., 2010; Nakamoto, 2006). It was recently found (Kudla et al., 2009) that synonymous mutations in the coding region of the *gfp* gene encoding Green fluorescence Protein (GFP) causing tight folding of the mRNA molecule in the ribosome binding site (including the UTR), correlated strongly with a reduction in the GFP protein production level. This indicated a context dependency between the UTR and the gene to be expressed. On the other hand, 5′-stabilizing elements in mRNAs which are usually located in the UTR and involve elements such as stem–loop structures, are believed to be critical for the mRNA stability because the first endonucleolytic cleavage seems to be the rate-limiting step in mRNA turnover (Condon, 2007; Picard et al., 2009). Thus, UTRs can presumably affect mRNA translation in several different ways.

Besides the well documented role of UTRs for translation it has recently been shown that the UTR DNA region can also influence the transcriptional process. Mutations near the transcriptional start site can affect the start site selection and thus influence the gene expression levels (Walker and Osuna, 2002; Lewis and Adhya, 2004). The about 25 bps 5′-part of the UTR DNA region (the initial transcribed region) has been shown to influence the promoter escape efficiency of the *Escherichia coli* RNA polymerase *in vitro* (Hsu et al., 2006). Also, sequences similar to the −10 promoter element (TATAAT) in UTRs have been demonstrated to induce a σ^{70}-dependent transcription pause after the promoter escape (Brodolin et al., 2004; Nickels et al., 2004; Hatoum and Roberts, 2008) and thus reduced transcription.

We previously integrated the inducible and σ^{32}/σ^{38}-dependent *Pm* promoter (Marqués et al., 1999) and its cognate UTR together with its positive regulator gene *xylS* into minimal replicon broad-host-range expression vectors (the pJBn vectors) based on the RK2 plasmid (Blatny et al., 1997a,b). These vectors have many favourable properties (for a review see Brautaset et al., 2009), and one of them (pJB658) has been used to produce industrial levels of several medically important recombinant proteins in *E. coli* under high-cell densities (Sletta et al., 2004, 2007). Surprisingly, we recently found that the expression of β-lactamase (encoded by *bla*) from wild-type *Pm* can be stimulated up to 20-fold by mutating the *Pm* UTR DNA region, and such effects appeared to be partly or for some UTR sequences mainly due to stimulated transcript

* Corresponding author. Tel.: +47 98 28 39 77; fax: +47 73 59 69 95.
E-mail address: trygve.brautaset@sintef.no (T. Brautaset).

0168-1656/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2011.07.012

production. This work demonstrated that UTR DNA coding sequences can have major effects on transcription level and that they should be regarded important mutational targets for improving bacterial expression systems (Berg et al., 2009). We later used an analogous approach to isolate UTR variants causing strongly reduced expression level of recombinant genes from *Pm*. The latter UTR variants represented valuable metabolic engineering tools for fine-tuning of the expression level of a complex biosynthetic pathway at physiologically relevant levels, enabling controlled heterologous production of the C_{50} carotenoid sarcinaxanthin in *E. coli* (Lale et al., 2011). A general observation made was that the induction ratio of the *Pm* promoter with the isolated UTR variants remained high, which is an advantage for any applications using this promoter system.

The purpose of this study was to further explore to what extent such improved *Pm* UTR DNA variants can lead to stimulated expression of any gene of interest (GOI) other than *bla*, and to investigate the sequence context dependency between the UTR DNA sequence and the concomitant recombinant gene to be expressed.

2. Materials and methods

2.1. Bacterial strains and growth media

E. coli DH5α (Bethesda Research Laboratories) was used as general cloning host and as host for recombinant expression of *celB* and *bla* genes. *E. coli* NovaBlue (Novagen) was used as host for the construction and screening of the UTR mutant library (LV). *E. coli* RV308 (ATCC 31608) was used as host for recombinant expression of *inf-α2b*. *E. coli* cells were generally grown in L broth (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) or on L agar (L broth with 15 g/L Agar), while HiYe medium was used for the *inf-α2b* gene expression analyses (Sletta et al., 2007). For general cloning experiments cells were grown at 37 °C. For expression experiments, recombinant cells were grown at 30 °C and induction of the *Pm/xylS* system was done by adding *m*-toluic acid as indicated. Kanamycin (50 μg/mL) was added as appropriate.

2.2. Vector constructions

Plasmid pLB11 is a pIB11 (Berg et al., 2009) derivative in which the *Nde*I/*Bam*HI fragment containing *bla* was substituted with the *Nde*I/*Bam*HI fragment containing the *celB* coding region from pJB658*celB* (Blatny et al., 1997b).

Plasmid pKO1 is a pIB11 derivative in which a 37 bp synthetic fragment containing a *Pac*I restriction site, and a translational stop overlapping the translational start of *bla* was inserted at the 5-terminal end of the *bla* gene. This insertion was done by amplifying the *bla* gene from pIB11 by PCR with the primer pair 5′-TCATGAACATATGTTATTTCTCTTTCTTAAAAATAT**TTAATTAA**A**TAATGA**GTATTCAACATTTCCGTGT-3′ and 5′-AGCTAGAGGATCCCCGGGTA-3′ and subsequent cloning of the *Nde*I/*Bam*HI digested (*Nde*I and *Bam*HI restriction sites are underlined in the primers, and the *Pac*I restriction site and overlapping stop/*bla* start codon are written in bold) PCR product into the corresponding sites of pIB11.

Plasmid pKO1*celB* is a pKO1 derivative with the *celB* coding region cloned downstream of *Pm*. The *celB* gene was amplified from pLB11 with the primer pair 5′-GAGTCATGAACATATGCCCAGCATAA-3′ and 5′-ATGGAATCATTTAATTAAATAGCCAGCGTT-3′, and end digested with *Nde*I and *Pac*I (restriction sites are underlined in the primers) prior to ligation into the corresponding sites of pKO1.

The *pelB-inf-α2b-c-myc-his6* coding region was PCR-amplified from plasmid pINF30SpelB (Sletta et al., 2007) by using the primer pair 5′-AAGAAGCGGATACAGGAGTG-3′ (binds upstream of the *Nde*I site) and 5′-CTTGGTACCTTGTTCGGCCGGAT-3′. The resulting DNA fragment was digested with *Nde*I and *Kpn*I (restriction site underlined in the primer), and ligated into the corresponding sites of plasmid pIB11 (substituting the *bla* coding region), yielding plasmid pVK1. A copy-up version of pVK1 (pVK1cop271) was constructed by substituting the *Sex*AI/*Pvu*II fragment containing the *trfA* gene in pVK1 respectively, with the corresponding fragment from plasmid pJBphOx-271 (Sletta et al., 2004).

2.3. Standard DNA manipulations

A modified RbCl protocol (www.promega.com) was used for the transformation of *E. coli*. Plasmids were isolated by using WizardPlus SV minipreps DNA purification kit (Promega) or Qiagen plasmid midi kit (Qiagen), and enzymatic manipulations were performed as described by the manufacturers. PCRs were performed using the Expand high fidelity PCR system kit (Boehringer Mannheim) for cloning purposes, and the polymerase DynazymeII (Finnzymes) for generation of templates for DNA sequencing. Prior to DNA sequencing the PCR products were treated with the enzyme mixture ExoSapIt (USB). DNA sequencing was performed with the ABI PRISM BigDye sequencing kit (Applied Biosystems), and analyzed using an Applied Biosystems 3130 XL DNA sequencing machine (Applied Biosystems). Alternatively, the sequencing analyses were purchased from Eurofins MWG operon (www.eurofinsdna.com).

2.4. Construction of the LV Pm UTR DNA mutant library

The pKO1*celB* LV *Pm* UTR DNA mutant library was constructed essentially by using the similar approach as for the LII *Pm* UTR DNA mutant library (Berg et al., 2009), except that the doping percentages of the nucleotide mixtures were modified. The numbers in the oligonucleotide sequence (5′-CATGT12114114114133134214311CA-3′) indicate nucleotide mixtures with the doping percentages: 1 = 79% A, 7% C, G, and T; 2 = 79% C, 7% A, G, and T; 3 = 79% G, 7% A, C, and T; 4 = 79% T, 7% A, C, and G. Approximately 300,000 *E. coli* transformants were obtained for the LV library.

2.5. Using the synthetic operon for the selection of optimized UTR variants

E. coli cells harboring the pKO1*celB* LV *Pm* UTR DNA mutant library was screened on solid agar plates containing different ampicillin concentrations as described previously (Winther-Larsen et al., 2000; Berg et al., 2009). The DNA sequence of the UTR variant LV-3 and LV-4 were determined by using the primer 5′-CTATCAAACCGGACACGTTTATCGTGGTTATGC-3′. PCR products were used as templates for the sequencing reactions generated with the primer pair 5′-GATGTAGAAAGGCGCCAAGTC-3′ and 5′-TGGTCAGCGAGGAACCACGATG-3′. The selected UTR variants were reconstructed by cloning annealed newly synthesized oligonucleotides with the specific mutations representing each UTR variant into plasmid pLB11. The reconstructed UTR variants were confirmed by DNA sequencing.

2.6. Quantification of recombinant proteins by enzyme assays (β-lactamase and CelB), SDS-PAGE (CelB) and Western blot (INF-α2b)

For the analysis of recombinant protein production levels by enzymatic assays and SDS-PAGE, exponentially growing cells were induced with *m*-toluic acid as described above (Section 2.1), cell growth continued for 4–6 h, and crude extracts were prepared and assayed for CelB and β-lactamase enzyme activities as described

previously (Winther-Larsen et al., 2000). All enzyme assays were repeated at least twice, and measurements were carried out with minimum three technical recurrences. For SDS-PAGE analysis cell extracts were prepared as described in Sletta et al. (2007), except for that the lysis of cell samples was done by sonication 4×90 s, with 30 s cooling periods (Branson Sonifier, 30% duty control, 3 output control). Total protein concentration was determined with the Bio-Rad Detergent-Compatible Protein Assay (Bio-Rad Laboratories, Hercules, CA, USA) as described by the manufacturers. 10 μg total protein was loaded in each well and separated under denaturing conditions using 12% SDS–PAGE gels followed by Coomassie blue staining. Qualitative detection of INF-α2b protein was performed by using Western blot as described in Sletta et al. (2007), and the cell extract were prepared as described above for SDS-PAGE analysis.

2.7. RNA isolation, cDNA synthesis, and transcript quantification by qRT-PCR

For the analysis of transcript production levels by quantitative real-time PCR (qRT-PCR), recombinant strains were cultivated as described in Section 2.6. RNA isolation, cDNA synthesis and two-step qRT-PCR were performed and analyzed as previously described (Berg et al., 2009). PCR cycles were 95 °C for 10 min, followed by 40 cycles of amplification (95 °C for 15 s; 60 °C for 1 min). Primer pairs used for transcript quantification were 5′-ACCAGCTTCAATGAAAACCACAT-3′ and 5′-CGCCCTTGCGGTAATCG-3′ for *celB*, 5'-ACGTTTTCCAATGATGAGCACTT-3′ and 5′-TGCCCGGCGTCAACAC-3′ for *bla*, and 5′-CGAGACCCCGCTGATGAA-3′ and 5′-CAGATACAGGGTGATACGCTGAAA-3′ for *inf-α2b*. The kanamycin resistance gene, which is present on all plasmids, was used as a normalizer by using primer pair 5′-TACCTTTGCCATGTTTCAGAAACA-3′ and 5′-AATCAGGTGCGACAATCTATCGA-3′.

2.8. In silico translation initiation rate predictions

For the prediction of translation initiation rates, the mRNA region −32/+57 (base + 1 is the adenine of the translation initiation codon) of the wild-type *Pm* UTR and the *Pm* UTR variants were analyzed using the reverse engineering tool of the RBS calculator website (http://voigtlab.ucsf.edu/software/) developed by the Voigt lab (Salis et al., 2009). The wild-type mRNA sequences were (translation initiation codon underlined): 5′-AACATGTACAATAATAATGGAGTCATGAACATATGCCCAGCATAAGCCCATTTGCCGGCAAGCCGGTCGATCCGGACCGTCTTGTCAAT-3′ for *celB* and 5′-AACATGTACAATAATAATGGAGTCATGAACATATGAAATCCCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCGGCCCAGCCG-3′ for *inf-α2b*. Extending the coding region sequence beyond the length described here did not have any effect on the predictions of the translation initiation rate.

3. Results and discussion

3.1. Previously isolated Pm UTR DNA variants can confer further increase of celB gene expression, but mainly at the transcriptional level

By using a combinatorial mutagenesis strategy, we previously isolated six different *Pm* UTR DNA sequence variants (UTR variants), LII-7 to LII-12, that strongly stimulated recombinant expression of β-lactamase (up to 20-fold) encoded by the *bla* gene (Berg et al., 2009). The UTR variants were in these cases selected directly on the basis of their ability to stimulate the expression of this particular gene. Interestingly, even though these UTR variants seemed to lead to a significant increase at the translational level (significantly higher stimulation at the protein product level compared to that of the accumulated transcript level), the SD sequence (GGAG) remained intact in all UTRs. The latter should indicate that the SD sequence is close to optimal for efficient translation. To further investigate the impact of these findings it was of interest to unravel if the UTR variants can also be used to increase the expression level of any gene of interest (GOI) from *Pm*, or whether there is a context dependency such that the stimulatory effects are selective for *bla* only. We have previously shown that the phosphoglucomutase gene, *celB*, originating from *Acetobacter xylinum*, can be expressed to very high levels (clearly visible bands on SDS–PAGE of crude cell extracts) from the wild-type *Pm/xylS* cassette in *E. coli* (Blatny et al., 1997b). Furthermore, high CelB levels determined by SDS-PAGE were correlated with corresponding high CelB activities, demonstrating that this protein can be expressed to very high levels in soluble and active form. We have therefore used *celB* as a valuable reporter gene in several studies aiming at developing expression tools for various biotechnological applications (Bakke et al., 2009; Berg et al., 2009; Lale et al., 2011).

Motivated by the challenge of improving an already established high expression level, we tested if any of the six LII UTR variants can be used to further improve *celB* expression from *Pm*. Plasmid pLB11 was constructed by substituting the *bla* coding region in plasmid pIB11 (Berg et al., 2009) with the *celB* coding region, and the wild-type *Pm* UTR DNA sequence in plasmid pLB11 was substituted with each of the LII-7 to LII-12 UTR variants (Fig. 1A). All plasmids were transformed to *E. coli* and the resulting recombinant strains were analyzed for *celB* expression at both the transcript (qRT-PCR) and protein levels (CelB activity assay) under induced (2 mM *m*-toluic acid) conditions. The six UTR variants all stimulated *celB* transcript accumulation (up to 3.0-fold, LII-10), relative to what was obtained with the corresponding wild-type *Pm* UTR (Fig. 1B). Interestingly, the transcript accumulation was not accompanied by a corresponding stimulation at the protein level, even though the LII-10 UTR mutant gave rise to an about 50% increase relative to wild-type UTR. These observations indicated to us that the UTR mutants may act in a gene-independent way, but that the *celB* gene is translated with a close to optimal efficiency, even based on a wild-type UTR from *Pm*.

Alternatively, there exist other UTR variants that are even more optimal for *celB* transcription and translation, but these were not found in previous experiments because they were selected on the basis of their efficiency to stimulate expression of *bla*.

3.2. Construction of a synthetic operon enabling selection of UTR variants optimized for efficient expression of any GOI

To experimentally investigate the alternative hypotheses described above, we needed to develop a new selection tool to identify *Pm* UTR variants adapted to any GOI. In the previous studies we found it very useful that there is a good correlation between the *bla* expression level and the corresponding ampicillin tolerance level in *E. coli* host cells (Berg et al., 2009; Bakke et al., 2009), and we therefore tried to make use of this feature also in the new screening system. A plasmid designated pKO1 was constructed (Fig. 2), and in its expression cassette the mechanism of translation reinitiation is utilized (Adhin and van Duin, 1990; André et al., 2000). In pKO1, *bla* is as before used as a reporter gene, but by inserting an upstream GOI, it is possible to select for increased expression level of in principle any gene. The GOI can be inserted at the ATG start downstream of *Pm* and the bases corresponding to the stop codon (TAA) of this gene overlaps the start codon ATG of the *bla* gene (generating the sequence TAATG). In this way, the GOI and *bla* are co-transcribed as one single mRNA molecule from *Pm* and translation initiation of *bla* relies on the translation efficiency of the upstream GOI. Any *de novo*

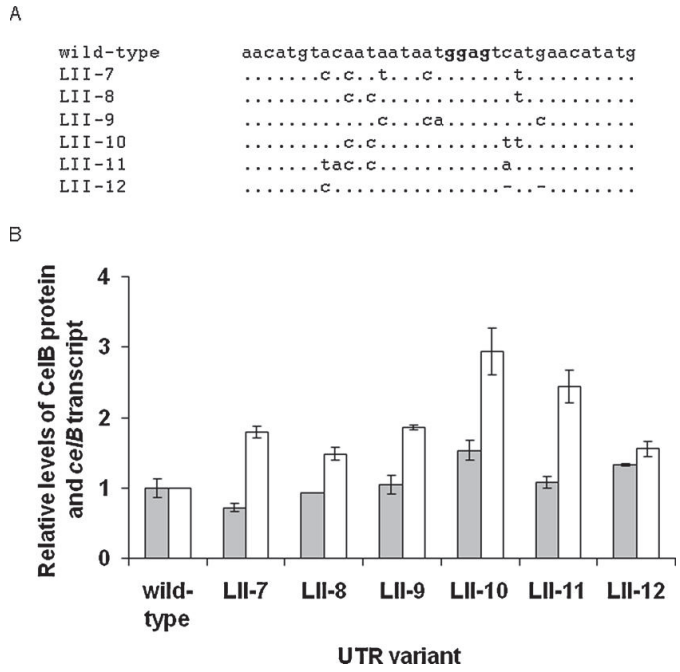


Fig. 1. *celB* transcript level and CelB activity level of recombinant *E. coli* cells as a function of different *Pm* UTR variants. (A) Sequence of the *Pm* UTR DNA sequence (including the translational start codon) and the previously identified UTR variants LII-7 to -12 (imported from Berg et al., 2009). The putative Shine–Dalgarno sequence is written in bold. Deletion mutations are indicated by short horizontal lines. (B) *celB* transcript (white) and CelB activity (grey) levels. The values are relative to that of the cells containing the wild-type *Pm* UTR DNA sequence (pLB11), arbitrarily set to 1, and are the average of at least two biological recurrences. The error bars show the deviation between them.

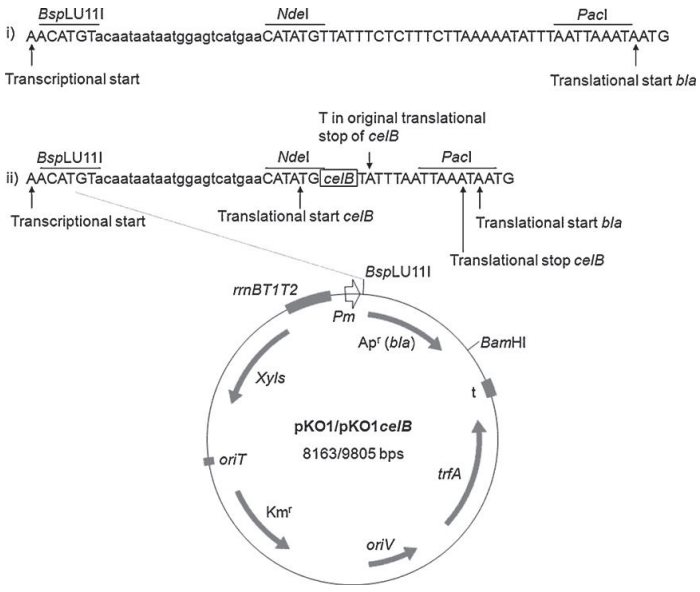


Fig. 2. Map of the synthetic operon plasmids pKO1 and pKO1*celB*. The restriction sites shown are unique. *bla*, gene encoding β-lactamase; *celB*, gene encoding phosphoglucomutase; Km^r, kanamycin resistance gene; *trfA*, gene encoding the essential replication protein; *xylS*, gene encoding the activator XylS; *oriV*, origin of vegetative replication; *oriT*, origin of transfer; *t*, bidirectional transcriptional terminator; *rrnBT1T2*, bidirectional transcriptional terminator. Details for the transcriptional and translational initiation regions of *Pm* are displayed above the plasmid map for pKO1 (i) and pKO1*celB* (ii). To obtain the overlapping translational stop and *bla* start, a region encoding four amino acid residues were introduced at the 3′-terminal end of *celB*. Nucleotides in lowercase were randomly mutagenized by the use of a mixed oligonucleotide solution. The transcriptional start site for the *Pm* promoter in pKO1 and pKO1*celB* is assumed to be identical to their parental plasmid pIB11 (Berg et al., 2009).

translation initiation of *bla* is minimized due to the lack of a competent SD sequence upstream of this gene (see Fig. 2). Accordingly, any genetic manipulations causing increased transcription and/or increased translation of the GOI should lead to increased transcription and/or increased translation of *bla*, with concomitant increased ampicillin tolerance level of the host cells. A similar approach using the tet^R gene for selection has been used to identify an efficient translation initiation site for expression of the human *RACK1* gene in *E. coli* (Zhelyabovskaya et al., 2004).

To test the functionality of this synthetic operon, *celB* was again chosen as the GOI, and the resulting vector pKO1*celB* and the control vector pKO1 (Fig. 2) were transformed into *E. coli* for expression analyses. The maximal ampicillin tolerance levels under induced (2 mM *m*-toluic acid) conditions of *E. coli* (pKO1) and *E. coli* (pKO1*celB*) were approximately 200 μg/mL and 900 μg/mL, respectively, confirming that the presence of *celB* caused a strong increase in the β-lactamase production level. Despite of some background *bla* expression from pKO1, it appeared likely that pKO1*celB* can be used as a tool to select for UTR variants specifically optimized for high-level expression of *celB*.

3.3. Screening of UTR libraries in the synthetic operon revealed new variants stimulating celB expression

A new *Pm* UTR mutant library (designated LV) was generated in pKO1*celB* by the use of randomly mutated synthetic oligonucleotides, and established in *E. coli.* The resulting library was screened with respect to increased ampicillin tolerance level relative to the control strain, *E. coli* (pKO1*celB*), under induced conditions (2 mM *m*-toluic acid). Two promising candidates with vectors carrying UTR variants LV-3 and LV-4 were selected and their UTR DNA sequences were determined (Table 1). To rule out that the observed improved phenotypes of the recombinant strains are caused by mutations outside the UTR DNA regions, LV-3 and LV-4 were resynthesized and cloned into pLB11. The *celB* expression levels of the corresponding recombinant strains were measured under induced (2 mM *m*-toluic acid) conditions (Table 1). The *celB* transcript production level was increased about 3-fold (LV-3) compared to the wild-type UTR, similar to the best result obtained with the LII UTR variants (LII-10, see above). At the protein level the stimulation level was also similar (1.5-fold) to that observed for the best UTR variant obtained previously (LII-10). These results confirmed that the synthetic operon can be used to select for UTR DNA variants that cause improved expression level of *celB* and probably also any other GOI, but also indicate that for *celB* the potential for improvement of the wild-type UTR is probably exhausted by the best of both the previously available and new UTR mutations. We noticed that for UTR variant LV-4 the CelB protein production level was reduced to 0.7-fold compared to when using the wild-type UTR, possibly indicating that a given UTR in the synthetic operon context in some cases may lead to a different effect on expression after transfer to the single GOI context. It is tempting to speculate that changes in mRNA structure play a role in this.

To verify that differences in measured CelB activities truly correspond to the respective CelB expression levels *in vivo*, cell extracts were prepared from the recombinant cultures and analyzed by SDS-PAGE. The results clearly showed that there is a good correlation between enzyme activity and expression level of CelB, thus confirming that this protein is produced into soluble active form in all strains tested (Fig. 3).

For reporter gene *celB*, the overall production results obtained with the most efficient UTR variants from the LV (LV-3) and from the LII (LII-10) libraries were similar, while the mutations in these two UTR variants were different. We believe that these results strengthen the hypothesis that the wild-type *Pm* UTR DNA sequence is close to optimal for the translation of *celB*, and that

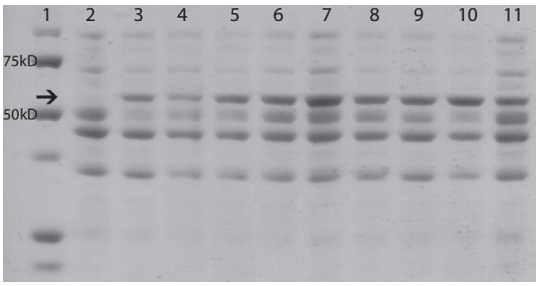


Fig. 3. SDS-PAGE analysis of recombinant strains expressing CelB from different UTR variants. Crude extracts were prepared from recombinant cell cultures as described in Section 2 and samples containing 10 μg protein were run on SDS–PAGE. Lanes: 1, molecular weight standard; 2, negative control (cells harboring plasmid pIB11); 3, UTR variant LII-7; 4, UTR variant LII-8; 5, UTR variant LII-9; 6, UTR variant LII-10; 7, UTR variant LII-11; 8, UTR variant LII-12; 9, UTR variant LV-3; 10, UTR variant LV-4; 11, wild-type UTR. The position of CelB (55 kDa) is indicated with an arrow.

changes in the UTR sequence that lead to enhanced transcription are likely to reduce the translational efficiency per transcript. The overall conclusion is still that the synthetic operon can be used as a screening tool for any GOI, but that the final outcome is dependent on gene-specific characteristics.

3.4. LV UTR variants can strongly stimulate bla expression levels independent of celB

The LV UTR variants affected *celB* expression mainly by causing increased transcript accumulation levels, and accordingly we predicted that they could likely be used to increase the transcript production level of any GOI. To test this, the wild-type *Pm* UTR DNA sequence in plasmid pIB11 (carrying *bla* under control of *Pm*) was substituted with the LV-3 and LV-4 UTR variants, and recombinant *E. coli* cells harboring the resulting plasmids were analyzed under induced (2 mM *m*-toluic acid) conditions with respect to *bla* expression (Table 1). The *bla* transcript levels were increased about 14- and 11-fold, respectively, and the corresponding β-lactamase production levels were increased about 13- to 14-fold, compared to the values obtained with the wild-type *Pm* UTR (pIB11). This observation is consistent with the assumption that LV-3 and LV-4 mainly act by stimulating transcription (as suggested above), in contrast to the LII-10 UTR variant, which to a greater extent stimulate translation. The more limited increase in accumulated transcript amounts for this mutant may instead be the result of mRNA stabilization due to increased translation efficiency.

It has been reported that sequences similar to the −10 promoter element (TATAAT) in UTRs can promote σ^{70}-dependent transcription pause after the promoter escape (Brodolin et al., 2004; Nickels et al., 2004; Hatoum and Roberts, 2008) and thus reduced transcription. Immediately upstream of the *Pm* SD sequence there is a similar motif (AATAATAAT) and several of the selected UTR mutations are in this region (see Fig. 1A and Table 1). It was thus plausible to suggest that the effects of such mutations might at least partly be due to reduced pausing and concomitant more efficient transcription.

3.5. Pm UTR variants can be used to further improve the high production level of the medically important cytokine interferon-α2b

We recently demonstrated that UTR variants specifically selected for low-level gene expression could be successfully applied to control heterologous production of the industrially important C_{50} carotenoid sarcinaxanthin in *E. coli* (Lale et al., 2011). To also address the applied potential of modifying the UTR DNA region

Table 1
Accumulated *celB* and *bla* transcript production levels (qRT-PCR) versus CelB and β-lactamase protein production levels (enzyme activity assays) by using different *Pm* UTR DNA variants.

Pm UTR DNA variant	DNA sequence (5′–3′)[a]	Relative accumulated celB transcript production level	Relative CelB protein/activity	Relative accumulated bla transcript production level	Relative Bla protein/activity
Wild-type	caacatgtacaataataatggagtcatgaacatatg	1.0 ± 0.0	1.0 ± 0.1	1.0 ± 0.0	1.0 ± 0.0
LII-10[b]	caacatgtac**c**a**c**aataatggagt**tt**tgaacatatg	3.0 ± 0.3	1.5 ± 0.1	5.6 ± 0.1	16.2 ± 0.9
LV-3	caacatgtacaataata**ca**gga**aca**atgaacatatg	2.9 ± 0.1	1.6 ± 0.1	13.7 ± 1.1	13.4 ± 0.3
LV-4	caacatgtacaat**c**ata**cca**gagt**a**atgaacatatg	1.6 ± 0.2	0.6 ± 0.2	11.4 ± 0.4	13.7 ± 1.2

[a] Mutations giving rise to the identified UTR DNA variants are written in bold and underlined.
[b] The relative accumulated *bla* transcript and β-lactamase protein production levels are collected from Berg et al. (2009).

for high-level gene expression, we chose to test the *ifn-α2b* gene encoding the medically important human cytokine interferon-α2b (IFN-α2b). We previously showed that the *Pm/xylS* system could be used to express high levels (0.6 g/l) of IFN-α2b in *E. coli* under high-cell density conditions (Sletta et al., 2007). This result relied on an *ifn-α2b* coding sequence that was genetically optimized for expression in *E. coli*, and fusion of the translocation signal sequence, *pelB*, at the 5′-end of the synthetic *ifn-α2b* gene. All detectable IFN-α2b was present as insoluble protein. Here we wanted to test if the *ifn-α2b* expression level from *Pm* could be increased by using improved UTR variants, and for this purpose plasmid pVK1 was constructed, which is identical to pLB11 except that *celB* is substituted with the *pelB-inf-α2b-c-myc-his6* gene fusion (Sletta et al., 2007).

In order to mimic previously optimized expression conditions for this particular protein, *E. coli* RV308 was chosen as the host and 0.5 mM toluic acid was used for the induction of *Pm*, as described by Sletta et al. (2007). To simplify Western blot analyses we also substituted the wild-type *trfA* gene of plasmid pVK1 with the mutant *trfAcop271C* gene (Blatny et al., 1997a), resulting in plasmid pVK1cop271 with an estimated 4–5-fold increased plasmid copy number compared to parental plasmid pVK1. Moreover, this is the same vector copy number as used for IFN-α2b production previously (Sletta et al., 2007). The wild-type *Pm* UTR DNA sequence in pVK1cop271 was then substituted with UTR variants LII-10, LII-12, LV-3, and LV-4, and all plasmids were transformed into *E. coli* for expression analyses. Interestingly, the LII-10 variant stimulated transcript accumulation from *inf-α2b* about 3.3-fold compared to that of the wild-type *Pm* UTR (pVK1cop271; Fig. 4A). Also, the LII-12 and to some degree the LV-3 variant, resulted in increased transcript accumulation, while this was not the case for the LV-4 variant. The LII-12 UTR variant stimulated INF-α2b production level about 3–4-fold compared to the wild-type *Pm* UTR under the conditions tested, as judged from Western blot (Fig. 4B). The remaining three tested UTR variants displayed only minor or no positive effects on the recombinant IFN-α2b production levels compared to the wild-type *Pm* UTR (data not shown). In all cases, all detectable IFN-α2b protein was produced as insoluble proteins similar as reported previously (Sletta et al., 2007). Summarized, these data clearly demonstrated that access to a small collection of stimulatory *Pm* UTR variants is sufficient to identify variants that improve production processes of presumably most proteins, including proteins of medical importance.

3.6. *In silico translation initiation rate predictions of the UTR variants correlate partly with the corresponding experimental data*

Based on the results presented here and elsewhere (see Section 1), it seems like the UTR or the corresponding DNA region can significantly influence both the transcript and protein production levels in *E. coli*. However, the underlying mechanisms of the effect displayed by the different *Pm* UTR variants on the *celB* and *ifn-α2b* gene expression remained unclear, and it was difficult to predict how the different UTR DNA variants affect the expression levels of different genes. It is not for example obvious why an increase in the *ifn-α2b* transcript production level lead to an increase in the INF-α2b protein production for the LII-12 UTR but not for the LII-10 UTR, while both these UTR variants led to an increase in both the transcript and protein production levels for the *celB* gene. Thus, there clearly exist a context dependency between the UTR DNA and the coding sequence of the gene to be expressed, and the strength of the secondary structures in the mRNA is believed to be a key determinant as well. To analyse this further, we therefore used the RBS calculator (Salis et al., 2009) to predict the translation initiation rate of an existing ribosome binding site sequence (including effects of any secondary structures in this region) for the UTR DNA variants combined with both the *celB* gene and the *ifn-α2b* gene (Table 2). Interestingly, the overall results of the translation initiation rate predictions correlated well with the experimentally determined protein product levels, except for the LII-10 and LV-3 DNA variant with *celB* and the LV-4 DNA variant with the *inf-α2b*

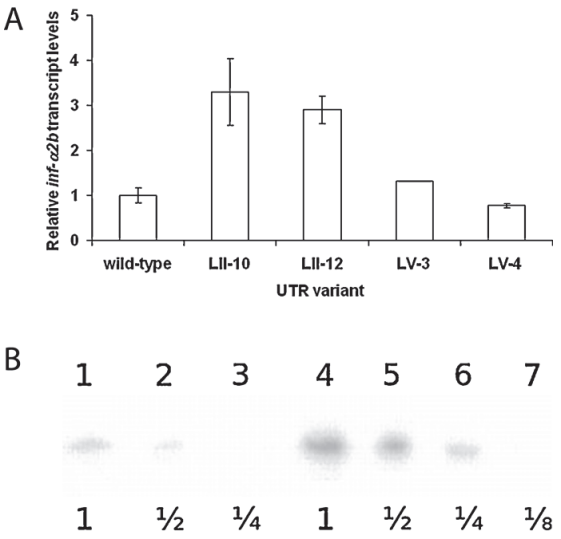


Fig. 4. *inf-α2b* transcript production levels (A) and accompanying IFN-α2b production levels (B) of recombinant *E. coli* cells as a function of different *Pm* UTR variants. (A) The values are relative to that of the cells containing pVK1cop271 (wild-type *Pm* UTR DNA sequence) arbitrarily set to 1, and are the average of at least two biological recurrences. The error bars show the deviation between them. (B) Western blot of total amount of INF-α2b protein produced by recombinant *E. coli* strains containing either the wild-type *Pm* UTR DNA sequence (pVK1cop271; lanes 1–3) or the LII-12 DNA variant (lanes 4–7). The relative amount of total protein loaded on the corresponding protein gel is displayed below the Western blot, for which a relative amount of total protein of 1 equals 60 μg.

Table 2
Relative *in silico* translation initiation rate prediction and its accuracy from the RBS Calculator (Salis et al., 2009) for the different UTR DNA variants combined with either the *celB* or the *ifn-α2b* gene.

Pm UTR DNA variant	*celB*		*inf-α2b*	
	Relative translation initiation rate prediction	Satisfactory accuracy[a]	Relative translation initiation rate prediction[b]	Satisfactory accuracy[a]
Wild-type	1.00[b]	Yes	1.00	Yes
LII-7	0.84	Yes	ND[c]	–
LII-8	1.00	No	ND	–
LII-9	1.79	Yes	ND	–
LII-10	0.67	No	0.64	Yes
LII-11	0.87	No	ND	–
LII-12	1.48	Yes	1.41	Yes
LV-3	0.41	Yes	0.73	Yes
LV-4	0.32	No	0.14	Yes

[a] Satisfactory accuracy signifies a successful analysis (Yes), or alternatively an unsatisfactory analysis (No), the latter means that the analysis might be of low quality because the mRNA sequence around this start codon is expected to fold slowly to its equilibrium (see Salis et al., 2009).
[b] Wild-type UTR is here set to 1.0, for further details see text.
[c] ND is not determined.

gene. Based on these results, it seems like even though both the LII-10 and LII-12 DNA variants lead to a significant stimulation of the *inf-α2b* transcript production level, the LII-10 variant is less favourable for efficient translation compared to both the wild-type and the LII-12 UTR.

4. Conclusions

In the present study we investigated and extended previous studies regarding the importance of the UTR DNA region as an important target for manipulations aiming at optimizing and maximizing heterologous protein production in *E. coli*. We managed to significantly increase the expression levels of two new and already well expressed genes from the strong *Pm* promoter by mutating its UTR DNA sequence. In particular, the new data supported our recent discovery that the UTR DNA region can have a strong impact on the transcript production level of recombinant genes in *E. coli*, and that there is a certain degree of context dependency between the UTR and the concomitant GOI coding sequence that determines the final GOI expression level. An important finding is that the improvements (fold relative to wild-type) of expression resulting from UTR modifications are highly gene-dependent, presumably due to some other limiting factors characteristic of each gene. Nevertheless, within these limitations a small set of UTR mutants appears to work for most genes, simplifying identification of sequences that work close to optimal for any GOI. Our results confirm that the UTR DNA region plays a very complex role in controlling the levels of gene expression, but also that in an applied context efficient screening systems can be used to eliminate the need for a full understanding of the underlying molecular mechanisms, similar to what has been observed in protein engineering. Further research should probably focus on the combination of this knowledge and a better understanding of the role of the gene coding sequences themselves in determining the absolute levels of expression of the final protein product.

Acknowledgements

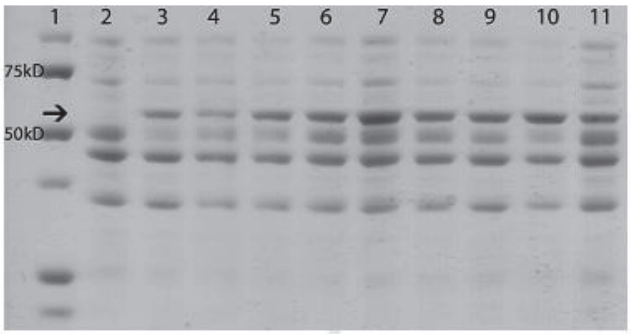
This work was supported by a grant from the Research Council of Norway.

References

Adhin, M.R., van Duin, J., 1990. Scanning model for translational reinitiation in Eubacteria. J. Mol. Biol. 213, 811–818.
André, A., Puca, A., Sansone, F., Brandi, A., Antico, G., Calogero, R.A., 2000. 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.
Bakke, I., Berg, L., Aune, T.E.V., Brautaset, T., Sletta, H., Tøndervik, A., Valla, S., 2009. Random mutagenesis of the *Pm* promoter as a powerful strategy for improvement of recombinant gene expression. Appl. Environ. Microbiol. 75, 2002–2011.
Berg, L., Lale, R., Bakke, I., Burroughs, N., Valla, S., 2009. The expression of recombinant genes in *Escherichia coli* can be strongly stimulated at the transcript production level by mutating the DNA-region corresponding to the 5′-untranslated part of mRNA. Microb. Biotechnol. 2, 379–389.
Blatny, J.M., Brautaset, T., Winther-Larsen, H.C., Haugan, K., Valla, S., 1997a. 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.
Blatny, J.M., Brautaset, T., Winther-Larsen, H.C., Karunakaran, P., Valla, S., 1997b. Improved broad-host-range RK2 vectors useful for high and low regulated gene expression levels in Gram-negative bacteria. Plasmid 38, 35–51.
Brautaset, T., Lale, R., Valla, S., 2009. Positively regulated bacterial expression systems. Minireview. Microb. Biotechnol. 2, 15–30.
Brodolin, K., Zenkin, N., Mustaev, A., Mamaeva, D., Heumann, H., 2004. The σ^{70} subunit of RNA polymerase induces *lac*UV5 promoter-proximal pausing of transcription. Nat. Struct. Mol. Biol. 11, 551–557.
Condon, C., 2007. Maturation and degradation of RNA in bacteria. Curr. Opin. Microbiol. 10, 271–278.
Gu, W., Zhou, T., Wilke, C.O., 2010. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. PLoS Comput. Biol. 6, e1000664.
Hatoum, A., Roberts, J., 2008. Prevalence of RNA polymerase stalling at *Escherichia coli* promoters after open complex formation. Mol. Microbiol. 68, 17–28.
Hsu, L.M., Cobb, I.M., Ozmore, J.R., Khoo, M., Nahm, G., Xia, L., Bao, Y., Ahn, C., 2006. Initial transcribed sequence mutations specifically affect promoter escape properties. Biochemistry 45, 8841–8854.
Kozak, M., 2005. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 361, 13–37.
Kudla, G., Murray, A.W., Tollervey, D., Plotkin, J.B., 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. Science 324, 255–258.
Lale, R., Berg, L., Stüttgen, F., Netzer, R., Stafsnes, M.H., Brautaset, T., Aune, T.E.V., Valla, S., 2011. Continuous control of the flow in biochemical pathways through 5′ untraslated region sequence modifications in mRNA expressed from the broad-host-range promoter *Pm*. Appl. Environ. Microbiol. 77, 2648–2655.
Lewis, D.E.A., Adhya, S., 2004. Axiom of determining transcription start points by RNA polymerase in *Escherichia coli*. Mol. Microbiol. 54, 692–701.
Marqués, S., Manzanera, M., González-Pérez, M., Gallegos, M., Ramos, J.L., 1999. 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.
Nakamoto, T., 2006. A unified view of the initiation of protein synthesis. Biochem. Biophys. Res. Commun. 341, 675–678.
Nickels, B.E., Mukhopadhyay, J., Garrity, S.J., Ebright, R.H., Hochschild, A., 2004. The σ^{70} subunit of RNA polymerase mediates a promoter-proximal pause at the *lac* promoter. Nat. Struct. Mol. Biol. 11, 544–550.
Picard, F., Dressaire, C., Girbal, L., Cocaign-Bousquet, M., 2009. Examination of post-transcriptional regulations in prokaryotes by integrative biology. C. R. Biol. 332, 958–973.
Salis, H., Mirsky, E., Voigt, C., 2009. Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. 27 (10), 946–950.
Shine, J., Dalgarno, L., 1974. The 3′-terminal sequence of *E* coli 16S ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U.S.A. 71, 1342–1346.
Simonetti, A., Marzi, S., Jenner, L., Myasnikov, A., Romby, P., Yusupova, G., Klaholz, B.P., Yusupov, M., 2009. A structural view of translation initiation in bacteria. Cell. Mol. Life Sci. 66, 423–436.
Sletta, H., Nedal, A., Aune, T.E.V., Hellebust, H., Hakvåg, S., Aune, R., 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.
Sletta, H., Tøndervik, A., Hakvåg, S., Aune, T.E.V., Nedal, A., Aune, R., 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.
Walker, K.A., Osuna, R., 2002. Factors affecting start site selection at the *Escherichia coli fis* promoter. J. Bacteriol. 184, 4783–4791.
Winther-Larsen, H.C., Blatny, J.B., Valand, B., Brautaset, T., Valla, S., 2000. Pm promoter expression mutants and their use in broad-host-range RK2 plasmid vectors. Metab. Eng. 2, 92–103.
Zhelyabovskaya, O.B., Berlin, Y.A., Birikh, K.R., 2004. Artificial genetic selection for an efficient translation initiation site for expression of human *RACK1* gene in *Escherichia coli*. Nucleic Acids Res. 32, e52.

Berg, L., V. Kucharova, I. Bakke, S. Valla, and T. Brautaset. 2012. Exploring the 5´-UTR DNA region as a target for optimizing recombinant gene expression from the strong and inducible *Pm* promoter in *Escherichia coli*. Journal of Biotechnology 158:224-230.

Correction to the legend of Figure 3



Correct Legend:

SDS-PAGE analysis of recombinant strains expressing CelB from different UTR variants. Crude extracts were prepared from recombinant cell cultures as described in Materials and Methods and samples containing 10 µg protein were run on SDS-PAGE. Lanes: 1, molecular weight standard; 2, negative control (cells harboring plasmid pIB11); 3, wild-type UTR; 4, UTR variant LII-7; 5, UTR variant LII-8; 6, UTR variant LII-9; 7, UTR variant LII-10; 8, UTR variant LII-11; 9, UTR variant LII-12; 10, UTR variant LV-3; 11, UTR variant LV-4. The position of CelB is indicated with an arrow.

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

Is not included due to copyright