Rahmi Lale

The *xylS/Pm* expression cassette:

New functional insights and application potentials

Thesis for the degree of Philosophiae Doctor

Trondheim, December 2009

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

NTNU
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 978-82-471-1952-5 (printed ver.)
ISBN 978-82-471-1953-2 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2009:266

Printed by NTNU-trykk

Acknowledgements

I am indebted to several people for their contributions to this PhD thesis.

Firstly, to my academic supervisor Svein Valla. I am grateful to him for his assistance and energetic spirit: your intellect and humour have been a constant inspiration. Thanks for the productive and inspiring conversations.

To the current and former members of the MOLGEN Gene Expression Group. A special thank you goes to Friederike, Ingrid, Laila, Nigel, Trond Erik and Trygve.

To the Faculty of Natural Science and Technology for granting the scholarship that I have received for this PhD study.

To the committee members for their efforts.

To the staff of Department of Biotechnology.

To Hanne for proof reading part of the thesis.

To my parents for letting me ramble with my life.

And finally, to Kerstin an uber thanks for providing love and care.

Trondheim, November 2009
Rahmi Lale

Abstract

Positively regulated expression systems are widely used to control gene expression for applied and basic studies in bacteria (**Paper I**). XylS/*Pm*, originating from the *Pseudomonas putida* TOL plasmid is an example of such a system which had previously been coupled with the minimal replicons of the plasmid RK2 to create broad-host-range expression vectors (Blatny *et al.* 1997a). These systems have been used as versatile expression tools which cover a wide range of properties and applications, and the wild type XylS/*Pm* system was proven to be useful for production of recombinant proteins in *Escherichia coli* at industrial levels (Sletta *et al.* 2004; 2007). RK2 replicons are functional in at least 58 Gram-negative (Toukdarian, 2004; Trond Erik Vee Aune, personal communication) and in at least one Gram-positive bacterium, *Clavibacter xyli* (Thomas & Helinski 1989). Both the replicon and the *xylS*/*Pm* expression cassette contributes to the broad-host-range properties of this system (Winther-Larsen *et al.* 2000b; Fu *et al.* 2008). Important characteristics of these vectors include adjustable copy number (between 4 and 100 per chromosome), several different and inexpensive benzoic acid derivatives such as *m*-toluic acid can be used as inducers. They act in a dose-dependent manner apparently not requiring an active uptake system (Lambert & Stratford 1999), further simplifying the use of the system across species barriers.

One of the aims of the work reported in this thesis was to further improve the expression levels from XylS/*Pm* and to gain new insights about the underlying mechanisms involved in determination of the levels of expression. For these studies *E. coli* was used as the main model system. The 5′-UTR region was the first DNA regulatory element to be studied. Large mutant libraries covering this region were created (using doped synthetic oligonucleotide mixtures) and screened with respect to expression levels, using *bla* (encoding β-lactamase) as reporter gene. This screening system has the important advantage that it allows very strong selection for enhanced expression levels by simply plating cells from the library on agar media supplied with a gradient of increasing levels of ampicillin (Winther-Larsen *et al.* 2000a). As an outcome of this work mutations in the 5′-UTR were found to represent an efficient tool for enhancing gene expression (up to over 20-fold compared to the wild type). Interestingly, some of the mutations were found to act by stimulating transcription, whereas the 5′-UTR has previously been considered to

be mainly involved in control of mRNA stability and translation. Further inspection on the variant 5′-UTR sequences revealed that certain nucleotides were repeatedly substituted, such as the A to C transversion at position +10 (11 out of 14) relative to the transcriptional start site (**Paper II**).

To investigate further the contribution of each base position, this region was also studied for identifying low-level expression variants. Mutant libraries covering the 5′-UTR were constructed with the same set-ups as described above but were, this time, screened to identify low-level expression variants. Sequencing of candidates revealed that they were not exhibiting any obvious positional hot-spots, but variants could be identifed which displayed a strongly reduced background expression in the absence of the inducer. Interestingly, these variants had not lost their inducibility properties, although the maximum level of induction was reduced relative to the wild type. Such variants will have a potential use in studies such as metabolic engineering and expression of toxic genes (**Paper III**). In a collaborative project the variant 5′-UTR sequences are currently being analysed by bionformatic tools, and the preliminary results indicate that generating *in silico* 5′-UTR sequences with predicted and desired properties may become possible (Skancke, unpublished).

In parallel to the work on the 5′-UTR, other conceptually similar activities have resulted in identification of many up-regulated variants of the *Pm* promoter and the positive regulator XylS sequences. Similar to the 5′-UTR studies it was found that the expression from *Pm* can be strongly stimulated by introducing mutations in the promoter region (Bakke *et al.* 2009) or in the XylS regulator coding sequence (Aune *et al.* 2009). Interestingly, when several variants from each element (XylS, *Pm* and 5′-UTR) were combined in a single construct, much stronger stimulation (both at the transcriptional and translational levels) was observed than that of the each variant element separately (**Paper IV**).

An important conclusion from the work presented in this thesis is, therefore, that expression levels can be improved by mutating the regulatory elements, even in a case where the wild type version of the cassette can express proteins at industrial levels. These findings are likely to be relevant also for other well established expression cassettes.

List of Publications

Publications included as part of the thesis

Publications listed will be referred by their Roman numerals in the text.

Paper I. Brautaset, T., **Lale, R.** and Valla, S. (2009). Positively regulated bacterial expression systems. *Microbial Biotechnol.* **2**, 15-30.

Paper II. Berg, L., **Lale, R.**, Bakke, I., Burroughs, N. and 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 Biotechnol.* **2**, 379-389.

Paper III. Lale, R., Berg, L., Stüttgen, F., Aune, T.E.V. and Valla, S. (2009). Fine-tuning of gene expression in *Escherichia coli* over five log-factors by using mutations in the 5′-UTR DNA region. *Manuscript*.

Paper IV. Lale, R., Stüttgen, F., and Valla, S. High-level heterologous protein expression in *Escherichia coli* by combining mutations in three different regulatory elements. *In prep*.

Contributions to related publications not included in the thesis

Aune, T.E.V., Bakke, I., Drabløs, F., **Lale, R.**, Brautaset, T. and Valla, S. Directed Evolution of the Transcription Factor XylS for Development of Improved Expression Systems. *Microbial Biotechnol.* **In press**.

UK Patent Application 'Enhanced Expression From The *Pm* Promoter' Valla, S. **Lale, R.**, Berg, L., Bakke, I., Aune, T.E.V. Pub. No.: WO/2008/075057, Int. Appl. No.: PCT/ GB2007/004903, Int. Filing Date: 20.12.2007

UK Patent Application 'Enhanced Expression Method' Valla, S. **Lale, R.**, Berg, L., Bakke, I. Pub. No.: WO/2008/015447, Int. Appl. No.: PCT/GB2007/002951, Int. Filing Date: 03.08.2007

Nomenclature

aa	amino acid
ATP	adenosine triphosphate
A-site	amino-tRNA acceptor site
bp	basepair
DNA	deoxyribonucleic acid
EF-G	elongation factor G
EF-Tu	elongation factor Tu
E-site	transfer RNA exit site
GDP	guanosine diphosphate
GTP	guanosine triphosphate
GTPase	guanosine triphosphatase
IF- 1, 2, 3	initiation factors
IPTG	isopropyl thio-ß-D-galactoside
mRNA	messenger RNA
nt	nucleotide
PCR	polymerase chain reaction
P-site	peptidyl-tRNA acceptor site
qRT-PCR	relative quantification real-time RT-PCR
RBS	ribosome binding site
RF- 1, 2, 3	release factors
RNA	ribonucleic acid
RNAP	RNA polymerase
rRNA	ribosome RNA
RT-PCR	reverse transcriptase PCR
sRNA	small non-coding RNA
TnSS	translational start site
tRNA	transfer RNA
$tRNA^{fMet}$	Met formylated initiator tRNA
TrSS	transcriptional start site
Ts	temperature sensitive
UTR	untranslated region

σ	sigma
α	alpha
β	beta

Contents

1

Introduction

1.1 Historical Background

The realisation of DNA as the genetic material in the 1940s and the development of cornerstone technologies in the 1970s have shaped the route to the emergence of the era of today's molecular biology. In the 1950s, proteins and the relationship between RNA and protein synthesis were known, but the key elements of information flow were still to be identified. In those 'germinal years' biological science was focused on two main problems: i) the nature of the gene and the processing of genetic information and ii) the mechanism of protein synthesis (Hoagland 2003).

The experimental demonstration proving DNA as the material of genes was achieved by Avery, McLeod and McCarty in 1944 and by Hershey and Chase in 1952. Their findings were proximal but not as complete as it was stated by Watson and Crick in 1953. In contrast to the independently suggested alternative models by Pauling (Pauling & Corey 1953), and Wilkins and Franklin (Crick 1988), Watson and Crick presented the structure of DNA as to be formed by two anti-parallel oriented deoxyribonucleic acid strands. Later Crick presented the 'sequence hypothesis' and the 'central dogma' revealing the relationship between proteins and DNA (Crick 1958). According to the hypothesis the specificity of the nucleic acids is solely dependent on their base sequences, which in turn determine the amino acid sequence of proteins.

Pauling and Corey (1950) proposed the alpha helix and beta sheet as the primary structural motifs in protein secondary structure. Insulin was the first protein to be sequenced, accomplished by Frederick Sanger. The primary structure provided an experimental basis for the hypothesis that protein sequences were essentially non-redundant and highly specific (Sanger 1952). Primary sequence determination of insulin was followed by solving the three-dimensional structures of myoglobin and hemoglobin by X-ray diffraction analysis by Sir John Cowdery Kendrew (1958) and by Max Perutz (1960), respectively.

Despite these developments two decades passed before yet another breakthrough revolutionised the field of molecular biology. In 1972, Jackson, Symons and Berg announced the first construction of 'a covalently closed-circular DNA molecule', a recombinant plasmid (Jackson *et al.* 1972). It consisted of DNA molecules from Simian Virus 40, lambda phage genes, and the galactose operon of *Escherichia coli*. In 1973, Cohen and colleagues presented the first genetically engineered organism: *E. coli* transformed with a plasmid conferring antibiotic resistance (Cohen *et al.* 1973). Following this advancement the first genetic engineering company, Genentech, was founded in 1976. One of the first economic successes based on utilising genetic engineering was the production of insulin. Scientists at Genentech and City of Hope inserted the genes carrying the genetic code for human insulin, along with the necessary control mechanism, into an *E. coli* strain (Crea *et al.* 1978). Clinical studies of recombinant insulin, humulin, began in 1980 and in 1982 it became the first genetically engineered drug approved by the U.S. Food & Drug Administration (FDA).

Another heterologous expression success story was the production of human interferon. It was discovered back in 1957 and was known to be produced by the human body in response to viral attacks (Isaacs & Lindenmann 1957). Despite its potential needs it could not be produced in large amounts. The human body produces this compound in such small quantities that it would take 90,000 donors to provide only one gram of interferon, and in addition the purity would be about one percent. Through genetic engineering, Swiss researchers introduced the gene for human interferon into *E. coli* (Nagata *et al.* 1980), and they were able to produce pure interferon at about one U.S. dollar per dose. Recombinant technology enables production high value proteins useful in research, therapy and diagnostics in a cost-effective manner. This is a very important aspect in recombinant protein production as many of the most interesting proteins are naturally available in very small quantities. Recombinant methods also allowed improvements in purity.

Over the years the field of molecular biology has expanded from the recombinant expression of insulin to genome sequencing, and nowadays it is moving in the direction of synthetic biology. From the recombinant expression point of view, tailoring an expression platform for specific purposes would be a great advancement. With these rapid developments it has become increasingly important to have a firm and detailed understanding of the basics of gene expression. This study aims to contribute to the understanding of the regulations involved in gene expression both from the research and applied perspective.

1.2 Expression Platforms for Recombinant Protein Production

A plethora of expression platforms have been developed, ranging from bacteria, yeasts and filamentous fungi to cells of higher eukaryotes. All known production platforms have favourable characteristics as well as drawbacks and limitations. There is no simple system that can serve as an optimal platform for all types of recombinant protein production. Due to historical developments, ease of genetic manipulations and easy access to process monitoring and validation microbial expression platforms are widely used. As it was stated previously, *E. coli* was the first organism to be employed for recombinant protein production. Single-subunit proteins can easily be produced in bacterial hosts, however the lack of glycosylation and limitations in secretion impose restrictions on the range of applications. Despite these drawbacks many more microbial platforms have been developed. The Gram-negative bacterium *Pseudomonas fluorescens* has for example been employed for antibiotic production due to its improved secretion capabilities (Chew *et al.* 2005) and the Gram-positive *Staphylococcus carnosus* is a good candidate to avoid proteolytic degradation, which is a problem for the widely used *Bacillus subtilis* (Gelissen *et al.* 2005). Members of the Gram-positive Actinomycetes (e.g. *Amycolatopsis*, *Nocardia*, and *Streptomyces*) are the producers of two thirds of all known bioactive compounds including antibacterials, antifungals, antiparasites and antivirals. They have also been found to synthesise compounds that can be used as immunosuppressants, and in anti-cancer (chemotherapeutic) treatment (Salas *et al.* 2001). In addition to drugs, *Streptomyces* enzymes like xylose isomerase, are also widely used in the food industry (Naeimpoor & Mavituna 2001).

For the production of proteins that require authentic and complex mammalian glycosylation or the presence of several disulfide bonds, higher eukaryote platforms are needed. These systems are also suitable for production of proteins that require multiple post-translational modifications. The baker's yeast *Saccharomyces cerevisiae* has been used for production of surface antigen of the Hepatitis-B-Virus (HBsAG) (Bitter & Egan 1984; Hadiji-Abbes *et al.* 2009) as well as insulin (Tøttrup & Carlsen 1990), both products are FDA-approved. However, also *Saccharomyces cerevisiae* can have limitations due to over-glycosylation of glycoproteins, which may cause allergic reactions (Young *et al.* 1998). *Kluveromyces lactis*, *Pichia pastoris*, *Hanensula polymorpha* are amongst the most commonly used yeast species (Gelissen *et al.* 2005).

Fungi exhibit combined features of microbial and eukaryotic systems, they are simple to ferment and simultaneously have secretion abilities comparable to general eukaryotic schemes. Filamentous fungi of the genus *Aspergillus* are widely used, and several food additives produced by these organism have obtained 'generally recognised as safe' (GRAS) status. One of the drawbacks of this systems is that production of spores is undesirable in production of pharmaceuticals (Heerikhuisen

et al. 2005).

Recombinant baculoviruses are platforms that are widely used to express heterologous genes in insect cells (Philipps *et al.* 2005). Most commonly used baculoviruses are *Autographa californica* nuclear polyhedrosis virus (AcNPV) and the *Bombyx mori* nuclear polyhedrosis virus (BmNPV) (Luckow 1993). AcNPV has been shown to be taken up by human hepatocytes and since then it has been heavily employed for protein expression (Hofmann *et al.* 1995). Several baculovirus platforms have been demonstrated to transduce a wide range of mammalian cell types *in vitro* and *in vivo* (Kost & Condreay 2002), and improvements of efficiency by modification of transduction parameters or the use of different promoters have been described (Wang *et al.* 2006).

Mammalian cell lines are another platform used especially for production of therapeutic proteins and antibodies. Chinese hamster ovary (CHO), baby hamster kidney (BHK), mouse myelome (NSO) and myeloma (Sp2/0) cells, are among the most widely used mammalian cell lines (Gelissen *et al.* 2005). Also human cell lines were created *in vitro* from primary cells, such as human embryonic kidney cells (HEK293) or human retinoblast derived (PER.C6) cells (Sandig *et al.* 2005). They are extensively used due to their high DNA uptake rates and capabilities of reaching extremely high cell densities (10^7 cells mL^{-1}), allowing accumulation of recombinant products to high titres. With the above mentioned human cell lines authentic glycosylation can be achieved. Disadvantages of mammalian systems are represented by the requirement for specialised expensive media, cells are extremely prone to virus contaminations and fermentation of human cell lines with high cell densities creates technical fermentation obstacles (Sandig *et al.* 2005). Despite these challenges, the expected market share is increasing dramatically. For 2004, the estimated production of mammalian cell-based therapeutic protein production was stated to be 59% followed by 27% for *E. coli*-based products (Sandig *et al.* 2005).

The last platform to be mentioned are plant-based systems. These systems exhibit several advantages in that they are inexpensive to maintain, the nutrient source is cheap and plants can be cultured under precise conditions in green houses. One striking feature is that they have the ability to synthesise, fold and assemble complex multi-subunit proteins and glycoproteins and they do not contain any endotoxins that are present in many bacteria. Restrictions are imposed by long development times and inconsistent product yields due to fertilisers, pests, diseases, weather, soil compositions, and interactions with other organisms. Especially environmental issues, for example the use of transgenic crops in open fields, are aspects inhibiting a wider use of these systems (Fischer *et al.* 2005).

Today many life forms have been used both by research and industry ranging from bacteria to mammalian cell lines. Examples include genetically modified corns, metabolically engineered strain of *E. coli* used for production of 1,3-propanediol (used as a building block in the production of polymers) and the

fungus *Ashbya gossypii* employed as a biocatalyst for production of more than 1000 tones of vitamin B2 per year (Frazzetto 2003). With recent technological developments revenues for this industry were reported to reach $2-billion during 2006-07, up from $1 billion in 2004-05. The industry is now aiming to reach revenues of $5 billion before 2010. The McKinsey consultancy predicts that by 2010 biotechnology could be applied in the production of between 10% and 20% of all chemicals sold (amounting to a value of $160 billion) and that up to 60% of all fine chemicals (medium-volume products used as intermediates in the manufacturing of products such as pharmaceuticals, flavours, fragrances, agro-chemicals and detergents) might be produced using advanced recombinant technology (Lorenz & Eck 2005).

Considering the increasing interest in recombinant gene expression by both research and industry, characterising new expression platforms and optimising existing systems will prove to be important. Thus, expanding the knowledge on basic mechanisms will be essential for understanding natural gene expression processes.

In this thesis two microbial organisms were used, *E. coli* and *P. fluorescens*. The following subsections describe their characteristics.

1.2.1 Escherichia coli as a host

The bacterium known as *E. coli* today was discovered by a German paediatrician and bacteriologist Theodor Escherich (1857-1911), and was named *Bacterium coli* by then. In 1919 the name of this organism was changed to *E. coli*. *E. coli* is regarded the most widely used prokaryotic organism for recombinant protein production (Altenbuchner & Mattes 2005). The 4,639,221 bp genome of *E. coli* K12 has been sequenced and 4288 protein-coding genes have been annotated (Blattner *et al.* 1997).

Different classes of proteins, ranging from full-length bacterial and human proteins to protein complexes, to even some human integral membrane proteins, can be produced in *E. coli*. According to TargetDB (a target registration database for structural genomics projects, http://targetdb.pdb.org/) various proteins can be expressed in soluble form in *E. coli*: up to 50% of proteins from prokaryotes (including archaea) and 10% of proteins from eukaryotes (Braun & LaBaer 2003). A limiting factor for successfully expressing a soluble protein is represented by its molecular weight, with expression decreasing considerably at molecular weights above ~60 kDa (Consortium *et al.* 2008).

The production of recombinant proteins in microbial cells represents a convenient and cost-effective method. *E. coli* has become the most extensively used bacterial host. It is a genetically and physiologically well-characterised organism that grows fast and performs well in simple media. Many regulatable expression systems have been developed for *E. coli*, and new ones are under development. These systems can be categorised based on the mode of activation: positively

regulated, negatively regulated or systems that encompasses both positive and negative regulation. The best known example of negative regulation in *E. coli* is the *lac* operon, and its regulatory elements are widely used in pET vectors (see subsection 2.4.3). Another frequently used negatively regulated system is the p_L promoter of phage λ. This promoter is very tightly regulated by the thermosensitive λ cI repressor variant and can be induced by increasing the growth temperature. Another commonly used system is the *trp* promoter derived from the *E. coli* tryptophan operon. The tryptophan repressor binds to the *trp* operator in presence of tryptophan and induction is achieved in the absence of tryptophan.

Based on the *trp* promoter other systems were developed with the aim to create stronger hybrid promoters, namely the *tac* and *trc* promoters. The *tac* promoter carries the consensus –35 sequence from the *trp* promoter and the consensus –10 sequence from the *lacUV5* promoter. This promoter has a spacer length of 16 nt between the –10 and –35 regions. *trc* is identical to *trp* except for a spacer length of 17 nt (Brosius *et al.* 1985). The main drawback of these systems is their leakiness. In general, negatively regulated systems are regarded as difficult to use if a stringent control of expression is desired. Once the system is induced the production fires and this often causes formation of inclusion bodies and a burden to the host metabolism (Wilms *et al.* 2001; Altenbuchner & Mattes 2005).

Examples of positively regulatable systems are listed in **Paper I**, Table 1. Positively regulated systems can confer certain advantages compared to those that are negatively regulated (Wilms *et al.* 2001, and references therein), particularly when very tight control (i.e. low background expression under non-induced conditions) is needed (Altenbuchner & Mattes 2005). Such properties are important for expression of host-toxic proteins, when running the production process under HCDC and also when the system is used as a tool for physiological studies.

Major drawbacks of the use of *E. coli* as host for recombinant protein production include vector instability, inefficient translation initiation and elongation, instability of mRNA, toxicity of gene products, codon usage, inappropriate protein folding with resulting inactive protein products and formation of inclusion bodies (Hannig & Makrides 1998). Despite these difficulties, Consortium *et al.* (2008) have listed a 'consensus' strategy which can serve as a guideline for soluble protein expression and purification. According to these suggestions, *E. coli* should be the first host choice for recombinant production of any protein regardless of the source and BL21 (DE3) with the bacteriophage T7 promoter based system are recommended as a strain and expression platform, respectively.

Strategies for improving expression systems in *E. coli* are based on increasing the stability of vectors, co-expression of rare tRNAs, tight regulation of promoters and construction of new host strains for special purposes, allowing e.g. disulfide bond formation in the cytoplasm or co-expression of chaperones to improve proper protein folding. Important developments have also been made in expressing glycoproteins in *E. coli*. In 2004, the successful expression of myoglobin containing beta-

N-acetylglucosamine (GlcNAc)-serine at a defined position was reported (Zhang *et al.* 2004). Furthermore, the *de novo* production of heterologous polyketides and nonribosomal peptides could successfully be established in *E. coli* (Watanabe & Oikawa 2007). The introduction of these highly complex pathways into *E. coli* was achieved by using 16 monocistronic gene cassettes, each containing a biosynthetic, self-resistance or post-translational modification gene with an upstream T7 promoter and ribosome binding site, a downstream T7 transcription terminator, and compatible flanking SpeI and XhoI restriction sites on three mutually compatible plasmids. These examples clearly demonstrate that *E. coli* can still serve as a suitable platform for heterologous protein expression in the future.

1.2.2 Pseudomonas fluorescens as a host

Strains of *Pseudomonas fluorescens* are commonly found on plant surfaces, as well as in decaying vegetation, soil and water. They are aerobic, saprophytic, and Gram-negative bacteria. The American Type Culture Collection lists *P. fluorescens* under Biosafety Level-1, which means that these bacteria are not known to cause diseases in humans or in animals (Chew *et al.* 2005).

Sequencing of the *P. fluorescens* SWB25 genome has been completed and the 6,722,540 bp sizing genome has an average G+C content of 60.5%. Shotgun sequencing data is available at the Sanger Institute's web site comprising at present 126,206 reads accounting for 57,041 Mb and thus representing a theoretical genome coverage of 99.98%. The annotation of the genome sequence is still ongoing (Sanger-Institute 2008) while three additional *P. fluorescens* genome sequences (of strains Pf-5, PfO-1, and MB214) have been completed (Paulsen *et al.* 2005).

P. fluorescens has been reported to handle high-level expression and to be suitable for fermentation under a wide range of conditions. When this bacterium was introduced for recombinant expression, its expression capabilities were compared to those from *E. coli* carrying the T7 expression system. While the production of human growth hormone (rhGH) in *E. coli* leads to formation of inclusion bodies, expression of this hormone in a *P. fluorescens* strain at a 20 L fermentation scale resulted in 1.6-fold more production of rhGH per gram dry biomass. Another impressive example was the comparison between *E. coli* and *P. fluorescens* regarding production of human gamma interferon (γ-IFN). Fermentation of *E. coli* with the T7 system at 20 L scale resulted in formation of 2-4 gL^{-1} of insoluble protein, whereas the *P. fluorescens* expression system produced $\sim$4 gL^{-1} of mostly soluble (>95%) protein. Comparative studies like these are available for several strains of *P. fluorescens*, most of them for the industrial strain Pf-MB101 (Chew *et al.* 2005).

Requirements for a rapid and efficient microbial producer make *P. fluorescens* a suitable candidate with advantageous characteristics include: i) its suitability for

high cell density fermentations without oxygen enrichment in a mineral salt medium without animal products, ii) a protein production processes regarded as simple, predictable and scalable and iii) provision of high volumetric yields of soluble and active target proteins.

1.3 Gene Expression in Escherichia coli

For most of the presented work, the microorganism *E. coli* was used and the following sections contain relevant information for bacterial regulation based on findings in *E. coli*.

1.3.1 The transcriptional process

Transcription is the primary common regulatory step in gene expression in all organisms ranging from higher eukaryotes to single cells. Transcriptional regulation is involved in many vital processes such as cellular metabolism and organ and tissue development. Development of tumours and cancers can also depend on changes in transcription (Greive & von Hippel 2005). In bacteria, a single RNAP transcribes DNA, unlike in eukaryotic systems, which harbour three distinct RNAP.

Transcription is a multi-phase process in bacterial gene regulation. It can be roughly divided into three major steps: the recognition of the promoter site by the RNAP complex and RNA chain *initiation*, processive chain *elongation* and interruption of elongation by *termination*. Before briefly presenting these steps, the elements involved in transcription will be addressed in the following subsections.

RNA Polymerase. *E. coli* RNAP consists of five subunits, $\alpha_2\beta\beta'\omega$, that make up the core enzyme (Fig. 1.1).

Core RNAP can synthesise RNA, but it must associate with an accessory σ subunit to form the RNAP holoenzyme which then binds to specific DNA sequences located at promoters. RNAP holoenzyme containing σ^{70}, the predominant σ factor in *E. coli*, binds to promoters of housekeeping genes. In addition to σ^{70}, *E. coli* also encodes six 'alternative' σ factors that allow the RNAP holoenzyme to associate with smaller subsets of promoters (see subsection *Sigma factors*) (Wade *et al.* 2006). Each alternative σ factor is required for expression of certain genes in response to a specific environmental stimulus. Most bacterial genomes encode multiple σ factors that are required for complex cellular processes such as stress response, morphogenesis and virulence. Alternative σ factors generally recognise different promoter sequences from the housekeeping σ factor. RNAP synthesises RNA in a processive manner at a rapid speed ranging from 20-80 nt s^{-1} up to 200 nt s^{-1} (Pan & Sosnick 2006). This feature is stated to be specific for RNAP. While other processive molecular motors such as DNAP, kinesin and myosin-V operate in a uniform speed, RNAP is known to display long and short pauses (Artsimovitch

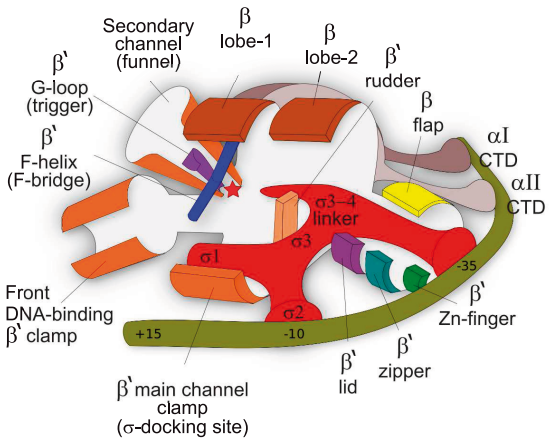

Figure 1.1: Schematic representation of RNAP. Simplified cartoon depicts the mobile elements and subunits of the RNAP, red start indicates the catalytic centre. Double stranded DNA is shown as flexible green cylinders. Adapted from Borukhov & Nudler (2008).

& Landick 2000). Several reasons have been reported for this pausing: stem-loop structures in the template that interact with RNAP; backtracking of the RNAP due to formation of weak DNA/RNA hybrids and post-translational modifications (Berghöfer-Hochheimer *et al.* 2005).

Promoters. Promoters are DNA sequences at the 5′ end of the coding sequence of a gene that direct RNAP binding and transcription initiation. Most of the bacterial promoters bear two conserved hexamers located around 35 and 10 nt upstream of the transcription start site at position +1 and separated by a spacer of 15-21 bp. Traditionally the optimum length of these spacer sequences is thought to be 17±1, but initiation over a range of 15-20 bases has been shown (Dombroski *et al.* 1996). *E. coli* has many classes of promoters and all are recognised by different σ factors, which are the exchangeable subunits of RNAP holoenzyme. Strong promoters of the σ^{70} type harbour the consensus sequences 5′-TTGACA-3′ and 5′-TATAAT-3′ in the regions –35 and –10, respectively. It has also been reported that recognition of the –10 region alone can be sufficient for initiation to occur, since a number of effector-independent promoters do not require −35 hexamer contacts for transcription initiation (Barne *et al.* 1997). Initiation of transcription at these promoters was found to require a so called 'extended' –10 element upstream of the region containing the consensus sequence 5′-TGnTATAAT-3′. This has been suggested to create alternative contact points for RNAP, most likely via the σ subunit (Barne *et al.* 1997).

It has been speculated that promoters play major roles in all steps of transcription. During growth, a cell faces continuously changing environments and needs

to respond accordingly to maintain cell function. Under such circumstances gene products are needed in different amounts depending on the conditions. To maintain a tight transcriptional control it is known that some promoters are recognised by different σ factors and promoter sequences can differ considerably from what is known as the consensus sequences. In fact according to one compilation study only a few of the *E. coli* σ^{70}-promoters fully match the consensus sequence in the –35 and –10 positions (Lisser & Margalit 1993). On average *E. coli* promoters preserve only 8 of the 12 canonical bases of the hexamers. To add more complexity to the topic, a third element an AT-rich region located upstream of the –35 promoter region was identified (Ross *et al.* 1993). This region was found to be recognised by the α-subunit of the RNAP. The so called UP element consists of a 11 nt distal region ranging from –57 to –47 and a 4 nt proximal region ranging from –44 to –41, and has been shown to increase promoter strength more than 300-fold (Estrem *et al.* 1998). Not all promoters are thought to contain a UP element, but one observation indicated that even if no UP elements can be identified at the mentioned positions, most promoters bear an AT rich sequence upstream of the –35 promoter region (Ozoline *et al.* 1997).

In addition to regulation of expression by σ factor affinity, promoters can be controlled by repressor and activator proteins. These regulatory proteins (also known as transcription factors, TFs) can be classified according to sequence similarity of their DNA binding motifs or by alignments of their amino acid sequences. Two large families of regulators have been identified for bacterial promoters, namely positive and negative regulators. Currently, about 50 families of bacterial TFs are reported (Rodionov 2007), and the largest of these families is represented by LysR (consisting mainly of negative regulators), followed by AraC-XylS (consisting mainly of positive regulators) (see www.bactregulators.org, **Paper I**). A common feature shared by many TFs is that they bind their ligands with high specificity and that this binding determines their activation state. TFs can be subjected to *in vitro* evolution strategies in order to change their ligand specificities (Galvão & de Lorenzo 2006; Aune *et al.* 2009). Interestingly, TFs have also been described as molecular targets for drugs against pathogenic bacteria (Bowser *et al.* 2007).

Sigma factors. σ factors are crucial as core RNAP cannot initiate transcription at promoters unless it is guided by a σ subunit. The σ subunit specifically binds to the –10 element located in the non-template DNA strand (Roberts & Roberts 1996).

As stated above σ^{70} is considered to be primarily responsible for transcription of most of the 'housekeeping' genes in a cell during exponential growth. Alternative σ factors are thought to be specialised for activation of expression under conditions such as stress, growth transitions, and morphological changes (Missiakas & Raina 1998). A total a pool of seven different σ factors (σ^{70}, σ^{N} (σ^{54}),

σ^S (σ^{38}), σ^H (σ^{32}), σ^F (σ^{28}), σ^E (σ^{24}), and σ^{FecI} (σ^{19})) have been found in *E. coli* and about 350 transcription factors (Pérez-Rueda & Collado-Vides 2000) to fine-tune transcriptional outputs. The collection of transcription factors comprise 35% activators, 43% repressors and 22% dual regulators. The affinity of σ factors for their DNA targets is crucial for their functions. Tight binding of σ to DNA can inhibit transcription: a perfect recognition sequence might lock RNAP in abortive cycling at the promoter or induce RNAP pausing when located downstream of the promoter (Artsimovitch 2008). Both types of inhibitory events have been reported (Hatoum & Roberts 2008) but their ubiquity *in vivo* has not been assessed.

The optimum length of the 17±1 spacing between the two hexamers in promoters was found to cause the two hexamer regions facing to the same side, ∼2 helical twists of B-form DNA. This was shown based on the crystal structure of a bacterial RNAP holoenzyme from *Thermus thermophilus* (Vassylyev *et al.* 2002). It was suggested that the RNAP has a DNA-structure-dependent contact which also demonstrates that the σ factor is sensitive to genomic structure.

According to the traditional σ cycle view the σ-subunit is thought to dissociate from elongating RNAP shortly after RNAP leaves the promoter. However, several reports exist describing a population of RNAP from *E. coli* that has been found to retain the contact with the σ^{70} throughout elongation for multiple rounds (Bar-Nahum & Nudler 2001; Mukhopadhyay *et al.* 2001; Mooney *et al.* 2005). These findings demonstrate that in addition to transcription initiation σ also involves in other steps of the transcription cycle.

Initiation

The first step of the multi-phase transcriptional process involves the recognition of the promoter sequence by RNAP holoenzyme. The holoenzyme slides along the DNA (Sakata-Sogawa & Shimamoto 2004) and locates promoter sequences. In this process the σ-subunit directs the core enzyme to a subset of σ-specific promoters. Once the holoenzyme-promoter complex has been formed, it causes a series of conformational changes that unwinds the DNA between base pairs −11 and +1 (Leibman & Hochschild 2007). This form, the so called open complex, exposes the transcription start site for base pairing with the incoming substrates. In the open complex, RNAP covers the region from −50 and +20 (Mooney *et al.* 1998). In subsequent steps the holoenzyme initiates transcription, typically yielding a series of short RNA transcripts of 9 to 11 nt, known as *abortive initiation* products (Carpousis & Gralla 1985; Hsu 2002; Hsu *et al.* 2006; Goldman *et al.* 2009). Eventually, after a number of abortive cycles, RNAP looses its interaction with the promoter sequence (*promoter escape*) and enters into processive synthesis of RNA as an RNAP-DNA elongation complex.

Two important regulation steps occur once the RNA chain synthesis has been initiated: i) promoter escape (as evidenced by abortive initiation) and ii) RNA

pausing due to the affinity of σ for its DNA target (Hatoum & Roberts 2008).

Abortive initiation remains to be a controversial mechanism and there are three models proposed for RNAP active-centre translocation: *transient excursions*, *inchworming*, and *scrunching* (Kapanidis *et al.* 2006).

> The first model, termed *transient excursions* (Fig. 1.2a), suggests transient cycles of forward and reverse translocation of RNAP. According to this model, in each cycle of abortive initiation, RNAP translocates forward as a unit, translocating 1 bp per phosphodiester bond formed (as in elongation); on release of the abortive RNA, RNAP reverse-translocates as a unit, regenerating the initial state. According to this model, the cycles of forward and reverse translocation are so short in duration and so infrequent in occurrence that, although they occur, they are not detected by a time-averaged, population-averaged method such as DNA-footprinting.
>
> The second model, termed *inchworming* (Fig. 1.2b), proposes a flexible element in RNAP. According to this model, in each cycle of abortive initiation, a module of RNAP containing the active centre detaches from the remainder of RNAP and translocates downstream, 1 bp per phosphodiester bond formed on release of the abortive RNA, this module of RNAP retracts, regenerating the initial state.
>
> The third model, termed *scrunching* (Fig. 1.2c), advocates a flexible element in DNA. According to this model, in each cycle of abortive initiation, RNAP pulls downstream DNA into itself, pulling in 1 bp per phosphodiester bond formed and accommodating the accumulated DNA as single-stranded bulges within the unwound region; on release of the abortive RNA, RNAP extrudes the accumulated DNA, which regenerates the initial state.

According to Kapanidis *et al.* (2006) these three models are not necessarily mutually exclusive. Combinations of all three mechanisms or different mechanisms at different stages of initial synthesis may also be used (e.g., one for synthesis of short RNA products, and another for synthesis of longer RNA products).

Abortive initiation is stated to be a critical determinant for promoter strength but up to now it was only observed under *in vitro* conditions (McClure & Cech 1978; Gralla *et al.* 1980). However, a recently published paper exhibited evidences that abortive initiation takes place also *in vivo* (Goldman *et al.* 2009). The authors claim that the generated short transcripts may play a functional role in gene regulation. An abortive transcript generated from a first promoter could serve as a primer for transcription initiation at a second promoter or could function as an antisense effector against a specific RNA.

As noted above, the rate of transcription by RNAP is not constant throughout transcription. This is generally considered to be due to RNAP's pausing feature. Pausing is thought to allow RNAP to synchronise events with the binding of the

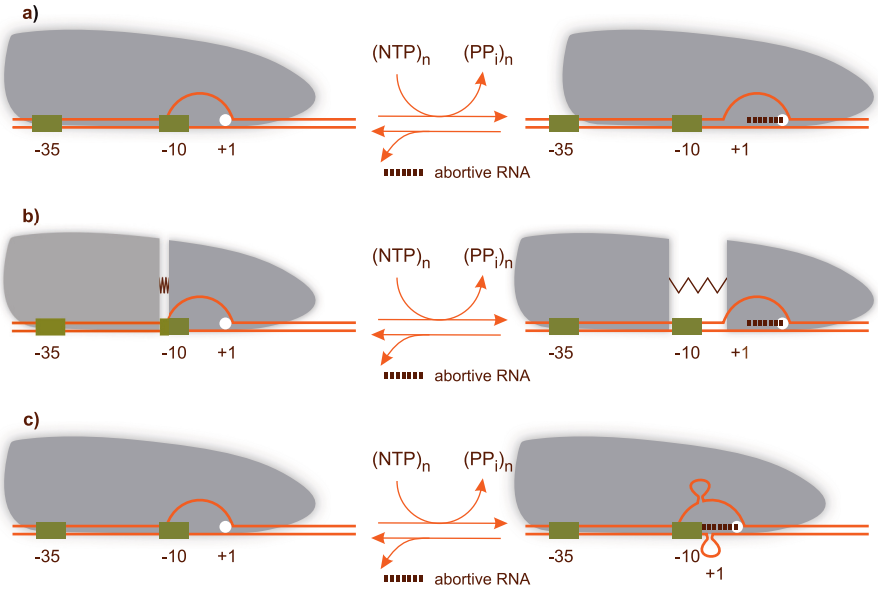

Figure 1.2: RNAP active-center translocation models. The ellipse represents RNAP, the dashed lines represents the abortive RNA, and the white circles represents the RNAP active centre. Promoter regions -35 and -10, and transcriptional start $+1$ are depicted. Adapted from Kapanidis *et al.* (2006).

regulatory factors and/or RNA folding. So far three general classes of RNAP pause sites have been described: promoter proximal, hairpin dependent, and hairpin independent (Yakhnin *et al.* 2006, and references therein). The details of these pauses will be addressed briefly in the following subsection.

Elongation

Traditionally it has been described that after successful initiation (*promoter escape*) the σ factor is released from the core enzyme and RNAP then will continue to elongate the RNA transcript. The first nucleotide in the RNA transcript is always a purine, either pppG or pppA. Synthesis of the RNA occurs in the $5' \rightarrow 3'$ direction, using the four ribonucleoside 5′-triphosphates (ATP, CTP, GTP, UTP) as precursors. The $3'-OH$ at the end of the growing RNA chain forms a $3'5'$ phosphodiester bond with α phosphate group of the incoming ribonucleoside 5′-triphosphate.

In *E. coli* the elongation complex consists of RNAP, the template DNA, and the nascent RNA that is being synthesised. The complex covers ∼35 nt of DNA sequence which protects the DNA from cleavage by nucleases (∼35 nt) or free radicals (∼25 nt). For this complex to proceed, further DNA needs to be unwound forming the so-called transcription bubble. After passage of the transcription

complex, DNA rewinds and forms again a double stranded molecule (Borukhov & Nudler 2008; Roberts *et al.* 2008).

Unlike DNA replication, synthesis of RNA can involve multiple RNAPs on a single DNA template and multiple rounds of transcription, so that several RNA molecules can be produced from a single copy of a gene (Epshtein & Nudler 2003). This step also involves a proofreading mechanism that can replace incorrectly incorporated bases (Park *et al.* 2002).

RNAP can adapt multiple conformational states as the elongation process involves tight regulations (Erie 2002). According to Erie (2002) the RNAP elongation complex has two states, activated and inactivated. RNAP is competent for synthesis in both states but it is faster in the activated state. The elongation complex undergoes a conformational change to the inactivated state in several instances like at pause or termination sites or at positions where RNAP is halted due to deprivation of NTPs or physical roadblocks.

As stated previously, three classes of pausing have been described: promoter proximal, hairpin dependent, and hairpin independent pausing. Promoter proximal pausing has been shown to occur before dissociation of the σ^{70} from the elongation complex *in vivo* (Roberts *et al.* 2008). σ^{70} recognises the near repeat of the -10 component which causes a pause in the order of seconds. In such cases antitermination takes place by the help of antiterminator bacteriophage λ Q protein (Ring *et al.* 1996; Roberts *et al.* 2008). Hairpin dependent pausing has been observed in the 5′-UTR of the *his* and *trp* operons in *E. coli* and *Salmonella* (Roberts *et al.* 2008). Strong secondary structures forming in the emerging RNA have been stated as reason for the pausing (Toulokhonov *et al.* 2001). Besides hairpin formation, the efficiency of pausing has been determined partially by sequence elements surrounding the hairpin-coding segment (Lee *et al.* 1990). Another study indicated that this type of pause is not backtracked. Instead, there is evidence that pausing occurs because a structural rearrangement of the active site prevents nucleotide addition to the RNA 3′-end (Toulokhonov *et al.* 2007). A recent study presented new findings regarding sequence-specific pausing by bacterial RNAPs. The study suggests that backtracking and translocation delay are not the main mechanisms for this type of pausing (Kireeva & Kashlev 2009). The group could show that NTP binds to the paused elongation complex as efficiently as the active elongation complex, suggesting that ubiquitous transcriptional pausing involves misalignment of the 3′ NMP and the substrate NTP in the catalytic centre of RNAP (see Fig. 1.1).

In its inactivated state, the elongation complex can induce cleavage of the nascent transcript. In these states, RNAP catalyses the hydrolysis of the RNA transcript leading to the rapid release of the 3′-end fragment, which can be as long as 17 nt. The resulting complexes containing the shortened transcripts can resume synthesis from the new 3′-end. Alternatively, the elongation complex can undergo a series of conformational changes to an arrest (or dead end) state. RNAP

in this state cannot resume elongation, even in the presence of high concentrations of all four NTPs and despite the elongation complex remaining intact and RNAP remaining active (Erie 2002). Such stalling can cause further accumulation of trailed elongation (Lewis *et al.* 2008), and can also block DNA replication fork complexes which may have detrimental effects (Borukhov *et al.* 2005). At this stage the protein GreB is required which activates the cleavage activity of RNAP to resume the synthesis (Borukhov *et al.* 1993; Erie 2002; Borukhov *et al.* 2005).

Other transcription factors involved in the elongation is Nus and Mfd. NusA has a role in defining the rate of transcription elongation, which has been reported to be important for efficient coupling of transcription and translation (Lewis *et al.* 2008; Roberts *et al.* 2008). NusG is considered to be essential in Gram-negative bacteria. It inhibits pausing and increases the rate of elongation through interactions with termination factor Rho (see below) during mRNA transcription. It also enhances Rho-dependent termination, particularly under suboptimal conditions for Rho (Lewis *et al.* 2008, and references therein). NusG binds both to Rho and RNAP and thus provides a possible link between them (Lewis *et al.* 2008). Mfd (*mutation frequency decline*) transcription factor from *E. coli* has been shown to use ATP hydrolysis to reactivate arrested elongation complexes without transcript cleavage (Park *et al.* 2002).

Termination

Transcription in bacteria employs two types of terminators. So called intrinsic terminators, also termed 'Rho-independent' terminators, mainly require elements located on mRNA, while 'Rho-dependent' terminators require mRNA elements and additional proteins (Banerjee *et al.* 2006). However, historically called Rho-dependent terminators are not restricted only to Rho, see below. According to Ciampi (2006) about half of the transcription terminators identified in *E. coli* are Rho-dependent.

Rho-independent terminators are characterised by GC-rich inverted repeat followed by a stretch of A residues on the template strand. The inverted repeat is believed to form a hairpin loop in the nascent RNA that induces RNAP to pause, destabilise and disengage followed by RNA release (Artsimovitch & Landick 1998). A sequence of U residues are believed to play a role in pausing of RNAP due to formation of a weak A:U hybrid which allows the hairpin to form (Gusarov & Nudler 1999). These terminators are considered to be economical as they function in the absence of any additional protein factors (Unniraman *et al.* 2002).

Rho-dependent terminators are characterised by two fundamental features: a proximal Rho binding site called the *rut* (*Rho utilization site*) and a distal sequence harbouring the termination zone. The 70-80 nt Rho binding site is rich in C and lacks any secondary structure. The termination zone begins about 60-90 nt downstream of the *rut* site (Banerjee *et al.* 2006).

Rho is a homohexameric protein. It is a global regulator of RNA polymerase that exhibits RNA-dependent ATPase activity. It binds to the nascent mRNA and translocates in $5' \rightarrow 3'$ direction along mRNA by using energy derived from the hydrolysis of ATP. At certain sites, Rho contacts RNAP and terminates the elongation complex by an unknown mechanism (Skordalakes *et al.* 2005; Banerjee *et al.* 2006). It has been speculated that Rho-dependent termination's primary role is the prevention of R-loop formation (see below) from occasional untranslated RNAs (rather than its regular function to terminate transcription at the end of genes or operons) (Gowrishankar & Harinarayanan 2004). Rho-dependent termination sites are rare outside the coding regions of genes, and are abundant in intragenic regions (Gowrishankar & Harinarayanan 2004). A recent study revealed previously unknown roles of Rho in termination (Peters *et al.* 2009). This study suggests that Rho involves in halting mRNA chain elongation for antisense transcription and terminating synthesis of stable RNAs, including tRNAs and sRNAs. These findings bring new perspectives on termination as Rho may play a crucial role in halting transcription at locations where intrinsic terminators could not readily evolve (e.g. horizontally transferred DNA and antisense transcripts).

R-loops are structures that form when RNA forms heteroduplexes with single-stranded DNA, and they are known to be lethal. R-loops have been claimed to mediate the aberrant initiation of chromosomal DNA replication (constitutive stable DNA replication) in bacteria. Prevention of accumulation of non-functional transcripts in the cytoplasm is thus considered to be one of the reasons for the coupling of transcription and translation in prokaryotes (Gowrishankar & Harinarayanan 2004).

1.3.2 The translational process

Translation is considered to be a complex, multistep, multicomponent process that is highly regulated (Laursen *et al.* 2005). It is divided into four conceptual phases: *initiation*, *elongation*, *termination* and *ribosome recycling*. As transcription and translation are tightly coupled in prokaryotes, the decoding of mRNA to proteins starts long before transcription is completed. Translation of the genetic information into proteins takes place on the ribosomes. During the initiation step, the ribosome and other complexes are assembled on the translation initiation region. In the following elongation phase, the mRNA slides through the ribosome and its decoding begins. The peptide elongation continues until the ribosome reaches a stop codon triggering it to enter the termination step. In the last step, the newly synthesised protein is released, the ribosome dissociates and the mRNA is liberated. The described steps are believed to be universally conserved, although differences exist in details. In bacteria, relatively few factors involve in translation. In the following sections some of the elements that are important in translation will be presented.

The Ribosome. The ribosome is a ribonucleoprotein complex that is composed of two subunits. A large unit (consisting of 34 proteins and 2 rRNAs, 5S and 23S rRNA) with a relative sedimentation rate of 50S (1.5 MDa), and a small unit (consisting of 21 proteins and the 16S rRNA) with a relative sedimentation rate of 30S (0.8 MDa) (Steitz & Moore 2003) (Fig. 1.3). The ribosome has two major tasks in protein synthesis: reading the genetic code and catalysing peptide bond formation. To decipher the genetic code, the ribosome employs aminoacyl-tRNAs via three unique tRNA-binding sites: aminoacyl (A-site, accepts the incoming aminoacylated tRNA), peptidyl (P-site, holds the tRNA with the nascent peptide chain) and exit (E-site, which holds the deacylated tRNA before it leaves the ribosome) (Marshall *et al.* 2008; Ramakrishnan 2008). *E. coli* has 41 different tRNA species with different anticodons which need to be recognised by the ribosome to reach the corresponding codon located on the mRNA. During peptide elongation, ribosomes were found to synthesise 6 to 20 peptide bonds s^{-1}, while misincorporating less then 1 in 1000 amino acids *in vivo* (Kurland 1992). The exact speed of the ribosome is a matter of debate. In a paper by Boni (2006) the avearage speed of the ribosome is stated to be 40 nt s^{-1} in *E. coli*. The high-resolution crystal structure of *Thermus thermophilus* ribosomal 30S at 3 Å resolution (Wimberly *et al.* 2000), *Haloarcula marismortui* 50S subunits at 2.4 Å resolution (Ban *et al.* 2000), the complete *Thermus thermophilus* 70S ribosome containing bound messenger RNA and transfer RNAs (tRNAs) at 5.5 Å resolution (Yusupov *et al.* 2001) as well as the structure of the 70S ribosome complexed with mRNA and tRNA at 2.8 Å resolution (Selmer *et al.* 2006) have been solved. In addition the *E. coli* ribosome structure at 3.5 Å resolution was also published (Schuwirth *et al.* 2005).

The Shine-Dalgarno Sequence. Fellner and colleagues published the partially sequenced 16S rDNA from *E. coli* in 1972 and it was found that rRNA genes contain regions of large repeats except for the sequence 5′-ACCUCC-3′ which was found to be non-repetitive (Ganley & Kobayashi 2007). Shine and Dalgarno (1974) later found that this region is complementary to 5′-GGAGGU-3′ found in coliphage mRNA species. By then it was already known that the 30S subunit of the ribosome is involved in translation initiation and in their corner-stone paper they postulated that this stretch of sequence may promote interaction with mRNA and the small subunit of the ribosome. Based on X-ray crystallographic studies, the interaction between SD and the anti-SD on the 30S ribosomal subunit has been experimentally confirmed at 7 Å resolution (Yusupova *et al.* 2001) and later on also at 3.3 Å resolution (Kaminishi *et al.* 2007). Since then the importance of this sequence has been stated in numerous papers. SD sequences have been observed in the majority of prokaryotic mRNAs where it is mainly located between the bases +7 and +12 and considered as 5′-GGAGG-3′ (Jiong *et al.* 2002; Chang *et al.* 2006).

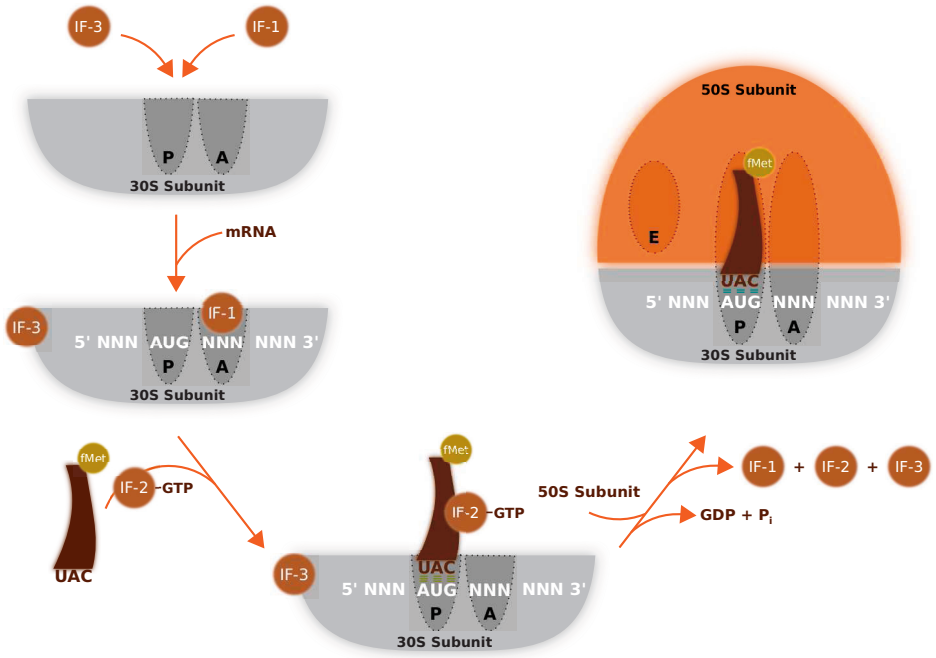

Figure 1.3: Formation of the initiation complex in bacteria. P designates the peptidyl site, A the aminoacyl site, and E the exit site. The coloured ribbon represent tRNA. Initiation factors (IF-1, -2, and -3), the mRNA coding for the polypeptide to be made, initiating $tRNA^{fMet}$ and GTP are depicted in the figures. The large subunit of the ribosome is shown in orange and the small subunit in grey.

Initiation

Translational initiation is believed to be the rate-limiting and most heavily regulated step (Laursen *et al.* 2005) among the four steps in translation. At initiation, the 30S subunit of the ribosome binds to the mRNA at a region called 'ribosome binding site', which is characterised as the region that is protected from ribonucleases within the 30S initiation complex (Steitz 1969) and that is near the 5′-end of a coding sequence. With this binding the ribosome covers the translational start site, the 5′-UTR (harbouring the SD sequence) and also a part of the coding sequence. The overall coverage is considered to span the region from −20 to +15 relative to the transcriptional start site (Boni 2006).

As noted previously, bacterial initiation involves the interaction of the 30S subunit with the SD sequence on the mRNA via base pairing with the anti-SD sequence at the 3′-end of 16S RNA (Shine & Dalgarno 1974). Initiation requires three initiation factors IF1, IF2, and IF3, and one GTP molecule mediating efficiency and fidelity of this process (Gualerzi & Pon 1990) (see Fig. 1.3). They modulate the affinity of the P-site for initiator $tRNA^{fMet}$. IF3 is known to bind to

the 30S subunit and prevent its association with the 50S subunit. IF2 is thought to act as a carrier for the formation of initiation complex (Mayer *et al.* 2003). IF1 promotes a more efficient binding of IF2 and IF3 to the 30S subunit, and cooperates with IF2 to ensure the correct location of the initiator in the P-site (Simonetti *et al.* 2008). It is also believed to play an 'initiation fidelity function' by blocking the A-site from access of aminoacyl-tRNA (Boelens & Gualerzi 2002).

The SD sequence is considered to be one of the conserved features in prokaryotic transcripts. With the availability of an increasing number of prokaryotic genome sequences, it became possible to apply computational approaches for identification of genes led by SD sequences (SD-led). A recent analysis of 162 completely sequenced prokaryotic genomes (with 141 of bacterial origin), revealed that a large percentage (46%) of mRNAs does not contain a SD sequence (including non-SD-led and leaderless transcripts), with the corresponding value for *E. coli* mRNA being 39% (Chang *et al.* 2006). It has been described that mRNAs lacking any obvious SD or SD-like sequences, including leaderless genes, can be translated with relatively high efficiencies (Chang *et al.* 2006; Brock *et al.* 2007; 2008). Translation initiation from leaderless mRNAs has been reported to involve 70S ribosomes (O'Donnell & Janssen 2001; Moll *et al.* 2002) or 30S subunits free of IF3 (Moll *et al.* 2002; Brock *et al.* 2008). Based on toeprint assays tRNA-dependent ribosome binding to leaderless mRNA and an AUG start codon is required for efficient ribosome binding and translation in *E. coli* (O'Donnell & Janssen 2001). However, the details on how initiation (ribosome association and P-site placement of the start codon) takes place with non-SD-led genes are not well understood.

Moreover, leaderless genes could be translated in all three domains of life. A study by Grill and colleagues (2000) reported that leaderless lambda *cI* mRNA is translated *in vitro* by both archaebacterial and eukaryotic translation systems. This supports the hypothesis that the translation initiation pathway is universally conserved. However, the SD dependent pathway does not appear to be the primary and universal mechanism for translation initiation.

Elongation

At the end of the initiation phase the 30S and 50S subunits associate. The P-site holds the initiator aminoacyl-tRNA while the A-site is empty and ready to receive aminoacyl-tRNA which is brought into the A-site as a ternary complex with elongation factor Tu (EF-Tu) and GTP (Kaczanowska & Ryden-Aulin 2007). After hydrolysis of GTP, EF-Tu releases the aminoacyl end of the A-site tRNA, allowing it to shift into the P-site, where the formation of peptide bonds takes place (Pape *et al.* 1998). The translocation of tRNAs is driven by GTP and the GTPase EF-G (Kiel *et al.* 2003). However, the canonical mechanism of translocation in elongation has also been challenged. Alternative models describe that i) EF-G

favours GDP instead of GTP so that EF-G - GDP drives the ribosome (from its relaxed state with full binding sites for three tRNAs to a twisted conformation with hybrid sites for two tRNAs) and ii) the switch of the EF-G conformation from GDP-bound to its GTP-bound structure occur on, rather than off, the ribosome (Zavialov *et al.* 2005; Agirrezabala *et al.* 2008).

Termination

Termination of translation begins when a stop codon in the mRNA enters to the A-site. The termination codon is recognised by either of the release factors (RFs) 1 or RF2. RF1 responds to UAG and UAA, and RF2 to UAA and UGA. Binding of either of the RFs triggers hydrolysis of the ester linkage between tRNA and the polypeptide on the ribosome (Kisselev *et al.* 2003). Upon the release of the completed polypeptide, the third release factor, RF3, which possesses GTPase activity, catalyses the dissociation of the RF1 and/or RF2 from the A-site (Freistroffer *et al.* 1997).

Ribosome recycling

After the release of the completed peptide, the ribosome is left with bound mRNA and a deacetylated tRNA in the P-site. In order to recycle the ribosome, yet another factor is needed. The so called ribosome recycling factor (RRF) along with EF-G ensures the process of subunit dissociation (Kiel *et al.* 2003). Finally, IF3 removes the deacylated tRNA from the 30S subunit and allows the mRNA to either detach from the complex or form a new stable ribosome binding (Karimi *et al.* 1999).

1.4 Tools for Regulating Gene Expression

Regulation occurs at many levels in natural biological systems, namely at the level of transcription, RNA processing, translation, protein-protein interactions and protein-substrate interactions (Boyle & Silver 2009). All these processes exert control on cellular processes and through genetic engineering some of these processes can be manipulated and directed e.g., recombinant protein production. The following subsections will briefly address methods for artificial control of transcription and translation.

Promoter engineering. One of the most commonly used methods for driving the expression of a desired gene is using native promoters or different variants thereof. Promoter variants with different strengths can be obtained by combining known promoters (Casadaban 1975), by screening synthetic promoter libraries (Jensen & Hammer 1998; Solem & Jensen 2002; Alper *et al.* 2005; Hammer *et al.*

2006; Mey *et al.* 2007; Bakke *et al.* 2009) or by constructing a combinatorial library of random promoter architectures (Cox *et al.* 2007). For creation of promoter variants both inducible and constitutive promoters can be used. In addition to the choice of a regulated promoter with varying strengths, expression levels can also be adjusted by altering the levels of inducer concentrations (Siegele & Hu 1997; Winther-Larsen *et al.* 2000b; Mnaimneh *et al.* 2004).

Global transcription machinery engineering. The Stephanopolus group has presented several methods for altering gene expression in biological systems. The approach "global transcription machinery engineering" is defined as 'an approach for reprogramming gene transcription to elicit cellular phenotypes important for technological applications'. The group could show in the yeast *Saccharomyces cerevisiae*, that through mutagenesis and selection of TATA-binding proteins, strains can be obtained that are tolerant to high concentrations of ethanol and glucose (Alper *et al.* 2006). In this approach the transcription factor Spt15p had been mutagenised by error-prone PCR leading to identification of the strains.

A second approach focussed on engineering of the σ factor. In this work, the *rpoD* gene, which encodes the main sigma factor, σ^{70}, was subjected to random mutagenesis and introduced into *E. coli* to search for varying cellular phenotypes (Alper & Stephanopoulos 2007).

In a third approach, the bacterial RNAP α-subunit was chosen as mutagenesis target . In this study Klein-Marcuschamer (2009) presented the possibility that by mutagenising the α-subunit of RNAP core enzyme itself, new variants with different strengths can be obtained. They presented this tool for improving tolerance of *E. coli* to butanol and other solvents and for increasing the titres of L-tyrosine and hyaluronic acid.

Riboswitches. Riboswitches are cis-acting genetic regulatory elements that are noncoding and found both in prokaryotic (Mandal & Breaker 2004) and eukaryotic (Sudarsan *et al.* 2003) mRNAs. The first natural riboswitch was reported by Winkler and colleagues (2002). They showed that mRNAs encoding enzymes involved in thiamine (vitamin B1) biosynthesis in *E. coli* can bind thiamine in their 5′-UTR's.

Isaacs and colleagues (2004) described that a small noncoding RNA expressed in trans, can be used both to silence and activate gene expression. They inserted a complementary cis sequence directly upstream of the ribosome binding site of a target gene. Upon transcription, this cis-repressive sequence would cause a stem-loop structure to form at the 5′-UTR of the mRNA. The stem-loop structure interferes with ribosome binding, silencing gene expression. The RNA that is expressed in trans targets the cis-repressed RNA with high specificity, causing an alteration in the stem-loop structure that activates expression.

Suess *et al.* (2004) demonstrated the use of riboswitches in translational control. In this study a theophylline-binding aptamer was integrated into the 5′-UTR of a reporter gene in such a way that binding of the molecule to the reporter mRNA resulted in repression of gene expression via helix slippage and subsequent RBS occlusion. Binding of the ligand theophylline induces a structural transition in the bridge helix leading to a 1 nt shift which moves the element away by exactly the critical distance to allow ribosome binding.

Designing proteins. Proteins are made from a set of 20 amino acids, each of which contain an amine and a carboxylic acid group flanking a central carbon atom. A recent study demonstrated that it is possible to design proteins by engineering a bacterial cell to work with amino-acid-like molecules called -hydroxy acids that have an alcohol group where the amine would normally be (Guo *et al.* 2008). During translation, instead of forming an amide bond to link polymer subunits, this -hydroxy acid forms an ester bond. By replacing a nitrogen and a hydrogen atom in a polymer chain the ester bond can easily be cut in basic solution.

Gene-regulatory circuits. Artificial gene-regulatory circuits can be considered as part of synthetic biology. These approaches can be used to create any desired artificial network with simple regulatory elements to obtain desired behaviours.

A recent rather elegant example demonstrated the use of synthesized regulatory promoter libraries to construct feed-forward loop networks with different predicted input-output characteristics. Ellis and colleagues (2009) presented a synthetic gene network acting as a predictable timer, modifiable by component choice. They use this network to control the timing of yeast sedimentation, illustrating how the plug-and-play nature of our design can be readily applied to biotechnology in *Saccharomyces cerevisiae*.

Genome reorganisation and synthetic cell. Researchers from the Craig Venter Institute have published several studies illustrating the possibility of reorganising entire bacterial chromosomes. Although these studies are still in their infancies, the challenging task of reorganising or designing a chromosome from scratch clearly illustrates technological progress in this field.

Their first study demonstrated whole genome transplantation in bacteria. In this work intact genomic DNA from *Mycoplasma mycoides* was transplanted into *Mycoplasma capricolum* cells by polyethylene glycol-mediated transformation, and reported to be successfully established (Lartigue *et al.* 2007). This method enables the transformation and establishment of entire genomes across species.

The second study represented the chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome (Gibson *et al.* 2008). This work exhibited ways

for constructing large DNA molecules from chemically synthesized pieces and also from combinations of natural and synthetic DNA segments.

The remaining work to be accomplished is the synthesis of a whole genome in one go, without assembling it from smaller DNA fragments. This last step, if succeeded, will certainly bring new challenging aspects to the field both from academic and applied perspectives.

2

Biological Tools Used in This Study

Our research group developed a system many years ago which harbours the inducible *Pm* promoter inserted into minimal replicons of the broad-host-range plasmid RK2 (Blatny *et al.* 1997a). In this system there are several DNA control elements that this thesis work is focussed on (both directly and indirectly): the *Pm* (promoter), *xylS* (gene encoding positive regulator of *Pm*, acts in the presence of inducer), the 5′-UTR (5′-untranslated region at the mRNA level) and *trfA* (gene encoding plasmid replication initiation and copy-number control protein). Throughout this and other studies, we have identified variants of these regulatory sequences that lead to changed expression levels (both increased and decreased compared to wild type forms) by using *E. coli* as host organism. In this chapter the origin of the above mentioned DNA regulatory and other elements used in this study will be addressed.

2.1 The RK2 Plasmid

RK2 was isolated in 1969 at a Birminhgam (England) hospital from the bacterium *Klebsiella aerogenes* (Ingram *et al.* 1973). It belongs to the incompatibility group IncPα and its complete sequence has been published (Pansegrau *et al.* 1994). IncP plasmids can replicate in diverse Gram-negative bacteria and can be transferred conjugatively to a large set of microorganisms (Toukdarian 2004). RK2 replicons have been shown to be functional in at least 58 Gram-negative bacteria (Toukdarian, 2004; Trond Erik Vee Aune, personal communication) and in at least one Gram-positive bacterium (*Clavibacter xyli*) (Thomas & Helinski 1989). RK2 requires two essential regions for replication: the *trfA* gene encoding the replication initiation protein, and *ori*V, the *cis*-acting origin for vegetative DNA replication (Perri *et al.* 1991) (see Fig. 2.1 for the minimal replicon of RK2).

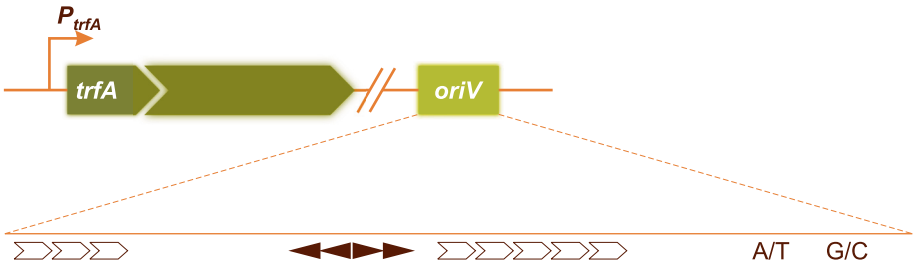

Figure 2.1: The minimal replicon of RK2. *trfA*, coding for replication initiation protein with alternative translational start and the origin of replication, and *oriV* are depicted. Zooming in on *oriV* (below the map) demonstrates the important features: eight 17 nt iterons (open arrow), four DnaA boxes (black arrowheads), AT-rich region, and GC-rich region (Toukdarian 2004).

trfA encodes two versions of the replication initiation protein as a consequence of an internal in-frame translation start. The small protein (TrfA-33) was shown to be sufficient for plasmid replication in many bacterial hosts, including *E. coli* and *Pseudomonas putida* (Durland & Helinski 1987; Shingler & Thomas 1989), while efficient replication in *P. aeruginosa* was found to require the larger protein (TrfA-44) (Fang & Helinski 1991) . The copy number of RK2 plasmids is 4 to 7 per chromosome in *E. coli* and 2 to 3 in *P. putida* (Thomas & Helinski 1989). The plasmid copy number is regulated by interactions between the origin and the TrfA protein in a process that has been termed as handcuffing (Blasina *et al.* 1996; Toukdarian & Helinski 1998). Based on *in vitro* studies, replication of RK2 requires DnaA (host initiation factor), DnaB (helicase), DnaC (helicase accecory protein), DnaG (primase), DNA polymerase III and DNA gyrase, in *E. coli* (Kline 1985; Pinkney *et al.* 1988).

The handcuffing model (Fig. 2.2) is based on three main postulates: *oriV* with TrfA bound to its iterons is replication proficient. When all of the TrfA-bound plasmid molecules are reversibly coupled at their replication origins, plasmid replication cannot take place. And finally, as cell mass increases, handcuffed structures fall apart and the initiation potential is restored (Krüger *et al.* 2004).

Iterons, 17 nt direct repeats, are regions where plasmid-encoded replication proteins (Rep, TrfA in RK2) bind and act as either initiators of plasmid replication or as inhibitors of replication (Kunnimalaiyaan *et al.* 2005). In RK2 iterons are grouped in two clusters, one with three and the other containing five iterons (Fig. 2.1). A study by Kittell & Helinski (1991) demonstrated that deletion of the upstream three-iteron cluster did not inactivate the origin in *E. coli* but resulted in an increase in copy-number and plasmid instability. This study also described further evidences supporting the handcuffing model. Copy-up TrfA variants are much less sensitive to iteron inhibition than replication by wild-type TrfA both *in vivo* and *in vitro*. The handcuffing model was further supported by the finding

that when the TrfA concentration was elevated over a 170-fold range *in vivo* , the copy-number of intact RK2 or of RK2 minireplicons did not increase as a result (Durland *et al.* 1990).

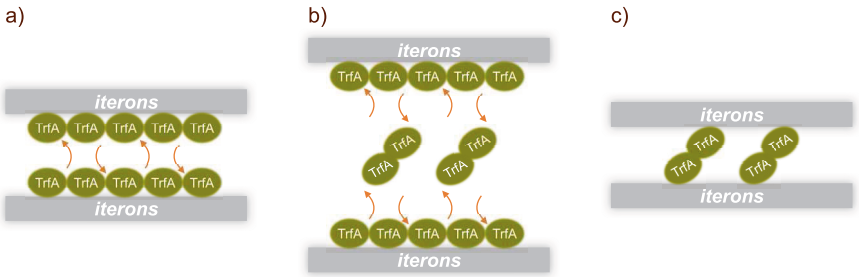

Figure 2.2: Proposed handcuffing model. (a) Two layers of monomers bound to the iterons of two individual plasmids are able to bring two *ori*'s together (curved small arrows), shutting down replication of both. (b) Dimers of the TrfA protein bridge two plasmids by associations with monomers bound to iterons. (c) A single layer of dimers binds the iterons of two *ori*'s (Krüger *et al.* 2004).

2.2 The TOL Plasmid

The toluene-degradative plasmid (TOL) pWWO, originates from the *P. putida* mt-2, which is self-transmissible, 117-kb in size and a member of the incompatibility group IncP9. The TOL plasmid encompasses a set of genes that are involved in the degradation pathway, which are composed in two segments, upper and lower pathways (Fig. 2.3). Upper and lower pathway genes do have two regulatory loops. When the cell grows on a substrate of the lower pathway, such as *m*-toluate, the *meta* loop (lower pathway) operates via XylS to express the lower pathway genes. In contrast when it grows on substrates for the upper pathway, such as xylene, the cascade operates via XylR and XylS, enabling the expression of both pathways (Marqués & Ramos 1993; Ramos *et al.* 1997).

In the lower pathway the *Pm* promoter requires the positive regulator XylS for activation. XylS expression is controlled by two sets of promoters, *Ps1*, and *Ps2*. *Ps2* is a $\sigma^{70}(\sigma^{D})$-dependent promoter and it yields low levels of XylS protein in their inactive forms. When substituted benzoates are present, inactive XylS interacts with the inducer molecules and is activated, which in turn triggers expression from *Pm*. *Pm* requires two different σ factors depending on the growth phase. σ^{32} (σ^{H}) is employed in the early exponential phase while σ^{38} (σ^{S}) is characteristic for the late exponential or stationary phase (Marqués *et al.* 1995). The master regulator XylR is a member of the XylR-NtrC family of transcription factors, and expression of the *xylR* gene is autoregulated by its own promoter *Pr*. XylR is inactive in the absence of upper pathway inducers (toluene/xylenes).

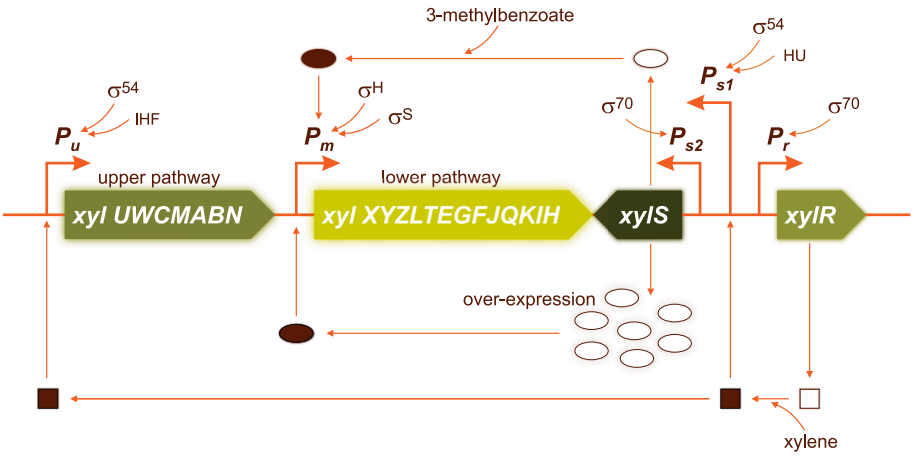

Figure 2.3: Regulatory circuits of the catabolic genes from the TOL plasmid pWW0. Squares: XylR; circles: XylS; open symbols: transcriptional regulator forms unable to stimulate transcription; closed symbols: forms able to stimulate transcription. See text for details.

In the presence of inducers, XylR becomes activated and stimulates transcription from the *xylS* gene promoter *Ps1*, and the upper pathway operon promoter *Pu* with σ^{54}-associated RNAP. *Pu* requires the DNA-binding protein integration host factor (IHF), while *Ps1* does not require IHF which in fact negatively affects it's promoter activity (Holtel *et al.* 1995). Activation of *Ps1* requires chromatin-associated DNA-binding protein HU. As a result, large amounts of XylS can be produced which teh subsequently activate *Pm* even in the absence of the inducer of the lower pathway (Inouye *et al.* 1987; Mermod *et al.* 1987).

2.2.1 Characteristics of the transcriptional activator XylS and the Pm promoter

The XylS protein belongs to the AraC-XylS transcription factor family. It is a 321 amino acid long modular protein with a conserved C-terminal domain for DNA binding and interactions with RNAP and a non-conserved N-terminal domain involved in effector binding and protein dimerization. In the C-terminus of XylS, a conserved stretch of about 100 amino acids is proposed to yield in total seven α-helices, including two HTH binding motifs (α2-T-α3 and α5-T-α6) connected by a linker α-helix (α4) and two flanking α-helices (α1 and α7). This structural organisation has also been confirmed experimentally for the monomeric MarA protein (Rhee *et al.* 1998) (see **Paper I**, Fig. 2). XylS binds to DNA as a dimer recognising two 15-bp direct repeats (TGCA-N6-GGNTA) which stretch over a 35-bp long region upstream of the *Pm* promoter (González-Pérez *et al.* 2002). It has been shown that XylS binds to 5′-box A (TGCA) and 3′-box B (GGNTA) in a

head-to-tail organisation: α-helix 3 recognises the A boxes, and α-helix 6 interacts with the B boxes (**Paper I**, Fig. 2). Due to the spatial arrangement of these two repeats the XylS binding site overlaps with the RNAP –35 binding by 2 bp (see Fig. 2.4).

Figure 2.4: *Pm* promoter sequence organisation. The bold letters indicate the two XylS binding sites (proximal and distal), each composed of conserved A and B boxes. The –10 and –35 hexamers are underlined. +1 indicates the transcription start site.

The formation of XylS dimers is a crucial step in activating transcription from the *Pm* promoter. A recent study by Domínguez-Cuevas *et al.* (2008) suggested a model for transcriptional activation by XylS. Under basal conditions, XylS DNA-binding domains are unable to make contact with DNA, while the presence of the effector causes a conformational change in XylS that results in dimerisation and binding of the dimer to DNA, activating transcription. This proposed model is in accordance with earlier findings indicating that when XylS is present in high concentrations in the cell, it can initiate transcription as a result of dimerisation even in the absence of the effector molecule. It has also been described that the C-terminal domain of XylS on its own is able to activate *Pm* in the absence of the effector (Mermod *et al.* 1987; Kaldalu *et al.* 2000). Based on these findings, the N-terminal domain appears to be acting as an intramolecular repressor. This repression phenomenon has been demonstrated for members of the NtrC family of regulators such as DmpR and XylR, in which deletion of the N-terminal effector binding domain resulted in an activator that mediated transcription constitutively. In the same paper by Domínguez-Cuevas *et al.* (2008) it has been speculated that the XylS dimer in the presence of the effector, recruits RNAP to the *Pm* promoter and this subsequently increases the rate of isomerisation of RNAP from closed to open complexes.

The *Pm* promoter does not exhibit conserved –35 and –10 regions (Marqués *et al.* 1999). The reason for the divergence in the –35 region was thought to be due to the fact that XylS and RNAP binding sites overlap in this region (González-Pérez *et al.* 2002). Divergence in –10 regions can be explained by the fact that expression from *Pm* can be mediated with two alternative σ-factors ($\sigma^{H}$and σ^{S}) so that both RNAP holoenzymes need to be able to bind to this region (Domínguez-Cuevas 2007).

2.2.2 The 5′-untranslated region of mRNA

The 5′-UTR used in this study originates from the TOL plasmid. It is the native DNA sequence that is found between the *Pm* promoter and upstream of the *xylX* gene. It consists of 32 nt between the transcriptional start and translational start site(Fig. 2.5). In this region the putative SD sequence is one base shorter than the proposed consensus sequence 5′-GGAGG-3′ for *E. coli* (Ma *et al.* 2002).

Figure 2.5: 5′-UTR sequence. Transcriptional and translational start sites are indicated. The Shine-Dalgarno sequence is shown as SD.

2.3 The Reporter Genes Used in This Study

2.3.1 The gene coding for β-lactamase (bla)

β-lactamase activity was first described in 1940, soon after the isolation of penicillin, by Abraham & Chain (1940). They reported that an extract from a crushed strain of *E. coli* inactivated penicillin. Later on, plasmid-bearing staphylococci carrying the gene for β-lactamase were characterised. β-lactamase (β-lactam hydrolase, EC 3.5.2.6) hydrolyses the β-lactam binding and thereby inactivates the antibiotic with no need for a cofactor (Uhlin & Nordström 1977). β-lactam antibiotics bind to primary receptors, membrane-associated penicillin-binding proteins (PBP's). These proteins perform a central role in synthesis of the cell wall component peptidoglycan. Thus, inactivation of PBP's through antibiotic binding has detrimental effects on their function. Peptidoglycan is responsible for maintaining the integrity of bacterial cell walls therefore disruption of its structure in growing cells leads to lysis (Tipper 1985). Gram-negative bacteria, in contrast to Gram-positives, have a selectively permeable cell wall component in form of the outer membrane, allowing them to control the composition of the periplasmic space located between this outer and the cytoplasmic membrane. Access to the periplasm, an environment rich in PBP's, is determined by the rate of passive diffusion through the outer membrane. The periplasm provides a small compartment in which secreted enzymes, such as β-lactamases, can reach high concentrations (Tipper 1985). Uhlin & Nordström (1977) reported a linear relation between the increased copy number of the R plasmid R1*drd-19* and the resistance to ampicillin observed on agar plates for *E. coli* K-12 (Fig. 2.6). Thus β-lactamase is suitable as a reporter gene, since increased expression can be monitored by increased resistance levels to ampicillin on agar media. It is known that the permeability of

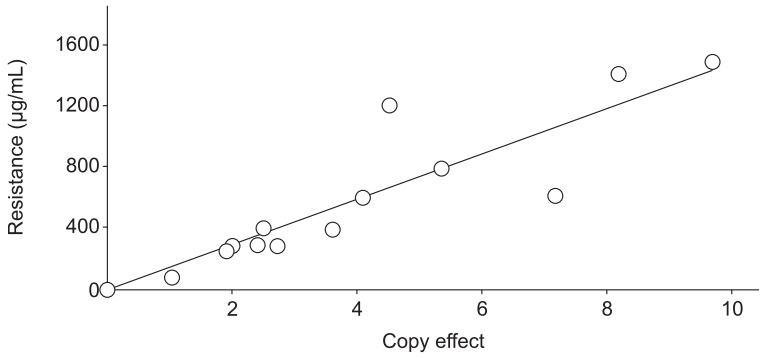

Figure 2.6: Effect of copy number of the R plasmid R1*drd-19* in *Escherichia coli* K-12 strain D1 on resistance to ampicillin (Uhlin & Nordström 1977).

the outer membrane varies among Gram-negative bacteria. Permeability is high in *Neisseria*, very low in *Pseudomonads* and intermediate in Enterobacteriaceae (Tipper 1985). Due to this host specific permeability the use of β-lactamase as a universal reporter gene has limitations.

2.3.2 The gene coding for luciferase (luc)

The light-producing chemical reactions of bioluminescent organisms are catalyzed by luciferases. The luciferase (*Photinus* luciferin:oxygen 4-oxidoreductase, EC 1.13.12.7) used in this study originates from the firefly *Photinus pyralis*. Insect luciferases require the organic molecule luciferin, 0_2, and ATP as substrates for light production. The requirement for ATP is characteristic for insect luciferases and it has been used to develop bioluminescence assays. The chemical reactions of this process are:

$$luciferase + luciferin + ATP \xrightarrow{Mg^{2+}} luciferase{\cdot}luciferyl - AMP + PP_i$$

$$luciferase{\cdot}luciferyl - AMP + O_2 \longrightarrow luciferase + oxyluciferin + AMP + CO_2 + hv$$

The first reaction results in formation of an enzyme-bound luciferyl-adenylate and the second reaction in the production of CO_2, oxyluciferin, AMP and light. When substrates are added in excess to firefly luciferase, the reaction produces a flash of light that is proportional to the quantity of luciferase in the reaction mixture (de Wet *et al.* 1987). The luciferase encoding gene was identified from cDNA libraries of *P. pyralis* and heterologous synthesis of active luciferase has been established in *E. coli* (de Wet *et al.* 1985). The luciferase gene has a clear advantage for use as a reporter gene in that the bioluminescence enzyme assay is very sensitive. Thus, expression can be accurately determined at the protein level over a very wide range.

2.3.3 The gene coding for phosphoglucomutase (celB)

The phosphoglucomutase (α-D-glucose 1,6-phosphomutase, EC 5.4.2.2), encoded by *celB*, used in this study originates from *Acetobacter xylinum*. It catalyzes the reversible transformation of α-D-glucose 1-phosphate to D-glucose 6-phosphate (Brautaset *et al.* 1994). Expression of the *celB* from the wild-type *Pm* promoter was shown to be very high, as CelB could directly be visualised in crude extracts by SDS-PAGE (Blatny *et al.* 1997b). Hence, *celB* is a suitable reporter to test the upper limits of gene expression.

2.4 Expression Vectors

For this study, several promoter/expression systems were used in order to gain a comparative insight into the XylS/*Pm* system. The following systems were chosen for comparison studies.

2.4.1 Expression system based on XylS/Pm

In the subsection 2.2.1 characteristics of the activator XylS and the inducible promoter *Pm* were described. In this subsection the features of the expression system based on XylS/*Pm* will be addressed. In our research group, the expression cassette consisting of the inducible XylS/*Pm* positive regulator/promoter was combined with minimal replicons of the broad-host-range plasmid RK2 (Blatny *et al.* 1997a) to create a broad-host-range expression vector (pJBn). This system has been used for the construction of multiple different expression tools which together represent a wide range of properties and applications. Utilisation of this system as an expression vector was first demonstrated by Mermod *et al.* (1986). The vectors based on the broad-host-range RSF1010 replicon were shown to be useful for various levels of regulated expression of the XylE reporter protein in a relatively large number of different Gram-negative species. These vectors were later modified by utilising a variant form of the *xylS* gene, designated XylStr6, with altered effector affinity leading to eightfold higher expression levels from the *Pm* promoter compared to the wild-type XylS (Ramos *et al.* 1988). These studies are a clear indication that the XylS/*Pm* system can function well for regulated expression of heterologous proteins in a wide range of Gram-negative bacterial species. In particular, the plasmid pJB658 which is a combined cloning and expression vector (Blatny *et al.* 1997b), has proven to be very useful for many different aspects of recombinant expression. Relevant characteristics of this vector include adjustable vector copy number (between 4 and 100 per chromosome), use of several cheap benzoic acid derivatives such as *m*-toluic acid acting in a dose-dependent manner and presumably passive uptake of the inducer (no transport system is required, further simplifying the use of the system across species barriers, see **Paper I**, Fig.

4) (Lambert & Stratford 1999).

2.4.2 Expression system based on AraC/P$_{BAD}$

The pBAD plasmid used in this work is a derivative of the pBR322 expression vector (pBAD/gIII, Invitrogen). The regulatory protein, AraC, allows positive regulation of the P_{BAD} promoter. AraC regulates transcription from promoters involved in catabolism of L-arabinose, which is also the inducer of the system (Englesberg *et al.* 1965). When L-arabinose is present, it positively activates transcription from the P_{BAD} promoters. In contrast, when L-arabinose is not present it produces very low levels of transcript from P_{BAD} (Englesberg *et al.* 1969). This promoter is catabolite repressed, so uninduced levels of transcription is further reduced by cell growth in the presence of glucose (Miyada *et al.* 1984). The AraC protein is a homodimeric protein, and each of its monomers contains a conserved C-terminal DNA binding domain made up of about 100 amino acids and an N-terminal non-conserved dimerisation domain. This domain also contains an L-arabinose binding pocket (Wilcox & Meuris 1976; Steffen & Schleif 1977; Soisson *et al.* 1997b;a). In the absence of L-arabinose, the AraC dimer binds to two DNA sites O_2 and I_1 separated by 210 base pairs (**Paper I**, Fig. 1). This generates a characteristic DNA loop (Schleif 1988) which in turn negatively interferes with the access of RNAP to the promoter in the looping region. Due to this repression characteristics, the AraC/P_{BAD} system is considered to be one of the tightest (low basal expression in the absence of inducer) expression system.

Despite its useful features, this system has disadvantages as well. Siegele and Hu (1997) have reported that these promoters exhibit a so called all-or-none expression phenomenon which was found to be due to dependency on the inducer uptake system of the cell (Khlebnikov *et al.* 2002; Megerle *et al.* 2008). Under conditions of subsaturating levels of inducer, while a subpopulation is fully induced the remaining cells can remain uninduced. This may cause serious problems, since the fraction of the cells that are induced may be subject to a higher metabolic burden and thus grows slower than those cells that are not induced. After some generations, this may reduce the relative number of induced cells in the population. In order to overcome this problem, several approaches have been proposed. One of them is to over-express the L-arabinose transporter AraE (Khlebnikov *et al.* 2002). This approach would eliminate the all-or-none problem, although it creates a restriction on the application range of the system since a modified host is required. Further details on the characteristics of the AraC/P_{BAD} system can be found in **Paper I**.

2.4.3 Expression system based on bacteriophage T7 promoter

The most commonly used expression system based on bacteriophage T7 promoter is considered to be represented by the pET vector family (Novagen). In these systems, expression of the desired gene is placed under control of the T7 RNAP promoter and *lac* operator. Since the T7 promoter is used, these systems require a host which harbours a genomic copy of the gene for T7 RNAP controlled by the *lac* repressor. The systems mostly use a derivative of the *lac* promoter, *lacUV5*. Expression of *lacUV5* can be induced by IPTG, in turn initiating expression of the T7 RNAP.

Natural T7 promoters contain highly conserved sequences spanning the region from -17 to $+6$ relative to the transcriptional start site ($+1$) (Dunn & Studier 1983). Based on compilation studies the existence of termination signals for T7 RNAP was found to be rare, which makes this enyzme capable of transcribing almost any DNA sequence. T7 RNAP has been reported to be so active that a target protein can accumulate to greater than 50% of the total cell protein in about three hours (Studier & Moffatt 1986).

Drawbacks of pET systems include: the *lac* promoter is considered to be leaky (Dubendorff & Studier 1991), the fully induced T7 system can create severe burden on the host, and induction with IPTG is economically costly (in big scale fermentations) and toxic to the cell. Leaky expression of T7 RNAP can be overcome by introducing plasmids that express T7 lyzozyme which is the natural inhibitor of T7 RNAP (Studier 1991). Expression from the T7 promoter is considered to be strongest among all known promoter systems. However, the high speed of processivity comes with a price. The host translating ribosome moves much slower compared to T7 RNAP, thus this de-synchronisation results in low overall protein production. This has been shown by utilisation of tRNA as a transcriptional reporter (Lopez *et al.* 1994). To avoid the use of IPTG, two alternative promoter systems have been used to regulate the chromosomally located T7 *gene 1* expression: (i) through temperature regulated λP_L and λP_R tandem promoters (Chao *et al.* 2002b), and (ii) through *araBAD* promoter (Chao *et al.* 2002a).

2.5 The P1 Promoter

The work conducted on the 5′-UTR revealed several successful outcomes. All 5′-UTR variants detected in this study were identified with the *Pm* promoter. In order to assess if the observed phenomena was not just due to inherent features of the XylS/*Pm* system, a constitutive promoter was chosen for comparison. The *P1* promoter was first described by Brosius *et al.* (1982). It was observed that a promoter in pBR322 transcribed towards β-lactamase from the 5′-end of the tetracyline gene *tetA* (transcribing away from *tetA*). Due to this feature *P1* is also known as $P_{antitet}$. *P1* is σ^{70}-dependent constitutive promoter and it was

artificially created by ligation of two different DNA fragments to create pBR322 (Gilbert 2002).

3

Aims of the Study

The main aim of the work reported in this thesis was to further improve the expression levels from XylS/*Pm* and to gain new insights about the underlying mechanisms involved in determination of the levels of expression. For this work *E. coli* was chosen as the model system.

The method applied to improve the expression levels consisted of construction and screening of complex libraries of the 5′-UTR sequence (at the DNA level). These libraries were constructed using doped synthetic oligonucleotide mixtures that replaced the corresponding wild-type sequence after incorporation. Screening of the libraries was performed using *bla* (encoding β-lactamase) as reporter gene to monitor expression from *Pm*. *bla* was chosen as reporter gene since cells exhibiting high-expression phenotypes could be directly selected based on their ability to grow on agar medium supplemented with high concentrations of ampicillin, and thus exploiting the correlation between *Pm* activity and host tolerance to ampicillin. In this way, low-frequency occurring up-regulated variants could identified.

To understand the nature of the phenotypes identified in the screening, it was important to distinguish between transcriptional and translational effects and changes in mRNA stability. Further mechanistic insights were addressed by techniques that appeared relevant for each specific phenotype.

Parallel studies of other parts of the expression cassette were already ongoing when this work was initiated. This work, therefore, was aimed to combine different types of identified variants from several studies to achieve high-level expression of proteins from a single or a few gene copies inserted into the host chromosome. This approach could potentially eliminate the need for plasmids in recombinant protein production

The final goal was to compare the performance of the *xylS*/*Pm* expression cassette with two of the most frequently used commercial expression systems based on the bacteriophage T7 and P_{BAD} promoters. This comparison took into account both maximum expression levels and the degree of control from the lowest possible background level in the absence of inducer.

4

Summary of Results and Discussion

In this chapter the work (both published and unpublished) that was carried out within the scope of the thesis will be presented.

4.1 Utilisation of 5′-UTR as a Tool in Gene Expression Studies

The 5′-UTR's role in gene expression was the initial focus of the thesis. Instead of studying this region via site-specific mutagenesis, large combinatorial mutant libraries of 5′-UTR nucleotide sequences were used in this work. In total three libraries were created: LI, LII, and DI. All three libraries were screened using the ampicillin resistance conferring gene (*bla*) as reporter to identify 5′-UTR variants with high- and low-level expression characteristics in *E. coli* DH5α.

4.1.1 Identification of high-level expression 5′-UTR variants

A first plasmid library was constructed by Laila Berg (designated LI, **Paper II**) where doped oligonucleotides were used to introduce mutations within the 5′-UTR. Preliminary screening of this library for ampicillin resistance on agar medium revealed candidates with up to 10 times increased expression levels compared to that of the wild type. The 11 high-level-expression 5′-UTR variants obtained had three to six points mutations (Fig. 4.1). None of them were identical and none possessed any changes in the putative SD sequence (GGAG). It was observed that most of the 11 identified variants carried point mutations close to the 5′-end, located just one base away from the transcriptional start site (TrSS). Due to the close proximity of the mutated area to the TrSS, another set-up was created where the mutated area started 7 nt downstream of the TrSS (see **Paper II**, Fig. 1).

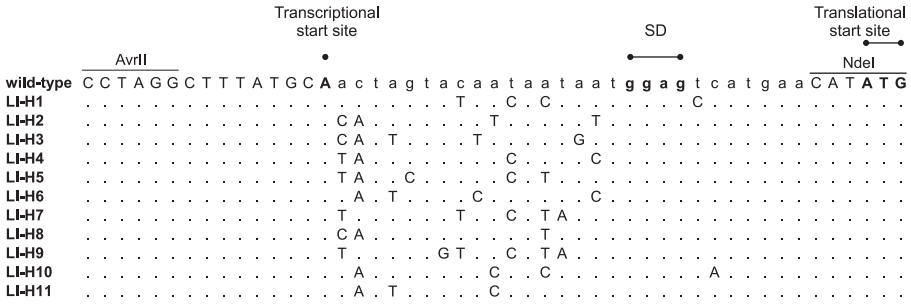

Figure 4.1: 5′-UTR sequences of variants identified in the library LI. Identical bases are depicted by dots. Transcriptional and translational start sites are indicated. Putative Shine-Dalgarno sequence is shown as SD. Bases subjected to mutagenesis are written in lower case. The indicated restriction enzyme sites are unique.

The rational behind this was to avoid alteration in the TrSS, as it has been stated in the literature that mutations near the TrSS may influence the start site selection (Walker & Osuna 2002; Lewis & Adhya 2004). It should also be noted that for none of the 11 identified variants, the TrSS was determined experimentally. The introduction of a new restriction site (BspLU11I) made it possible to combine the identified 5′-UTR's with different *Pm* variants (see section 4.2) for combinatorial studies (see section 4.3, **Paper IV**).

For construction of the second library (designated LII, **Paper II**) the same doped oligonucleotide technology was used as for library LI, although with different pool percentage (see **Paper II**, *Materials and Methods* for details). Library LII was also screened for clones with high-level-expression of *bla*, and also this screening resulted in identification of solely 5′-UTR variants conferring increased protein expression compared to that of the wild-type (**Paper II**, Fig. 1). For all identified 5′-UTR variants β-lactamase activities and *bla* transcript amounts were determined. This study revealed the 5′-UTR variant LII-11 which exhibited the highest observed increase both in protein production (20 times) and transcript level (seven times) (**Paper II**, Fig. 2). The observation of increased transcript levels was surprising as it is known that the role of the 5′-UTR is mainly attributed to translation due to the presence of the SD sequence in this region. In order to investigate the increased transcription further, an effort was made by Berg to identify 5′-UTR variants showing a higher increase in transcript than in protein levels. By screening artificial operon libraries (data not shown) two 5′-UTR variants (LV-1 and LV-2) were obtained (**Paper II**, Fig. 1). Using these variants expression plasmids were constructed with *bla* as reporter gene, as before. *bla* transcript levels were measured and it was observed that the enhancement at the transcript level was similar to that of the protein level. This indicates that while transcription was elevated, the efficiency of β-lactamase translation remained similar to the wild type.

Based on this finding the observed rise in the transcript could be the outcome of i) increased transcript stability, as it is known that translational activity protects the transcript from degradation (Joyce & Dreyfus 1998; Briani *et al.* 2008) or ii) increased transcription rate. In order to investigate these two possibilities 5′-UTR variants LV-1 and LV-2 were used in a stability study. This study relies on the fact that tRNA is virtually stable, and when it is transcriptionally fused to a gene it can reflect its transcript formation. Based on the results of this study the observed transcriptional stimulation was not due to increased stability but due to increased transcript formation, since the amount of tRNA was higher in both of the LV constructs compared to that of the wild-type. In an additional experimental approach, a transcription kinetics study was conducted for the variant LV-2 (**Paper II**, Fig. 5) and the wild-type. The transcript accumulation was monitored by qRT-PCR upon induction, which revealed that the transcript accumulation rate was significantly higher in the LV-2 variant compared to that of the wild-type. This outcome was further investigated using a fitting model for transcript accumulation by Nigel Burroughs (see **Paper II**, Fig. S1) which also supported the conclusions based on the tRNA studies (**Paper II**, Fig. S1b).

It is known that the 5′-UTR harbours features that affect translational and post-translational processes. These features include the presence of weak or strong secondary structures (Pfleger *et al.* 2006), target sequences for small non-coding RNAs (sRNAs) (Urban & Vogel 2007) and binding regions for S1 protein (Briani *et al.* 2008). However, hitherto no study indicated involvement of the 5′-UTR in the transcriptional processes. The study reported here, therefore, reveals a critical role of the 5′-UTR in determining transcription rates and provides new insights in gene expression processes with biotechnological and biological importance.

4.1.2 Identification of low-level expression 5′-UTR variants

In addition to studies investigating high-level-expression 5′-UTR variants it was also decided to study variants conferring low levels of protein expression. Such variants were envisioned to have potential applications in pathway engineering studies, where control of expression at low levels is often essential.

Library LI was screened for clones that were not able to grow on medium with ampicillin concentrations exceeding 400 µg/mL. Fifty-two clones with the corresponding phenotypes were selected for sequencing. Thirty-three unique 5′-UTR variants with two to seven point/deletion mutations were identified. Interestingly, 31 out of these 33 variants carried point mutations within the SD sequence (Fig. 4.2a). This finding is in agreement with previous findings showing that mutations within this region reduce the affinity to the ribosome and thus result in reduced expression (Jiong *et al.* 2002). However, the remaining two variants LI-19 and LI-50, (Fig. 4.2b) also exhibited down-regulated expression while having no changes within the SD sequence.

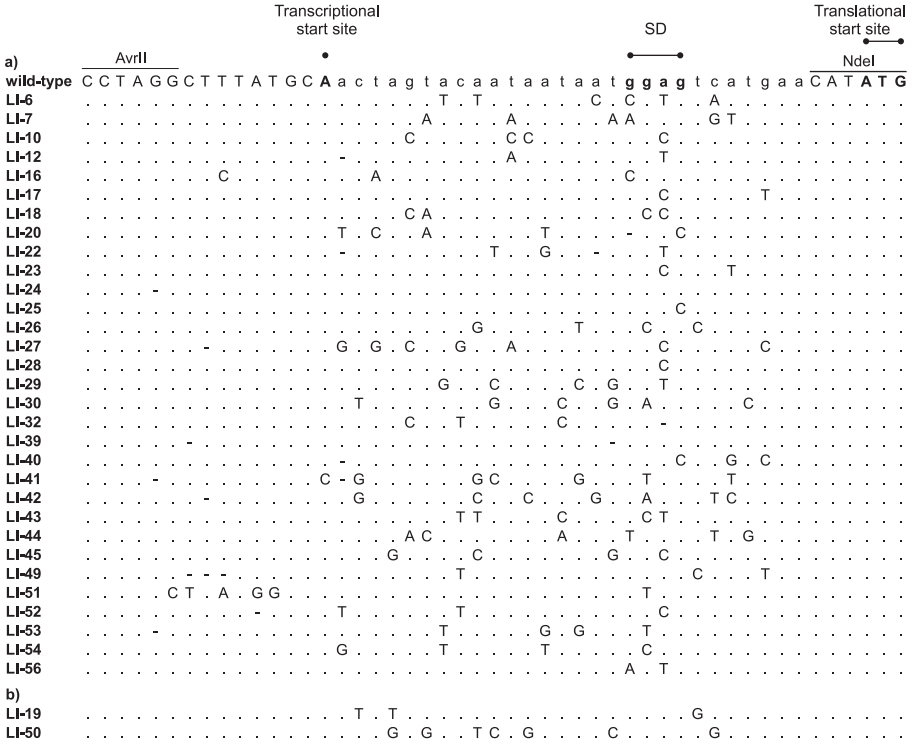

Figure 4.2: 5′-UTR sequences of variants identified in the library LI. Identical bases are indicated by dots, and deletions with dash. Transcriptional and translational start sites are indicated. Putative Shine-Dalgarno sequence is shown as SD. Bases subjected to mutagenesis are written in lower case. The indicated restriction enzyme sites are unique.

In order to further investigate the role of the 5′-UTR, particularly of regions flanking the SD region, a new library was constructed (designated DI, **Paper III**, Fig. 1b) in which the bases close to the TrSS (due to the same reason as described previously) and the SD sequences were excluded from mutagenesis. Screening of this library identified 38 clones and five low-level expression variants were selected randomly for further studies. Each variant carried two to five point mutations in the 5′-UTR (**Paper III**, Fig. 1). Ampicillin resistance phenotype of these five variants varied from 10 µg/mL (DI-3) to 60 µg/mL (DI-8) ampicillin under induced conditions. In the absence of induction none of the corresponding clones grew in the presence of more than 5 µg/mL ampicillin in agar medium (**Paper III**, Fig. 2a). *bla* transcript levels of the five selected variants were measured by qRT-PCR (**Paper III**, Fig. 2a) and revealed that *bla* transcript levels varied from 43% (DI-7) to 79% (DI-8) relative to the wild type. The presence of substantial levels of transcript but simultaneous loss of the final protein product indicated translational inefficiency due to the mutations within the 5′-UTR. Enzyme measurements could

not be performed for these variants since the β-lactamase enzyme assay is not sensitive enough to detect such low levels of expression. Therefore it was wondered whether the ratio of active vs inactive protein was affected as this would result in reduced ampicillin resistance. To assess the β-lactamase protein production, a western blot analysis was performed which revealed that ampicillin resistance levels were consistent with band intensities observed for the different variants (**Paper III**, Fig. 2b).

Since the loss of translation was only shown using *bla* as reporter gene it was questioned whether the identified 5′-UTR low-expression variants would exhibit the same phenotype for other coding sequences as well. To test this idea the *bla* gene was substituted with *luc* and *celB* and enzyme activities were measured for the corresponding five 5′-UTR variants. Interestingly, these variants also exhibited low-level expression phenotypes (**Paper III**, Fig. 3).

However, a dramatic difference was observed when the transcript levels were measured both for the *luc* and the *celB* gene. The relative *luc* transcript levels were found to be similar to those of the constructs with the *bla* gene. In contrast, *celB* transcript levels were barely detectable (**Paper III**, Fig. 4a). To characterise this reduction in transcript level further, the same tRNA approach that was described previously was used. Based on these experiments it became evident that transcription of the *celB* gene was terminated before the polymerase reached the fused tRNA gene. The measured level of tRNA was much lower than that of the construct with the wild-type 5′-UTR (**Paper III**, Fig. 4b). Intriguingly, transcriptional loss was observed for all five 5′-UTRs in combination with the *celB* gene, however, this phenomenon was not visible when the wild-type 5′-UTR was used. To investigate if transcription termination could be the reason for the loss of *celB* transcript BCM, a natural inhibitor of Rho, was used for transcription termination studies. BCM blocks Rho-dependent termination by binding to a region close to the ATPase of Rho and thus prevents the ATP hydrolysis (Skordalakes *et al.* 2005) which is crucial for Rho activity (see section *Termination*). High concentrations of BCM are reported to be lethal in *E. coli* K-12 as Rho is an essential protein for cell function. However, sub-lethal doses of BCM are enough to prevent Rho-termination *in vivo* (Skordalakes *et al.* 2005; Peters *et al.* 2009). Clones carrying the wild-type and the 5′-UTR construct DI-7 with *celB* were grown up to OD_{600} 0.1 in liquid media. Before induction samples were drawn from both cultures and used as reference in the qRT-PCR analysis. After reference sampling the bacterial cultures were split in two and while the first culture was treated with the inducer, the second culture was treated with the inducer plus BCM (100 µg/mL final concentration). Cultures were grown for 30 minutes and samples were collected at the end of the incubation period. All samples were analysed by qRT-PCR (Table 4.1a). The inhibition of the Rho activity resulted in a drastic increase in *celB* transcript from the DI-7 construct compared to that of the wild-type. It was, however, questioned whether such levels of transcriptional recovery

would also be valid for other genes. Therefore this study was repeated for both of the constructs (with the wild-type and variant DI-7) with the *bla* gene (Table 4.1b). The high transcriptional recovery that was observed for the *celB* transcript was not seen for the *bla* gene. Based on this observation it was concluded that Rho leads to premature transcription termination of *celB* mRNA synthesis, and in addition this effect is strongly stimulated when translation is inefficient (in the presence of the DI 5′-UTR sequences).

Plasmid	Condition		Constructs	
			wild-type	DI-7
a) pLB11	uninduced	−BCM	1	1
	induced	−BCM	94± 6	16± 1
	induced	+BCM	1165± 115	1942± 142
b) pIB11	uninduced	−BCM	1	1
	induced	−BCM	44± 5	9± 1
	induced	+BCM	472± 18	226± 13

Table 4.1: Relative *celB* transcript amounts for the constructs with the wild-type and DI-7 5′-UTR sequences. Transcript amounts are relative to the uninduced samples which are set as 1. Respective standard deviations are shown.

Both phospoglucomutase and β-lactamase enzyme essays are not sensitive enough to precisely determine low levels of protein, particularly in the uninduced stage. However, the enzyme assay for luciferase is very sensitive and enzyme activities can be accurately monitored at both very low and high levels of protein. Therefore, luciferase enzyme activities of the cells carrying the plasmids with the three DI 5′-UTR constructs (DI-3, -7, and -8) were determined, both under induced and uninduced conditions (**Paper III**, Fig. 5). In addition to low-level-basal-expression from these three variants, it was observed that the variants still exhibited high inducibility characteristics (**Paper III**, Fig. 5). Even with strongly reduced background expression, the level of induction was not compromised.

To assess the application potentials of the identified low-level expression 5′-UTRs, the variants DI-3 and DI-8 were used in an industrial project. The aim of this project has been to heterologously express and produce carotenoid (xanthophyll) in *E. coli*. Six genes are involved in the biosynthesis of xanthophyll. Three of these genes are organised in an operon and for this project they were placed under control of the XylS/*Pm* system in its wild type form and expressed in *E. coli*. The remaining three genes were co-expressed from their natural promoters. Roman Netzer and Marit Stafsnes showed that the wild type system is capable of producing about 2mg/L of carotenoid under 0.5 mM induced conditions. However, leakage levels were found to be high under uninduced conditions (see Fig. 4.3). In order to decrease basal expression levels the two 5′-UTR variants DI-3 and DI-8 were introduced to the wild type system and the final carotenoid production was

measured. It was found that the system containing the DI-3 variant exhibited a phenotype that was tight in background expression while production levels under induced conditions were not satisfactory. In contrast, the system carrying the DI-8 variant yielded low background expression under uninduced and high levels of production under induced conditions (Fig. 4.3). The production level under induced conditions was a bit lower than that of the wild type but due to the drastic reduction in background expression this was found to be satisfactory.

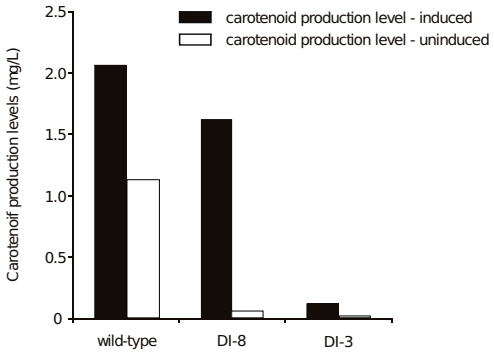

Figure 4.3: Carotenoid production levels (mg/L) under induced (0.5 mM) (black bars) and uninduced (open bars) conditions.

The inducer-concentration-dependent-induction feature of *Pm* has been reported (Ramos *et al.* 1986; 1997; Winther-Larsen *et al.* 2000a). These studies indicated that expression from *Pm* can be regulated by different inducers and concentrations thereof. In this work the inducibility features of the low-expression 5′-UTR variants in response to different inducer concentrations was also studied. For this study, two constructs, wild-type and DI-7, were selected and their inducibility was assessed at inducer levels varying from 10 nM to 2 mM. It was observed that the XylS/*Pm* system can be induced with inducer levels as low as 1 µM (**Paper III**, Fig. 6) and still exhibited induction ratios of about two. Whereas, induction with 2 mM inducer resulted in induction ratios of about 97, and 147 for the cells carrying the constructs with the wild-type, and the variant DI-7, respectively.

For many applications achieving the highest possible expression at the transcript level may not always lead to high levels of protein production. Problems that are associated with such approaches has been briefly mentioned in *Introduction* of **Paper III**. Tight controlled expression with no or as little as possible expression in the absence of induction is known to be critical for cloning and expression of toxic genes and regulation of different genes in metabolic pathways or in cascade systems. Based on the findings in this work low levels of basal expression with

wide-range induction characteristics and apparent gene independent feature of the 5′-UTRs makes these variants suitable for a wide array of applications.

4.2 Identification of Pm Variants with Low Basal-Level Expression

The *Pm* promoter has relatively low levels of background expression in the absence of induction. Despite the already low leakage expression It was aimed to identify *Pm* variants with even tighter expression characteristics by using a synthetic oligo library where the –10 region of *Pm* was mutagenised. Also in this screening *bla* served as a reporter gene and *E. coli* DH5α as a host. The variant library was constructed by randomly mutagenising the coding strand using an oligonucleotide mixture: 5′-CTAGAA133224122224413324441432AA-3′. The numbers in the oligonucleotide indicate the doping percentages of the nucleotides: 1= 88% A, 4% C, 4% G, 4% T; 2= 4% A, 88% C, 4% G, 4% T; 3= 4% A, 4% C, 88% G, 4% T; 4= 4% A, 4% C, 4% G, 88% T. The created library consisted of about 13,000 clones. In total 1000 clones were screened for identification of *Pm* variants that were not be able to grow in agar medium supplemented with more than 100 µg/mL ampicillin. Eleven of a total of 48 identified clones were selected randomly and subjected to sequencing. Sequence analysis revealed six unique *Pm* variants conferring a down regulated phenotype (Fig. 4.4).

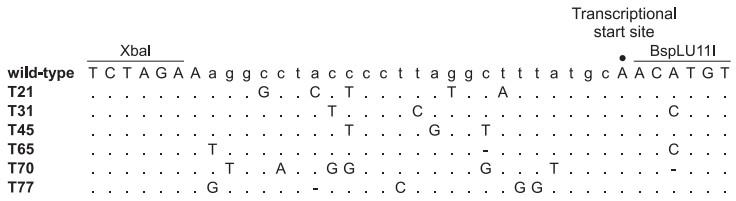

Figure 4.4: Identified low-level expression *Pm* promoter sequences. Bases subjected to mutagenesis are written in lower case. Identical bases are indicated by dots, and deletions with dash. Transcriptional start site and unique restriction enzyme sites are indicated.

In addition, the sequencing results revealed that the BspLU11I restriction site that is located within the part of the sequence that was not mutagenised, carried mutations. These were, point mutations for the variants T31 and T65, and a deletion mutation in the variant T70. A similar observation (mutations outside the mutagenised region) was made in a parallel study conducted by Ingrid Bakke (Bakke *et al.* 2009) aiming at identification of high-expression *Pm* variants. The introduction of these mutations may be due to either technical failures during the synthesis of the oligos or processes in the cells after transformation.

To detect low levels of expression, an exchange of *bla* with the *luc* gene was

planned. However, the variants with the highest induction ratios, T65 and T70, were carrying the destructed BspLU11I restriction site, and, therefore, the idea was not further pursued. Instead, an ampicillin plate assay was performed using agar medium containing as low as 2 µg/mL ampicilin to characterise precise induced (2mM) and uninduced expression levels (the host DH5α does not grow in the presence of 1 µg/mL ampicillin) (Fig. 4.5).

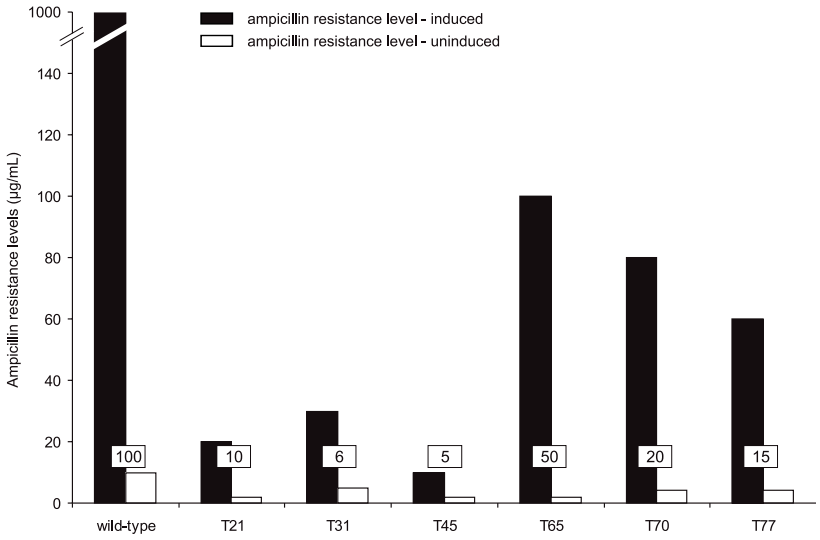

Figure 4.5: Ampicillin resistance levels (µg/mL) under induced (black bars) and uninduced (open bars) conditions. The ratios between the induced and uninduced levels are displayed directly on the bars.

The ratios between induced and uninduced levels were calculated and it was found that the identified *Pm* variants lost their high induction capabilities. While the reduction in ampicillin resistance levels under uninduced conditions was about five times less for the variants T21, T45, and T65 compared to the wild type, the reduction at the induced level was about 10 times for the variant T65 and 100 times for the variant T45. These *Pm* promoter variants represent promising candidates for studies where less leaky (compared to the wild type) and moderate induction rates are required. In studies such as pathway engineering or gene-circuit design possessing tools that allows wide range of selection is always beneficial.

Studies on the 5′-UTR and the *Pm* promoter –10 region led to identification of variants exhibiting low levels of expression in the absence of induction. Considering the induction ratios observed among the 5′-UTR and *Pm* variants, utilising the 5′-UTR as a tool for decreasing basal-level expression appears to be a better choice. It is important to note, however, that different ampicillin concentrations were used in identification of *Pm* and 5′-UTR variants. For *Pm* variants, 100 µg/mL ampicillin were chosen, whereas 400 µg/mL was used for the 5′-UTR variants.

4.3 Combining mutations in three different regulatory elements of the XylS/Pm system

The studies described so far were concentrated on identification of individual variants of the 5′-UTR and the *Pm* region of the XylS/*Pm* expression system. As a result of these and other studies on this system, a large collection of variants (*xylS*, *Pm*, and 5′-UTR) with different expression characteristics has been identified (Aune *et al.* 2009; Bakke *et al.* 2009; **Paper II**; **Paper III**). To identify variants of these regulatory DNA elements the libraries were created focusing on one element at a time, i.e. for identification of *xylS* variants, both the *Pm* and the 5′-UTR were kept in the wild-type sequence. Having studied all these elements separately it was wondered if the system could be further improved by combining different variants of these elements. For this purpose a set of already known variants of the *xylS*, *Pm*, and 5′-UTR (See **Paper IV**, Fig. 1 for sequences) were chosen. The aim was to establish a *xylS*/*Pm* expression cassette allowing highest possible levels of expression per gene copy, which subsequently might be incorporated into the chromosome of *E. coli*.

Initially, three different variants of each element were selected that separately conferred high, wild-type (moderate), and low levels of expression. In a first screening it was observed that independent of the particular type of combination tested, elements which separately (in an otherwise wild type background) reduced expression had a negative effect in all combinations tested. Therefore these combinations were excluded, and only the combinations that stimulated the expression levels (both at the level of transcription and translation) were studied further. A total of eight such constructs were then selected; wild-type, three constructs containing one variant element (C2, C4, C10), three constructs with each possible combination of two variant elements (C6, C11, C13), and one construct (C15) in which all three elements were combined (**Paper IV**, Table 1).

To monitor expression from the different *xylS*/*Pm* variants, *bla* was used as reporter. The *bla* gene was in all cases expressed from the relevant variant version of plasmid pTA16, such that the only differences between each construct was in the variant control elements, and the combinations of elements could easily be constructed by using unique restriction sites specific for each element (**Paper IV**, Fig. 2). The three variant control elements used were StEP-13, which carries five substitution mutations in *xylS* (Aune *et al.* 2009), a promoter variant designated ML2-5 which carries five point mutations in *Pm* (Bakke *et al.* 2009), and a 5′-UTR variant designated H39 with two point mutations.

In the initial tests of the different constructs, a low concentration of inducer (0.1 mM) was used to ensure that potential host toxicity would not occur due to extremely elevated expression levels. First, different constructs were compared by plating the corresponding cells on agar medium supplemented with varying concentrations of ampicillin. Under these conditions the upper tolerance level under

induced conditions was approximately 1 (C2), 1 (C10), and 2 mg/mL ampicillin (C4) (**Paper IV**, Table 1). Cells carrying the wild-type construct were shown to tolerate only 0.2 mg/mL. Interestingly, all three combinations of the variant elements, *Pm* with 5′-UTR (C6), *xylS* with 5′-UTR (C13), and *xylS* with *Pm* (C11) resulted in enhanced expression levels, concluding from the maximum ampicillin tolerance levels of 5, 12, and 3 mg/mL, respectively. This effect was further confirmed by analysing the phenotype of a strain containing a plasmid in which all three elements were combined (C15). This strain tolerated 15 mg/mL of ampicillin. It should be noted that the uninduced resistance levels also increased for all strains (up to 3 mg/mL for the triple combination).

All the eight constructs discussed above were compared by measuring the amounts of *bla* transcript by qRT-PCR and the corresponding β-lactamase enzyme activities. Based on the results (**Paper IV**, Fig. 3a), it was observed that a combination of two elements in all cases stimulated expression at the protein level over a wide range, spanning from about eight-fold for the *xylS*-*Pm* combination to about 50-fold for the *xylS*-5′-UTR pair (all compared to the wild type). As expected, the constructs with the 5′-UTR variant (H39) displayed the highest increase both at translational and transcriptional level. Furthermore, combination of all three variant elements led to the strongest stimulation, about 80-fold at the translational and about 65-fold at the transcript level. As in other expression systems, individual proteins are expressed at quite varying levels from *Pm*, and β-lactamase is not among the most highly expressed proteins. However, it was surprising to see that the system's expression capabilities could be improved by 80-fold by simply combining the individually identified high-level variants of the control elements.

It was wondered whether the identified mutations would stimulate genes other than *bla*, which had been used as a reporter to identify the variants. Therefore *bla* was exchanged with the *celB* gene which, in contrast to *bla*, is known to be very efficiently expressed from the wild type *xylS*/*Pm* (Blatny *et al.* 1997b). Relative *celB* transcript levels were determined by qRT-PCR, and CelB enzyme activities were measured (**Paper IV**, Fig. 3b). This work was conducted by Friederike Stüttgen. Interestingly, the *celB* expression was also significantly stimulated (about six-fold at the protein level) by the cassettes containing the variant control elements, although not to the same extent as for *bla* expression. The latter finding was expected due to the demonstrated efficiency of the wild type system for this particular protein (Blatny *et al.* 1997b). Another interesting observation was that the highest level of stimulation at the protein level (six-fold) was achieved by three of the constructs, possibly indicating that some kind of upper-limit had been reached. Strikingly, *celB* transcription tended to be more stimulated than the corresponding protein level. The transcript amount was highest in the strain containing the plasmid with three variant elements (C15) (about 15-fold). Thus, it appears as if the proposed upper limitation may be caused by the inability of the

cell to translate all the transcripts produced. It is tempting to speculate that the apparent overload of transcripts that cannot be efficiently translated might lead to reduced transcript stability, possibly explaining the even stronger stimulation of *bla* transcript accumulation.

The results described here clearly show that it is possible to use apparently context-independent combinations of mutations in the expression cassette to strongly stimulate protein production. A system designed using this strategy out-competes by 80-times the current industrially used wild type system. For many proteins the combination approach will probably be sufficient to reach the maximum capacity of the host. To what extent these methods can be used to eliminate the need for plasmids in recombinant protein production is currently under investigation. For this purpose, integration of the identified triple combination construct into the *E. coli* chromosome is in preparation.

4.4 Comparison of the xylS/Pm expression cassette performance with the bacteriophage T7 and P_{BAD} promoters

A wide range of positively regulated commercial expression systems are available for bacterial recombinant gene expression (**Paper I**). These expression systems commonly consist of an inducible promoter controlling the recombinant gene encoded on a multicopy plasmid vector. Ideally, the expression systems will permit a tight control of expression at all relevant levels. However, the promoter and vector characteristics may vary considerably. For bacterial recombinant gene expression two systems are most commonly utilised; the pET system, containing the versatile T7 phage promoter; and the P_{BAD} system containing the P_{BAD} promoter regulated by the AraC protein. These two commercial systems were compared to the XylS/*Pm* system.

The pET system requires the T7 RNAP to be expressed in the host and the *E. coli* BL21 (DE3) lysogen was therefore chosen as expression strain fro the comparison study. Each system requires different inducer molecules. *m*-toluic acid, IPTG, and arabinose are used for the XylS/*Pm*, pET and P_{BAD} systems, respectively.

The XylS/*Pm* system is located on a plasmid harbouring the minimal replicons of the broad-host-range plasmid RK2. The plasmid with the wild-type *trfA* gene has a copy-number of about 4-7 copy per chromosome in *E. coli*. To have a common platform for performance comparison it was decided to test the systems initially with a low copy-number backbone. For this reason the RK2 replicon was chosen. First, the two selected system's expression cassettes, composed of the regulator gene and the promoter (*lacI* with T7, and *araC* with P_{BAD}), were PCR amplified from pET and P_{BAD} plasmids, respectively, with primers flanking unique

restriction sites. The corresponding PCR fragments were cloned into the RK2 plasmid. Hence, all constructed plasmids, pTA16-XylS/*Pm*, -T7, and -P_{BAD}, carried the same backbone. For all three constructs, the *bla* gene was under the control of the respective promoter. These three plasmids were transferred and stably established in the BL21 strain.

In an initial study to characterise the expression profiles of the systems, strains carrying the cognate pTA16 plasmid were plated on agar medium supplied with varying concentrations of ampicillin. Based on the maximum resistance levels observed the XylS/*Pm* system was found to be the tightest at the uninduced, and strongest at the induced level (Table 4.2a).

	Constructs	Maximum ampicillin resistance	
		uninduced (mg/mL)	induced (mg/mL)
	pTA16-XylS/*Pm*	0.02	0.75
a)	pTA16-T7	0.05	0.1
	pTA16-P_{BAD}	0.05	0.5
	pJPK12-XylS/*Pm*	0.1	2
b)	pJPK12-T7	0.1	0.2
	pJPK12-P_{BAD}	0.1	1

Table 4.2: Expression profiles of *E. coli* BL21 cells containing the indicated constructs. Phenotypes are given as maximum ampicillin resistance levels (mg/mL) under induced and uninduced conditions.

After observing the resistance levels of the constructs it was decided to compare them based on *bla* transcript levels. The strains containing the plasmids were grown up to OD_{600} 0.1 and the cultures were then split into two. One of the cultures was induced with the cognate inducer (XylS/*Pm*, and P_{BAD} at the 1mM, and T7 at 10 mM levels) and the second culture was left uninduced. Both cultures were then grown for four hours and harvested for transcription analysis (conducted by Hege Sletvold). The results of the qRT-PCR analysis are shown in Fig. 4.6. For transcriptional comparison the XylS/*Pm* values were set to 1. Based on this study the background expression from the T7 system was found to be three times higher than that of the XylS/*Pm* system, while the P_{BAD} system's background expression was about half of that of the XylS/*Pm* system. The P_{BAD} system is known for its low background expression and this study also confirmed that. However, the low background expression was not reflected on the resistance level observed on ampicillin plates. Based on the transcript measurements at induced levels, the transcript amount of the *bla* gene was highest under the control of the XylS/*Pm* system whereas it was about 8 and 10 times less under control of the T7 and P_{BAD} promoters, respectively. The induction rate and hence stable expression was found to be the highest from the XylS/*Pm* compared with the T7 and the P_{BAD} promoter systems.

It is known that the strength of the T7 and the P_{BAD} systems lies, to some

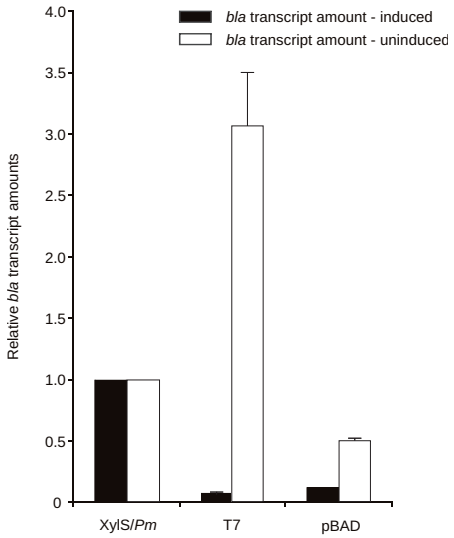

Figure 4.6: Relative *bla* transcript levels under induced (black bars) and uninduced (open bars) conditions in plasmid pTA16. Transcript amounts are normalized against the XylS/*Pm* values which is arbitrarily set to 1.

extent, in their high copy-number backbones. For additional comparison a high copy number replicon, SC101, was selected (Cohen & Chang 1977). This replicon has a copy-number of 40 per chromosome in *E. coli*. All three expression cassettes were separately constructed with the SC101 replicon, resulting in plasmids pJPK12-XylS/*Pm*, -T7, and -P_{BAD}. The expression levels were monitored by detecting resistance levels on ampicillin media (Table 4.2b) and *bla* transcripts were measured by qRT-PCR (Fig. 4.7). Concerning ampicillin resistance and *bla* transcript levels the outcome was similar to the previous observations where the low-copy number replicon was used. The XylS/*Pm* system in its wild type form was found to confer the lowest level of background expression under undinduced and the highest level of protein production under induced conditions.

So far these systems have only been investigated using the *bla* reporter gene and two different replicons. In all tested set-ups the XylS/*Pm* system out-competed the two commercial systems with higher protein production under induced and lower protein production under uninduced conditions. However, the P_{BAD} expression system exhibited tighter transcript levels in the absence of induction compared to that of the XylS/*Pm* system. Ongoing studies investigate the behaviour of these systems using different genes and in response to different concentrations of inducers.

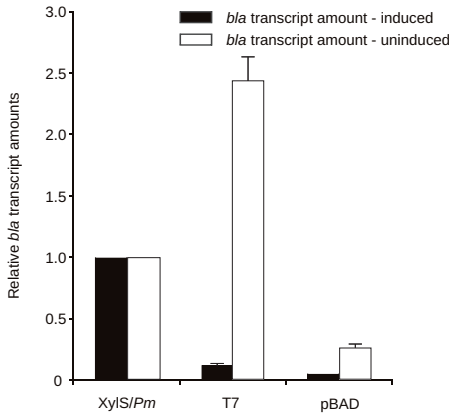

Figure 4.7: Relative *bla* transcript levels under induced (black bars) and uninduced (open bars) conditions in plasmid pJPK12. Transcript amounts are normalized against the XylS/*Pm* values which is arbitrarily set to 1.

4.5 Performance of 5′-UTR and Pm Variants in Pseudomonas fluorescens

It is known that different hosts exhibit varying expression characteristics. As it has been mentioned throughout this thesis, the broad-host-range feature of the XylS/*Pm* system makes it suitable for use in across species. In order to investigate the host dependency of *Pm* and 5′-UTR elements it was decided to test variants of these elements in *Pseudomonas fluorescens* in addition to *E. coli*. For this study the reporter gene was exchanged with *luc*, as the *bla* gene is not suitable for use in *P. fluorescens*.

Two high- (LII-9 and LII-12) and two low-level expression 5′-UTR variants (DI-1 and DI-8) in addition to two high-level-expression *Pm* variants (ML1-16 and ML2-5) were selected for this study. The luciferase activities of the corresponding clones carrying these variants were measured at 1 mM induced level (Fig. 4.8). Wild-type activities were set to 1. It was observed that the 5′-UTR variants presented both their high- and low-level expression phenotypes in *P. fluorescens*. However, the *Pm* variants conferring high-level-expression in *E. coli* did not exhibit their high-expression features in *P. fluorescens*. This finding indicates that 5′-UTR's might represent a host independent feature. Ongoing studies aim to identify *Pm* variants in *P. fluorescens* and compare their phenotypic features in *E. coli*. These studies also include Gram-positive hosts to assess both functionality of the XylS/*Pm* system and whether the observed host independency of the 5′-UTR is valid in distantly related hosts as well.

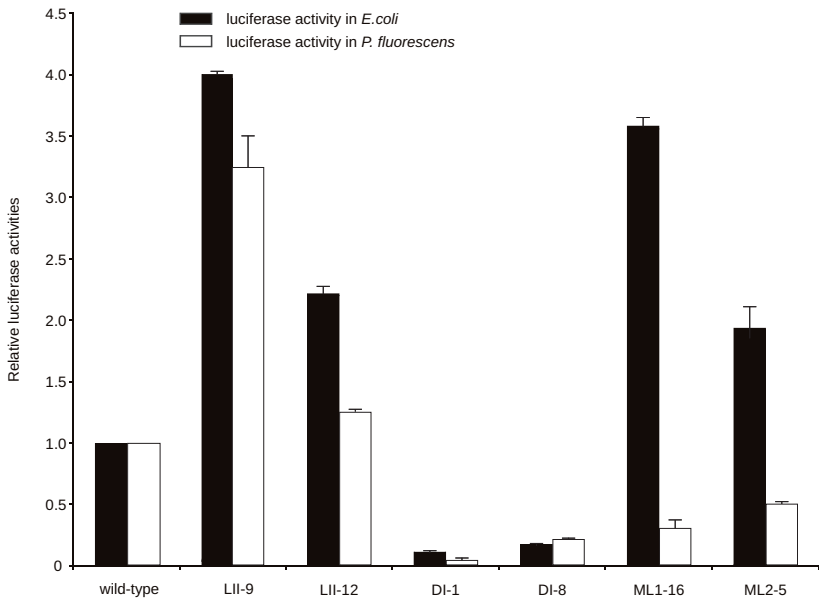

Figure 4.8: Relative luciferase activities of constructs in *E. coli* (black bars), and in *P. fluorescens* (open bars). Activity measurements are normalized against the wild- type value, which is arbitrarily set to 1.

5

Future Perspectives

This chapter gives an outlook on possible further studies based on findings presented in this thesis.

5.1 Artificial Operon Libraries

Throughout the thesis study, synthetic oligonucleotide libraries were used for identification of 5′-UTR variants. For instance, in identification of 5′-UTR high-expression variants libraries were plated on agar medium containing ampicillin at different concentrations. Colonies growing on plates with high ampicillin concentrations were characterised as translationally enhanced. Comparison of these clones against the wild type system revealed that stimulation at the protein production level was higher than stimulation at the transcript level. This indicates that these 5′-UTR variants confer increased translatability, but not to the same extend transcriptability.

The 5′-UTR variants can be categorised into three classes based on the stimulation that they confer, i.e. those that are (i) translationally stimulated (**Paper II**, LII variants), (ii) transcriptionally stimulated or (iii) stimulated both at the transcription and translation (**Paper II**, LV variants). Based on these observations it was concluded that the 5′-UTR possesses characteristics that influence both transcriptional and translational processes (**Paper II**). The current screening strategy does not allow identification of a 5′-UTR variant which is stimulated only at the transcriptional or only at translational level. Therefore, for selection of variants stimulating either of the two processes, two operon constructs were created (Fig. 5.1). For transcriptional selection (Fig. 5.1a), the 5′-UTR library was placed immediately upstream of the first gene, *celB*, and a wild-type 5′-UTR was placed upstream of the second gene, *bla*. Clones growing on high ampicillin concentrations would then be expected to carry a 5′-UTR variant with high transcriptability due to formation of high levels of transcript generated up-

stream of the *celB* gene. Since the 5′-UTR that is selected for in this screening will not mirror its translatability, i.e. high levels of phosphoglucomutase production, the highest β-lactamase production could only be achieved if high levels of transcript were formed.

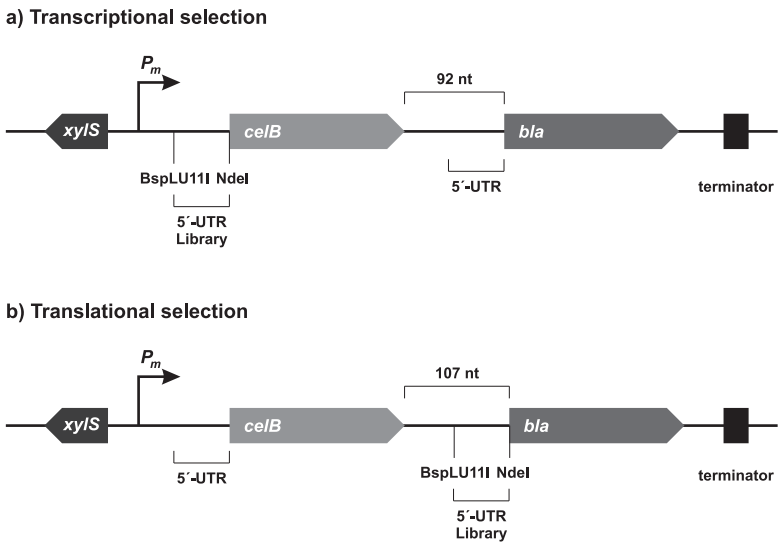

Figure 5.1: Schematic representation of the artificial operons. Operon systems designed for transcriptional selection (a) and for translational selection (b) are depicted. The *Pm* promoter is indicated with an arrow, the gene coding for the positive regulator of *Pm*, *xylS*; gene coding for phosphoglucomutase, *celB*; and the gene coding for β-lactamase, *bla* are shown with pointed boxes indicating transcriptional direction. Positioning of the 5′-UTRs and the 5′-UTR libraries and the unique restriction sites are indicated.

For translational selection, the same operon structure was used. However, for this construct the 5′-UTR library was placed immediately upstream of the second gene (Fig. 5.1b). The transcript would be generated depending on the context of upstream of the *celB* gene. Clones growing on high ampicillin concentrations would then be expected to carry a 5′-UTR variant with high translatability as the identified 5′-UTR would not be involved in transcript formation. This way of screening would also identify variants yielding highest protein production for a given amount of transcript.

The outcome of these artificial operon studies will hopefully enhance our understanding of the dual role of the 5′-UTR in gene expression.

5.2 Computational Analysis of the 5′-UTR Variants

The work that was presented so far resulted in a data set consisting of over 100 5′-UTR DNA sequences with a range of phenotypic characteristics. Although a number of variants were used in the studies presented, the digital data generated has not been analysed. Computational studies provide different approaches to identify sequence features that are otherwise not immediately apparent. Examples include correlation of expression levels with folding energy of 5′-UTRs (Ringnér & Krogh 2005), correlation of mRNA expression levels with folding free energy and codon usage bias (Jia & Li 2005) and correlation of gene expression, codon usage bias and GC content with mRNA stability (Eck & Stephan 2008). However, most of these studies are carried out using genome wide data and the finding of these studies are generally not confirmed by *in vivo* studies.

Each identified 5′-UTR sequence corresponds to different levels of expression. By categorising the sequences based on features like folding energy vs. expression levels, it might be possible to identify certain correlations. Based on these correlations, random sequences can then be generated containing the identified feature and expression levels of the generated sequence can be measured *in vivo*. To optimise the outcome, this work could be carried out iteratively and such studies may provide a more detailed understanding of the correlation between sequences and observed phenotypes.

5.3 Studies on Abortive Initiation

The finding that the 5′-UTR determines transcription rates provides new insights into gene expression processes. However, the biological mechanisms underlying these observations are not known. In transcription initiation RNAP covers the region from –50 and +20 (Mooney *et al.* 1998) and it is known that the initial transcribed region +1 to +20 can play a role in promoter escape (Hsu *et al.* 2006). The study by Hsu presents findings that increased number of purines in the initially transcribed region results in increased promoter escape (see section *Initiation*) rates. An introduction of purines, however, was not observed for the 5′-UTR variants with enhanced expression profiles (**Paper II**, Fig. 1). Considering the observed increased transcript formation, abortive initiation may occur to lesser extent among the high-expression variants. This idea will be tested for the identified 5′-UTR variants using both *in vivo* (Goldman *et al.* 2009) and *in vitro* (Hsu 2009) abortive initiation studies.

References

Abraham, E. & Chain, E. (1940). An Enzyme from Bacteria able to Destroy Penicillin. *Nature*, **146**, 837.

Agirrezabala, X., Lei, J., Brunelle, J.L., Ortiz-Meoz, R.F., Green, R. & Frank, J. (2008). Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol. Cell*, **32**, 190–197.

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

Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. (2005). Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. USA*, **102**, 12678–12683.

Alper, H., Moxley, J., Nevoigt, E., Fink, G.R. & Stephanopoulos, G. (2006). Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science*, **314**, 1565–1568.

Altenbuchner, J. & Mattes, R. (2005). *Escherichia coli*. In G. Gelissen, ed., *Production of Recombinant Proteins. Novel Microbial and Eukaryotic Expression Systems.*, chap. 2, 7–43, Wiley-VCH, Weinheim.

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

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

Artsimovitch, I. & Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. USA*, **97**, 7090–7095.

Aune, T.E.V., Bakke, I., Drabløs, F., Lale, R., Brautaset, T. & Valla, S. (2009). Directed evolution of the transcription factor XylS for development of improved expression systems. *Microbial Biotechnol.*, **In pres.**

Avery, O.T., MacLeod, C.M. & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus type III. 1944. *J. Exp. Med.*, **79**, 137–158.

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.

Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*, **289**, 905–920.

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

Bar-Nahum, G. & Nudler, E. (2001). Isolation and characterization of sigma(70)-retaining transcription elongation complexes from *Escherichia coli*. *Cell*, **106**, 443–451.

Barne, K.A., Bown, J.A., Busby, S.J. & Minchin, S.D. (1997). Region 2.5 of the *Escherichia coli* RNA polymerase sigma70 subunit is responsible for the recognition of the 'extended−10' motif at promoters. *EMBO J.*, **16**, 4034–4040.

Berghöfer-Hochheimer, Y., Lu, C.Z. & Gross, C.A. (2005). Altering the interaction between sigma70 and RNA polymerase generates complexes with distinct transcription-elongation properties. *Proc. Natl. Acad. Sci. USA*, **102**, 1157–1162.

Bitter, G.A. & Egan, K.M. (1984). Expression of heterologous genes in *Saccharomyces cerevisiae* from vectors utilizing the glyceraldehyde-3-phosphate dehydrogenase gene promoter. *Gene*, **32**, 263–274.

Blasina, A., Kittell, B., Toukdarian, A. & Helinski, D. (1996). Copy-up mutants of the plasmid RK2 replication initiation protein are defective in coupling RK2 replication origins. *Proc. Natl. Acad. Sci. USA*, **93**, 3559–3564.

Blatny, J.B., Brautaset, T., Winther-Larsen, H.C., Haugan, K. & Valla, S. (1997a). Construction of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. *Appl. Environ. Microbiol.*, **63**, 370.

Blatny, J.B., 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.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. & Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453–1474.

Boelens, R. & Gualerzi, C.O. (2002). Structure and function of bacterial initiation factors. *Curr. Protein Pept. Sci.*, **3**, 107–119.

Boni, I. (2006). Diverse molecular mechanisms of translation initiation in prokaryotes. *Mol. Biol.*, **40**, 587.

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

Borukhov, S., Sagitov, V. & Goldfarb, A. (1993). Transcript cleavage factors from *Escherichia coli*. *Cell*, **72**, 459–466.

Borukhov, S., Lee, J. & Laptenko, O. (2005). Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol. Microbiol.*, **55**, 1315–1324.

Bowser, T.E., Bartlett, V.J., Grier, M.C., Verma, A.K., Warchol, T., Levy, S.B. & Alekshun, M.N. (2007). Novel anti-infection agents: small-molecule inhibitors of bacterial transcription factors. *Bioorg. Med. Chem. Lett.*, **17**, 5652–5655.

Boyle, P.M. & Silver, P.A. (2009). Harnessing nature's toolbox: regulatory elements for synthetic biology. *J. R. Soc. Interface*, **6 Suppl 4**, S535–S546.

Braun, P. & LaBaer, J. (2003). High throughput protein production for functional proteomics. *Trends Biotechnol.*, **21**, 383–388.

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

Briani, F., Curti, S., Rossi, F., Carzaniga, T., Mauri, P. & Dehò, G. (2008). Polynucleotide phosphorylase hinders mRNA degradation upon ribosomal protein S1 overexpression in *Escherichia coli*. *RNA*, **14**, 2417–2429.

Brock, J.E., Paz, R.L., Cottle, P. & Janssen, G.R. (2007). Naturally occurring adenines within mRNA coding sequences affect ribosome binding and expression in *Escherichia coli*. *J. Bacteriol.*, **189**, 501–510.

Brock, J.E., Pourshahian, S., Giliberti, J., Limbach, P.A. & Janssen, G.R. (2008). Ribosomes bind leaderless mRNA in *Escherichia coli* through recognition of their 5′-terminal AUG. *RNA*, **14**, 2159–2169.

Brosius, J., Cate, R.L. & Perlmutter, A.P. (1982). Precise location of two promoters for the b-lactamase gene of pBR322. *J. Biol. Chem.*, **257**, 9205–9210.

Brosius, J., Erfle, M. & Storella, J. (1985). Spacing of the −10 and −35 regions in the tac promoter. Effect on its in vivo activity. *J. Biol. Chem.*, **260**, 3539–3541.

Carpousis, A.J. & Gralla, J.D. (1985). Interaction of RNA polymerase with lacUV5 promoter DNA during mRNA initiation and elongation. Footprinting, methylation, and rifampicin-sensitivity changes accompanying transcription initiation. *J. Mol. Biol.*, **183**, 165–177.

Casadaban, M.J. (1975). Fusion of the *Escherichia coli* lac genes to the ara promoter: a general technique using bacteriophage Mu-1 insertions. *Proc. Natl. Acad. Sci. USA*, **72**, 809–813.

Chang, B., Halgamuge, S. & Tang, S.L. (2006). Analysis of SD sequences in completed microbial genomes: non-SD-led genes are as common as SD-led genes. *Gene*, **373**, 90–99.

Chao, Y.P., Chiang, C.J. & Hung, W.B. (2002a). Stringent regulation and high-level expression of heterologous genes in *Escherichia coli* using T7 system controllable by the araBAD promoter. *Biotechnol. Prog.*, **18**, 394–400.

Chao, Y.P., Law, W., Chen, P.T. & Hung, W.B. (2002b). High production of heterologous proteins in *Escherichia coli* using the thermo-regulated T7 expression system. *Appl. Microbiol. Biotechnol.*, **58**, 446–453.

Chew, L.C., Ramseier, T.M., Retallack, D.M., Schneider, J.C., Squires, C.H. & Talbot, H.W. (2005). *Pseudomonas fluorescens*. In G. Gelissen, ed., *Production of Recombinant Proteins. Novel Microbial and Eucaryotic Expression Systems.*, chap. 3, 45–66, Wiley-VCH, Weinheim.

Ciampi, M. (2006). Rho-dependent terminators and transcription termination. *Microbiology*, **152**, 2515.

Cohen, S.N. & Chang, A.C.Y. (1977). Revised interpretation of the origin of the pSC101 plasmid. *J. Bacteriol.*, **132**, 734–737.

Cohen, S.N., Chang, A.C., Boyer, H.W. & Helling, R.B. (1973). Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. USA*, **70**, 3240–3244.

Consortium, S.G., Consortium, C.S.G., Consortium, N.S.G., Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., dhe Paganon, S., won Park, H., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R., Kim, S.H., Rao, Z., Shi, Y., Terwilliger, T.C., Kim, C.Y., Hung, L.W., Waldo, G.S., Peleg, Y., Albeck, S., Unger, T., Dym, O., Prilusky, J., Sussman, J.L., Stevens, R.C., Lesley, S.A., Wilson, I.A., Joachimiak, A., Collart, F., Dementieva, I., Donnelly, M.I., Eschenfeldt, W.H., Kim, Y., Stols, L., Wu, R., Zhou, M., Burley, S.K., Emtage, J.S., Sauder, J.M., Thompson, D., Bain, K., Luz, J., Gheyi, T., Zhang, F., Atwell, S., Almo, S.C., Bonanno, J.B., Fiser, A., Swaminathan, S., Studier, F.W., Chance, M.R., Sali, A., Acton, T.B., Xiao, R., Zhao, L., Ma, L.C., Hunt, J.F., Tong, L., Cunningham, K., Inouye, M., Anderson, S., Janjua, H., Shastry, R., Ho, C.K., Wang, D., Wang, H., Jiang, M., Montelione, G.T., Stuart, D.I., Owens, R.J., Daenke, S., Schütz, A., Heinemann, U., Yokoyama, S., Büssow, K. & Gunsalus, K.C. (2008). Protein production and purification. *Nat. Methods*, **5**, 135–146.

Cox, R.S., Surette, M.G. & Elowitz, M.B. (2007). Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.*, **3**, 145.

Crea, R., Kraszewski, A., Hirose, T. & Itakura, K. (1978). Chemical synthesis of genes for human insulin. *Proc. Natl. Acad. Sci. USA*, **75**, 5765–5769.

Crick, F. (1958). On Protein Synthesis. *Symp. Soc. Exp. Biol.*, **XII**, 139–163.

Crick, F. (1988). *What Mad Pursuit–A Personal View of Scientific Discovery*. New York: Basic Books.

de Wet, J.R., Wood, K.V., Helinski, D.R. & DeLuca, M. (1985). Cloning of firefly luciferase cDNA and the expression of active luciferase in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **82**, 7870–7873.

de Wet, J.R., Wood, K.V., DeLuca, M., Helinski, D.R. & Subramani, S. (1987). Firefly luciferase gene: structure and expression in mammalian cells. *Mol. Cell. Biol.*, **7**, 725–737.

Dombroski, A.J., Johnson, B.D., Lonetto, M. & Gross, C.A. (1996). The sigma subunit of *Escherichia coli* RNA polymerase senses promoter spacing. *Proc. Natl. Acad. Sci. USA*, **93**, 8858–8862.

Domínguez-Cuevas, P. (2007). *Interacciones moleculares del regulador XylS y la ARN polimerasa con el promoter Pm del plásmido TOL pWWO de* Pseudomonas putida. Ph.D. thesis, Universidad de Granada.

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

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

Dunn, J.J. & Studier, F.W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.*, **166**, 477–535.

Durland, R.H. & Helinski, D.R. (1987). The sequence encoding the 43-kilodalton trfA protein is required for efficient replication or maintenance of minimal RK2 replicons in *Pseudomonas aeruginosa*. *Plasmid*, **18**, 164–169.

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

Eck, S. & Stephan, W. (2008). Determining the relationship of gene expression and global mRNA stability in *Drosophila melanogaster* and *Escherichia coli* using linear models. *Gene*, **424**, 102–107.

Ellis, T., Wang, X. & Collins, J.J. (2009). Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.*, **27**, 465–471.

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

Englesberg, E., C., S. & F., M. (1969). The L-arabinose operon in *Escherichia coli* B/r: a genetic demonstration of two functional states of the product of a regulator gene. *Proc. Natl. Acad. Sci. USA*, **62**, 1100.

Epshtein, V. & Nudler, E. (2003). Cooperation between RNA polymerase molecules in transcription elongation. *Science*, **300**, 801–805.

Erie, D. (2002). The many conformational states of RNA polymerase elongation complexes and their roles in the regulation of transcription. *Biochim. Biophys. Acta, Gene Struct. Expression*, **1577**, 224–239.

Estrem, S.T., Gaal, T., Ross, W. & Gourse, R.L. (1998). Identification of an UP element consensus sequence for bacterial promoters. *Proc. Natl. Acad. Sci. USA*, **95**, 9761–9766.

Fang, F. & Helinski, D. (1991). Broad-host-range properties of plasmid RK2: Importance of overlapping genes encoding the plasmid replication initiation protein TrfA. *J. Bacteriol.*, **173**, 5861–5868.

Fellner, P., Ehresmann, C., Stiegler, P. & Ebel, J.P. (1972). Partial nucleotide sequence of 16S ribosomal RNA from *E. coli*. *Nat. New Biol.*, **239**, 1–5.

Fischer, R., M., T.R., Drossard, S., J. an Hellwig & Schillberg, S. (2005). Plant cells. In G. Gelissen, ed., *Production of Recombinant Proteins. Novel Microbial and Eucaryotic Expression Systems.*, chap. 12, 253–272, Wiley-VCH, Weinheim.

Frazzetto, G. (2003). White biotechnology. *EMBO Rep.*, **4**, 835–837.

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

Fu, J., Wenzel, S.C., Perlova, O., Wang, J., Gross, F., Tang, Z., Yin, Y., Stewart, A.F., Müller, R. & Zhang, Y. (2008). Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Res.*, **36**, e113.

Galvão, T.C. & de Lorenzo, V. (2006). Transcriptional regulators à la carte: engineering new effector specificities in bacterial regulatory proteins. *Curr. Opin. Biotechnol.*, **17**, 34–42.

Ganley, A.R.D. & Kobayashi, T. (2007). Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res*, **17**, 184–191.

Gelissen, G., Strasser, A.W.M. & Suckow, M. (2005). Key and criteria to the selection of an expression platform. In G. Gelissen, ed., *Production of Recombinant Proteins. Novel Microbial and Eucaryotic Expression Systems.*, chap. 1, 1–5, Wiley-VCH, Weinheim.

Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A., Merryman, C., Young, L., Noskov, V.N., Glass, J.I., Venter, J.C., Hutchison, C.A. & Smith, H.O. (2008). Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, **319**, 1215–1220.

Gilbert, W. (2002). Obtained from VecBase 3.0. **Ref. 23**, http://www.ncbi.nlm.nih.gov/nuccore/208958?report=genbank.

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

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

Gowrishankar, J. & Harinarayanan, R. (2004). Why is transcription coupled to translation in bacteria? *Mol. Microbiol.*, **54**, 598–603.

Gralla, J.D., Carpousis, A.J. & Stefano, J.E. (1980). Productive and abortive initiation of transcription in vitro at the lac UV5 promoter. *Biochemistry*, **19**, 5864–5869.

Greive, S. & von Hippel, P. (2005). Thinking quantitatively about transcriptional regulation. *Nat. Rev. Mol. Cell Biol.*, **6**, 221–232.

Grill, S., Gualerzi, C.O., Londei, P. & Bläsi, U. (2000). Selective stimulation of translation of leaderless mRNA by initiation factor 2: evolutionary implications for translation. *EMBO J.*, **19**, 4101–4110.

Gualerzi, C.O. & Pon, C.L. (1990). Initiation of mRNA translation in prokaryotes. *Biochemistry*, **29**, 5881–5889.

Guo, J., Wang, J., Anderson, J. & Schultz, P. (2008). Addition of an alpha-Hydroxy Acid to the Genetic Code of Bacteria. *Angewandte Chemie*, **120**, 734–737.

Gusarov, I. & Nudler, E. (1999). The Mechanism of Intrinsic Transcription Termination. *Mol. Cell*, **3**, 495.

Hadiji-Abbes, N., Borchani-Chabchoub, I., Triki, H., Ellouz, R., Gargouri, A. & Mokdad-Gargouri, R. (2009). Expression of HBsAg and preS2-S protein in different yeast based system: A comparative analysis. *Protein Expr. Purif.*, In press.

Hammer, K., Mijakovic, I. & Jensen, P.R. (2006). Synthetic promoter libraries–tuning of gene expression. *Trends Biotechnol.*, **24**, 53–55.

Hannig, G. & Makrides, S.C. (1998). Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends Biotechnol.*, **16**, 54.

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

Heerikhuisen, M., Van den Hondel, C. & Punt, P. (2005). *Aspergillus sojae*. In G. Gellissen, ed., *Production of Recombinant Proteins. Novel Microbial and Eucaryotic Expression Systems.*, chap. 9, 191–214, Wiley-VCH, Weinheim.

Hershey, A.D. & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.*, **36**, 39–56.

Hoagland, M.B. (2003). Celebrating complementarity. *Ann. Intern. Med.*, **138**, 583–586.

Hofmann, C., Sandig, V., Jennings, G., Rudolph, M., Schlag, P. & Strauss, M. (1995). Efficient gene transfer into human hepatocytes by baculovirus vectors. *Proc. Natl. Acad. Sci. USA*, **92**, 10099–10103.

Holtel, A., Goldenberg, D., Giladi, H., Oppenheim, A. & Timmis, K. (1995). Involvement of IHF protein in expression of the Ps promoter of the *Pseudomonas putida* TOL plasmid. *J. Bacteriol.*, **177**, 3312–3315.

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

Hsu, L.M. (2009). Monitoring abortive initiation. *Methods*, **47**, 25–36.

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.

Ingram, L.C., Richmond, M.H. & Sykes, R.B. (1973). Molecular characterization of the R factors implicated in the carbenicillin resistance of a sequence of *Pseudomonas aeruginosa* strains isolated from burns. *Antimicrob. Agents Chemother.*, **3**, 279–288.

Inouye, S., Nakazawa, A. & Nakazawa, T. (1987). Overproduction of the xylS gene product and activation of the xylDLEGF operon on the TOL plasmid. *J. Bacteriol.*, **169**, 3587–3592.

Isaacs, A. & Lindenmann, J. (1957). Virus interference. I. The interferon. *Proc. R. Soc. Lond. B. Biol. Sci.*, **147**, 258–267.

Isaacs, F.J., Dwyer, D.J., Ding, C., Pervouchine, D.D., Cantor, C.R. & Collins, J.J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.*, **22**, 841–847.

Jackson, D.A., Symons, R.H. & 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*. *Proc. Natl. Acad. Sci. USA*, **69**, 2904–2909.

Jensen, P.R. & Hammer, K. (1998). Artificial promoters for metabolic optimization. *Biotechnol. Bioeng.*, **58**, 191–195.

Jia, M. & Li, Y. (2005). The relationship among gene expression, folding free energy and codon usage bias in *Escherichia coli*. *FEBS Lett.*, **579**, 5333–5337.

Jiong, M., Allan, C. & Samuel, K. (2002). Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol.*, **184**, 5733–5745.

Joyce, S.A. & Dreyfus, M. (1998). In the absence of translation, RNase E can bypass 5′ mRNA stabilizers in *Escherichia coli*. *J. Mol. Biol.*, **282**, 241–254.

Kaczanowska, M. & Ryden-Aulin, M. (2007). Ribosome Biogenesis and the Translation Process in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.*, **71**, 477–494.

Kaldalu, N., Toots, U.D.E., Lorenzo, V. & Ustav, M. (2000). Functional domains of the TOL plasmid transcription factor XylX. *J. Bacteriol.*, **182**, 1118.

Kaminishi, T., Wilson, D.N., Takemoto, C., Harms, J.M., Kawazoe, M., Schluenzen, F., Hanawa-Suetsugu, K., Shirouzu, M., Fucini, P. & Yokoyama, S. (2007). A snapshot of the 30S ribosomal subunit capturing mRNA via the Shine-Dalgarno interaction. *Structure*, **15**, 289–297.

Kapanidis, A.N., Margeat, E., Ho, S.O., Kortkhonjia, E., Weiss, S. & Ebright, R.H. (2006). Initial Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism. *Science*, **314**, 1144–1147.

Karimi, R., Pavlov, M.Y., Buckingham, R.H. & Ehrenberg, M. (1999). Novel roles for classical factors at the interface between translation termination and initiation. *Mol. Cell*, **3**, 601–609.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. & Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, **181**, 662–666.

Khlebnikov, A., Skaug, T. & Keasling, J.D. (2002). Modulation of gene expression from the arabinose-inducible araBAD promoter. *J. Ind. Microbiol. Biotechnol.*, **29**, 34–37.

Kiel, M.C., Raj, V.S., Kaji, H. & Kaji, A. (2003). Release of ribosome-bound ribosome recycling factor by elongation factor G. *J. Biol. Chem.*, **278**, 48041–48050.

Kireeva, M.L. & Kashlev, M. (2009). Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proc. Natl. Acad. Sci. USA*, **106**, 8900–8905, Published online before print.

Kisselev, L., Ehrenberg, M. & Frolova, L. (2003). Termination of translation: interplay of mRNA, rRNAs and release factors? *EMBO J.*, **22**, 175–182.

Kittell, B.L. & Helinski, D.R. (1991). Iteron inhibition of plasmid RK2 replication in vitro: evidence for intermolecular coupling of replication origins as a mechanism for RK2 replication control. *Proc. Natl. Acad. Sci. USA*, **88**, 1389–1393.

Klein-Marcuschamer, D., Santos, C.N.S., Yu, H. & Stephanopoulos, G. (2009). Mutagenesis of the bacterial RNA polymerase alpha subunit for improving complex phenotypes. *Appl. Environ. Microbiol.*, **75**, 2705–2711.

Kline, B.C. (1985). A review of mini-F plasmid maintenance. *Plasmid*, **14**, 1–16.

Kost, T.A. & Condreay, J.P. (2002). Recombinant baculoviruses as mammalian cell gene-delivery vectors. *Trends Biotechnol.*, **20**, 173–180.

Krüger, R., Rakowski, S.A. & Filutowicz, M. (2004). Participatin Elements in the Replication of Iteron-Containing Plasmids. In B.A. Funnel & J.G. Phillips, eds., *Plasmid biology*, chap. 2, 25–45, ASM Press, Washington, D. C.

Kunnimalaiyaan, S., Inman, R.B., Rakowski, S.A. & Filutowicz, M. (2005). Role of pi dimers in coupling ("handcuffing") of plasmid R6K's gamma ori iterons. *J. Bacteriol.*, **187**, 3779–3785.

Kurland, C.G. (1992). Translational accuracy and the fitness of bacteria. *Annu. Rev. Genet.*, **26**, 29–50.

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

Lartigue, C., Glass, J.I., Alperovich, N., Pieper, R., Parmar, P.P., Hutchison, C.A., Smith, H.O. & Venter, J.C. (2007). Genome transplantation in bacteria: changing one species to another. *Science*, **317**, 632–638.

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

Lee, D.N., Phung, L., Stewart, J. & Landick, R. (1990). Transcription pausing by *Escherichia coli* RNA polymerase is modulated by downstream DNA sequences. *J. Biol. Chem.*, **265**, 15145–15153.

Leibman, M. & Hochschild, A. (2007). A sigma-core interaction of the RNA polymerase holoenzyme that enhances promoter escape. *EMBO J.*, **26**, 1579–1590.

Lewis, D.E.A. & Adhya, S. (2004). Axiom of determining transcription start points by RNA polymerase in *Escherichia coli*. *Mol. Microbiol.*, **54**, 692–701.

Lewis, P.J., Doherty, G.P. & Clarke, J. (2008). Transcription factor dynamics. *Microbiology*, **154**, 1837–1844.

Lisser, S. & Margalit, H. (1993). Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Res.*, **21**, 1507–1516.

Lopez, P.J., 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 Res.*, **22**, 1186–1193.

Lorenz, P. & Eck, J. (2005). Metagenomics and industrial applications. *Nat. Rev. Microbiol.*, **3**, 510–516.

Luckow, V.A. (1993). Baculovirus systems for the expression of human gene products. *Curr. Opin. Biotechnol.*, **4**, 564 – 572.

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

Mandal, M. & Breaker, R.R. (2004). Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.*, **5**, 451–463.

Marqués, S. & Ramos, J.L. (1993). Transcriptional control of the *Pseudomonas putida* TOL plasmid catabolic pathways. *Mol. Microbiol.*, **9**, 923–929.

Marqués, S., Gallegos, M.T. & Ramos, J.L. (1995). Role of sigma S in transcription from the positively controlled Pm promoter of the TOL plasmid of *Pseudomonas putida*. *Mol. Microbiol.*, **18**, 851.

Marqués, S., Manzanera, M., González-Pérez, M.M. & Ramos, J.L. (1999). The XylS-dependent Pm promoter is transcribed in vivo by RNA polymerase with sigma 32 or sigma 38 depending on the growth phase. *Mol. Microbiol.*, **31**, 1105.

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

Mayer, C., Köhrer, C., Kenny, E., Prusko, C. & RajBhandary, U.L. (2003). Anticodon sequence mutants of *Escherichia coli* initiator tRNA: effects of overproduction of aminoacyl-tRNA synthetases, methionyl-tRNA formyltransferase, and initiation factor 2 on activity in initiation. *Biochemistry*, **42**, 4787–4799.

McClure, W.R. & Cech, C.L. (1978). On the mechanism of rifampicin inhibition of RNA synthesis. *J. Biol. Chem.*, **253**, 8949–8956.

Megerle, J.A., Fritz, G., Gerland, U., Jung, K. & Rädler, J.O. (2008). Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys. J.*, **95**, 2103–2115.

Mermod, N., Ramos, J.L., Lehrbach, P.R. & Timmis, K.N. (1986). Vector for regulated expression of cloned genes in a wide range of gram-negative bacteria. *J. Bacteriol.*, **167**, 447.

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

Mey, M.D., Maertens, J., Lequeux, G.J., Soetaert, W.K. & Vandamme, E.J. (2007). Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnol.*, **7**, 34.

Missiakas, D. & Raina, S. (1998). The extracytoplasmic function sigma factors: role and regulation. *Mol. Microbiol.*, **28**, 1059–1066.

Miyada, C.G., Stotzfus, L. & Wilcox, G. (1984). Regulation of the araC gene of *Escherichia coli*, catabolite repression, autoregulation and effect on araBAD expression. *Proc. Natl. Acad. Sci. USA*, **81**, 4120.

Mnaimneh, S., Davierwala, A.P., Haynes, J., Moffat, J., Peng, W.T., Zhang, W., Yang, X., Pootoolal, J., Chua, G., Lopez, A., Trochesset, M., Morse, D., Krogan, N.J., Hiley, S.L., Li, Z., Morris, Q., Grigull, J., Mitsakakis, N., Roberts, C.J., Greenblatt, J.F., Boone, C., Kaiser, C.A., Andrews, B.J. & Hughes, T.R. (2004). Exploration of essential gene functions via titratable promoter alleles. *Cell*, **118**, 31–44.

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

Mooney, R.A., Artsimovitch, I. & Landick, R. (1998). Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. *J. Bacteriol.*, **180**, 3265–3275.

Mooney, R.A., Darst, S.A. & Landick, R. (2005). Sigma and RNA polymerase: an on-again, off-again relationship? *Mol. Cell*, **20**, 335–345.

Mukhopadhyay, J., Kapanidis, A.N., Mekler, V., Kortkhonjia, E., Ebright, Y.W. & Ebright, R.H. (2001). Translocation of sigma(70) with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. *Cell*, **106**, 453–463.

Naeimpoor, F. & Mavituna, F. (2001). Metabolic flux analysis in *Streptomyces coelicolor*: Effect of nitrogen source. In *Novel Frontiers in the Production of Compounds for Biomedical Use.*, 131–145, Hingham, MA, USA: Kluwer Academic Publishers.

Nagata, S., Taira, H., Hall, A., Johnsrud, L., Streuli, M., Ecsödi, J., Boll, W., Cantell, K. & Weissmann, C. (1980). Synthesis in *E. coli* of a polypeptide with human leukocyte interferon activity. *Nature*, **284**, 316–320.

O'Donnell, S.M. & Janssen, G.R. (2001). The initiation codon affects ribosome binding and translational efficiency in *Escherichia coli* of cI mRNA with or without the 5′ untranslated leader. *J. Bacteriol.*, **183**, 1277–1283.

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

Pan, T. & Sosnick, T. (2006). RNA folding during transcription. *Annu. Rev. Biophys. Biomol. Struct.*, **35**, 161–175.

Pansegrau, W., Lanka, E., Barth, P.T., Figurski, D.H., Guiney, D.G., Haas, D., Helinski, D.R., Schwab, H., Stanisich, V.A. & Thomas, C.M. (1994). Complete Nucleotide Sequence of Birmingham IncPα Plasmids Compilation and Comparative Analysis. *J. Mol. Biol.*, **239**, 623–663.

Pape, T., Wintermeyer, W. & Rodnina, M.V. (1998). Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the *E. coli* ribosome. *EMBO J.*, **17**, 7490–7497.

Park, J.S., Marr, M.T. & Roberts, J.W. (2002). *Escherichia coli* Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell*, **109**, 757–767.

Pauling, L. & Corey, R.B. (1950). Two hydrogen-bonded spiral configurations of the polypeptide chain. *J. Am. Chem. Soc.*, **72**, 5349.

Pauling, L. & Corey, R.B. (1953). A Proposed Structure For The Nucleic Acids. *Proc. Natl. Acad. Sci. USA*, **39**, 84–97.

Paulsen, I.T., Press, C.M., Ravel, J., Kobayashi, D.Y., Myers, G.S.A., Mavrodi, D.V., DeBoy, R.T., Seshadri, R., Ren, Q., Madupu, R., Dodson, R.J., Durkin, A.S., Brinkac, L.M., Daugherty, S.C., Sullivan, S.A., Rosovitz, M.J., Gwinn, M.L., Zhou, L., Schneider, D.J., Cartinhour, S.W., Nelson, W.C., Weidman, J., Watkins, K., Tran, K., Khouri, H., Pierson, E.A., Pierson, L.S., Thomashow, L.S. & Loper, J.E. (2005). Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat. Biotechnol.*, **23**, 873–878.

Perri, S., Helinski, D. & Toukdarian, A. (1991). Interactions of plasmid-encoded replication initiation proteins with the origin of DNA replication in the broad host range plasmid RK2. *J. Biol. Chem.*, **266**, 12536–12543.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Murihead, H., Will, G. & North, A.C. (1960). Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature*, **185**, 416–422.

Peters, J.M., Mooney, R.A., Kuan, P.F., Rowland, J.L., Keles, S. & Landick, R. (2009). Rho directs widespread termination of intragenic and stable RNA transcription. *Proc. Natl. Acad. Sci. USA*.

Pfleger, B.F., Pitera, D.J., Smolke, C.D. & Keasling, J.D. (2006). Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.*, **24**, 1027–1032.

Philipps, B., Rotmann, D., Wicki, M., Mayr, L.M. & Forstner, M. (2005). Time reduction and process optimization of the baculovirus expression system for more efficient recombinant protein production in insect cells. *Protein Expr. Purif.*, **42**, 211 – 218.

Pinkney, M., Diaz, R., Lanka, E. & Thomas, C. (1988). Replication of mini RK2 plasmid in extracts of *Escherichia coli* requires plasmid-encoded protein TrfA and host-encoded proteins DnaA, B, G DNA gyrase and DNA polymerase III. *J. Mol. Biol.*, **203**, 927–938.

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

Ramakrishnan, V. (2008). What we have learned from ribosome structures. *Biochem. Soc. Trans.*, **36**, 567–574.

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

Ramos, J.L., Gonzäles-Carrero, M. & Timmis, K.N. (1988). Broad-host range expression vectors containing manipulated meta-cleavage pathway regulatory elements of the TOL plasmid. *FEBS Lett.*, **226**, 241.

Ramos, J.L., Marquésand, S. & Timmis, K.N. (1997). Transcriptional control of the *Pseudomonas* TOL plasmid catabolite operons is achieved through an interplay of host factors and plasmid-encoded regulators. *Annu. Rev. Microbiol.*, **51**, 341.

Rhee, S., Martin, R.G., Rosner, J.L. & Davies, D.R. (1998). A novel DNA-binding motif in MarA: The first structure ofor an AraC family transcriptional activator. *Proc. Natl. Acad. Sci. USA*, **95**, 10413.

Ring, B.Z., Yarnell, W.S. & Roberts, J.W. (1996). Function of *Escherichia coli* RNA polymerase sigma factor sigma 70 in promoter-proximal pausing. *Cell*, **86**, 485–493.

Ringnér, M. & Krogh, M. (2005). Folding free energies of 5′-UTRs impact post-transcriptional regulation on a genomic scale in yeast. *PLoS Comput. Biol.*, **1**, e72.

Roberts, C.W. & Roberts, J.W. (1996). Base-specific recognition of the nontemplate strand of promoter DNA by *E. coli* RNA polymerase. *Cell*, **86**, 495–501.

Roberts, J.W., Shankar, S. & Filter, J.J. (2008). RNA polymerase elongation factors. *Annu. Rev. Microbiol.*, **62**, 211–233.

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

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

Sakata-Sogawa, K. & Shimamoto, N. (2004). RNA polymerase can track a DNA groove during promoter search. *Proc. Natl. Acad. Sci. USA*, **101**, 14731.

Salas, J.A., Blanco, G., Braña, A.F., Fernandez, E., a J. Fernandez, M., Bernardo, J.G., Gonzalez, A., Lombo, F., Prado, L., Quiros, L.M., Sanchez, C. & Mendez, C. (2001). Towards the generation of novel antitumour agents from actinomycetes by combinatorial biosynthesis. In *Novel Frontiers in the Production of Compounds for Biomedical Use.*, 383–399, Hingham, MA, USA: Kluwer Academic Publishers.

Sandig, W., Rose, T., Winkler, K. & Brecht, R. (2005). Mammalian cells. In G. Gelissen, ed., *Production of Recombinant Proteins. Novel Microbial and Eucaryotic Expression Systems.*, chap. 11, 233–252, Wiley-VCH, Weinheim.

Sanger, F. (1952). The arrangement of amino acids in proteins. *Adv. Protein. Chem.*, **7**, 1–67.

Sanger-Institute (2008). http://www.sanger.ac.uk/Projects/P_fluorescens/.

Schleif, R. (1988). DNA looping. *Science*, **240**, 127.

Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M. & Cate, J.H.D. (2005). Structures of the bacterial ribosome at 3.5 A resolution. *Science*, **310**, 827–834.

Selmer, M., Dunham, C.M., Murphy, F.V., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. & Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935–1942.

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

Shingler, V. & Thomas, C.M. (1989). Analysis of nonpolar insertion mutations in the trfA gene of IncP plasmid RK2 which affect its broad-host-range property. *Biochim. Biophys. Acta*, **1007**, 301–8.

Siegele, D.A. & Hu, J.C. (1997). Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl. Acad. Sci. USA*, **94**, 8168–8172.

Simonetti, A., Marzid, S., Jenner, L., Myasnikov, A., Romby, P., Yusupov, G., Klaholz, B. & Yusupov, M. (2008). A structural view of translation initiation in bacteria. *Cell. Mol. Life Sci.*.

Skordalakes, E., Brogan, A.P., Park, B.S., Kohn, H. & Berger, J.M. (2005). Structural mechanism of inhibition of the Rho transcription termination factor by the antibiotic bicyclomycin. *Structure*, **13**, 99–109.

Sletta, H., Nedal, A., Aune, T.E.V., Hellebust, H., Hakvåg, S., Aune, R., Ellingsen, T.E., Valla, S. & Brautaset, T. (2004). Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host-toxc single-chain antibody fragment in *Escherichia coli*. *Appl. Environ. Microbiol.*, **70**, 7033.

Sletta, H., Tøndervik, A., Hakvåg, S., Vee Aune, T.E., Nedal, A., Aune, R., Evensen, G., Valla, S., Ellingsen, T.E. & 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*. *Appl. Environ. Microbiol.*, **73**, 906.

Soisson, S., MacDougall-Shackleton, B., Schleif, R. & Wolberger, C. (1997a). Structural basis for ligand-regulated oligomerization of AraC. *Science*, **276**, 421.

Soisson, S., MacDougall-Shackleton, B., Schleif, R. & Wolberger, C. (1997b). The 1.6 Å crystal structure of the AraC sugar-binding and dimerization domain complexed with D-fucose. *J. Mol. Biol.*, **273**, 226.

Solem, C. & Jensen, P.R. (2002). Modulation of gene expression made easy. *Appl. Environ. Microbiol.*, **68**, 2397–2403.

Steffen, D. & Schleif, R. (1977). Overproducing araC protein with lambda-arabinose transducing phage. *Molec. Gen.*, **157**, 133.

Steitz, J.A. (1969). Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature*, **224**, 957–964.

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

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

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

Sudarsan, N., Barrick, J.E. & Breaker, R.R. (2003). Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA*, **9**, 644–647.

Suess, B., Fink, B., Berens, C., Stentz, R. & Hillen, W. (2004). A theophylline responsive riboswitch based on helix slipping controls gene expression *in vivo*. *Nucleic Acids Res.*, **32**, 1610–1614.

Thomas, C.M. & Helinski, D.R. (1989). Vegetative replication and stable inheritance of IncP plasmids. In C.M. Thomas, ed., *Promiscuous Plasmids of Gram-Negative Bacteria*, 1–25, Academic Press, San Diego, Calif.

Tipper, D. (1985). Mode of action of β-lactam antibiotics. *Pharmacol. Ther.*, **27**, 1–35.

Toukdarian, A. (2004). Plasmid strategies for broad-host-range replication in gram-negative bacteria. In B.A. Funnel & J.G. Phillips, eds., *Plasmid biology*, chap. 11, 1259–1270, ASM Press, Washington, D. C.

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

Toulokhonov, I., Artsimovitch, I. & Landick, R. (2001). Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science*, **292**, 730–733.

Toulokhonov, I., Zhang, J., Palangat, M. & Landick, R. (2007). A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol Cell*, **27**, 406–419.

Tøttrup, H.V. & Carlsen, S. (1990). A process for the production of human proinsulin in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.*, **35**, 339–348.

Uhlin, B. & Nordström, K. (1977). R plasmid gene dosage effects in *Escherichia coli* K-12: copy mutants of the R plasmic R1*drd-19*. *Plasmid*, **1**, 1–7.

Unniraman, S., Prakash, R. & Nagaraja, V. (2002). Conserved economics of transcription termination in eubacteria. *Nucleic Acids Res.*, **30**, 675–684.

Urban, J.H. & Vogel, J. (2007). Translational control and target recognition by *Escherichia coli* small RNAs in vivo. *Nucleic Acids Res.*, **35**, 1018–1037.

Vassylyev, D.G., ichi Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S. & Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature*, **417**, 712–719.

Wade, J., Roa, D., Grainger, D., Hurd, D., Busby, S., Struhl, K. & Nudler, E. (2006). Extensive functional overlap between σ factors in *Escherichia coli*. *Nat. Struct. Mol. Biol.*, **13**, 806–814.

Walker, K.A. & Osuna, R. (2002). Factors affecting start site selection at the *Escherichia coli* fis promoter. *J. Bacteriol.*, **184**, 4783–4791.

Wang, C.Y., Li, F., Yang, Y., Guo, H.Y., Wu, C.X. & Wang, S. (2006). Recombinant baculovirus containing the diphtheria toxin A gene for malignant glioma therapy. *Cancer Res.*, **66**, 5798–5806.

Watanabe, K. & Oikawa, H. (2007). Robust platform for de novo production of heterologous polyketides and nonribosomal peptides in *Escherichia coli*. *Org. Biomol. Chem.*, **5**, 593–602.

Watson, J.D. & Crick, F.H.C. (1953). Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.

Wilcox, G. & Meuris, P. (1976). Stabilization and size of AraC protein. *Mol. Gen. Genet.*, **145**, 97.

Wilms, B., Hauck, A., Reuss, M., Syldatk, C., Mattes, R., Siemann, M. & Altenbuchner, J. (2001). High-cell-density fermentation for production of L-N-carbamoylase using an expression system base on the *Escherichia coli* rhaBAD promoter. *Biotechnol. Bioeng.*, **73**, 95.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.

Winkler, W., Nahvi, A. & Breaker, R.R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952–956.

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

Winther-Larsen, H.C., Josefsen, K.D., Brautaset, T. & Valla, S. (2000b). Parameters affecting gene expression from the Pm promoter in gram-negative bacteria. *Metab. Eng.*, **2**, 79.

Yakhnin, A.V., Yakhnin, H. & Babitzke, P. (2006). RNA polymerase pausing regulates translation initiation by providing additional time for TRAP-RNA interaction. *Mol. Cell*, **24**, 547–557.

Young, M., Davies, M.J., Bailey, D., Gradwell, M.J., Smestad-Paulsen, B., Wold, J.K., Barnes, R.M. & Hounsell, E.F. (1998). Characterization of oligosaccharides from an antigenic mannan of *Saccharomyces cerevisiae*. *Glycoconj J*, **15**, 815–822.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H. & Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883–896.

Yusupova, G.Z., Yusupov, M.M., Cate, J.H. & Noller, H.F. (2001). The path of messenger RNA through the ribosome. *Cell*, **106**, 233–241.

Zavialov, A.V., Hauryliuk, V.V. & Ehrenberg, M. (2005). Guanine-nucleotide exchange on ribosome-bound elongation factor G initiates the translocation of tRNAs. *J. Biol.*, **4**, 9.

Zhang, Z., Gildersleeve, J., Yang, Y.Y., Xu, R., Loo, J.A., Uryu, S., Wong, C.H. & Schultz, P.G. (2004). A new strategy for the synthesis of glycoproteins. *Science*, **303**, 371–373.

Paper I

Is not included due to copyright

Paper II

Is not included due to copyright

Paper III

Is not included due to copyright

Paper IV

Is not included due to copyright