Trond Erik Vee Aune

High level recombinant protein production in *Escherichia coli* by engineering broad-host-range plasmid vectors containing the *Pm/xylS* expression cassette

Thesis for the degree philosophiae doctor

Trondheim, March 2008

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

NTNU

NTNU
Norwegian University of Science and Technology

Thesis for the degree philosophiae doctor

Faculty of Natural Sciences and Technology
Department of Biotechnology

ISBN 978-82-471-7359-6 (printed version)
ISBN 978-82-471-7362-6 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2008:70

Printed by NTNU-trykk

ACKNOWLEDGEMENTS

The scientific work presented in this thesis was carried out at the Department of Biotechnology at the Norwegian University of Science and Technology (NTNU). The financial support was provided by the Norwegian Research Council, Affitech AS, and Alpharma AS, and is highly appreciated.

The work has been supervised by Professor Svein Valla at the Department of Biotechnology at NTNU. I would like to thank him for his great help during my work, never-lasting optimism and ability to always see the possibilities in every situation; but most of all I would like to thank him for all these inspiring years which I always will remember with happiness.

I would also like to thank all my other colleagues whom I have benefited greatly from, both professionally and socially, especially Sven Even and Magnus.

Finally I would like to thank good fortune for meeting Marie and thank Marie for always being there for me.

Optimism is the foundation of courage ~ Nicholas Murray Butler

Trondheim, December 2007
Trond Erik Vee Aune

ABSTRACT

The use of recombinant gene expression for industrial protein production has since the early 1970s become a multi-billion industry. Despite many years of intensive research, no universal expression system exists that can be used for high-level industrial production of any given recombinant gene. Because of this, much work must be done prior to production to identify the ideal production organism for the gene of interest, usually followed by optimization of an expression system useful in this organism.

One of the most popular production organisms is the gram-negative bacterium *Escherichia coli*, which is regularly used for high-level production of both prokaryotic and eukaryotic proteins. Despite its popularity there are some well-known drawbacks of using *E. coli* for recombinant expression, including lack of some post-translational modifications of recombinant proteins, and few and inefficient systems for translocation and secretion.

In the present work selected key processes in *E. coli* recombinant protein production has been studied with focus on exploring methods to improve protein production in this organism. The key processes include the use of signal peptides for efficient translocation of eukaryotic proteins to the periplasm, directed evolution of the transcription regulators for increased expression from their cognate promoters, and the use of codon-optimized synthetic genes.

The use of signal peptides to achieve efficient translocation of recombinant genes is a well-known and common method. By testing the signal peptides sequences *ompA*, *pelB* and the novel, designed consensus signal peptide sequence *CSP*, in combination with the eukaryotic cDNA for the proteins hGM-CSF, scFv-phOx and hIFN-α2b, it was shown that the effect of signal peptides are gene-specific and, more importantly, that the presence of signal peptides not only directs the gene products to the periplasm but also increases the expression of the recombinant genes. Under high cell density culture (HCDC) expression production volumes of 2.3, 1.7 and 0.6 g/L of scFv-phOx, hGM-CSF and hIFN-α2b, respectively, were obtained.

The use of directed evolution on transcription regulators featured in expression systems is a novel approach for improving recombinant protein production. By evolving the transcriptional activator XylS with error-prone PCR and subsequent DNA shuffling it was possible to obtain evolved activators that increased induced expression from their cognate promoter *Pm* at least 10-fold, both with the prokaryotic *bla* gene and the eukaryotic gene for the antibody fragment scFv-phOx as reporters. The basal expression only increased about four-fold, and hence the induction ratio was increased. By predicting the 3D structure of the N-terminal of XylS it was shown that the obtained substitutions were positioned in different structural sub-domains. Based on this it was proposed that the substitutions probably affect different functions of XylS.

By analyzing *xylS* mutants that caused both high induced and basal levels of expression from *Pm*, the mutation G33A was shown to form a new promoter reverse oriented within the coding region of *xylS*. This novel promoter is the origin of a transcript that traverses the more than 800 base pair region between the *xylS* gene and the *bla* gene, resulting in increased amounts of *bla* transcripts available for translation and an apparent, though false, increase in transcription from *Pm*.

In order to produce XylS in large quantities for structural studies it was discovered that the *xylS* gene contains unfavorable codons that may limit its expression in *E. coli*. To circumvent this problem, a synthetic *de novo* version of the *xylS* gene, *syn-xylS*, was prepared. *syn-xylS* was *in silico* optimized to give optimal translation of its corresponding mRNA in *E. coli*, containing only frequently used codons and forming no known mRNA secondary structures that would inhibit translation. Surprisingly, the translation of *syn-xylS* was significantly reduced due to changes in the 5' region of the gene (relative to wild type), with a single change in position 42 (C→42) alone being able to cause the entire reduction.

The findings presented in this work contribute to our overall understanding of recombinant expression and open up possibilities for development of improved expression systems useful for industrial scale protein production. Although the focus of this work has been the optimization of recombinant expression systems in *E. coli*, the results presented are generally applicable to other bacterial production organisms and can potentially be used to improve any such system, probably including eukaryotic

expression systems. In addition the findings expand our understanding of the specific *Pm*/*xylS* expression system. Although the work has been done in *E. coli*, this system is known to function in many organisms and the findings can hence be used for improvement of broad-host-range applications.

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Recombinant protein production

In 1973 Stanley Cohen and Herbert Boyer pioneered the use of recombinant DNA technology for cloning and expression of genes in foreign organisms. They cloned DNA from the *Salmonella typhimurium* streptomycin resistance plasmid RSF1010 into the *Escherichia coli* plasmid pSC101 and observed tolerance to streptomycin among the transformants (Cohen et al. 1973). The first reported production of a human recombinant protein took place a few years later when the then newly started biotech company Genentech announced that they had managed to express the gene encoding human somatostatin in *E. coli* (Itakura et al. 1977). The value of the resulting bioactive substance was similar to that of somatostatin extracted from the brains of 500.000 sheep. In 1982 Genentech followed up this success with the product humulin, a recombinant insulin produced in *E. coli* and the first recombinant biotech drug to be accepted for market by the Food and Drug Administration. Today the production of recombinant proteins has become a huge global industry with an annual market volume exceeding $50 billion (Schmidt 2004).

At the start of the recombinant protein expression era the bacteria *Escherichia coli* and *Bacillus* spp. dominated as hosts for recombinant expression, but the realization that a protein may require a specific host physiology and biochemistry for optimal production stimulated a search for new hosts, both prokaryotic and eukaryotic. Parallel to this quest, recombinant DNA technology advanced tremendously thereby opening up possibilities for the use of novel organisms. As a consequence, many different expression systems for use in many different hosts are now available, including systems for use in yeasts (Gellissen et al. 2005), filamentous fungi (Nevalainen et al. 2005), insect and animal cell cultures (Wurm 2004; Kost et al. 2005), gram-positive bacteria like *Bacillus* (Westers et al. 2004) and *Streptomyces* (Binnie et al. 1997), and gram-negative bacteria like *Escherichia coli* (Jana and Deb 2005). Success in gene expression projects is therefore

both a question of identifying the optimal host for the specific gene and then choosing and optimizing an expression system in this host.

1.1.1. Recombinant protein production in gram-negative bacteria, with focus on *Escherichia coli*

Bacterial expression systems are the preferred choice for production of many prokaryotic and eukaryotic proteins. The reasons for this lie in the cost-effectiveness of bacteria, their well-characterized genetics, and the availability of many different bacterial expression systems. Among the hosts available for recombinant expression, *Escherichia coli* is in an exceptional position. This stems from the many decades of intense research on its genetics as well as the broad scope of biotechnological tools available for genetic engineering of this organism. As a host for recombinant expression, *E. coli* is especially valued because of its rapid growth rate, capacity for continuous fermentation, low media costs and achievable high expression levels (Yin et al. 2007). One consequence of this popularity is that about 80% of all proteins used to solve three-dimensional structures submitted to the protein data bank (PDB) in 2003 were prepared in *E. coli* (Sørensen and Mortensen 2005) and during 2003 and 2006, nine out of 31 approved therapeutic proteins were produced in *E. coli* (Walsh 2006), among them important growth factors, insulins and interferons (Schmidt 2004).

The major drawbacks of using *E. coli* for recombinant protein production are its lack of secretion systems for efficient release of proteins to the growth medium, limited ability to facilitate extensive disulfide-bond formation and other posttranslational modifications, inefficient cleavage of the amino terminal methionine which can result in lowered protein stability and increased immunogenicity, and occasional poor folding due to lack of specific molecular chaperones (Makrides 1996; Yin et al. 2007).

Most expression systems for *E. coli* are plasmid-based. This ensures, in most cases, a relatively easy cloning of the recombinant gene into the host cell (no chromosomal integration of the recombinant gene is necessary) and it facilitates many copies of the gene in a single cell. On the other hand, recombinant gene expression from

plasmids demands high plasmid stability which is often achieved by the use of antibiotic resistance markers. This represents a potential problem when the produced protein are to be used for therapeutic purposes since any traces of antibiotics, especially β-lactams, may cause allergic reactions among patients (Williams et al. 1998; Swartz 2001). In addition there is the potential possibility for transfer of drug resistance to environmental organisms, with especial concern for pathogens (Williams et al. 1998).

The most popular commercial expression system for *E. coli* is the Novagen pET expression system based on the T7 promoter (Yin et al. 2007). It represents more than 90% of the 2003 PDB protein preparation systems (Sørensen and Mortensen 2005). The pET system was first described in 1990, and has been developed for a variety of expression applications (Studier et al. 1990; Dubendorff and Studier 1991). Other popular *E. coli* expression systems include the Invitrogen pBAD system based on the arabinose operon promoter (P_{BAD}), the Amersham Biosciences pGEX system based on the *tac* promoter, and the Qiagen pQE system based on the *lac*/T5 promoter.

Gram-negative alternatives to *E. coli* include the *Ralstonia eutrophus* system which has been reported to be useful for high-level production of the soluble protein organophospohydrolase caused by its low inclusion body forming tendencies (Srinivasan et al. 2002; Srinivasan et al. 2003), the *Caulobacter crescentus* system which is promising for high-level secretion of many recombinant proteins (Umelo-Njaka et al. 2001), and *Pseudomonas fluorescens* which has been used for high-level production of α-amylase (Landry et al. 2003).

1.2. Maximizing recombinant protein production in Escherichia coli

Despite all the recombinant expression technology that is available for *E. coli*, successful protein production in this organism still frequently relies on carefully choosing and optimizing the expression system for the gene of interest. There is no single system available that is optimal for any given recombinant gene. The most important approaches to maximize recombinant protein production in *E. coli* involve choosing or designing the optimal promoter, engineering the transcription regulator, adjusting the recombinant gene

copy number, engineering the 5' untranslated region, increasing messenger RNA longevity, transport to specific cellular compartments, choosing and designing the ideal production host, and optimization of fermentation conditions (Makrides 1996; Baneyx 1999; Schumann and Ferreira 2004; Jana and Deb 2005; Sørensen and Mortensen 2005; Choi et al. 2006; Terpe 2006). All these approaches will be discusse in the following chapters. Downstream processes, including purification and refolding of inclusions bodies, will not be covered.

Although this is written with *E. coli* in mind, the presented topics are in general applicable to other bacteria while restrictions in available genetic tools as well as fundamental differences in physiology and biochemistry between the organisms, may limit their relevance.

1.2.1. Choosing and designing the optimal promoter

1.2.1.1. Promoter structure and transcription initiation

The promoter is an important target for increasing recombinant protein production because of its role in controlling the transcription initiation of associated genes. The promoter consists of DNA sequences crucial for recognition by the RNA polymerase. In *E. coli* these sequences are compromised by the core sequences, usually centered around positions -10 and -35, an extended -10 element and occasionally a UP element (Browning and Busby 2004). Consensus sequences for the σ^{70}-recognized -35 and -10 sequences in *E. coli* are shown in Table 1.

Transcription initiation requires that the core RNA polymerase (consisting of the subunits $\beta\beta'\alpha_2\omega$) associates with a specific σ initiation factors to form the holoenzyme. The holoenzyme will bind promoters differentially as specified by its σ factor. *E. coli* encodes in total seven different σ factors. The predominant σ^{70} is responsible for controlling transcription of the housekeeping genes whereas the six alternative factors are required for transcription of smaller subsets of promoters (Gruber and Gross 2003). The association between the holoenzyme and the promoter is stabilized by the differentiating contacts between the σ factor and the core sequences of the target promoter and the

extended -10 element, and contacts between the α subunit and the UP element (Wösten 1998). Subsequent unwinding of the duplex DNA from approximately the -10 to the +2 position facilitates contact between the template strand and the active site of the RNA polymerase essential for the following mRNA synthesis (de Haseth et al. 1998; Tomsic et al. 2001; Tsujikawa et al. 2002). The initiation step is characterized by many rounds of abortive synthesis, but once the RNA polymerase has synthesized a polynucleotide of 13-15 base pairs length, the transcription complex undergoes promoter clearance and the elongation step begins (Carpousis and Gralla 1980; Krummel and Chamberlin 1989; Hsu et al. 2003).

Table 1: DNA sequences of some promoters used in *Escherichia coli* expression vectors recognized by the housekeeping sigma factor σ^{70}

Promoter	-35 region	Spacer	-10 region
lac	TTtACA	18 bp	TATgtT
*lac*UV5	TTtACA	18 bp	TATAAT
trp	TTGACA	17 bp	TtaAcT
tac	TTGACA	17 bp	TATAAT
trc	TTGACA	18 bp	TATAAT
λ P_L	TTGACA	17 bp	gATAcT
λ P_R	TTGACt	17 bp	gATAAT
lacI	gcGcaA	17 bp	cATgAT
lacI^q	gTGcaA	17 bp	cATgAT
lacI^{q1}	TTGACA	18 bp	cATgAT
Consensus	**TTGACA**	**17 bp**	**TATAAT**

1.2.1.2. Promoters suitable for recombinant expression and their features

A promoter used for recombinant protein production should primarily give enough mRNA so that its level is not the limiting factor for the overall protein yield (i.e. it should be *strong*). Still, an increase in mRNA may not lead to more recombinant protein if other parts are limiting, e.g. if the translation machinery already works at maximum capacity. Secondly, the promoter should be controllable since constitutive promoters will not allow efficient production of toxic proteins and even some native

proteins which are deleterious to the cell when over-expressed. One example is membrane proteins which when overproduced may cause cell death (O'Connor and Timmis 1987), possibly by jamming the inner membrane (Schumann and Ferreira 2004). Thirdly, promoters used for recombinant expression should have a low basal transcription rate (i.e. they should have *low leakage*). The low leakage reduces pre-induction strain on the host from the metabolic burden of recombinant protein production and expression of host-toxic proteins.

Some of the most commonly used promoters for recombinant protein production in *E. coli* are presented in Table 2. A few key promoters will be discussed with emphasis on how they have been optimized through rational engineering for maximized protein production.

Table 2: Examples of commonly used promoters for recombinant expression in *E. coli*

Promoter	Regulators	Induction[a]	Original reference
lac	LacI, LacIts	IPTG, thermal	(Gronenborn 1976)
trc and *tac*	LacI, LacIts	IPTG, thermal	(Brosius et al. 1985)
λP_L	λcIts 857	Thermal	(Elvin et al. 1990)
T7	LacI, LacIts	ITPG, thermal	(Studier and Moffatt 1986)
phoA	PhoB, PhoR	Phosphate starvation	(Miyake et al. 1985)
P_{BAD}	AraC	L-arabinose	(Guzman et al. 1995)
cspA	unknown	Thermal	(Mujacic et al. 1999)
$rhaP_{BAD}$	RhaR, RhaS	L-rhamnose	(Haldimann et al. 1998)
tetA	TetR	Tetracyclin	(Skerra 1994)
Pm	XylS	*m*-toluic acid	(Blatny et al. 1997)

[a]*IPTG: Isopropyl-β-D-thiogalactoside*

1.2.1.3. The *lac* promoter and its derivatives

One of the earliest operons to be studied in detail was the *E. coli* lactose utilization operon (the *lac* operon), and the classic model of gene regulation by Jacob and Monod was based on its properties (Jacob and Monod 1961). Hence, the *lac* operator with its repressor, LacI, were quickly adopted for use in recombinant expression (Polisky

Bonehi et al. 1998). Attempts at circumventing the all-or-nothing phenomenon for P_{BAD} have resulted in host strains with engineered transporter systems for L-arabinose (Khlebnikov et al. 2000) or with systems for facilitated diffused transport of L-arabinose (Morgan-Kiss et al. 2002).

1.2.1.6. Randomized engineering of promoters

In addition to the aforementioned examples of rational engineering of promoters for improved recombinant expression, the engineering of promoters through the process of randomized engineering has gained recent popularity. One example of randomized engineering of promoters is the construction of a library of synthetic stationary-phase and stress promoters, recognized by σ^{38}, based on randomization of the sequence in the -37 to -14 region in a set of such promoters (Miksch et al. 2005). The resulting promoters exhibited activities (measured in specific fluorescence units, sFU) from 670 to 13,380 with the strongest synthetic promoters having a three- to four-fold higher activity than the natural promoters which were used as references.

Another recent example is the construction of a promoter mutant library based on randomization of a *E. coli* consensus promoter sequence (de Mey et al. 2007). In this study the -35 and -10 region were kept constant and the flanking regions were randomized. The activity of isolated promoters varied tremendously, from 27.5 fold higher than the *lac* promoter, to almost 7 fold lower.

1.2.3. Engineering transcription regulators

1.2.3.1. Modifying the expression of transcription regulators

Another potential target for maximizing recombinant gene transcription in *E. coli* is the transcription regulator which together with the promoter controls the crucial step of transcription initiation. This can be achieved either by engineering its function or adjusting its expression. The latter has been demonstrated with the reduction of *lac*

promoter leakiness through increase of LacI expression. One example is the *PlacI*^{q} mutant which contains a C → T mutation in the -35 region leading to 5-10-fold increase in intracellular LacI concentration (Müller-Hill et al. 1968; Calos 1978) (Table 2). The *lacI* gene under transcriptional control of *PlacI*^{q} mutant integrated in the host's chromosome is sufficient to efficiently repress transcription on medium-copy number plasmids (Baneyx 1999). For repression of recombinant expression from high-copy number plasmids, the *lacI* gene with *PlacI*^{q} is typically cloned onto the plasmid itself, or provided *in trans* from a second plasmid (Baneyx 1999). Another *PlacI* mutant is *PlacI*^{q1} for which the entire -35 region is exchanged with the consensus sequence for σ^{70}-dependent promoters (Table 2). The resulting promoter is 170-times stronger than *PlacI* and can be integrated on the host chromosome and still efficiently repress expression from high-copy number plasmids (Glascock and Weickert 1998).

Another example of transcription regulator expression optimization comes from the transcription activator XylS. By over-expressing XylS from the strong *tac* or *tet* promoters, and hence increase its intracellular concentration, the basal transcription from its cognate *Pm* promoter has been shown to be increased substantially (Inouye et al. 1987; Mermod et al. 1987; Spooner et al. 1987). There is an equilibrium between active XylS (which stimulates transcription) and inactive XylS. Higher production of XylS in the absence of effectors will thus result in more inactive XylS proteins, but because of the equilibrium also more active XylS proteins. The role of effector molecules are to shift this equilibrium towards more active XylS (Inouye et al. 1981; Franklin et al. 1983; Ramos et al. 1986; Mermod et al. 1987; Ramos et al. 1987; Marqués et al. 1994).

1.2.3.2. Optimizing the function of transcription regulators

Another approach for optimizing expression of recombinant genes is to engineer the function of the transcription regulator. Again *lacI* provides an example with the gene mutants *lacI*ts and *lacI*^{q}ts which encode the temperature sensitive LacIts repressor. LacIts can be used to evade the need for chemical induction of *lac*-based promoters (Bukrinsky

et al. 1988; Hasan and Szybalski 1995). LacIts contains a glycine to serine substitution at position 187 and full induction is achieved at 42°C (Hasan and Szybalski 1995).

Efforts have also been made to create structural XylS mutant proteins which are more efficient at stimulating transcription from *Pm* in the presence of different inducers (Ramos et al. 1986; Ramos et al. 1990; Michán et al. 1992). A mutation in XylS which increases its ability to stimulate transcription from *Pm* in the presence of its most potent effector, *m*-toluic acid, the R45T mutation, has also been documented (Michán et al. 1992).

Systems that utilize two promoters for the simultaneous expression of two recombinant proteins are sometimes desired, and in such cases independent control of each promoter may be important. It has been shown that IPTG is an inhibitor of AraC, thus rendering the simultaneous use of P_{BAD} and *lac*-based promoters impractical (Lee et al. 2007). To overcome this problem directed evolution of AraC was initiated with the goal to obtain an AraC mutant protein which is not inhibited by IPTG. This resulted in the isolation of an AraC mutant protein that was able to stimulate transcription from P_{BAD} in the presence of 2.5 mM IPTG at approximately the same level as parental AraC in the absence of IPTG (Lee et al. 2007).

1.2.3. Adjusting the gene copy number

The copy number of the expression vector is an important parameter for optimizing recombinant expression. This number is determined by the specific origin of replication and sometimes associated *cis* acting elements like rep proteins (del Solar et al. 1998). It is also affected by plasmid and host genetics and cultivation conditions such as growth rates, media and temperature (del Solar et al. 1998). Most commercial expression vectors are based on replicons with moderate copy numbers (15-60 copies per cell) (Baneyx 1999; Jana and Deb 2005) and thus it is often possible to achieve an increase in recombinant protein expression by increasing the expression vector's copy number and hence the copy number of the recombinant gene.

Increase of plasmid copy numbers should be used with caution since many studies have shown that the metabolic burden on the host cell from elevated copy numbers can be detrimental to the host cells (Bentley et al. 1990; Birnbaum and Bailey 1991). It is also known that high-copy vectors may suffer from segregational and structural instability (Jones and Melling 1984; O'Connor et al. 1989). Segregational instability is defined as the loss of plasmid from one of the daughter cells during division because of defective partitioning. Structural instability is attributed to deletions, insertions and rearrangements in the plasmid structure, resulting in the loss of the desired gene function (Rai and Padh 2001).

1.2.3.1. Adjusting plasmid copy-number by engineering ColE1 type replication

The majority of vectors used for recombinant expression replicate by replicons derived from naturally occurring plasmids of the ColE1 type (Sørensen and Mortensen 2005). The ColE1 replicon featured in the majority of expression vectors is derived from the pBR322 plasmid (Bolivar et al. 1977) which has a copy number of 15-20 per cell (Muesing et al. 1981; Lin-Chao and Bremer 1986; Lee et al. 2006). Replication of ColE1-type replicons are initiated by the replicator primer RNA II which hybridizes to the replication origin and, upon cleavage by RNase H to release the 3'OH, is elongated by the *E. coli* DNA polymerase I (Tomizawa et al. 1974; Tomizawa et al. 1975). Cleavage of RNA II is inhibited by prior hybridization with antisense RNA I (Tomizawa 1984; Hjalt and Wagner 1992). The interaction between RNA II and RNA I is stabilized by dimers of the Rop (Rom) protein (Tomizawa and Som 1984; Brenner and Tomizawa 1989; Castagnoli et al. 1989) (Figure 4).

1.2.3.3. Benefits from use of chromosomal integration and very-low copy number plasmids

Some of the problems observed with elevated copy numbers may be magnified problems arising from the general presence of plasmids inside the host. Many studies have focused on the effects of plasmids on the host cell and demonstrated perturbations in normal replication, transcription and translation, changes in nutrient uptake and decreased growth rates (Diaz Ricci and Hernández 2000; Grabherr and Bayer 2002; Grabherr et al. 2002). A recent study showed that key pathways like glycolysis, the tricarboxylic acid cycle and the pentose phosphate pathway are perturbed by the presence of both low (50 copies per cell) and high (400 copies per cell) copy number plasmids (Wang et al. 2006). Thus, in some cases very-low copy number plasmids are desired (Jones and Keasling 1998), an example of which is the pETcoco vector (Novagen) which has a copy number of one per cell. A completely different strategy is chromosomal integration of expression cassettes featuring the recombinant gene and the systems necessary for recombinant expression (Olson et al. 1998) (Jana and Deb 2005).

In addition to circumventing problems emerging from the presence of plasmids in the cell, very-low copy vectors and chromosomal integration usually eliminate the need for antibiotics selection markers as discussed in Chapter 1.1.1.

1.2.4. Increasing messenger RNA longevity

Degradation of messenger RNA is an important regulatory mechanism limiting the number of times a messenger can be translated into a protein. This form of regulation of gene expression is important in virtually all organisms (Hall et al. 1982; Marcaida et al. 2006). In *E. coli* the average messenger decays with a half-life of about five minutes, but the half-life of individual messengers can be as short as a few seconds or as long as an hour (Baumeister et al. 1991; Bernstein et al. 2002). A wide array of enzymes is responsible for the degradation including the endonucleases RNase E, RNase K and

RNase III, as well as the 3' exonucleases RNase II and polynucleotide phosphorylase (PNPase) (Jana and Deb 2005; Carpousis 2007).

mRNA stability is mainly influenced by sequences in the 5' UTR which can provide protective 5'-terminal hair-pin structures or high-affinity ribosome binding sites, and to a lesser degree by sequences in the 3' region (Bouvet and Belasco 1992; Emory et al. 1992; Arnold et al. 1998; Baker and Mackie 2003). Increasing the longevity of a messenger can thus be achieved by replacing the native 5' UTR with 5' UTR from messengers with known high stability, like the *E. coli ompA* transcript (Belasco et al. 1986; Hansen et al. 1994). In addition, 3' UTR hair-pin structures can stabilize mRNA by blocking exonucleotic degradation from the 3' terminus (Newbury et al. 1987; Massé et al. 2003). The use of inserting stabilizing hairpin structures to both the 3' and 5' ends of transcripts have been reported (Smolke et al. 2000).

A completely different approach is to use engineered strains with deficiencies in the RNase machinery. One such example is the Invitrogen BL21 star strain, which carries a mutation in the gene encoding RNase E (*rne*131 mutation) (Lopez et al. 1999).

1.2.5. Engineering of the 5' untranslated region for increased translation initiation efficiency

Regulation of translation is a very important level of gene expression control. The sequence and structure of the messenger RNA determine its interaction with the translational machinery and hence the efficiency and frequency of translation, and so both sequence and structure are possible targets for optimizing recombinant gene expression. The rate-limiting step for translation is generally considered to be the initiation of translation (Draper 1996; Skorski et al. 2006). The *E. coli* mRNA region important for initiation usually spans a few tens of nucleotides bracketing the translation start, thus including parts of the 5' untranslated region (5' UTR, also called "leader") (Laursen et al. 2005) and it is generally understood that the wide range of translation efficiencies observed for different mRNA molecules is predominantly due to unique structural features in this region (Makrides 1996). A highly translated mRNA contains some or all

of the following elements: (I) a Shine-Dalgarno (SD) sequence which interacts with the more-or-less complementary 3' end sequence (…ACCUCCUUA-3') of 16S rRNA (ASD) (Shine and Dalgarno 1974), (II) the translation initiation codon AUG (the frequency of *E. coli* start codons are AUG (83%), GUG (14%), or UUG (3%) (Ma et al. 2002)), (III) a pyrimidine tract that interacts with S1 ribosomal protein (Boni et al. 1991; Zhang and Deutscher 1992; Sengupta et al. 2001) and (IV) base-specific enhancer elements upstream (O'Connor et al. 1999) or downstream (Sprengart et al. 1996) of the initiation codon.

The importance of the SD sequence has been documented in studies where the expression of specific messengers is first reduced through mutations in the SD sequence and then restored through compensatory mutations in the ASD (Hui and de Boer 1987; Jacob et al. 1987). However, the strength of the SD sequence is only weakly correlated to translation levels (Lee et al. 1996), and the existence of natural messengers completely lacking SD sequences (Boni et al. 1991) and leaderless mRNAs (Moll et al. 2002) suggest that the other elements must be important as well.

Another important determinant for translation initiation besides primary sequences is the structure of the mRNA molecule. The structure of the mRNA determines translation efficiency by inhibiting or permitting 30S ribosomal subunit access (de Smit and van Duin 1994; Ramesh et al. 1994), for instance can occlusion of the SD sequence and/or the initiation codon inhibit translation (Cruz-Vera et al. 2004).

Several different strategies have been attempted to maximize translation initiation efficiency and to minimize mRNA secondary structure. A random alteration of the translation initiation region resulted in a library that when screened exposed mutants with a 10-fold difference in translational strengths (Simmons and Yansura 1996). The enrichment of adenine and thymidine nucleotides in the leader has been shown to enhance the expression of some genes (Chen et al. 1994). Mutations of positions upstream and downstream of the SD sequence have been shown to suppress formation of secondary structures and thereby enhance the translation efficiency (Coleman et al. 1985; Gross et al. 1990; Jiang et al. 2001). Translational enhancers have been utilized for improved recombinant expression. One example is the U-rich region upstream of the SD sequence in the *E. coli atpE* gene (McCarthy et al. 1985). This 30 base-pair sequence has

been used to successfully over-express the human interleukin-2 and interferon β genes (McCarthy et al. 1986).

1.2.6. Optimizing codon usage

Each amino acid is encoded by more than one codon and each organism has its own bias in the usage of the 61 available codons. The intracellular tRNA populations are correlated to the codon bias of the mRNA population (Dong et al. 1996). Highly expressed genes tend to contain codons for which the cell has abundant tRNA whereas genes that are expressed at low levels tend to include rare codons. Such codon biases can be a big problem for high-level expression of recombinant genes (Clark 1998). If the mRNA from the recombinant gene contains rare codons, or if the amino acid composition of the correlated protein is skewed compared to typical *E. coli* proteins (Table 3), it is possible that translational problems – including translational stalling, premature translation termination, translation frameshift and amino acid mis-incorporation – may occur, leading to reduction in quantity or quality of the protein synthesized (Goldman et al. 1995; Kane 1995; Kurland and Gallant 1996).

General and unambiguous "rules" which predict whether the content of low-usage codons in a specific gene might adversely affect the efficiency of its expression in *E. coli*, have not been discovered (Zhang et al. 1991). The reason for this may be that several variables, such as positional effects, the clustering or interspersion of rare codons, mRNA secondary structure, and other effects, confound the experimental data, as shown for *Plasmodium falciparum* (Peixoto et al. 2004).

The presence of the most infrequent *E. coli* codons, AGG and AGA for arginine, has been shown to be a limitation for expression of several mammalian genes in *E. coli* (Kane et al. 1992; Hua et al. 1994). To circumvent this problem, the *argU* (*dnaY*) gene which codes for the tRNA for AGG and AGA has been co-expressed, and this resulted in high-level production of the target proteins (Chen and Inouye 1994; Hua et al. 1994). Such experiments validate the construction of strains that over-express tRNAs which will alleviate problems connected to infrequent codon usage in recombinant genes. An

example of such engineered strains is the Rosetta strain (Novagen) which carries tRNA genes for the rare codons AGG, AGA, AUA, CUA, CCC, and GGA. Derivatives of Rosetta have been constructed which harbors additional tRNA genes (Table 4). Another example is the BL21-CodonPlus-RP strain (Stratagene) which carries the *argU* gene. Additional versions of this strain have also been constructed which include more genes for rare tRNAs.

Table 3: Rare codon usage in the *Escherichia coli* strains K12 and B. The codon frequencies are based on codons from sequences submitted to the GenBank (Nakamura et al. 2000)

Rare codons	Amino acid	Frequency (K12)	Frequency (B)
AGG	Arg	0.11	0.21
AGA	Arg	0.20	0.24
CGA	Arg	0.35	0.24
CUA	Leu	0.39	0.34
AUA	Ile	0.43	0.50
UGU	Cys	0.51	0.42
CCC	Pro	0.54	0.24
CGG	Arg	0.54	0.50
UGC	Cys	0.64	0.58
ACA	Thr	0.70	0.61
CCU	Pro	0.70	0.58
UCA	Ser	0.71	0.61
GGA	Gly	0.79	0.82
CCA	Pro	0.84	0.74
UCC	Ser	0.86	0.80
AGU	Ser	0.87	0.90
UCG	Ser	0.89	1.14

Another approach to circumvent adverse codon bias is to chemically redesign the gene of interest, or parts thereof, thereby replacing all problematic codons with synonymous, frequently used codons (Gustafsson et al. 2004). Such *de novo* gene synthesis has traditionally largely been avoided due to high costs (Wu et al. 2004). Recently the synthesis of genes has gained more popularity as the technology has matured and the costs have dropped comparably to what has been observed for synthetic

oligonucleotides during the last two decades (Gustafsson et al. 2004). Such *de novo* gene synthesis services can also attempt to optimize the gene for increased mRNA stability by reducing stable structures, i.e. 5' stem-loop formations, and, if desired, modulate the GC content according to the host.

The results from codon-optimizing mammalian genes for expression in *E. coli* frequently results in five to 15-fold increase in protein production, sometimes corresponding to about 5% of total soluble protein in the cell (Gustafsson et al. 2004), and has also been reported to reach between 10% and 20% in specific cases (Deng 1997; Feng et al. 2000). Still, *de novo* gene synthesis does not always lead to an increase in the expression of problematic heterologous genes, even when all the changes are synonymous, and in a few cases the reported expression levels decreased (Biemelt et al. 2003; Warzecha et al. 2003).

1.2.7. Optimizing recombinant protein quality

Rapid production of recombinant proteins can lead to macromolecular crowding with unfavorable folding conditions, resulting in formation of structurally complex aggregates (inclusion bodies) (Betts and King 1999; Villaverde and Carrío 2003). Although inclusion body formation greatly simplifies protein purification, there is no guarantee that the necessary *in vitro* refolding will yield large amounts of correctly folded protein (Clark 2001). The yield after refolding is usually between 10 and 50% of the initial inclusion body volume (Carrío and Villaverde 2002). To prevent inclusion body formation when expressing recombinant proteins many approaches are possible: co-expression of molecular chaperones, fusion with solubilization partners, lowering of temperature, use of weak promoters, addition of sucrose and glycyl betaine to the growth medium, translocation to the periplasm, fermentation at extreme pH values, and use of low copy number vectors (Makrides 1996; Schumann and Ferreira 2004; Terpe 2006).

1.2.7.1. Co-expression of molecular chaperones

Cells have evolved an elaborate protein quality control system which includes the activity of molecular chaperones, proteins that bind the nascent polypeptide and prevent off-pathway aggregation reactions by transiently binding hydrophobic domains (Gottesman et al. 1997). Many different studies have documented the benefits on protein folding from co-expression of different chaperones (Amrein et al. 1995; Thomas and Baneyx 1996; Nishihara et al. 1998; Nishihara et al. 2000) Unfortunately, only the chaperone involved with the folding of the specific protein will have a beneficial effect on the protein's folding, so identification and co-expression of the correct chaperone is essential (Baneyx 1999).

1.2.7.2. Fusion partners for increased solubility and stability

Expressing recombinant genes with protein fusion partners may enhance the solubility of the corresponding proteins, thereby preventing formation of inclusion bodies (Davis et al. 1999; Kapust and Waugh 1999; Sørensen et al. 2003). Fusion partners may also protect the recombinant protein from intracellular proteolysis (Martinez et al. 1995; Jacquet et al. 1999). Common fusion partners for increased solubility are *E. coli* maltose-binding protein (MBP), glutathione-S-transferase and thioredoxin. Of these three, the maltose-binding protein has been demonstrated to be the most efficient (Kapust and Waugh 1999). Another good solubilization partner is *E. coli* N-utilizing substance A (NusA). Both MBP and NusA have been used for solubilization of highly insoluble single-chain antibodies in the cytoplasm of *E. coli* (Bach et al. 2001; Zheng et al. 2003).

Fusion partners usually need to be removed after purification *in vitro*, and this is usually achieved by using cleavage sites between the fusion partner and the recombinant protein which facilitate site-specific proteolysis (Sørensen and Mortensen 2005).

1.2.7.3 Translocation to the periplasmic space or secretion to the extracellular growth media

Another method to circumvent problems with inclusion body formation is to translocate the recombinant proteins to the oxidizing environment of the periplasm. In addition to minimizing inclusion body formation, the periplasm contains less proteases than the cytoplasm (Gottesman 1996), thereby increasing the biological activity, stability and solubility of translocated proteins. In addition, transport to periplasm can simplify downstream processes and secure N-terminal authenticity of the produced protein (Makrides 1996; Cornelis 2000).

Another possibility is to secrete the recombinant protein into the extracellular growth medium. One advantage of secretion is that *E. coli* does not normally secrete proteins (Pugsley et al. 1997; Hannig and Makrides 1998) and thus contamination of the product by host proteins can be minimized. As for the periplasm, the proteolytic activity is much lower in the culture medium (Gottesman 1996).

Protein translocation and secretion in *E. coli* are complex processes and several factors can influence the success of translocating and secreting recombinant proteins (Mergulhão et al. 2005), including the size of the recombinant protein (Baneyx 1999; Palacios et al. 2001), the amino acid composition of the leader peptide as well as the target protein (Kajava et al. 2000; Belin et al. 2004), and the expression levels (Simmons and Yansura 1996). The last effect is probably a consequence of the limited secretion capacity of the *E. coli* transport machinery, and when overwhelmed the excess of expressed recombinant protein is likely to accumulate in cytoplasmic inclusion bodies (Mergulhão et al. 2004).

E. coli does not naturally secrete large amounts of proteins (Sandkvist and Bagdasarian 1996) but there are a few systems available for recombinant translocation with possibilities for subsequent secretion: The HlyA system is a type I system and as such transports protein in one step across the two cellular membranes, without a periplasmic intermediate (Binet et al. 1997). The SecB-dependent pathway, the signal recognition (SRP) and the twin-arginine translocation (TAT) pathways, are all type II systems which transport the protein to the periplasmic space (Mergulhão et al. 2005). The

Table 4: Examples of some *E. coli* strains available for recombinant protein production

Strain	Derivation	Key features	Reference or source
BL21	B834	*lon* and *ompT*	Novagen
BL21 trxB	BL21	*trxB*; *lon* and *ompT*.	Novagen
Bl21 CodonPlus®-RIPL	BL21	Contains additional codons rarely used in *E. coli*; *lon* and *ompT*	Stratagene
BL21 (DE3)	BL21	Contains the T7 RNA polymerase necessary for T7 promoter activity.	Novagen
BL21 (DE3) pLysS	BL21	Contains the T7 RNA polymerase necessary for T7 promoter activity; contains the T7 lysozyme which degrades T7 RNA polymerase.	Novagen
BLR	BL21	*recA*; stabilizes tandem repeats; *lon* and *ompT*.	Novagen
HMS174	K-12	*recA*; rifampicin resistant.	Novagen
JM 83	K-12	Usable for translocating recombinant proteins into the periplasm.	(High et al. 1993)
Origami	K-12	*trxB* and *gor*.	Novagen
Origami B	BL21	*trxB and gor*; *lon* and *ompT*.	Novagen
Rosetta	BL21	Contains additional codons rarely used in *E. coli; lon* and *ompT*.	Novagen
Rosetta-gami	BL21	*trxB* and *gor*; contains additional codons rarely used in *E. coli; lon* and *ompT*.	Novagen
BL21 star	BL21	*rne*131; *lon* and *ompT*.	Invitrogen
MM-294	K-12	Contains mutations in the acetate formation pathway.	(Bauer et al. 1990)

One disadvantage of producing recombinant proteins in *E. coli* is its inability to form disulfide bonds in the cytoplasm. This inability is caused by the reducing environment of the cytoplasm as well as the activity of at least five proteins (thioredoxins 1 and 2, and glutaredoxins 1,2 and 3 which are involved in the reduction of disulfide bridges that transiently arise in cytoplasmic enzymes (Åslund and Beckwith 1999). Key enzymes in disulfide bond formation regulation are thioredoxin reductase and glutaredoxin reductase (encoded by the genes *trxB* and *gor*, respectively), which reduces

oxidized thioredoxins and thereby recycles them (Prinz et al. 1997). As a consequence of this, hosts have been designed that are *trxB* and *gor* negative, resulting in increased cytoplasmic disulfide bond formation, e.g. the BL21 trxB, Origami, Origami B and Rosetta-gami strains.

Another important host modification is adjusted tRNA pools for increased expression of genes with rare codons, as was discussed in Chapter 1.2.6.

1.2.9. Optimizing growth conditions

Although the specific recombinant gene expression is reduced at high cell concentrations (Yoon et al. 2003) high cell density cultures (HCDC) of *E. coli* is commonly used for increased recombinant protein yield since the densities can reach in excess of 100 g/L (Lee 1996; Jeong and Lee 1999; Gerigk et al. 2002; Jeong et al. 2004). In addition, HCDC is a cost-effective methods, where the cost reduction is achieved due to reduction in size of fermentation equipment and upstream and downstream utilities (Shiloach and Fass 2005; Choi et al. 2006).

Care must be taken when optimizing HCDC since studies have shown that nutrient composition, nutrient feeding strategy and fermentation variables such as temperature, pH, acetate accumulation and dissolved oxygen can affect among others transcription, translation, proteolytic activities and secretion (Lee 1996). In addition, problems with heat and mass transfer are common. To solve these issues a variety of techniques have been developed, including the use of different nutrient feeding strategies, specific fermentation strains and the use of oxygen enriched air or pure oxygen (Lee 1996).

Selection of nutrient feeding strategy is crucial because it affects the metabolic pathway fluxes and consequently affects the maximum attainable cell concentration, the specific productivity of recombinant protein expression and formation of by-products (Lee 1996). The simple feeding strategies such as rate feeding, increased feeding, or exponential feeding are commonly used (Qiu et al. 1998; Wong et al. 1998; Babu et al. 2000). More advanced feeding strategies are those with feedback control (Shin et al. 1997;

Chao et al. 2002; Kleist et al. 2003; Hu et al. 2004; Yuan et al. 2004). These are based on the observation that pH and dissolved oxygen changes when the principal carbon source is depleted, and will use this as a cue for the addition of concentrated feeding solution to the bioreactor. Feedback systems prevent overfeeding of nutrients.

Different engineered strains of *E. coli* have been used in HCDC, mainly to alleviate problems arising from accumulated acetate which inhibits cell growth because of change in pH and anion-specific effects (Warnecke and Gill 2005). Some of these strains have defects in the acetate biosynthesis pathway, e.g. the MM-294 strains (Bauer et al. 1990; Farmer and Liao 1997). It has been shown that *E. coli* strain B has the advantage of producing much less acetic acid than K-12 strains even in the presence of excess glucose (Shiloach et al. 1996; Noronha et al. 2000).

1.3. Recombinant protein production with the Pm promoter and its cognate transcription activator XylS

1.3.1. The *Pm* promoter from the *Pseudomonas putida* TOL plasmid pWW0

The *Pm* promoter originates from the archetypal TOL plasmid pWW0 from *P. putida* (Williams and Murray 1974) which contains genes for toluene catabolism organized in two operons, the upper and meta pathway operons (Worsey et al. 1978; Ramos et al. 1997; Greated et al. 2002). The genes encoding the enzymes in the upper pathway, which oxidize toluene to benzoate and xylenes to toluates, are under transcriptional control of the *Pu* promoter (Harayama et al. 1989) whereas the second operon (the *meta* operon) contains the genes for the *meta*-cleavage pathway responsible for further oxidation and decarboxylation of the carboxylic acids to Krebs cycle substrates, and are under transcriptional control of the *Pm* promoter (Worsey and Williams 1975; Franklin et al. 1981; Harayama et al. 1984; Ramos et al. 1987).

Transcription from *Pu* is controlled by the positive regulator XylR which is activated by substrates for the upper pathway (Abril et al. 1989; Delgado and Ramos 1994), while transcription from *Pm* is controlled by the positive regulator XylS which is activated by substrates for the meta pathway, e.g. *m*-toluic acid (Inouye et al. 1981; Franklin et al. 1983; Ramos et al. 1986; Marqués et al. 1994). The basal expression from *Ps2*, one of *xylS'* two promoters, establishes a basal, low expression level of XylS in the cells (Marqués et al. 1998). In addition, activated XylR stimulates transcription from *Ps1*, the second *xylS* promoter, leading to elevated levels of XylS and transcription stimulation from *Pm*. (Inouye et al. 1987; Marqués et al. 1998).

1.3.1.1. Architecture of the *Pm* promoter

The *Pm* promoter extends to about 70 base pairs upstream of the transcription start site. It can be divided into two distinguishable regions: the XylS interaction region (the *meta* operator, *Om*), which extends from -34 to -69, and the downstream RNA polymerase recognition region (Franklin et al. 1981; Harayama et al. 1984; Kessler et al. 1993; Kessler et al. 1994; Gallegos et al. 1996; González-Pérez et al. 1999; González-Pérez et al. 2002) (Figure 7).

Om is composed of two 15-bp imperfect direct repeated motifs (5'-TGCAAPuAAPyGGNTA-3') extending from -69 to -55 and from -48 to -34. Mutation mapping has revealed that the nucleotides located at -48 to -45 and at -58, -59, -61, and -69 are the most critical bases for appropriate XylS-*Pm* interactions (Kessler et al. 1993; Gallegos et al. 1996; González-Pérez et al. 1999).

in other organisms than *Pseudomonas* was examined, and it was shown that regulated expression from *Pm* was obtainable in many organisms, but not in *Proteus mirabilis* and *Rhizobium meliloti*. High expression levels were reached in many *Pseudomonas* subspecies, *Aeromonas hydrophila*, *Aerobacter aerogenes*, *Serratia marcescens*, *Erwinia carotovora* and importantly, *Escherichia coli* (Mermod et al. 1986). In 1997 a new set of expression vectors based on *Pm* and XylS, the pJB vectors, was reported (Blatny et al. 1997; Blatny et al. 1997). An expression cassette featuring *Pm* and *xylS* was combined with a minimal replicon based on the broad-host range RK2 plasmid. The minimal RK2 replicon consisted of the origin of vegetative replication (*oriV*) and the *trfA* gene (Perri et al. 1991). By substituting the wild-type *trfA* gene with copy-up mutant genes the copy number of the pJB vectors can be increased 24-fold (Haugan et al. 1992) from its natural number of 5 to 7 per chromosome, although the maximum tolerance level of elevated copy numbers differs from species to species (Haugan et al. 1995). The strong correlation between inducer concentration and resulting transcription level from *Pm* has been used, together with *Pm* promoter mutants and *trfA* gene mutants, to obtain regulated recombinant protein expression levels from JB vectors covering a ten-thousand- to a hundred-thousand-fold continuous range, in several Gram-negative bacteria, and this expression has been shown to be useful for both metabolic engineering and maximized protein production (Blatny et al. 1997; Brautaset et al. 2000; Winther-Larsen et al. 2000; Gimmestad et al. 2003; Bakkevig et al. 2005).

2. AIMS OF THE STUDY

The main goal of this study was to gain a deeper understanding of recombinant protein production in *Escherichia coli* with expression vectors based on *Pm* and *xylS*, the JB-vectors. As elaborated on in the preceding chapters there are many possible targets for maximizing recombinant protein production in *E. coli*. In this work it was decided to focus on some selected targets which were believed to yield insights into how the JB-vectors can be applied for industrial scale production of proteins and illuminate fundamental mechanisms behind gene expression. These key targets were:

- Studies on how the plasmid copy number of JB-vectors can be adjusted to maximize recombinant protein expression in high cell density cultures (HCDC). The human antibody fragment scFv-phOx was chosen as the model protein.

- The use of different signal peptides for efficient translocation of recombinant proteins. It is known that different signal peptides can be used for translocation of recombinant proteins, but their relative efficiencies and their gene-specificities are largely unexplored. This research would in addition include the design of a novel signal peptide sequence based on a consensus sequence. The industrially relevant proteins scFv-phOx, hGM-CSF and hIFN-α2b were chosen for this study.

- Directed evolution of the transcription activator XylS for increased induced stimulation of transcription from *Pm*. This would be the first time a bacterial transcription activator would be subjected to directed evolution for the purpose of increasing recombinant protein production.

- Use results from the screening of the XylS mutant library to shed light on the structure/function relationship of XylS, and, if possible, use this information to *in silico* predict the XylS N-terminal domain structure.

- *De novo* synthesis of the IFN-α2b and *xylS* genes for circumventing low expression due to unfavorable codon usages.

Investigations into these topics would hopefully result in a broader knowledge of this specific expression system which would greatly benefit its use in applied biotechnology and would be likely to have a general value applicable to other expression systems, both for use in *E. coli* and other bacterial hosts.

In parallel with this project, other members of our research group have focused on other targets, including the DNA and RNA sequences affecting transcription and translation, respectively.

3. LIST OF PAPERS

Paper I: Sletta, H.; Nedal, A.; Aune T.E.V; Hellebust H.; Hakvag S.; Aune R.; Ellingsen T.E.; Valla S.; Brautaset T. Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host-toxic single-chain antibody fragment in *Escherichia coli*. *Appl. Environ. Microbiol*. **70**(12):7033-9 (2004).

Paper II: Sletta, H.; Tondervik, A.; Hakvag, S.; Aune, T.E.V; Nedal, A.; Aune, R.; Evensen, G.; Valla, S.; Ellingsen, T.E.; Brautaset, T. The presence of N-terminal secretion signal sequences leads to strong stimulation of the total expression levels of three tested medically important proteins during high-cell-density cultivations of *Escherichia coli*. *Appl. Environ. Microbiol*. **73**(3):906-12 (2007).

Paper III: Aune, T. E. V.; Bakke, I.; Drabløs, F.; Lale, R.; Brautaset, T.; Valla, S. Directed Evolution of the Transcription Factor XylS for Development of Improved Expression Systems. *Manuscript in preparation*.

Paper IV: Aune, T.E.V.; Stüttgen, F.; Lale, R.; Valla, S. A synthetic version of the *Pseudomonas putida xylS* gene designed for optimal expression in *Escherichia coli* has a strongly reduced translation efficiency caused by mutations in the 5' part of the coding sequence. *Manuscript in preparation.*

4. SUMMARY OF RESULTS & DISCUSSION

4.1. High-level production of human proteins through use of signal peptides for translocation in high-cell density cultures

(paper I and II)

4.1.1 Selection of model proteins and construction of expression vectors for production of therapeutic proteins

A derivative of the expression vector pJB658 (Blatny et al. 1997) was chosen for the production of the human antibody fragment scFv-phOx, transcribed from the *Pm* promoter. scFv-phOx requires disulfide bond formation for correct folding, and therefore translocation to the *E. coli* periplasm is necessary. To facilitate efficient translocation, a cassette system was developed where sequences for different N-terminal signal peptides could easily be exchanged. To be able to quantify the production and purify the protein products, a c-*myc* and a *his6* tag were included so as to be C-terminally fused to the expressed recombinant protein (Figure 9).

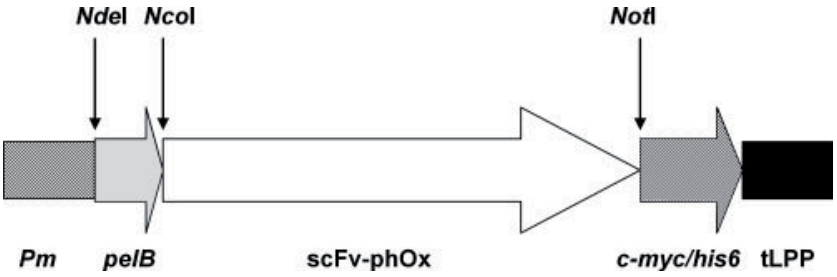

FIGURE 9: Map of the cloning and expression cassette vector featured on the pJB658 derivate used for expression of therapeutic proteins. The map shows restriction sites useful for gene cloning with and without a fused signal peptide sequence, as well as for exchanging the signal peptide sequence. The map starts with the *Pm* promoter and ends with the transcription terminator tLPP. *pelB* encodes an N-terminal signal peptide that directs translocation of scFv-phOx to the *E. coli* periplasm.

4.1.2. The highest production of scFv-phOx was reached with medium plasmid copy number

The production yield of scFv-phOx from HCDC with varying copy number vectors was determined. As anticipated the scFv-phOx production increased when the wild type copy number (5-7 copies per chromosome) (vector pJBphOx) was elevated three-to-fourfold by use of the *trfA* mutant gene cop271C (vector pJBphOx-cop271), from about 0.8 g/L to 2.3 g/L with 3 mM *m*-toluic acid (Figure 10). Increasing the copy number further by use of the *trfA* mutant gene cop251 (vector pJBphOx-cop251), which raises the copy number eightfold, caused extensive plasmid loss before the time of induction, with resulting low scFv-phOx production.

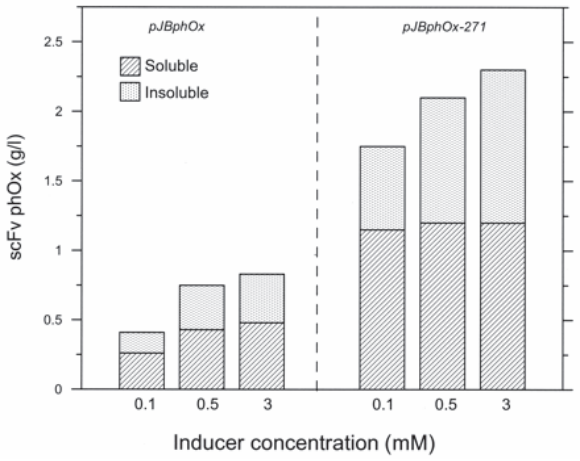

Figure 10: Total volumetric yields of scFv-phOx (soluble and insoluble) at different induction levels in HCDC fermentations with *E. coli* RV308 harboring pJBphOx and pJBphOx-271.

By deleting a region of pJBphOx-cop251 encompassing essential parts of the *xylS* gene the expression of scFv-phOx was greatly reduced, but the plasmid was then 100% stable. We therefore conclude that the reason for the observed plasmid loss is high basal expression from *Pm* caused by increased XylS levels resulting from elevated plasmid copy number, and not the high copy number per se. The increased basal expression of

scFv-phOx reduces the growth rate of the plasmid-containing cells and cells without plasmid thus take over the culture. Ampicillin in the broth will not counter this process since it is quickly degraded under HCDC conditions.

The results indicate that it might have been possible to produce scFv-phOx from pJBphOx-cop251 if the basal expression from *Pm* could be kept low. This could perhaps be accomplished by controlling the expression of XylS from a tight promoter *in trans*. If so, it is possible that the production of scFv-phOx, or another protein of interest, may be increased.

4.1.3. Signal peptides affect both translocation and expression of recombinant genes in a gene-specific fashion

It is possible that signal peptides not only affect the translocation efficiency of the recombinant protein, but also the expression rates. To investigate this expression from the most productive vector, pJBphOx-cop271, but with *pelB* exchanged with two alternative signal peptides sequences, *ompA* and CSP, yielding the vectors pJBphOx-271ompA and pJBphOx-271CSP, were tested. The signal peptide sequence CSP was designed for this work and is based on the consensus sequence for signal peptides in gram negative bacteria. Both the productions of soluble and insoluble scFv-phOx from these new vectors were lower than with *pelB* (totaling 0.25 and 0.60 g/L with *ompA* and *CSP*, respectively) (Figure 11). Based on this it was concluded that the choice of signal peptide sequence has a strong impact on both the translocation and the expression level of recombinant scFv-phOx.

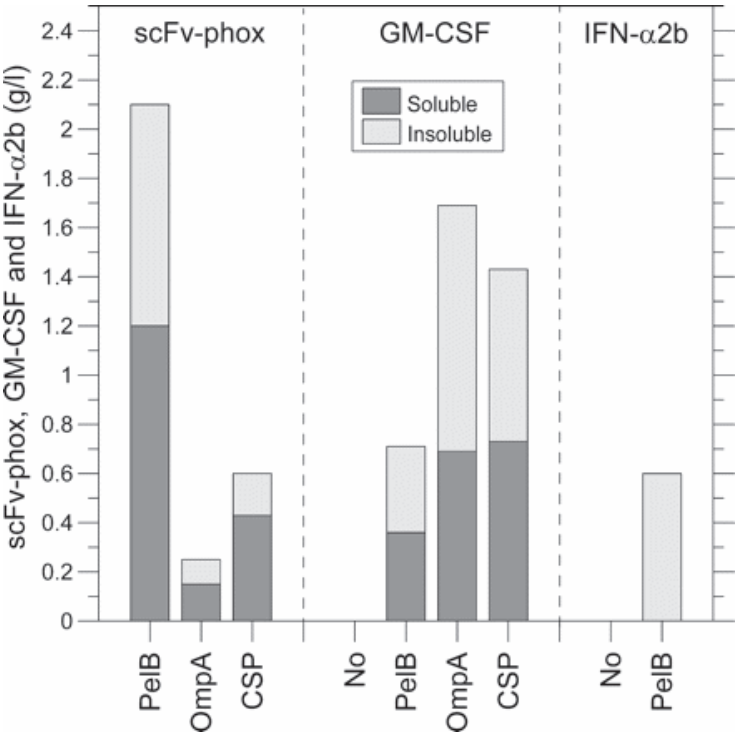

Figure 11: Production levels of scFv-phOx and GM-CSF with different signal peptide sequences, and no signal sequence, during HCDC. Both soluble and insoluble fractions are given for all recombinant strains tested. Production of synthetic IFN-α2b was not tested with *ompA* and CSP. Production of scFv-phOx without signal peptide in shake flask experiment gave no detectable product and was therefore not tested in HCDC.

This was further confirmed by expressing GM-CSF with and without the PelB signal peptide (vector pGM29 with derivatives). Although GM-CSF does not need disulfide bond formation, and thus translocation to the periplasm, its expression was greatly enhanced when fused to PelB. The *pelB* sequence was then exchanged with *ompA* and *CSP* and it was observed that the OmpA and CSP signal peptides cause an up to twofold higher production of GM-CSP than PelB (Table 5 and Figure 11), indicating that the effect of different signal peptides is highly gene-specific.

Table 5: Expression data of GM-CSF-producing strains during HCDC

Plasmid	Signal peptide	Protein product		mRNA
		Soluble	Insoluble	
pGM29	None	BD	BD	1.0x ± 0.1x
pGM29pelB	PelB	0.39 ± 0.03	0.4 ± 0.05	3.0x ± 0.5x
pGM29CSP	CSP	0.80 ± 0.11	0.7 ± 0.15	3.4x ± 0.3x
pGM29ompA	OmpA	0.70 ± 0.09	1.0 ± 0.25	8.0x ± 0.9x

BD: Below detection limit (below 1 mg/L). The values for soluble and insoluble protein yield are in g/L.

To investigate whether the effect of signal peptides are at the level of transcription, translation, or both, quantitative PCR was used to measure GM-CSF mRNA levels (Table 5). These results show that the mRNA levels of strains with *pelB* (pGM29pelB) and *CSP* (pGM29CSP) were only about three times higher than for the strain without signal peptide (pGM29), although the increases in total protein are much higher. This implies that the low GM-CSF expression level of pGM29 is largely due to inefficient translation and not poor transcription.

The finding that signal peptides seem to have a general positive effect on expression of eukaryotic recombinant genes might have big implications for recombinant protein production in bacteria. The low translation of recombinant genes that are evolutionary distant from the host cell is possibly caused by sequences in the start of the coding region that lower the translation initiation efficiency, perhaps by forming secondary structures that shield the translation machinery from the translation initiation region of the mRNA (Simmons and Yansura 1996; Puri et al. 1999). This phenomenon will be further studied in Chapter 4.4. By separating the recombinant gene from the 5' UTR with a signal peptide sequence known to be efficiently translated such inhibitory secondary structures are less likely to form. If so, substituting signal peptide sequences with other well-expressed sequences should also overcome this problem.

The observation that the effect of signal peptides is gene-specific represents an obstacle to optimal utilization of signal peptides for recombinant protein production. The reason for this gene-specificity might be interaction(s) between the signal peptide sequence and the start of the recombinant gene at the transcriptional and/or translational level, causing bottlenecks in expression. In that case, no single signal peptide may be

found that can lead to high level expression of any recombinant gene, but if the interactions can be understood it might be possible to engineer signal peptides for each recombinant gene of interest, resulting in maximized expression. Another point is that maximized expression may not be compatible with maximized translocation efficiency, and thus compromises between the two may be necessary.

4.1.4. The expression level of human IFN-α2b is low partly because of unfavorable codon usage

Production of IFN-α2b was below the detection limit both with and without signal peptides. Detection of product from *E. coli* strain BL21-CodonPlus(DE3)-RIPL indicated that this was caused by unfavorable codon usage. To obtain good production of this protein the IFN-2αb gene was *de novo* synthesized to contain optimal *E. coli* codon usage. The newly synthesized IFN-α2b gene was then expressed fused to the pelB signal peptide sequence and a yield of about 0.6 g/L was obtained (Figure 11). All of the produced protein was found as insoluble aggregates.

The starting point for *de novo* synthesizing of the IFN-α2b was no detectable production. This must not be understood to imply that low initial production is a prerequisite for using synthetic recombinant genes; it may still be possible to obtain an increase in production from synthesized genes that initially have a detectable production, although the rationale for such an approach may be correspondingly weaker. As the costs of gene synthesis continue to drop, synthesis of genes may become popular even for genes that are already expressed relatively efficient, even if the obtained production increase is modest because of limitations posed by other bottlenecks in expression.

communication between the N- and C-terminal domain, RNA polymerase contact and thermodynamic stability.

The structure compose of the first 39 residues of XylS were not predictable. In AraC the first 15 residues constitute an arm which in the uninduced state represses the DNA-binding efficiency of the regulator and closes the β-barrel in the induced state, thereby freeing the DNA-binding domain (Ghosh and Schleif 2001; Wu and Schleif 2001). If this "light-switch mechanism" was true for XylS as well, then deletion of the first 39 residues of the N-terminal region should give a constitutive phenotype, but such a XylS deletion mutant protein has been shown to be almost completely inactive (Kaldalu et al. 2000).

The aggregation of XylS and other AraC/XylS family members when overproduced is well-known (Lorenzo et al. 1993; Michán et al. 1995; Bhende and Egan 2000; Timmes et al. 2004). It is tempting to suggest that this phenomenon prevents adverse effects from supernormal concentrations of these transcription activators, such as stimulation of non-cognate promoters. Recently it has been shown that tyrosine 31 of AraC is at least partly responsible for aggregation when the protein is over-expressed (Weldon et al. 2007). Sequence alignment of AraC and XylS indicates that tyrosine 31 from AraC might be conserved in position 61 in XylS, suggesting that an identical mechanism may be at work for XylS.

The XylS NTD is connected to the XylS CTD through a flexible linker in the region of residues 200 to 211, which has not been possible to model. No substitutions affecting XylS function has been found in this region so the role of the linker is probably restricted to supplying a connection between the two domains, as shown for RhaS and RhaH (Kolin et al. 2007), and possibly provide the flexibility necessary for normal function of the activator, as shown for AraC (Eustance and Schleif 1996).

4.2.3. Combinations of mutations cause apparently unpredictable induction properties of the resulting mutant XylS activators

To test whether different XylS substitutions can be combined to yield XylS activators that further increase transcription from *Pm*, activators containing two or three substitutions were designed rationally via site-specific mutagenesis of *xylS*. Both substitutions identified after error-prone PCR and the R45T substitution which has previously been shown to increase XylS activity (Ramos et al. 1990), were combined. Eighteen different combinations were tested, and all of them gave their hosts higher induced ampicillin tolerance than the cells with XylS proteins with only single substitutions (Figure 15), verifying that the XylS substitutions can work in an additive manner. The XylS protein Syn-15 (containing the three substitutions F3Y, I50T and F97L) gave 5.9 to 6.2 times higher induced ampicillin tolerance than wild type, and only 1.4 to 1.7 fold higher basal ampicillin tolerance, thus increasing the induction ratio from 17 for wild type XylS to at least 56 with Syn-15.

It was not possible to predict the tolerance levels of cells with different *xylS* mutations because no specific correlation between combinations of *xylS* mutations and tolerance levels of the cells were discernable, although there was a general trend that combination of many mutations increased the tolerance. This observation suggests that optimal combinations of *xylS* mutations can be identified only from screening a randomly shuffled *xylS* mutant library.

4.2.4. DNA shuffling of mutant *xylS* genes and selection of XylS mutants proteins with further improved activation properties

To circumvent the unpredictable ampicillin tolerance levels observed upon the rational combinations of useful *xylS* mutations, a staggered extension process (StEP) DNA shuffling strategy for random recombination of mutant *xylS* genes was employed. A

4.3. Selection for constitutive XylS mutant proteins resulted in isolation of xylS mutant genes containing a novel promoter that increase expression from Pm independent of XylS

(unpublished results)

4.3.1. Selection for constitutive activity from the *Pm* promoter resulted in the isolation of a *xylS* mutant gene with the silent mutation G33A and the stop codon mutation C454T

It was decided to screen the XylS mutant library for mutants with enhanced ability to stimulate transcription from *Pm* in the absence of inducers (constitutive XylS mutant proteins). Such proteins can be useful for recombinant expression when controlled expression is unnecessary and they may help to elucidate the mechanisms behind XylS' stimulation of transcription from *Pm*.

The screening resulted in identification of 14 mutated *xylS* genes conferring a 25 to 250-fold increase in basal expression from *Pm*. The induced expression was also increased, although not to the same degree, and thus the isolated XylS mutant proteins were still regulatable (results not shown), at least to some extent.

Sequencing showed that 10 of the *xylS* mutant genes contained the silent point mutation G33A in addition to other mutations which varied among the mutant genes. Surprisingly, one of the *xylS* mutant genes contained the point mutation C454T in addition to G33A. C454T results in a stop codon (UGA) which should truncate the gene product at residue 152. The resulting truncated XylS mutant protein would lack the entire C-terminal domain assumed to be essential for DNA-binding and hence transcription stimulation.

4.3.2. A truncated version of *xylS* when combined with the silent G33A mutation is able to increase transcription from *Pm*

To verify that the observed high basal ampicillin tolerance level resulted from the *xylS* mutation G33A, site-specific mutagenesis was used to introduce this mutation in a *xylS* gene (*xylS-G33A*) and it was observed that the host cell's basal ampicillin tolerance increased from 25 (with wild type *xylS*) to about 450 μg/mL. Quantitative real-time PCR verified that this increase in basal ampicillin tolerance was in fact a result of increased transcription of *bla*.

Addition of the C454T mutation (generating the UGA stop codon) did not change the basal tolerance level, but a *xylS* gene with only the C454T mutation gave a tolerance resistance level of 25 μg/mL, indicating that the UGA is read as a stop signal. A *xylS* gene with the point mutations C454T, G455A and A456G, generating the strongest stop codon UAG, was then constructed. Again, in combination with the G33A point mutation the cells were able to grow to high ampicillin concentrations in the absence of *m*-toluic acid, indicating that the G33A point mutation is not somehow stimulating translational read-through of the UGA stop codon. Then *xylS* genes, with or without the G33A point mutation, with the 3' end starting from nucleotide 455 removed, were generated (*xylS-Δ455* and *xylS-G33A,Δ455*, respectively). The corresponding cells behaved as their stop codon counterparts confirming that the stop codons are read as translation termination signals (Figure 16).

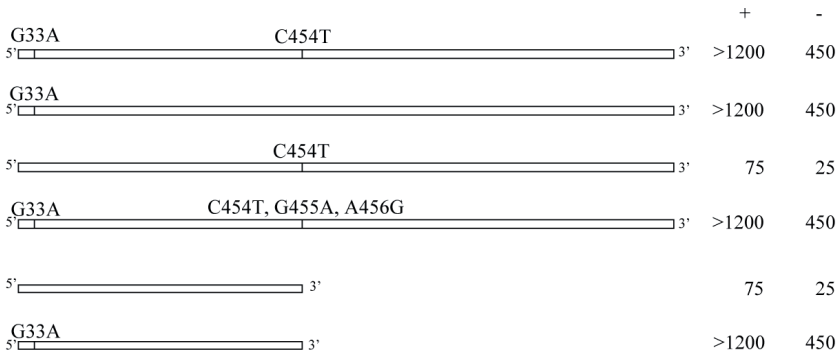

Figure 16: Different *xylS* mutant genes and their host cell's upper ampicillin tolerance limit (given in µg/mL) in the presence (+) of 1 mM *m*-toluic acid and its absence (-). The G33A mutation is silent (Q11Q); C454T generates an UGA stop codon and the C454T, G455A, A456G generate an UAG stop codon. The stop codons correspond to position 152 at the protein level.

4.3.3. *xylS-G33A* is dependent upon *in cis* configuration to increase expression from *Pm*

To investigate whether the *xylS-G33A* was dependent upon an *in cis* configuration to increase expression of *bla*, a set of vectors based on a pET16 replicon with regions from the pTA13 vectors were generated. The regions contained the *xylS* promoters *Ps1* and *Ps2* and the entire *xylS* gene coding region. To be able to select for these plasmids, and to avoid interference with the ampicillin selection system, the vectors contained the tetracycline resistance marker. The wild type *xylS* gene featured on pTA13 was deleted resulting in the vector pTA13Δ*xylS*. Each pET16-tc derivate could now be used together with pTA13Δ*xylS* to investigate the individual *xylS* mutant genes' ability to stimulate transcription from *Pm* when positioned *in trans* (Table 6).

The results showed that *xylS-G33A* leads to a two-fold decrease in induced ampicillin tolerance when *in trans* compared to wild type *xylS*. The truncated *xylS* gene with G33A (*xylS-G33A,Δ455*) was also not able to stimulate expression of *bla*, and behaved like a truncated *xylS* gene without G33A. For the truncated *xylS* genes an inducer effect of about two to three (from 60-65 to 100-200 µg/mL) was observed, but this is

explained by *m*-toluic acid stimulating the heat shock response and thus leading to more transcription from *Pm* through elevated levels of σ^{38} (Marqués et al. 1999). It was concluded that *xylS-G33A* is not able to increase transcription from *Pm* when *in trans*, with or without truncation.

Table 6: Upper ampicillin tolerance levels of cells with *xylS* mutant genes *in trans*

Vectors	+ (µg/mL)	- (µg/mL)
pTA13Δ*xylS*	100-200	60-70
pTA13Δ*xylS*; pET16-tc:*Ps1Ps2-xylS*	1600-1650	60-65
pTA13Δ*xylS*; pET16-tc:*Ps1Ps2-xylS-G33A*	800-900	60-65
pTA13Δ*xylS*; pET16-tc:*Ps1Ps2-xylSΔ455*	100-200	60-65
pTA13Δ*xylS*; pET16-tc:*Ps1Ps2-xylS-G33A,Δ455*	100-200	60-65

(+ denotes the upper ampicillin tolerance limit in the presence of 1 mM *m*-toluic acid and - in its absence)

4.3.4. The G33A point mutation apparently generates a new promoter causing *Pm*-independent transcription of the *bla* gene

As a result of the finding that *xylS-G33A* only increases the basal expression from *Pm* when *in cis*, the genetic context of the mutation was scrutinized and it was discovered that it generates a possible new -35 region reverse positioned relative to transcription of *xylS*, in the 5' end of the *xylS* coding region (Figure 17). If so, the putative transcript traverses the more than 800 base pair span between the start of the *xylS* coding region and the *bla* gene resulting in increased levels of *bla* transcript available for translation. To test this hypothesis the *Pm* promoter in a pTA13 derivate was exchanged with the 5' region of *xylS* either containing the wild type sequence or the sequence with the G33A point mutation. The ampicillin tolerance levels of the cells were significantly increased with the putative promoter compared to the wild type region (results not shown), proving that G33A creates a new promoter. This also proves that the observed increase in *bla* transcript with the G33A point mutation is not caused by a change in XylS function or concentration, but an increase in *bla* transcript originating from a new promoter more than 800 base pairs distant from the *bla* gene.

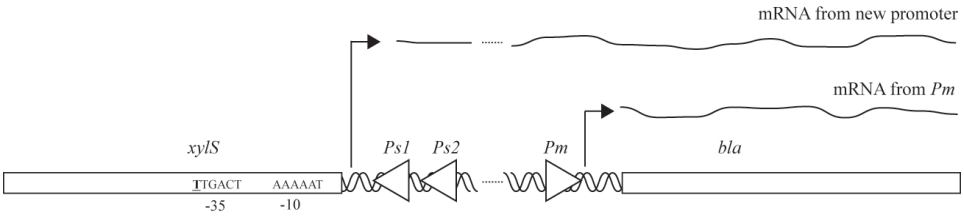

Figure 17: Sketch of the area encompassed by the *xylS* gene coding region and the *bla* gene coding region. The putative -10 and -35 regions of the new promoter are shown within the *xylS* gene. The G33A point mutation, relative to the *xylS* gene, is shown in bold and underscored. mRNAs generated from the *Pm* promoter and the new promoter, are indicated. The figure is not drawn to scale, and the distance between the start of the *xylS* gene and the start of the *bla* gene is approximately 800 base pairs.

These results suggest that to successfully isolate constitutive XylS mutant proteins the pTA13 vector is not optimal. The addition of a transcription terminator between the *xylS* and the *bla* gene would exclude any point mutations like G33A and thus increase the possibility of selecting constitutive XylS mutant proteins.

4.4. De novo synthesized xylS is not properly translated

(paper IV)

4.4.1. The expression level of wild type XylS in *E. coli* may be limited by unfavorable codon usage

In an attempt to over-express XylS for structure determination, the vector pET16b was used to drive *xylS* expression from the strong T7 promoter. By testing out different *E. coli* expression strains it was observed that XylS could only be detected from the BL21-CodonPlus® (DE3)-RIPL strain which contains additional copies of tRNA genes (results not shown). Based on this result the codon frequencies of *xylS* were analyzed and compared to *E. coli* strain K12 and B (Table 7). The analysis showed that *xylS* contains the arginine codons AGG, CGA, CGG, the leucine codon CTA, the isoleucine codon ATA, the proline codon CCC, and the cysteine codon TGC in frequencies at least 2.5 times higher than in *E. coli* strain B and K12. It therefore seemed possible that the specific codon usage in the *xylS* gene is a bottleneck for its efficient expression in *E. coli*.

Table 7: Rare codons in *E. coli* strain B and K12 (Nakamura et al. 2000) with frequencies and occurrences in the *xylS* gene

Rare codons in *E. coli* K12 and B	Encoded amino acid	Frequency per 1000 codons (K12/B)	Frequency in *xylS* per 1000 codons	Occurrence in *xylS*
AGG	Arg	1.1/2.1	12.8	4
AGA	Arg	2.0/2.4	3.1	1
CGA	Arg	3.5/2.4	9.4	3
CTA	Leu	3.9/3.4	9.4	3
ATA	Ile	4.3/5.0	12.8	4
TGT	Cys	5.1/4.2	9.4	3
CCC	Pro	5.4/2.4	12.8	4
CGG	Arg	5.4/5.0	12.8	4
TGC	Cys	6.4/5.8	21.8	7
ACA	Thr	7.0/6.1	3.1	1

4.4.2. The translation efficiency of a *de novo* synthesized and codon-optimized *xylS* gene is severely reduced compared to the wild type gene

To overcome reduced expression of the *xylS* gene caused by its frequent use of rare codons, a synthetic version of *xylS*, *syn-xylS*, was synthesized *de novo*. The *syn-xylS* gene was designed to have a codon usage as similar to *E. coli* as possible, and the codons AGG, CGA, CTA, ATA, CCC and CGG were removed altogether from the gene. All of the changes in *syn-xylS* were synonymous, securing that the gene product from *xylS* and *syn-xylS* are identical. The *xylS* gene in the vector pTA13 (Figure 12) was exchanged with the *syn-xylS* gene, creating the vector pTA13:*syn-xylS*. Induced *E. coli* cells harboring pTA13:*syn-xylS* could not grow on media containing more than 100 μg/ml ampicillin, indicating a reduction in stimulation of transcription from *Pm*.

Real time PCR experiments with primers directed against *syn-xylS* showed that the amount of *syn-xylS* mRNA were similar to the amount of wild type *xylS* mRNA from pTA13 (data not shown). Based on this it was concluded that the synthetic *xylS* gene, although it has been codon optimized, has acquired some elements that strongly reduce the translational efficiency of its mRNA.

4.4.3. Mutations in the 5' 44 base pairs of synthetic *xylS* reduce efficient translation of its messenger

To locate the region within the *syn-xylS* responsible for the reduced translation, a *Psh*AI restriction site in *xylS* (Figure 12) was used to construct two *xylS* chimera genes, one based on wild type *xylS* with the 5' 164 base pairs from *syn-xylS*, designated *syn[1-164]*; and one based on *syn-xylS* with the 5' 164 base pairs from wild type *xylS*, designated *syn[165-966]*. The ampicillin tolerance of cells containing the plasmid pTA13-*syn[1-164]* was measured to be between 0.13 and 0.15 of the tolerance of cells harboring wild type *xylS*, while cells containing the plasmid pTA13-*syn[165-966]* displayed wild type ampicillin tolerance (580 μg/mL).

Gene sowing was then used to exchange parts of the 5' 164 base pairs of wild type *xylS* with the corresponding sequences from *syn-xylS*. All the cells containing novel chimeric *xylS* genes had an ampicillin tolerance level of 0.13 to 0.15, compared to cells harboring wild type *xylS*, including cells containing the chimera with the shortest sequence from *syn-xylS*, *syn[1-44]*, containing only eight mutations. The mutations in *syn[1-44]* are shown in Figure 18. From these results it was concluded that the region responsible for the reduced translation is located within the first 44 base pairs of *syn-xylS*.

	6										13	15	16													27			30							36						42	
xylS	A	T	G	G	A	T	T	T	T	T	G	C	T	T	A	T	T	G	A	A	C	G	A	G	A	A	A	A	G	T	C	A	G	A	T	C	T	T	C	G	T	C	1.00
syn[1-44]	.	.	.	.	.	C	.	.	.	.	.	.	C	.	G	C	.	.	.	.	.	.	.	.	.	.	G	.	.	C	.	.	.	.	.	T	.	.	.	.	.	T	0.13 - 0.15
syn[1-6]	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0.31 - 0.33
syn[1-13]	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0.53 - 0.55
syn[1-15]	.	.	.	.	.	C	.	.	.	.	.	.	C	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0.25 - 0.26
syn[1-16]	.	.	.	.	.	C	.	.	.	.	.	.	C	.	G	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0.25 - 0.26
syn[1-30]	.	.	.	.	.	C	.	.	.	.	.	.	C	.	G	C	.	.	.	.	.	.	.	.	.	.	G	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	0.38 - 0.40
syn[1-36]	.	.	.	.	.	C	.	.	.	.	.	.	C	.	G	C	.	.	.	.	.	.	.	.	.	.	G	.	.	C	.	.	.	.	.	T	.	.	.	.	.	.	0.23 - 0.25

Figure 18: Schematic overview of the 5' 42 base pairs end of the *xylS* coding region in different *xylS* mutant genes. The sequence of wild type *xylS* is shown on top. For the different *xylS* mutant genes only the mutations are indicated. The base pair positions of the eight different mutations are shown above the *xylS* sequence. The induced (1 mM m-toluic acid) ampicillin tolerance levels for cells harboring these *xylS* gene derivatives on the pTA13 plasmid are shown relative to the resistance level of cells with wild type *xylS,* set to 1.00.

Under the assumption that the reduction in *syn-xylS* translation might be caused by a single mutation within the first 44 nucleotides of the *syn-xylS* gene, each of the first eight mutations in *syn-xylS* were reversed by site-specific mutagenesis, and the ampicillin tolerance levels of cells containing the corresponding pTA13 derivatives were measured. None of the cells had gained ampicillin tolerance levels similar to cells containing wild type *xylS* (data not shown).

A set of new *xylS* genes was designed with site-specific mutagenesis with increasing numbers of mutations from *syn-xylS* (Figure 18), were then constructed. The results showed that the full reduction in ampicillin tolerance was only reached for the *xylS* gene containing all the eight mutations from the 5' 44 base pairs of *syn-xylS*.

4.4.4. Most of the mutations in the 5' 44 base pairs of synthetic *xylS* apparently reduce its translation, but the full reduction can be achieved by a single mutation at position 42

Since many of the mutations in *syn[1-44]* affect expression of the *xylS* gene, it was decided to investigate the individual effect of each mutation by constructing, with site-specific mutagenesis, eight *xylS* genes each featuring one of the eight mutations found within *syn[1-44]*. The results of the corresponding phenotype analyses showed that every mutation except the 5th (*syn27*) diminished the ampicillin tolerance. In addition, the 8th mutation (*syn42*) alone caused the full reduction in ampicillin tolerance observed for *syn-xylS* (Figure 19). From this it was concluded that many positions at the 5' end of the *xylS* gene are necessary for normal translation efficiency of the *xylS* mRNA, and that one specific mutation (at position 42) has a particularly strong negative effect.

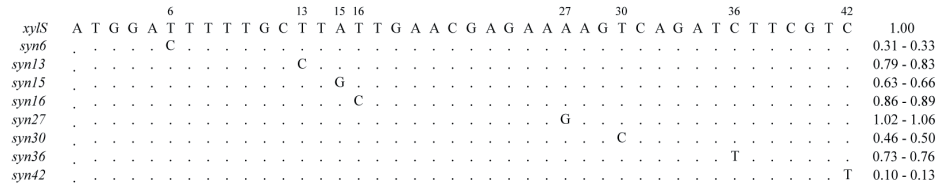

Figure 19: Schematic view of the 5' 42 base pairs end of different *xylS* mutant genes. See legend of Figure 18 for further details.

The finding that a synthetic version of *xylS* is apparently not properly translated was surprising, especially considering that it was designed for optimal translation. Since the region responsible for the reduced translation is at the beginning of the coding sequence, it is tempting to hypothesize that the cause is reduced translation initiation, and not elongation or termination, presumably caused by inhibitory mRNA secondary structures. If so, the algorithms for optimizing mRNA secondary structure, as implemented in the design of the synthetic gene, are far from optimal.

The observation that apparently neutral base substitutions can have drastic effects on gene expression is obviously of significant interest for the entire technology of using

synthetic genes to improve gene expression. The observation may also have relevance for the understanding of expression of natural genes, and an understanding of the underlying mechanisms may therefore be of broad interest.

5. CONCLUSIONS

The work presented in this thesis has focused on broadening our knowledge of recombinant expression in *E. coli* with emphasis on high-level protein production. Key topics have included the use of signal peptides for increased recombinant expression and efficient translocation of the gene product, directed evolution of a transcription regulator for tightly regulated and increased induced expression, and the use of synthetic genes to overcome unfavorable codon usage.

During studies of recombinant expression of human proteins, it was discovered that signal peptides, which are thought to chiefly enable translocation over the inner membrane, have a profound positive effect on the recombinant gene's overall expression. Unfortunately, there was clear gene-specificity regarding which signal peptide would yield the strongest effect on expression effect. To try to overcome this problem with natural occurring signal peptides, a synthetic signal peptide (CSP) based on signal peptide consensus sequence was rationally designed, but again the gene-specificity was observed. Hence, no tested signal peptide was optimal for any given recombinant gene. Therefore, it was concluded that signal peptides have the potential to improve expression of recombinant genes, but that some effort must be undertaken to identify the optimal signal peptide for the specific recombinant gene of interest. Further work must be done to elucidate this phenomenon and hopefully gain an understanding broad enough so that specific signal peptides can be designed for each and every recombinant gene to be expressed.

The importance of tight recombinant protein expression was clearly observed in the study of scFv-phOx with elevated vector copy numbers. At high copy numbers the leakage, apparently caused by excessive amounts of XylS, resulted in extensive plasmid loss and subsequent drop in production. Work is at hand to obtain *Pm* promoter mutants with reduced leakage.

Improving transcription regulators through directed evolution is a new approach for optimizing expression systems with respect to maximized recombinant expression. In this work it has been shown that this is a feasible method at least for XylS and that the

gain can be a 10-fold increase in induced expression of both bacterial (e.g. *bla*) as well as eukaryotic genes (i.e. scFv-phOx), without a proportional increase in the basal expression. It is reasonable to assume that this approach can be used with success on other transcription activators since these are presumably not evolved for maximum transcription stimulation. Many transcription factors used in expression systems are repressors, and their usefulness as templates for directed evolution in effort to obtain higher recombinant expression, is limited.

During this work the N-terminal structure of XylS was *in silico* predicted and a 3D structural model was built. The 3D model suggested that the individual mutations found to cause elevated expression from *Pm* were localized in varying sub-domains of the XylS structure, indicating that methods focusing on mutating smaller parts of a regulator may not be optimal. Many underlying mechanisms are responsible for the overall function of a transcription regulator and all of these might potentially be targets for improvement.

There is no reason to assume that the 10-fold improvement that was obtained represents a maximum, it is probably possible to increase XylS' ability to stimulate transcription from *Pm* even further through additional evolutionary steps, including the introduction of new mutations in a second round of error-prone PCR. Still, the presented results indicate that mutations that improve the induced expression also increase the system's basal expression, although at a lesser degree. Hence, if low basal expression is considered essential further evolution might be counter-productive. In addition to proving that directed evolution is a feasible approach for improving XylS for recombinant expression, it seems reasonable to believe that this general approach is applicable also to other transcriptional activators, both prokaryotic and eukaryotic.

In an attempt to overcome unfavorable codon usage in the *xylS* gene and the gene for IFN-2αb, synthetic versions were designed for optimized translation in *E. coli*. For IFN-2αb this resulted in high-level production. For *xylS*, on the other hand, the synthetic gene was, surprisingly, not properly translated. The reason for this reduced translation was found to be changes in the 5' coding region. One of these changes, in position 42, was sufficient to reduce its translation (presumably), as measured by a reduction in *bla* expression from *Pm*, down to about 10% (relative to wild type *xylS*), although many other

changes in the 5’ region also reduced *bla* expression. Future work will focus on revealing the mechanisms behind this reduction as well as quantifying the reduction of XylS proteins from *syn-xylS* relative to wild type *xylS*.

Through this work the current knowledge of recombinant protein expression has been expanded. The results both broaden the understanding of this fundamental process as well as shed light on methods useful for improving recombinant expression at industrial scales.

6. REFERENCES

Abril, M.-A., C. Michan, K. N. Timmis and J. L. Ramos (1989). "Regulator and enzyme specificities of the TOL plasmid-encoded upper pathway for degradation of aromatic hydrocarbons and expansion of the substrate range of the pathway." J. Bacteriol. **171**(12): 6782-6790.

Amann, E., J. Brosius and M. Ptashne (1983). "Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*." Gene. **25**(2-3): 167-178.

Amann, E., B. Ochs and K.-J. Abel (1988). "Tightly regulated *tac* promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*." Gene. **69**(2): 301-315.

Amrein, K. E., B. Takacs, M. Stieger, J. Molnos, N. A. Flint and P. Burn (1995). "Purification and characterization of recombinant human $p50^{csk}$ protein-tyrosine kinase from an *Escherichia coli* expression system overproducing the bacterial chaperones GroES and GroEL. ." Proc. Nat. Acad. Sci. USA. **92**(4): 1048-1052.

Arnold, F. H., L. Yuan and J. G. Belasco (1998). "mRNA stabilization by the *ompA* 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation." RNA. **4**: 319-330.

Babu, K. R., S. Swaminathan, S. Marten, N. Khanna and U. Rinas (2000). "Production of interferon-alpha in high cell density cultures of recombinant *Escherichia coli* and its single step purification from refolded inclusion body proteins." Appl. Microbiol. Biotechnol. **53**(6): 655-660.

Bach, H., Y. Mazor, S. Shaky, A. Shoham-Lev, Y. Berdichevsky, D. L.-. Gutnick and I. Benhar (2001). "*Escherichia coli* maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies." J. Mol. Biol. **312**(1): 79-93.

Bagdasarian, M., R. Lurz, B. Rückert, F. C. H. Franklin, M. M. Bagdasarian, J. Frey and K. N. Timmis (1981). "Specific purpose plasmid cloning vectors. II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas*." Gene. **16**(1-3): 237-247.

Baker, K. E. and G. A. Mackie (2003). "Ectopic RNase E sites promote bypass of 5'-end-dependent mRNA decay in *Escherichia coli*." Mol. Microbiol. **47**(1): 75-88.

Bakkevig, K., H. Sletta, M. Gimmestad, R. Aune, H. Ertesvåg, K. Degnes, B. E. Christensen, T. E. Ellingsen and S. Valla (2005). "Role of the Pseudomonas fluorescens alginate lyase (AlgL) in clearing the periplasm of alginates not exported to the extracellular environment." J. Bacteriol. **187**(24): 8375-8384.

Baneyx, F. (1999). "Recombinant protein expression in *Escherichia coli*." Curr. Opin. Biotechnol. **10**: 411-421.

Bass, S., Q. Gu and A. Christen (1996). "Multicopy suppressors of Prc mutant *Escherichia coli* include two HtrA (DegP) protease homologs (HhoAB), DksA, and a truncated RlpA." J. Bacteriol. **178**(4): 1154-1161.

Bauer, K. A., A. Ben-Bassat, M. Dawson, V. T. d. l. Puente and J. O. Neway (1990). "Improved expression of human interleukin-2 in high-cell-density fermentor cultures of *Escherichia coli* K-12 by a phosphotransacetylase mutant." Appl. Env. Microbiol. **56**(5): 1296-1302.

Baumeister, R., P. Flache, Melefors, v. Gabain and W. Hillen (1991). "Lack of a 5' non-coding region in Tn*1721* encoded *tetR* mRNA is associated with a low efficiency of translation and a short half-life in *Escherichia coli*." Nucleic Acids Res. **19**(17): 4595-4600.

Belasco, J. G., Nilsson, v. Gabain and Cohen (1986). "The stability of *E. coli* gene transcripts is dependent on determinants localized to specific mRNA segments." Cell. **46**: 245-251.

Belin, D., L.-M. Guzman, S. Bost, M. Konakova, F. Silva and J. Beckwith (2004). "Functional activity of eukaryotic signal sequences in *Escherichia coli*: the ovalbumin family of serine protease inhibitors." J. Mol. Biol. **335**(2): 437-453.

Bentley, W. E., N. Mirjalili, D. C. Andersen, R. H. Davies and D. S. Kompala (1990). "Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria." Biotechnol. Bioeng. **35**(7): 668-681.

Bernstein, Khodursky, Lin, Lin-Chao and Cohen (2002). "Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays." Proc. Nat. Acad. Sci. USA. **99**(15): 9697-9702.

Betts, S. and J. King (1999). "There's a right way and a wrong way: *in vivo* and *in vitro* folding, misfolding ad subunit assembly of the P22 tailspike." Structure. **7**(6): 131-139.

Bhende, P. M. and S. M. Egan (2000). "Genetic evidence that transcription activation by RhaS involves specific amino acid contactis with sigma 70." J. Bacteriol. **182**(17): 4959-4969.

Biemelt, S., U. Sonnewald, P. Galmbacher, L. Willmitzer and M. Müeller (2003). "Production of human papillomavirus type 16 virus-like particles in transgenic plants." J. Virol. **77**(17): 9211-9220.

Binet, R., S. Létoffe, J. M. Ghigo, P. Delepelaire and C. Wandersman (1997). "Protein secretion by Gram-negative bacterial ABC exporters - a review." Gene. **192**(1): 7-11.

Binnie, C., J. D. Cossar and D. I. H. Stewart (1997). "Heterologous biopharmaceutical protein expression in *Streptomyces*." Trends Biotechnol. **15**(8): 315-320.

Birnbaum, S. and J. E. Bailey (1991). "Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*." Biotechnol. Bioeng. **37**(8): 737-745.

Blatny, J. M., T. Brautaset, H. C. Winther-Larsen, K. Haugan and S. Valla (1997). "Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon." Appl. Environ. Microbiol. **63**(2): 370-379.

Blatny, J. M., T. Brautaset, H. C. Winther-Larsen, P. Karunakaran and S. Valla (1997). "Improved broad-host-range RK2 vectors useful for high and low regulated gene expression levels in gram-negative bacteria." Plasmid. **38**: 35-51.

Bolivar, F., R. Rodriguez, M. C. Betlach and H. W. Boyer (1977). "Construction and characterization of new cloning vehicles. I. Ampicillin-resistant derivatives of the plasmid pMB9." Gene. **2**(2): 75-93.

Boni, I. V., D. M. Isaeva, M. L. Musychenko and N. V. Tzareva (1991). "Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1." Nucleic Acids Res. **19**(1): 155-162.

Bouvet, P. and J. G. Belasco (1992). "Control of RNase E-mediated RNA degradation by 5'-terminal base pairing in *E. coli*." Nature. **360**: 488-491.

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

Brenner, M. and J.-I. Tomizawa (1989). "*Rom* transcript of plasmid ColE1." Nucleic Acids Res. **17**(11): 4309-4326.

Brosius, Erfle and Storella (1985). "Specing of the -10 and -35 regions in the *tac* promoter. Effects on its *in vivo* activity." J. Biol. Chem. **260**(6): 3539-3541.

Browning, D. F. and S. J. W. Busby (2004). "The regulation of bacterial transcription initiation." Nat. Rev. Microbiol. **2**(1): 57-65.

Bukrinsky, M. I., E. V. Barsov and A. A. Shilov (1988). "Multicopy expression vector based on temperature-regulated *lac* repressor: expression of human immunodefienciency virus *env* gene in *Escherichia coli*." Gene. **70**(2): 415-417.

Calos, M. P. (1978). "DNA sequence for a low-level promoter of the *lac* repressor gene and an 'up' promoter mutation." Nature. **274**: 762-769.

Carpousis (2007). "The RNA degradosome of *Escherichia coli*: A multiprotein mRNA-degrading machine assembled on RNase E." Annu. Rev. Microbiol. **61**: 71-87.

Carpousis, A. J. and J. D. Gralla (1980). "Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the *lac* UV5 promoter." Biochemistry. **19**(14): 3245-3253.

Carrío, M. M. and A. Villaverde (2002). "Construction and deconstruction of bacterial inclusion bodies." J. Biotechnol. **96**(1): 3-12.

Castagnoli, L., M. Scarpa, M. Kokkindis, D. W. Banner, D. Tsernoglou and G. Cesareni (1989). "Genetic and structural analysis of the ColE1 Rop (Rom) protein." EMBO. J. **8**(2): 621-629.

Chamberlin, M., J. McGrath and L. Waskell (1970). "New RNA polymerase from *Escherichia coli* infected with bacteriophage T7." Nature. **228**: 227-231.

Chamberlin, M. and J. Ring (1973). "Characterization of T7-specific ribonucleic acid polymerase. I. General properties of the enzymatic reaction and the template specificity of the enzyme." J. Biol. Chem. **248**(6): 2235-2244.

Chao, Y.-P., W.-S. Law and W.-B. Hung (2002). "High production of heterologous proteins in *Escherichia coli* using the thermo-regulated T7 expression system. ." Appl. Microbiol. Biotechnol. **58**(4): 446-453.

Chen and Inouye (1994). "Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*." Genes. Dev. **8**(21): 2641-2652.

Chen, H., L. Pomeroy-Cloney, M. Bjerknes, J. Tam and E. Jay (1994). "The influence of adenine-rich motifs in the 3' portion of the ribosome binding site on human IFN-gamma gene expression in *Escherichia coli*." J. Mol. Biol. **240**(1): 20-27.

Choi, J. H., K. C. Keum and S. Y. Lee (2006). "Production of recombinant proteins by high cell density culture of *Escherichia coli*." Chem. Eng. Sci **61**(3): 876-885.

Choi, J. H. and S. Y. Lee (2004). "Secretory and extracellular production of recombinant proteins using *Escherichia coli*." Appl. Microbiol. Biotechnol. **64**(5): 625-635.

Clark, E. D. B. (1998). "Refolding of recombinant proteins." Curr. Opin. Biotechnol. **9**(2): 157-163.

Clark, E. D. B. (2001). "Protein refolding for industrial processes." Curr. Opin. Biotechnol. **12**(2): 202-207.

Cohen, S. N., A. C. Y. Chang, H. W. Boyer and R. B. Helling (1973). "Construction of biologically functional bacterial plasmids *in vitro*." Proc. Nat. Acad. Sci. USA. **70**(11): 3240-3244.

Coleman, J., K. Nakamura and M. Inouye (1985). "Mutations upstream of the ribosome-binding site affect translational efficiency." J. Mol. Biol. **181**(1): 139-143.

Cornelis, P. (2000). "Expressing genes in different *Escherichia coli* compartments." Curr. Opin. Biotechnol. **11**(5): 450-454.

Cronan, J. E. (2006). "A family of arabinose-inducible *Escherichia coli* expression vectors having pBR322 copy control." Plasmid. **55**(2): 152-157.

Cruz-Vera, L. R., M. A. Magos-Castro, E. Zamora-Romo and G. Guarneros (2004). "Ribosome stalling and peptidyl-tRNA drop-off during translational delay at AGA codons." Nucleic Acids Res. **32**(15): 4462-4468.

Davis, G., C. Elisee, D. M. Newham and R. G. Harrison (1999). "New fusion protein systems designed to give soluble expression in *Escherichia coli*." Biotechnol. Bioeng. **65**(4): 382-388.

de Boer, H., L. J. Comstock and M. Vasser (1983). "The *tac* promoter: A functional hybrid derived from the *trp* and *lac* promoters." Proc. Nat. Acad. Sci. USA. **80**: 21-25.

de Haseth, P. L., M. L. Zupancic and J. M. Thomas Record (1998). "RNA polymerase-promoter interactions: the comings and going of RNA polymerase." J. Bacteriol. **180**(12): 3019-3025.

de Mey, M., J. Maertens, G. J. Lequeux, W. K. Soetaert and E. J. Vandamme (2007). "Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering." BMC Biotechnol. **17**: 7-34.

de Smit, M. and J. van Duin (1994). "Control of translation by mRNA secondary structure in *Escherichia coli*." J. Mol. Biol. **244**(2): 144-150.

del Solar, G., R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa and R. Díaz-Orejas (1998). "Replication and control of circular bacterial plasmids." Microbiol. Mol. Biol. Rev. **62**(2): 434-464.

Delgado, A. and J.-L. Ramos (1994). "Genetic evidence fo activation of the positive transcriptional regulator XylS, a member of the NtrC family of regulators, by effector binding." J. Biol. Chem. **269**(11): 8059-8062.

Deng, T. (1997). "Bacterial expression and purification of biologically active mouse c-Fos proteins by selective codon optimization." FEBS Lett. **409**(2): 269-272.

Diaz Ricci, J. C. and M. E. Hernández (2000). "Plasmid effects on *Escherichia coli* metabolism." Crit. Rev. Biotechnol. **20**(2): 79-108.

Dickson, R. C., J. Abelson, W. M. Barnes and W. S. Reznikoff (1975). "Genetic regulation: the Lac control region." Science. **187**(4171): 27-35.

Domínguez-Cuevas, P., P. Marín, J. L. Ramos and S. Marqués (2005). "RNA polymerase holoenzymes can share a single transcription start site for the Pm promoter." J. Biol. Chem. **280**(50): 41315-41323.

Dong, Nilsson and Kurland (1996). "Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates." J. Mol. Biol. **260**(5): 649-663.

Draper, D. E. (1996). Translational initiation. *Escherichia coli* and *Salmonella*: cellular and molecular biology. F. C. Neidhardt, R. C. III, J. L. Ingrahamet al. Washington, D. C., ASM Press.**:** 902-908.

Dubendorff, J. W. and F. W. Studier (1991). "Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with *lac* repressor." J. Mol. Biol. **219**(1): 45-59.

Dunn, J. J. and F. W. Studier (1981). "Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the beginning of gene *4*." J. Mol. Biol. **148**(4): 303-330.

Durland, R. H., A. Toukdarian, F. Fang and D. R. Helinski (1990). "Mutations in the *trfA* replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy numbers." J. Bacteriol. **172**(7): 3859-3867.

Ederth, J., L. A. Isaksson and F. Abdulkarim (2002). "Origin-specific reduction of ColE1 plasmid copy number due to mutations in a distinct region of the *Escherichia coli* RNA polymerase." Mol. Genet. Genomics **267**(5): 587-592.

Egan, S. M. (2002). "Growing repertoire of AraC/XylS activators." J. Bacteriol. **184**(20): 5529-5532.

Elvin, Thompson, Argall, Hendry, Stamford, Lilley and Dixon (1990). "Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*." Gene. **87**(1): 123-126.

Emory, S., P. Bouvet and J. G. Belasco (1992). "A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*." Genes. Dev. **6**: 135-148.

Eustance, R. J. and R. F. Schleif (1996). "The linker region of AraC protein." J. Bacteriol. **178**(24): 7025-7030.

Farmer, W. R. and J. C. Liao (1997). "Reduction of aerobic acetate production by *Escherichia coli*." Appl. Env. Microbiol. **63**(8): 3205-3210.

Fekkes, P. and A. J. M. Driessen (1999). "Protein targeting to the bacterial cytoplasmic membrane." Microbiol. Mol. Biol. Rev. **63**(1): 161-173.

Feng, L., W. W. Chan, S. L. Roderick and D. E. Cohen (2000). "High-level expression and mutagenesis of recombinant human phosphatidylcholine transfer protein using a synthetic gene: Evidence for a C-terminal membrane binding domain." Biochemistry. **39**(50): 15399-15409.

Figge, J., C. Wright, C. J. Collins, T. M. Roberts and D. M. Livingston (1988). "Stringent regulation of stably integrated chloramphenicol acetyl transferase genes by *E. coli lac* repressor in monkey cells." Cell. **52**(5): 713-722.

Figurski, D. H., R. J. Meyer and D. R. Helinski (1979). "Suppression of ColE1 replication properties by the Inc P-1 plasmid RK2 in hybrid plasmids constructed *in vitro*." J. Mol. Biol. **133**(3): 295-318.

Francetic, O., D. Belin, C. Badaut and A. P. Pugsley (2000). "Expression of the endogenous type II secretion pathway in *Escherichia coli* leads to chitinase secretion." EMBO. J. **19**(24): 6697-6703.

Franklin, F. C. H., M. Bagdasarian, M. M. Bagdasarian and K. N. Timmis (1981). "Molecular and functional analysis of the TOL plasmid pWW0 from *Pseudomonas putida* and cloning of genes for the entire regulated aromatic ring *meta* cleavage pathway." Proc. Natl. Acad. Sci. USA. **78**(12): 7458-7462.

Franklin, F. C. H., P. R. Lehrbach, R. Lurz, B. Rueckert, M. Bagdasarian and K. N. Timmis (1983). "Localization and functional analysis of transposon mutations in regulatory genes of the TOL catabolic pathway." J. Bacteriol. **154**(2): 676-685.

Franklin, F. C. H., P. R. Lehrback, R. Lurz, B. Rueckert, M. Bagdasarian and K. N. Timmis (1983). "Localization and functional analysis of transposon mutations in regulatory genes of the TOL catabolic pathway.." J. Bacteriol. **154**(2): 676-685.

Gallegos, M.-T., C. Michán and J. L. Ramos (1993). "The XylS/AraC family of regulators." Nucleic Acids Res. **21**(4): 807-810.

Gallegos, M.-T., R. Schleif, A. Bairoch, K. Hofmann and J. L. Ramos (1997). "AraC/XylS family of transcriptional regulators." Microbiol. Mol. Biol. Rev. **61**(4): 393-410.

Gallegos, M. T., S. Marqués and J. L. Ramos (1996). "The $TACAN_4TGCA$ motif upstream from the -35 region in the $sigma^{70}$-$sigma^{S}$-dependent Pm promoter of the TOL plasmid is the minimum DNA segment required for transcription stimulation by XylS regulators." J. Bacteriol. **178**(22): 6427-6434.

Gellissen, G., G. Kunze, C. Gaillardin, J. M. Cregg, E. Berardi, M. Veenhuis and I. v. d. Klei (2005). "New yeasts expression platforms based on methylotrophic *Hansenula polymorpha* and *Pichia pastoris* and on dimorphic *Arxula adeninivorans* and *Yarrowia lipolytica* - A comparison." FEMS Yeast Res. **5**(11): 1079-1096.

Gerigk, M., R. Bujnicki, E. Ganpo-Nkwenkwa, J. Bongaerts, G. Sprenger and R. Takors (2002). "Process control for enhanced L-phenylalanine production using different recombinant *Escherichia coli* strains." Biotechnol. Bioeng. **80**(7): 746-754.

Ghosh, M. and R. F. Schleif (2001). "Biophysical evidence of arm-domain interactions in AraC." Anal. Biochem. **295**: 101-112.

Gimmestad, M., H. Sletta, H. Ertesvåg, K. Bakkevig, S. Jain, S.-j. Suh, G. Skjåk-Bræk, T. E. Ellingsen, D. E. Ohman and S. Valla (2003). "The *Pseudomonas fluorescens* AlgG protein, but not its mannuronan C-5-epimerase activity, is needed for alginate polymer formation." J. Bacteriol. **185**(12): 3515-3523.

Glascock, C. and M. J. Weickert (1998). "Using chromosomal $lacI^{Q1}$ to control expression of genes on high-copy-number plasmids in *Escherichia coli*." Gene. **223**(1-2): 221-231.

Goldman, Rosenberg, Zubay and Studier (1995). "Consecutive low-usage leucine codons block translation only when near the 5' end of a message in *Escherichia coli*." J. Mol. Biol. **245**(5): 467-473.

Golomb, M. and M. Chamberlin (1974). "Characterization of T7-specific ribonucleic acid polymerase. IV. Resolution of the major *in vitro* transcripts by gel electrophoresis." J. Biol. Chem. **249**(9): 2858-2863.

González-Pérez, M. M., S. Marqués, P. Domínguez-Cuevas and J. L. Ramos (2002). "XylS activator and RNA polymerase binding sites at the Pm promoter overlap." FEBS Lett. **519**(1-3): 117-122.

González-Pérez, M. M., J. L. Ramos, M.-T. Gallegos and S. Marqués (1999). "Critical nucleotides in the upstream region of the XylS-dependent TOL *meta*-cleavage pathway operon promoter as deduced from analysis of mutants." J. Biol. Chem. **274**(4): 2286-2290.

Gottesman, S. (1996). "Proteases and their targets in *Escherichia coli*." Annu. Rev. Genet. **30**: 465-506.

Gottesman, S., S. Wickner and M. R. Maurizi (1997). "Protein quality control: triage by chaperones and proteases." Genes. Dev. **11**(7): 815-823.

Grabherr, R. and K. Bayer (2002). "Impact of targeted vector design on *ColE1* plasmid replication." Trends. Biotechnol. **20**(6): 257-260.

Grabherr, R., E. Nilsson, G. Striedner and K. Bayer (2002). "Stabilizing plasmid copy number to improve recombinant protein production." Biotechnol. Bioeng. **77**(2): 142-147.

Greated, A., L. Lambertsen, P. A. Williams and C. M. Thomas (2002). "Complete sequence of the IncP-9 TOL plasmid pWW0 from *Pseudomonas putida*." Env. Microbiol. **4**(12): 856-871.

Gronenborn, B. (1976). "Overproduction of phage lambda repressor under control of the *lac* promotor of *Escherichia coli*." Mol. Gen. Genet. **148**: 243-250.

Gross, G., C. Mielke, I. Hollatz, H. Blöcker and R. Frank (1990). "RNA primary sequence or secondary structure in the translational initiation region controls expression of two variant interferon-beta genes in *Escherichia coli*." J. Biol. Chem. **265**(29): 17627-17636.

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

Gustafsson, C., S. Govindarajan and J. Minshull (2004). "Codon bias and heterologous protein expression." Trends Biotechnol. **22**(7): 346-353.

Guzman, L.-M., D. Belin, M. J. Carson and J. Beckwith (1995). "Tight regulation, modulation, and high-level expression by vectors containing the P_{BAD} promoter." J. Bacteriol. **177**(14): 4121-4130.

Haldimann, A., L. L. Daniels and B. L. Wanner (1998). "Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the *Escherichia coli* phosphate regulon." J. Bacteriol. **180**(5): 1277-1286.

Hall, Gabay, Débarbouille and Schwartz (1982). "A role for mRNA secondary structure in the control of translation initiation." Nature. **295**(5850): 616-618.

Hannig, G. and S. C. Makrides (1998). "Strategies for optimizing heterologous protein expression in *Escherichia coli*." Trends. Biotechnol. **16**(2): 54-60.

Hansen, Chen, Fejzo and J. G. Belasco (1994). "The *ompA* 5' untranslated region impedes a major pathway for mRNA degradation in *Escherichia coli*." Mol. Microbiol. **12**(5): 707-716.

Harayama, S., P. R. Lehrback and K. N. Timmis (1984). "Transposon mutagenesis analysis of *meta*-cleavage pathway operon genes of the TOL plasmid of *Pseudomonas putida* mt-2." J. Bacteriol. **160**(1): 251-255.

Harayama, S., M. Rekik, M. Wubbolts, K. Rose, R. A. Leppik and K. N. Timmis (1989). "Characterization of five genes in the upper-pathway operon of TOL plasmid pWW0 from *Pseudomonas putida* and identification of the gene products." J. Bacteriol. **171**(9): 5048-5055.

Hasan and Szybalski (1995). "Construction of *lacI*ts and *lacI*^{q}ts expression plasmids and evaluation of the thermosensitive *lac* repressor." Gene. **163**(1): 35-40.

Hasenwinkle, D., E. Jervis, O. Kops, C. Liu, G. Lesnicki, C. A. Haynes and D. G. Kilburn (1997). "Very high-level production and export in *Escherichia coli* of a cellulose binding domain for use in generic secretion-affinity fusion system." Biotechnol. Bioeng. **55**(6): 854-863.

Hashemzadeh-Bonehi, Mehraein-Ghomi, Mitsopoulos, Jacob, Hennessey and Broome-Smith (1998). "Importance of using *lac* rather than *ara* promoter vectors for modulating the levels of toxic gene products in *Escherichia coli*." Mol. Microbiol. **30**(3): 676-678.

Haugan, K., P. Karunakaran, J. M. Blatny and S. Valla (1992). "The phenotypes of temperature-sensitive mini-RK2 replicons carrying mutations in the replication control gene *trfA* are suppressed nonspecifically by intragenic *cop* mutations." J. Bacteriol. **174**(21): 7026-7032.

Haugan, K., P. Karunakaran, A. Tøndervik and S. Valla (1995). "The host range of RK2 minimal replicon copy-up mutants is limited by species-specific differences in the maximum tolerable copy number." Plasmid. **33**(1): 27-39.

Hendrickson, W., C. Stoner and R. Schleif (1990). "Characterization of the *Escherichia coli ara*FGH and *ara*J promoters." J. Mol. Biol. **215**(4): 497-510.

High, A. S., S. D. Torosian and F. G. Rodgers (1993). "Cloning, nucleotide sequence and expression in *Escherichia coli* of a gene (*ompM*) encoding a 25 kDa major outer-membrane protein (MOMP) of *Legionella pneumophila*." J. Gen. Microbiol. **139**(8): 1715-1721.

Hjalt, T. and E. G. H. Wagner (1992). "The effect of loop size in antisense and target RNAs on the efficiency of antisense RNA control." Nucleic Acids Res. **20**(24): 6723-6732.

Hsu, L. M., N. V. Vo, C. M. Kane and M. J. Chamberlin (2003). "In vitro studies of transcript initiation by *Escherichia coli* RNA polymerase. 1. RNA chain initiation, abortive initiation, and promoter escape at three bacteriophage promoters." Biochemistry. **42**(13): 3777-3786.

Hu, S.-Y., J.-L. Wu and J.-H. Huang (2004). "Production of tilapia insulin-like growth factor-2 in high cell density cultures of recombinant *Escherichia coli*." J. Biotechnol. **107**(2): 161-171.

Hua, Z., H. Wang, D. Chen, Y. Chen and D. Zhu (1994). "Enhancement of expression of human granulocyte-macrophage colony stimulating factor by *argU* gene product in *Escherichia coli*." Biochem. Mol. Biol. Int. **32**(3): 537-543.

Hui, A. and H. de Boer (1987). "Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*." Proc. Nat. Acad. Sci. USA. **84**(14): 4762-4766.

Ingram, L. C., M. H. Richmond and R. B. Sykes (1973). "Molecular characterization of the R factors implicated in the carbenicillin resistance of a sequence of *Pseudomonas aeruginosa* strains isolated from burns." Antimicro. Agents. Chemoother. **3**(2): 279-288.

Inouye, S., A. Nakazawa and T. Nakazawa (1981). "Molecular cloning of gene *xylS* of the TOL plasmid: Evidence for positive regulation of the *xylDEGF* operon by *xylS*." J. Bacteriol. **148**(2): 413-418.

Inouye, S., A. Nakazawa and T. Nakazawa (1987). "Expression of the regulatory gene *xylS* on the TOL plasmid is positively controlled by the *xylR* gene product." Proc. Nat. Acad. Sci. USA. **84**(15): 5182-5186.

Inouye, S., A. Nakazawa and T. Nakazawa (1987). "Overproduction of the *xylS* gene product and activation of the *xylSDLEGF* operon on the TOL plasmid." J. Bacteriol. **169**(8): 3587-3592.

Itakura, K., T. Hirose, R. Crea and A. D. Riggs (1977). "Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin." Science. **198**(4321): 1056-1063.

Jacob, F. and J. Monod (1961). "Genetic regulatory mechanisms in the synthesis of proteins." J. Mol. Biol. **3**: 318-356.

Jacob, W. F., M. Santer and A. E. Dahlberg (1987). "A single base change in the Shine-Dalgarno region of 16S rRNA of *Escherichia coli* affects translation of many proteins." Proc. Nat. Acad. Sci. USA. **84**(14): 4757-4761.

Jacquet, A., V. Daminet, M. Haumont, L. Garcia, S. Chaudoir, A. Bollen and R. Biemans (1999). "Expression of a recombinant *Toxoplasma gondii* ROP2 fragment as a fusion protein in bacteria circumvents insolubility and proteolytic degradation." Protein. Expr. Purif. **17**(3): 392-400.

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

Jeong, K. J., J. H. Choi, W. M. Yoo, K. C. Keum, N. C. Yoo, S. Y. Lee and M.-H. Sung (2004). "Constitutive production of human leptin by fed-batch culture of recombinant *rpoS*$^{-}$ *Escherichia coli*. ." Protein Expr. Purif. **36**(1): 150-156.

Jeong, K. J. and S. Y. Lee (1999). "High-level production of human leptin by fed-batch cultivation of recombinant *Escherichia coli* and its purification." Appl. Env. Microbiol. **65**(7): 3027-3032.

Jiang, X., H. Nakano, T. Kigawa, T. Yabuki, S. Yokoyama, D. S. Clark and T. Yamane (2001). "Dosage effect of minor arginyl- and isoleucyl-tRNAs on protein synthesis in an *Escherichia coli in vitro* coupled transcription/translation system." J. Biosci. Bioeng. **91**(1): 53-57.

Jones, K. L. and J. D. Keasling (1998). "Construction and characterization of F plasmid-based expression vectors." Biotechnol. Bioeng. **59**(6): 659-665.

Jones, S. A. and J. Melling (1984). "Persistence of pBR322-related plasmids in *Escherichia coli* grown in chemostat cultures." FEMS Microbiol. Lett. **22**(3): 239-243.

Kajava, A. V., S. N. Zolov, A. E. Kalinin and M. A. Nesmeyanova (2000). "The net charge of the first 18 residues of the mature sequence affects protein translocation across the cytoplasmic membrane of gram-negative bacteria." J. Bacteriol. **182**(8): 2163-2169.

Kaldalu, N., T. Mandel and M. Ustav (1996). "TOL plasmid transcription factor XylS binds specifically to the *Pm* operator sequence." Mol. Microbiol. **20**(3): 569-579.

Kaldalu, N., U. Toots, V. d. Lorenzo and M. Ustav (2000). "Functional domains of the TOL plasmid transcription factor XylS." J. Bacteriol. **182**(4): 1118-1126.

Kane, J. F. (1995). "Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*." Curr. Opin. Biotechnol. **6**(5): 494-500.

Kane, J. F., B. N. Violand, D. F. Curran, N. R. Staten, K. L. Duffin and G. Bogosian (1992). "Novel in-frame two codon translational hop during synthesis of bovine placental lactogen in a recombinant strain of *Escherichia coli*." Nucleic Acids Res. **20**(24): 6707-6712.

Kapust, R. B. and D. S. Waugh (1999). "*Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused." Protein Sci. **8**(8): 1668-1674.

Keasling, J. D. (1999). "Gene-expression tools for the metabolic engineering of bacteria." Trends. Biotechnol. **17**(11): 452-460.

Kessler, B., V. de Lorenzo and K. N. Timmis (1993). "Identification of a *cis*-acting sequence within the *Pm* promoter of the TOL plasmid which confers XylS-mediated responsiveness to substituted benzoates." J. Mol. Biol. **230**: 699-703.

Kessler, B., M. Herrero, K. N. Timmis and V. de Lorenzo (1993). "Genetic evidence that the XylS regulator of the *Pseudomonas* TOL meta operon controls the *Pm* promoter through weak DNA-protein interactions." J. Bacteriol. **176**(11): 3171-3176.

Kessler, B., K. N. Timmis and V. de Lorenzo (1994). "The organization of the *Pm* promoter of the TOL plasmid reflects the structure of its cognate activator XylS." Mol. Gen. Genet. **244**(6): 596-605.

Khlebnikov, A., Ø. Risa, T. Skaug, T. A. Carrier and J. D. Keasling (2000). "Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture." J. Bacteriol. **182**(24): 7029-7034.

Kittell, B. L. and D. R. Helinski (1991). "Iteron inhibition of plasmid RK2 replication *in vitro*: evidence for intermolecular coupling of replication origins as a mechanism for RK2 replication control." Proc. Nat. Acad. Sci. USA. **88**(4): 1389-1393.

Kleist, S., G. Miksch, B. Hitzmann, M. Arndt, K. Friehs and E. Flaschel (2003). "Optimization of the extracellular production of a bacterial phytase with *Escherichia coli* by using different fed-batch fermentation strategies." Appl. Microbiol. Biotechnol. **61**(5-6): 456-462.

Kolin, A., V. Jevtic, L. Swint-Kruse and S. M. Egan (2007). "Linker regions of the RhaS and RhaR proteins." J. Bacteriol. **189**(1): 269-271.

Kost, T., J. Condreay and D. Jarvis (2005). "Baculovirus as versatile vectors for protein expression in insect and mammalian cells." Nat. Biotechnol. **23**(5): 567-575.

Krummel, B. and M. J. Chamberlin (1989). "RNA chain initiation by *Escherichia coli* RNA polymerase. Structural transitions of the enzyme in early ternary complexes." Biochemistry. **28**(19): 7829-7842.

Kurland and Gallant (1996). "Errors of heterologous protein expression." Curr. Opin. Biotechnol. **7**(5): 489-493.

Kwon, H., M. Bennik, B. Demple and T. Ellenberger (2000). "Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA." Nat. Struct. Biol. **7**(5): 424-430.

Landry, T., L. Chew, J. Davis, N. Frawley, H. Foley, S. Stelman, J. Thomas, J. Wolt and D. Hanselman (2003). "Safety evaluation of a αλπηα-amylase enzyme preparation derived from the archaeal order *Thermococcales* as expressed in *Pseudomonas fluorescens* biovar I." Regul. Toxicol. Pharmacol. **37**(1): 149-168.

LaRonde-LeBlanc, N. and C. Wolberger (2000). "Characterization of the oligomeric states of wild type and mutant AraC." Biochemistry. **39**: 11593-11601.

Laursen, B. S., H. P. Sørensen, K. K. Mortensen and H. U. Sperling-Petersen (2005). "Initiation of protein synthesis in bacteria." Microbiol. Mol. Biol. Rev. **69**(1): 101-123.

Lee, C., J. Kim, S. G. Shin and S. Hwang (2006). "Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*." J. Biotechnol. **123**(3): 273-280.

Lee, K., C. A. Holland-Staley and P. R. Cunningham (1996). "Genetic analysis of the Shine-Dalgarno interaction: selection of alternative functional mRNA-rRNA combinations." RNA. **2**(12): 1270-1285.

Lee, S. K., H. H. Chou, B. F. Pfleger, J. D. Newman, Y. Yoshikuni and J. D. Keasling (2007). "Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters." Appl. Env. Microbiol. **73**(18): 5711-5715.

Lee, S. Y. (1996). "High cell-density culture of *Escherichia coli*." Trends Biotechnol. **14**(3): 98-105.

Lin-Chao, S. and H. Bremer (1986). "Effect of the bacterial growth rate on replication control of plasmid pBR322 in *Escherichia coli*." Mol. Gen. Genet. **203**(1): 143-149.

Lin-Chao, S., W.-T. Chen and T.-T. Wong (1992). "High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II." Mol. Microbiol. **6**(22): 3385-3393.

Lobell, R. B. and R. F. Schleif (1990). "DNA looping and unlooping by AraC protein." Science. **250**(4980): 528-532.

Lopez, P. J., I. Marchand, S. A. Joyce and M. Dreyfus (1999). "The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not mRNA processing *in vivo*." Mol. Microbiol. **33**(1): 188-199.

Lopilato, J., S. Bortner and J. Beckwith (1986). "Mutations in a new chromosomal gene of *Escherichia coli* K-12, *pcnB*, reduce plasmid copy number of pBR322 and its derivatives." Mol. Gen. Genet. **205**(2): 285-290.

Lorenzo, V. d., L. Eltis, B. Kessler and K. N. Timmis (1993). "Analysis of *Pseudomonas* gene products usign *lacI*^{q}/*Ptrp-lac* plasmids and transposons that confer conditional phenotypes." Gene. **123**: 17-24.

Lory, S. (1998). "Secretion of proteins and assembly of bacterial surface organelles: shared pathways of extracellular protein targeting." Curr. Opin. Microbiol. **1**(1): 27-35.

Ma, J., A. Campbell and S. Karlin (2002). "Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures." J. Bacteriol. **184**(20): 5733-5745.

Maizels, N. M. (1973). "The nucleotide sequence of the lactose messenger ribonucleic acid transcribed from the UV5 promoter mutant of *Escherichia coli*. ." Proc. Nat. Acad. Sci. USA. **70**(12 (pt. 1)): 3585-3589.

Makrides, S. C. (1996). "Strategies for achieving high-level expression of genes in *Escherichia coli*." Microbiol. Rev. **60**(3): 512-538.

Maloney, P. C. and B. Rotman (1973). "Distribution of suboptimally induced beta-D-galactosidase in *Escherichia coli*." J. Mol. Biol. **73**(1): 77-91.

Mandecki, W., R. A. Goldman, B. S. Powell and M. H. Caruthers (1985). "*lac* up-promoter mutants with increased homology to the consensus promoter sequence." J. Bacteriol. **164**(3): 1353-1355.

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

Marcaida, DePristo, Chandran, Carpousis and Luisi (2006). "The RNA degradosome: life in the fast lane of adaptive molecular evolution." Trends Biochem. Sci. **31**(7): 359-365.

Marqués, S., M.-T. Gallegos, M. Manzanera, A. Holtel, K. N. Timmis and J. L. Ramos (1998). "Activation and repression of transcription at the double tanden divergent promoters for the *xylR* and *xylS* genes of the TOL plasmid of *Pseudomonas putida*." J. Bacteriol. **180**(11): 2889-2894.

Marqués, S., M. T. Gallegos and J. L. Ramos (1995). "Role of the σ^{S} in transcription from the positively controlled Pm promoter of the TOL plasmid of *Pseudomonas putida*." Mol. Microbiol. **18**(5): 851-857.

Marqués, S., A. Holtel, K. N. Timmis and J. L. Ramos (1994). "Transcriptional induction kinetics from the promoters of the catabolic pathways of TOL plasmid pWW0 of *Pseudomonas putida* for metabolism of aromatics." J. Bacteriol. **176**(9): 2517-2524.

Marqués, S., M. Manzanera, M.-M. González-Pérez, M.-T. Gallegos and J. L. Ramos (1999). "The XylS-dependent Pm promoter is transcribed *in vivo* by RNA polymerase with sigma32 or sigma38 depending on the growth phase." Mol. Microbiol. **31**(4): 1105-1113.

Martin, K., L. Huo and R. F. Schleif (1986). "The DNA loop model for *ara* repression: AraC protein occupies the proposed loop sites *in vivo* and repression-negative mutations lie in these same sites." Proc. Nat. Acad. Sci. USA. **83**(11): 3654-3658.

Martinez, A., P. M. Knappskog, S. Olafsdottir, A. P. Døskeland, H. G. Eiken, R. M. Svebak, M. Bozzini, J. Apold and T. Flatmark (1995). "Expression of recombinant human phenylalanine hydroxylase as fusion protein in *Escherichia coli* circumvents proteolytic degradation by host cell proteases." Biochem. J. **306**(Pt. 2): 589-597.

Massé, Escorcia and Gottesman (2003). "Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*." Genes. Dev. **17**(19): 2375-2383.

McCarthy, J. E. G., H. U. Schairer and W. Sebald (1985). "Translational initiation frequency of *atp* genes from *Escherichia coli*: identification of an intercistronic sequence that enhances translation." EMBO. J. **4**(2): 519-526.

McCarthy, J. E. G., W. Sebald, G. Gross and R. Lammers (1986). "Enhancement of translational efficiency by the *Escherichia coli atpE* translational initiation region: its fusion with two human genes." Gene. **41**(2-3): 201-206.

Mergulhão, F. J. M., D. K. Summers and G. A. Monteiro (2005). "Recombinant protein secretion in *Escherichia coli*." Biotechnol. Adv. **23**(3): 177-202.

Mergulhão, F. J. M., M. A. Taipa, J. M. S. Cabral and G. A. Monteiro (2004). "Evaluation of bottlenecks in proinsulin secretion by *Escherichia coli*." J. Biotechnol. **109**(1-2): 31-43.

Mermod, N., J. L. Ramos, A. Bairach and K. N. Timmis (1987). "The *xylS* gene positive regulator of TOL plasmid pWW0: identification, sequence analysis and overproduction leading to constitutive expression of *meta* cleavage operon." Mol. Gen. Genet. **207**(2-3): 349-354.

Mermod, N., J. L. Ramos, P. R. Lehrbach and K. N. Timmis (1986). "Vector for regulated expression of cloned genes in a wide range of Gram-negative bacteria." J. Bacteriol. **167**(2): 447-454.

Michán, C., B. Kessler, V. d. Lorenzo, K. N. Timmis and J. L. Ramos (1992). "XylS domain interactions can be deduced from intraallelic dominance in double mutants of *Pseudomonas putida*." Mol. Gen. Genet. **235**: 406-412.

Michán, C., L. Zhou, M.-T. Gallegos, K. N. Timmis and J. L. Ramos (1992). "Identification of critical amino-terminal regions of XylS." J. Biol. Chem. **267**(32): 22897-22901.

Michán, C. M., S. J. W. Busby and E. I. Hyde (1995). "The *Escherichia coli* MelR transcription activator: production of a stable fragment containing the DNA-binding domain." Nucleic Acids Res. **23**(9): 1518-1523.

Miksch, G., F. Bettenworth, K. Friehs, E. Flaschel, A. Saalbach, T. Twellmann and T. W. Nattkemper (2005). "Libraries of synthetic stationary-phase and stress promoters as a tool for fine-tuning of expression of recombinant proteins in *Escherichia coli*." J. Biotechnol. **120**(1): 25-37.

Miroux, B. and J. E. Walker (1996). "Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels." J. Mol. Biol. **260**(3): 289-298.

Miyake, T., T. Oka, T. Nishizawa, F. Misoka, T. Fuwa, K. Yoda, M. Yamasaki and G. Tamura (1985). "Secretion of human interferon-alpha induced by using secretion vectors containing a promoter and signal sequence of alkaline phosphatase gene of *Escherichia coli*." J. Biochem. **97**: 1429-1436.

Moll, I., S. Grill, C. O. Gualerzi and U. Bläsi (2002). "Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control." Mol. Microbiol. **43**(1): 239-246.

Morgan-Kiss, R. M., C. Wadler and J. John E. Cronan (2002). "Long-term and homogeneous regulation of the *Escherichia coli araBAD* promoter by use of a lactose transporter of relaxed specificity." Proc. Nat. Acad. Sci. USA. **99**(11): 7373-7377.

Moser, D. R., D. Ma, C. D. Moser and J. L. Campbell (1984). "cis-acting mutations that affect rop protein control of plasmid copy number." Proc. Nat. Acad. Sci. USA. **81**(14): 4465-4469.

Muesing, M., J. Tamm, H. M. Shepard and B. Polisky (1981). "A single base-pair alteration is responsible for the DNA overproduction phenotype of a plasmid copy-number mutant." Cell. **24**(1): 235-242.

Mujacic, M., K. W. Cooper and F. Baneyx (1999). "Cold-inducible cloning vectors for low-temperature protein expression in *Escherichia coli*: application to the production of a toxic and proteolytically sensitive fusion protein." Gene. **238**(2): 325-332.

Müller-Hill, B., L. Crapo and W. Gilbert (1968). "Mutants that make more Lac repressor." Proc. Nat. Acad. Sci. USA. **59**(4): 1259-1264.

Nakamura, Y., T. Gojobori and T. Ikemura (2000). "Codon usage tabulated from the international DNA sequence databases: status for the year 2000." Nucleic Acids Res. **28**(1): 292.

Nevalainen, K. M. H., V. S. J. Te'o and P. L. Bergquist (2005). "Heterologous protein expression in filamentous fungi." Trends Biotechnol. **23**(9): 468-474.

Newbury, S. F., N. H. Smith, E. C. Robinson, I. D. Hiles and C. F. Higgins (1987). "Stabilization of translationally active mRNA by prokaryotic REP sequences." Cell. **48**(2): 297-310.

Nishihara, K., M. Kanemori, M. Kitagawa, H. Yanagi and T. Yura (1998). "Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*." Appl. Env. Microbiol. **64**(5): 1694-1699.

Nishihara, K., M. Kanemori, H. Yanagi and T. Yura (2000). "Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*." Appl. Env. Microbiol. **66**(3): 884-889.

Noronha, S. B., H. J. C. Yeh, T. F. Spande and J. Shiolach (2000). "Investigation of the TCA cycle and the glyoxylate shunt in *Escherichia coli* BL21 and JM109 using ^{13}C-NMR/MS." Biotechnol. Bioeng. **68**(3): 316-327.

Novick, A. and M. Weiner (1957). "Enzyme induction as an all-or-none phenomenon." Proc. Nat. Acad. Sci. USA. **43**(7): 553-566.

O'Connor, C. D. and K. N. Timmis (1987). "Highly repressible expression system for cloning genes that specify potentially toxic proteins." J. Bacteriol. **169**(10): 4457-4462.

O'Connor, M., T. Asai, C. L. Squires and A. E. Dahlberg (1999). "Enhancement of translation by the downstream box does not involve base pairing of mRNA with the penultimate stem sequence of 16S rRNA." Proc. Nat. Acad. Sci. USA. **96**(16): 8973-8978.

O'Connor, M., M. Peifer and W. Bender (1989). "Construction of large DNA segments in *Escherichia coli*." Science. **244**(4910): 1307-1312.

Olson, P., Y. Zhang, D. Olsen, A. Owens, P. Cohen, K. Nguyen, J.-J. Ye, S. Bass and D. Mascarenhas (1998). "High-level expression of eukaryotic polypeptides from bacterial chromosomes." Protein Expr. Purif. **14**(2): 160-166.

Pabo, C. and R. Sauer (1992). "Transcription factors: structural families and pronciples of DNA recognition." Annu. Rev. Biochem. **61**: 1053-1095.

Palacios, J. L., I. Zaror, P. Martínez, F. Uribe, P. Opazo, T. Socías, M. Gidekel and A. Venegas (2001). "Subset of hybrid eukaryotic proteins is exported by the type I secretion system of *Erwinia chrysanthemi*." J. Bacteriol. **183**(4): 1346-1358.

Peixoto, L., V. Fernández and H. Musto (2004). "The effect of expression levels on codon usage in *Plasmodium falciparum*." Parasitology. **128**(3): 245-251.

Perri, S., D. R. Helinski and A. Toukdarian (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**(19): 12536-12543.

Polisky, B., R. J. Bishop and D. H. Gelfand (1976). "A plasmid cloning vehicle allowing regulated expression of eukaryotic DNA in bacteria." Proc. Nat. Acad. Sci. USA. **73**(11): 3900-3904.

Prinz, W. A., F. Åslund, A. Holmgren and J. Beckwith (1997). "The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm." J. Biol. Chem. **272**(25): 15661-15667.

Pugsley, A. P., O. Francetic, O. M. Possot, N. Sauvonnet and K. R. Hardie (1997). "Recent progress and future directions in studies of the main terminal branch of the general secretory pathway in Gram-negative bacteria - a review." Gene. **192**(1): 13-19.

Puri, N., K. B. C. Appa Rao, S. Menon, A. K. Panda, G. Tiwari, L. C. Gang and S. M. Totey (1999). "Effect of the codon following the ATG start site on the expression of ovine growth hormone in **Escherichia coli**." Protein Expr. Purif. **17**(2): 215-223.

Qiu, J., J. R. Swartz and G. Georgiou (1998). "Expression of active human tissue-type plasminogen activator in *Escherichia coli*." Appl. Env. Microbiol. **64**(12): 4891-4896.

Rai, M. and H. Padh (2001). "Expression systems for production of heterologous proteins." Curr. Sci. **80**(9): 1121-1128.

Ramesh, V., d. Amitabha and V. Nagaraja (1994). "Engineering hyperexpression of bacteriophage Mu C protein by removal of secondary structure at the translation initiation region." Protein Eng. **7**(8): 1053-1057.

Ramos, J. L., M. Gonzales-Carrero and K. N. Timmis (1988). "Broad-host range expression vectors containing manipulated *meta*-cleavage pathway regulatory elements of the TOL plasmid." FEBS Lett. **226**(2): 241-246.

Ramos, J. L., S. Marqués and K. N. Timmis (1997). "Transcriptional control of the *Pseudomonas* TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators." Annu. Rev. Microbiol. **51**: 342-372.

Ramos, J. L., N. Mermod and K. N. Timmis (1987). "Regulatory circuits controlling transcription of TOL plasmid operon encoding *meta*-cleavage pathway for degradation of alkylbenzoates by *Pseudomonas*." Mol. Microbiol. **1**(3): 293-300.

Ramos, J. L., C. Michan, F. Rojo, D. Dwyer and K. Timmis (1990). "Signal-regulator interactions. Genetic analysis of the effector binding site of *xylS*, the benzoate-activated positive regulator of *Pseudomonas* TOL plasmid meta-cleavage pathway operon." J. Mol. Biol. **211**: 373-382.

Ramos, J. L., F. Rojo, L. Zhou and K. N. Timmis (1990). "A family of positive regulators related to the *Pseudomonas putida* TOL plasmid XylS and the *Escherichia coli* AraC activators." Nucleic Acids Res. **18**(8): 2149-2152.

Ramos, J. L., A. Stolz, W. Reineke and K. N. Timmis (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. Nat. Acad. Sci. USA. **83**: 8467-8471.

Rhee, S., R. G. Martin, J. L. Rosner and D. R. Davies (1998). "A novel DNA-binding motif in MarA: The first structure for an AraC family transcriptional activator." Proc. Natl. Acad. Sci. USA. **95**(10413-10418).

Ruíz, R., S. Marqués and J. L. Ramos (2003). "Leucines 193 and 194 at the N-terminal domain of the XylS protein, the positive transcriptional regulator of the TOL *meta*-cleavage pathway, are involved in dimerization." J. Bacteriol. **185**(10): 3036-4041.

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

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

Russell, D. R. and G. N. Bennett (1982). "Construction and analysis of in vivo activity of *E. coli* promoter hybrids and promoter mutants that alter the -35 to -10 spacing." Gene. **20**(2): 231-243.

Sandkvist, M. (2001). "Biology of type II secretion." Mol. Microbiol. **40**(2): 271-283.

Sandkvist, M. and M. Bagdasarian (1996). "Secretion of recombinant proteins by Gram-negative bacteria." Curr. Opin. Biotechnol. **7**(5): 505-511.

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

Schmidt, F. R. (2004). "Recombinant expression systems in the pharmaceutical industry." Appl. Microbiol. Biotechnol. **65**: 363-372.

Schumann, W. and L. C. S. Ferreira (2004). "Production of recombinant proteins in *Escherichia coli*." Gen. Mol. Biol. **27**(3): 442-453.

Sengupta, J., R. K. Agrawal and J. Frank (2001). "Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA." Proc. Nat. Acad. Sci. USA. **98**(21): 11991-11996.

Shiloach, J. and R. Fass (2005). "Growing *E. coli* to high cell density - a historical perspective on method development." Biotechnol. Adv. **23**(5): 345-357.

Shiloach, J., J. Kaufman, A. S. Guillard and R. Fass (1996). "Effect of glucose supply strategy on acetate accumulation, growth, and recombinant protein production by *Escherichia coli* BL21 (lamdaDE3) and *Escherichia coli* JM109." Biotechnol. Bioeng. **49**(4): 421-428.

Shin, C. S., M. S. Hong, C. S. Bae and J. Lee (1997). "Enhanced production of human mini-proinsulin in fed-batch cultures at high cell density of *Escherichia coli* BL21(DE3)[pET-3aT2M2] " Biotechnol. Prog. **13**(3): 249-257.

Shine, J. and L. Dalgarno (1974). "The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites." Proc. Nat. Acad. Sci. USA. **71**(4): 1342-1346.

Shokri, A., A. M. Sandén and G. Larsson (2003). "Cell and process design for targeting of recombinant protein into the culture medium of *Escherichia coli*." Appl. Microbiol. Biotechnol. **60**(6): 654-664.

Siegele and Hu (1997). "Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations." Proc. Nat. Acad. Sci. USA. **94**(15): 8168-8172.

Simmons, L. C. and D. G. Yansura (1996). "Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*." Nat. Biotechnol. **14**(5): 629-634.

Skerra, A. (1994). "Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*." Gene. **151**(1-2): 131-135.

Skorski, P., P. Leroy, O. Fayet, M. Dreyfus and S. H.-L. Denmat (2006). "The highly efficient translation initiation region from the *Escherichia coli rpsA* gene lacks a Shine-Dalgarno element." J. Bacteriol. **188**(17): 6277-6285.

Slos, P., D. Speck, N. Accart, H. V. J. Kolbe, D. Schubnel, B. Bouchon, R. Bischoff and M.-P. Kieny (1994). "Recombinant cholera toxin B subunit in *Escherichia coli*: high-level secretion, purification and characterization." Protein Expr. Purif. **5**(5): 518-526.

Smolke, Carrier and J. D. Keasling (2000). "Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures." Appl. Env. Microbiol. **66**(12): 5399-5405.

Soisson, S., B. MacDougall-Shackleton, R. Schleif and C. Wolberger (1997). "Structural basis for ligand-regulated oligomerization of AraC." Science. **278**(5311): 421-425.

Som, T. and J.-I. Tomizawa (1983). "Regulatory regions of ColE1 that are involved in determination of plasmid copy number." Proc. Nat. Acad. Sci. USA. **80**(11): 3232-3236.

Spooner, R. A., M. Bagdasarian and F. C. H. Franklin (1987). "Activation of the *xylSDLEGF* promoter of the TOL toluene-xylene degradation pathway by overproduction of the *xylS* regulatory gene product. ." J. Bacteriol. **169**(8): 3581-3586.

Sprengart, M. L., E. Fuchs and A. G. Porter (1996). "The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*." EMBO. J. **15**(3): 665-674.

Srinivasan, S., G. C. Barnard and T. U. Gerngross (2002). "A novel high-cell-density protein expression system based on *Ralstonia eutropha*." Appl. Env. Microbiol. **68**(12): 5925-5932.

Srinivasan, S., G. C. Barnard and T. U. Gerngross (2003). "Production of recombinant proteins using multiple-copy gene integration in high-cell-density fermentations of *Ralstonia eutropha*." Biotechnol. Bioeng. **84**(1): 114-120.

Studier, F., A. Rosenberg, J. Dunn and J. Dubendorff (1990). "Use of T7 RNA polymerase to direct expression of cloned genes." Methods Enzymol. **185**: 60-89.

Studier, F. W. and B. A. Moffatt (1986). "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes." J. Mol. Biol. **189**(1): 113-130.

Swartz, J. R. (2001). "Advances in *Escherichia coli* production of therapeutic proteins." Curr. Opin. Biotechnol. **12**(2): 195-201.

Sørensen, H. P. and K. K. Mortensen (2005). "Advanced genetic strategies for recombinant protein expression in *Escherichia coli*." J. Biotechnol. **115**(2): 113-128.

Sørensen, H. P., H. U. Sperling-Petersen and K. K. Mortensen (2003). "A favorable solubility partner for the recombinant expression of streptavidin." Protein Expr. Purif. **32**(2): 252-259.

Tabor, S. and C. C. Richardson (1985). "A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes." Proc. Nat. Acad. Sci. USA. **82**(4): 1074-1078.

Terpe, K. (2006). "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems." Appl. Microbiol. Biotechnol. **72**(2): 211-222.

Thomas, J. G. and F. Baneyx (1996). "Protein folding in the cytoplasm of *Escherichia coli*: requirements for the DnaK-DnaJ-GrpE and GroEL-GroES molecular chaperone machines." Mol. Microbiol. **21**(6): 1185-1196.

Timmes, A., M. Rodgers and R. Schleif (2004). "Biochemical and physiological properties of the DNA binding domain of AraC protein." J. Mol. Biol. **340**: 731-738.

Tobes, R. and J. L. Ramos (2002). "AraC-XylS database: a family of positive transcriptional regulators in bacteria." Nucleic Acids Res. **30**(1): 318-321.

Tomizawa, J.-I. (1984). "Control of ColE1 plasmid replication: the process of binding of RNA I to the primer transcript." Cell. **38**(3): 861-870.

Tomizawa, J.-I., Y. Sakakibara and T. Kakefuda (1974). "Replication of colicin E1 plasmid DNA in cell extracts. Origin and direction of replication." Proc. Nat. Acad. Sci. USA. **71**(6): 2260-2264.

Tomizawa, J.-I., Y. Sakakibara and T. Kakefuda (1975). "Replication of colicin E1 plasmid DNA added to cell extracts." Proc. Nat. Acad. Sci. USA. **72**(3): 1050-1054.

Tomizawa, J.-I. and T. Som (1984). "Control of ColE1 plasmid replication: enhancement of binding of RNA I to the primer transcript by the Rom protein." Cell. **38**(3): 871-878.

Tomsic, M., L. Tsujikawa, G. Panaghie, Y. Wang, J. Azok and P. L. de Haseth (2001). "Different roles for basic and aromatic amino acids in conserved region 2 of *Escherichia coli* sigma^{70} in the nucleation and maintenance of the signle-stranded DNA bubble in open RNA polymerase-promoter complexes." J. Biol. Chem. **276**(34): 31891-31896.

Tsujikawa, L., O. V. Tsodikov and P. L. de Haseth (2002). "Interaction of RNA polymerase with forked DNA: evidence for two kinetically significant intermediates on the pathway to the final complex." Proc. Nat. Acad. Sci. USA. **99**(6): 3493-3498.

Twigg, A. J. and D. Sherratt (1980). "Trans-complementable copy-number mutants of plasmid ColE1." Nature. **283**(5743): 216-218.

Umelo-Njaka, E., J. F. Nomellini, H. Yim and J. Smit (2001). "Development of small high-copy-number plasmid vectors for gene expression in *Caulobacter crescentus*." Plasmid. **46**(1): 37-46.

Valešová, R., L. Hollerová-Sobotková, V. Stepánek and P. Kyslík (2004). "Optimization of the host-plasmid interaction in the recombinant *Escherichia coli* strains overproducing penicillin G acylase." Enzyme Microbial. Technol. **35**(1): 74-80.

Villaverde, A. and M. M. Carrío (2003). "Protein aggregation in recombinant bacteria: biological role of inclusion bodies." Biotechnol. Lett. **25**(17): 1385-1395.

Walsh, G. (2006). "Biopharmaceutical benchmarks in 2006." Nat. Biotechnol. **24**(7): 769-776.

Wang, Z., L. Xiang, J. Shao, A. Wegrzyn and G. Wegrzyn (2006). "Effects on the presence of ColE1 plasmid DNA in *Escherichia coli* on the host cell metabolism." Microb. Cell Fact. **5**: 34.

Warnecke, T. and R. T. Gill (2005). "Organic acid toxicity, tolerance, and production in *Escherichia coli* biorefining applications." Microb. Cell Fact. **4**: 25.

Warzecha, H., H. S. Mason, C. Lane, A. Tryggvesson, E. Rybicki, A.-L. Williamson, J. D. Clements and R. C. Rose (2003). "Oral immunogenicity of human papillomavirus-like particles expressed in potato." J. Virol. **77**(16): 8702-8711.

Weldon, J., M. Rodgers, C. Larkin and R. Schleif (2007). "Structure and properties of a truely apo form of AraC dimerization domain." Proteins. **66**(3): 646-654.

Westers, L., H. Westers and W. J. Quax (2004). "*Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism." Biochim. Biophys. Acta. **1694**(1-3): 299-310.

Williams, P. A. and K. Murray (1974). "Metabolism of benzoate and methylbenzoates by *Pseudomonas putida* (*arvilla*) mt-2: Evidence for the existence of a TOL plasmid." J. Bacteriol. **120**(1): 416-423.

Williams, S. G., R. M. Cranenburgh, A. M. E. Weiss, C. J. Wrighton, D. J. Sherratt and J. A. J. Hanak (1998). "Repressor titration: a novel system for selection and stable maintenance of recombinant plasmids." Nucleic Acids Res. **26**(9): 2120-2124.

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

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

Wong, H. H., Y. C. Kim, S. Y. Lee and H. N. Chang (1998). "Effect of post-induction nutrient feeding strategies on the production of bioadhesive protein in *Escherichia coli*." Biotechnol. Bioeng. **60**(3): 271-276.

Worsey, M. J., F. C. H. Franklin and P. A. Williams (1978). "Regulation of the degradative pathway enzymes coded for by the TOL plasmid (pWW0) from *Pseudomonas putida* mt-2." J. Bacteriol. **134**(3): 757-764.

Worsey, M. J. and P. A. Williams (1975). "Metabolism of toluene and xylenes by *Pseudomonas* (*putida* (*arvilla)* mt-2: evidence for a new function of the TOL plasmid. ." J. Bacteriol. **124**(1): 7-13.

Wu, M. and R. Schleif (2001). "Mapping arm-DNA-binding domain interactions in AraC." J. Mol. Biol. **307**: 1001-1009.

Wu, X., H. Jörnvall, K. D. Berndt and U. Oppermann (2004). "Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance." Biochem. Biophys. Res. Commun. **313**(1): 89-96.

Wurm, F. (2004). "Production of recombinant protein therapeutics in cultivated mammalian cells." Nat. Biotechnol. **22**(11): 1393-1398.

Wösten, M. (1998). "Eubacterial sigma-factors." FEMS Microbiol. Rev. **22**(3): 127-150.

Yanisch-Perron, C., J. Vieira and J. Messing (1985). "Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors." Gene. **33**(1): 103-119.

Yin, J., G. Li, X. Ren and G. Herrler (2007). "Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes." J. Biotechnol. **127**(3): 335-347.

Yoon, S. H., M.-J. Han, S. Y. Lee, K. J. Jeong and J.-S. Yoo (2003). "Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture." Biotechnol. Bioeng. **81**(7): 753-767.

Yuan, H., X. Yang and Z.-C. Hua (2004). "Optimization of expression of an Annexin V-Hirudin chimeric protein in *Escherichia coli*." Microbiol. Res. **159**(2): 147-156.

Zhang, Zubay and Goldman (1991). "Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates." Gene. **105**(1): 61-72.

Zhang, J. and M. P. Deutscher (1992). "A uridine-rich sequence required for translation of prokaryotic mRNA:." Proc. Nat. Acad. Sci. USA. **89**(7): 2605-2609.

Zheng, L., U. Baumann and J.-L. Reymond (2003). "Production of a functional catalytic antibody scFv-NusA fusion protein in bacterial cytoplasm." J. Biochem. **133**(5): 577-581.

Åslund, F. and J. Beckwith (1999). "The thioredoxin superfamily: redundancy, specificity, and gray-area genomics." J. Bacteriol. **181**(5): 1375-1379.

Paper I

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Dec. 2004, p. 7033–7039 Vol. 70, No. 12
0099-2240/04/$08.00+0 DOI: 10.1128/AEM.70.12.7033–7039.2004

Broad-Host-Range Plasmid pJB658 Can Be Used for Industrial-Level Production of a Secreted Host-Toxic Single-Chain Antibody Fragment in *Escherichia coli*

H. Sletta,[1] A. Nedal,[2] T. E. V. Aune,[2] H. Hellebust,[3] S. Hakvåg,[2] R. Aune,[1] T. E. Ellingsen,[1] S. Valla,[2] and T. Brautaset[2]*

SINTEF Materials and Chemistry, Department of Biotechnology, SINTEF,[1] and Department of Biotechnology, Norwegian University of Science and Technology,[2] Trondheim, and Affitech AS, Oslo,[3] Norway

Received 11 March 2004/Accepted 4 July 2004

In industrial scale recombinant protein production it is often of interest to be able to translocate the product to reduce downstream costs, and heterologous proteins may require the oxidative environment outside of the cytoplasm for correct folding. High-level expression combined with translocation to the periplasm is often toxic to the host, and expression systems that can be used to fine-tune the production levels are therefore important. We previously constructed vector pJB658, which harbors the broad-host-range RK2 minireplicon and the inducible *Pm*/*xylS* promoter system, and we here explore the potential of this unique system to manipulate the expression and translocation of a host-toxic single-chain antibody variable fragment with affinity for hapten 2-phenyloxazol-5-one (phOx) (scFv-phOx). Fine-tuning of scFv-phOx levels was achieved by varying the concentrations of inducers and the vector copy number and also different signal sequences. Our data show that periplasmic accumulation of scFv-phOx leads to cell lysis, and we demonstrate the importance of controlled and high expression rates to achieve high product yields. By optimizing such parameters we show that soluble scFv-phOx could be produced to a high volumetric yield (1.2 g/liter) in high-cell-density cultures of *Escherichia coli*.

The application range of antibodies in medicine and biotechnology is broad, and there has been great progress in the design and selection of new variants with novel affinities. In particular, there has been a growing interest in the development of antibody fragments comprising the V_H and V_L domains connected to each other as a single chain (scFv) (4). The small size of scFv proteins (about 250 amino acids) compared to native antibodies confers certain therapeutic advantages because of their shorter half-life (rapid blood clearance) and faster tissue penetration. scFv molecules can be used as selective carriers for delivering radionucleids, toxins, or cytotoxic drugs to malignant cell populations, as well as providing valuable tools for studying antibody-antigen interactions in detail (14). In addition, this feature makes it possible to construct and screen large scFv libraries by using phage display approaches (for a review, see reference 23).

For medical applications scFvs are needed in large amounts, and the ability to produce high yields in *Escherichia coli* has gained considerable interest (14). Native scFv proteins have two disulfide bonds and require oxidative conditions to fold correctly. Although expression of native scFv proteins without disulfide bridge formation has been reported (9), cytoplasmic production in bacteria typically results in aggregation of scFv polypeptides into insoluble inclusion bodies (14). Therefore, in *E. coli* it is usually desirable to express scFvs as fusion proteins targeted for translocation to the oxidative periplasm to obtain functional products (3, 26). Various vector systems for recombinant scFv expression in *E. coli* have been reported (11, 15, 19), but the experiments were typically performed in shake-flask cultures with product yields of 10 to 30 mg/liter. Plasmid-based gene expression under high-cell-density cultivation may dramatically promote problems related to plasmid instability and lysis of recombinant cells (16, 28). Therefore, to obtain industrially relevant yields of soluble scFv, the expression tools should preferentially allow tight control of the expression level and the plasmids must be kept stable under high-cell-density cultivation.

We previously developed broad-host-range expression vectors based on the minimal replication elements *oriV* and *trfA* of the natural RK2 plasmid (5, 6). The *trfA* gene encodes a replication initiation protein also exerting negative control on replication from *oriV* (10, 12), and inducible expression of recombinant genes is governed by the flexible *Pm*/*xylS* promoter/regulator system. Relevant characteristics of these vectors include adjustable vector copy number, the *Pm* promoter is very strong and active in many hosts, several cheap and harmless (like benzoic acid) inducers can be used and they act in a dose-dependent manner, and inducer uptake is based on passive diffusion which further simplifies the use of the system across species barriers. With one such vector, pJB658, we have demonstrated regulated high and low level recombinant gene expression in several gram-negative bacterial species (6, 7, 29).

In the present study we utilize the unique properties of pJB658 to express a host-toxic antibody fragment, scFv-phOx, in high-cell-density cultivation of *E. coli*. By optimizing the vector copy number, induction conditions and choice of signal sequence, we produce recombinant scFv-phOx to very high volumetric yield (2.3 g/liter) of which 60% is represented as soluble and active product.

* Corresponding author. Mailing address: Department of Biotechnology, Norwegian University of Science and Technology, N-7034 Trondheim, Norway. Phone: 47 73 59 86 89. Fax: 47 73 59 12 83. E-mail: Trygve.Brautaset@nt.ntnu.no.

TABLE 1. Bacterial strains and plasmids

Strain or plasmid	Description[a]	Reference or source
E. coli		
DH5α	General cloning host	BRL
RV308	Production strain	ATCC 31608
Plasmids		
pHOG21-phOx	ColE1 replicon harboring scFv-phOx fused to *pelB,* c-*myc,* and his6, Ap^r	15
pHKK	ColE1 replicon with *lacP/O* promoter-operator for recombinant transcription, *hok-sok* stabilization elements, Ap^r	13
pBR322	ColE1 replicon, copy number of 30 per genome, Ap^r Tc^r	NEB
pBR322Δamp	pBR322 with deleted *bla* gene, Tc^r	This study
pJB658	RK2-based expression vector harboring *Pm/xylS* regulatory promoter system for expression of cloned genes, Ap^r	1
pJB655cop271Cluc	Harbors *trfA* with *cop-271* mutation	2
pJB655cop251Mluc	Harbors *trfA* with *cop-251* mutation	2
pJBphOx	pJB658 with *pelB*-scFv-*phOx-c-myc*-his6 fusion gene downstream of *Pm*, harbors *hok-sok* suicide system (see text)	This study
pJBphOx-271	pJBphOx with 1,853-bp ClaI/XmaI region substituted with corresponding region from pJB655cop271Cluc	This study
pJBphOx-251	pJBphOx with 1,853-bp ClaI/XmaI region substituted with corresponding region from pJB655cop251Mluc	This study
pJB-251E	8,350-bp pJBphOx-251 NdeI/NotI fragment blunt-ended and religated	This study
pJBphOx-271ompA	Similar to pJBphOx-271 but with *pelB* substituted with *ompA* (see text)	This study
pJBphOx-271CSP	Similar to pJBphOx-271 but with *pelB* substituted with consensus signal peptide (see text)	This study
pJBphOx-271ΔxylS	pJBphOx-271 was digested with AgeI/HpaI, cohesive ends of the large fragment were blunt-ended and religated (*xylS*)	This study
pJBphOx-251ΔxylS	pJBphOx-271 was digested with AgeI/HpaI, cohesive ends of the large fragment were blunt-ended and religated (*xylS*)	This study

[a] Some of the plasmids are described in the text. Ap, ampicillin; Tc, tetracycline.

MATERIALS AND METHODS

Strains, plasmids, and DNA manipulations. The strains and plasmids used in this study are listed in Table 1. Standard DNA manipulations were performed as described elsewhere (25).

Construction of vector pJBphOx. The scFv–2-phenyloxazol-5-one (phOx) coding sequence fused to c-*myc* was PCR amplified from vector pHOG21-phOx with the following primer pair: scFv-F1, 5′-TTACTCGC*GGCCCAGCCGGCC*ATG GCGCAGGTGCAGCTGGTGCAGTCT-3′, and scFv-R1, 5′-GTGATC*GGCC CCCGAGGCC*TTTAGGTCTTCTTCTGAGATCAGCTTTTGTTC-3′.

The PCR was performed on the Perkin Elmer GeneAmp PCR system 2400. The resulting 843-bp PCR product was end digested with SfiI (recognition sites are italic in the primer sequences) and cloned between the two SfiI sites of plasmid pHKK. From this construct, the 2,420-bp NdeI/EcoRI fragment was excised and cloned into the corresponding sites of plasmid pJB658, yielding vector pJBphOx (Fig. 1).

Construction of the *ompA* and consensus signal peptide signal sequences. The DNA molecules encoding OmpA and consensus signal peptide were made synthetically by annealing of the following single-stranded oligonucleotide pairs: ompA-F, 5′-*TA*TGAAAAAAACTGCTATCGCTATCGCTGTTGCTCTGGCT GGTTTCGCTACTGTTGCTCAGGCGGCGGC-3′, and ompA-R, 5′-*CATG*G CCGCCGCCTGAGCAACAGTAGCGAAACCAGCCAGAGCAACAGCGA TAGCGATAGCAGTTTTTTTCA-3′, and CSP-F, 5′-*TA*TGAAAAAAAAATT ATTGGCGTTAGCCTTGTTAGCGTTATTGTTTAACGGCGCGCAGGC-3′, and CSP-R, 5′-*CATG*GCCTGCGCGCCGTTAAACAATAACGCTAACAA GGCTAACGCCAATAATTTTTTTTTCA-3′.

Each pair of oligonucleotides were annealed as described previously (30) and the NdeI (2 bp) and NcoI (4 bp) sticky ends formed at the ends of the resulting double-stranded DNA fragments are in italic. The double-stranded DNA molecules obtained could be directly cloned into the NdeI/NcoI sites of any plasmid.

Media, feeding solutions, and preparation of inoculum to high-cell-density cultivation fermentations. For routine cloning experiments and estimation of plasmid copy number, *E. coli* cells were grown at 37°C in liquid Luria broth (LB) or on solid Luria agar (LA) plates (25). The defined preculture medium and the main culture medium used were composed as described by others (13), whereas feeding solution 1 was slightly modified and contained glucose (480 g/kg solution) and $MgSO_4 \cdot 7H_2O$ (15.8 g/kg solution). All media, unless otherwise stated in the text, were supplemented with ampicillin (0.1 g/liter), and tetracycline (12 mg/liter) was used when appropriate.

The fermentation inoculum was prepared as follows. *E. coli* was grown overnight at 30°C on LA plates, and single colonies were transferred to 100 ml of LB medium in shake flasks (500 ml baffled) and incubated for 7 h at 30°C (200 rpm; orbital movement amplitude, 2.5 cm). Cells were diluted 200-fold into 100 ml of prewarmed preculture medium, and growth was continued at 30°C for 16 h to an optical density at 600 nm (OD_{600}) of 4 to 5. From these cultures, cells were used to inoculate 0.75 liter of prewarmed main culture medium in the fermentor (3-liter Applikon) to an initial OD_{600} of 0.05. Inducer solution (0.5 M) was prepared by dissolving *m*-toluate in ethanol.

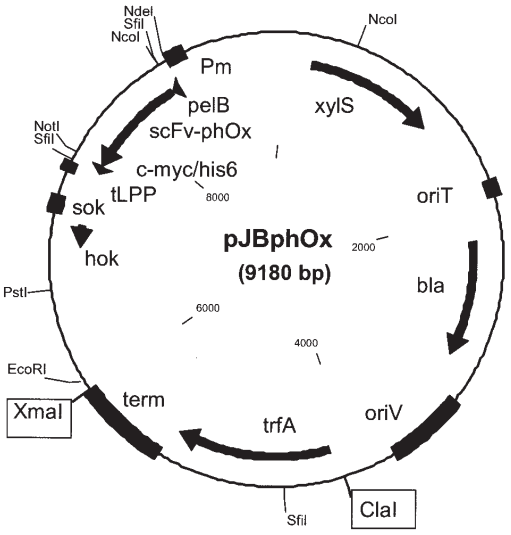

FIG. 1. Physical map of the expression plasmid pJBphOx. Plasmid pJBphOx is based on the broad-host-range plasmid pJB658 (Table 1) and harbors the regulatory *Pm/xylS* promoter system for expression of scFv-phOx. Relevant genetic elements: *tLPP*, transcription terminator; *term*, bidirectional transcription terminator; *bla*, β-lactamase; *oriT*, origin for conjugative transfer; *hok-sok*, suicide system. For remaining elements, see the text.

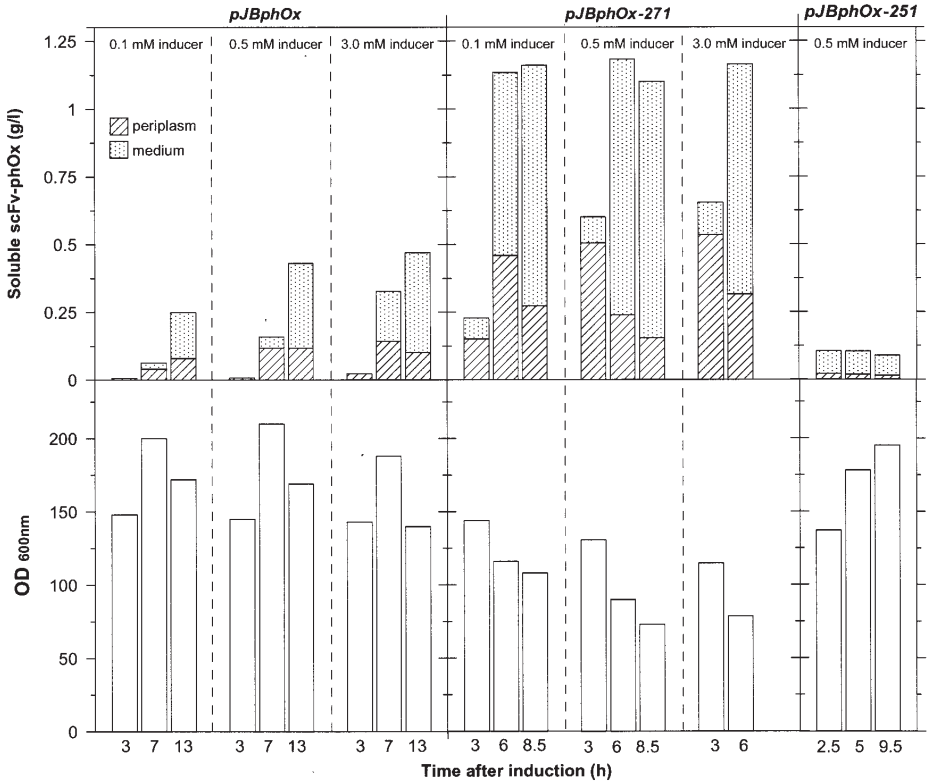

FIG. 2. Volumetric yield of soluble scFv-phOx (top) and cell densities (bottom) at different induction levels (*m*-toluate) in fermentations with *E. coli* RV308 strains harboring pJBphOx (low copy number) and its derivatives pJBphOx-271 (medium copy number) and pJBphOx-251 (high copy number). In all cases inducer was added at an OD_{600} of 90. The fermentation with RV308(pJBphOx-271) with 3 mM inducer was terminated 6 h after induction due to foaming and control difficulties caused by cell lysis.

High-cell-density cultivation. All fermentations were performed at 30°C, and the pH was maintained at 6.8 by addition of 12.5% (vol/vol) NH_3. Antifoam (Adecanol LG-109; Asahi Denka Kogyo, Japan) was added from the start (40 μl/liter) and thereafter when needed. The dissolved oxygen was maintained at 20% saturation by automatic adjustments of the stirrer speed or by controlled feeding of feeding solution 1. The airflow rate was initially 0.25 liter/liter of medium per min and thereafter increased to 1.1 liter/liter of medium per min throughout the batch phase (see below). Inducer (*m*-toluate) was added to the fermentors when cell growth reached an OD_{600} of 90.

The fermentations were divided into three different phases as follows. (i) Batch phase. Glucose (22.5 g/liter) was added to the freshly inoculated main culture medium, and cell growth was continued until all the sugar had been consumed. (ii) Exponential feeding phase. Growth was controlled to a specific growth rate (μ) of 0.2 h^{-1} by feeding with feeding solution 1. Initially the feed rate of feeding solution 1 was set to 17 g/liter/h and thereafter increased exponentially. This exponential feeding was continued until the stirrer speed reached 1,900 rpm, which is close to the maximal speed of the reactor (2,100 rpm). (iii) Induction phase. Bacterial growth was controlled by automatic adjustments of the feeding solution 1 feeding rate (flexible feeding control) so that the dissolved oxygen and stirring speed were maintained at 20% saturation and 1,900 rpm, respectively. Throughout the fermentations, pH, dissolved oxygen, feeding rate, airflow, and the molar fraction of CO_2 in the exhausted gas were monitored and recorded.

Determination of the fraction of plasmid-free cells. Cell samples were collected from each high-cell-density cultivation culture and diluted in LB medium to approximately 1,000 CFU per ml and grown on LA plates without antibiotic selection at 30°C overnight. From these plates 192 single colonies were picked with the Genetic QpixII colony picker and individually transferred to 96-well plates loaded with 140 μl of LB medium supplemented with ampicillin (100 μg/ml). In addition, each colony was similarly transferred to 96-well plates without antibiotic selection. The well plates were incubated at 30°C overnight, and cell growth in individual wells was analyzed by visual inspection. Plasmid-free cells were detected based on the inability to grow in the selective medium (observed as clear medium).

Preparation of cell samples for protein product analyses by ELISA and Western blotting. Cell samples (1 ml) were harvested by centrifugation (6,000 × g, 10 min) and the supernatants (medium) were collected. Periplasmic extraction was performed essentially as described elsewhere (21) by resuspension of the pellets in 0.5 ml of 50 mM Tris-HCl buffer, pH 8.0, addition of 0.5 ml of sucrose solution (40% sucrose containing 2 mM EDTA in 50 mM Tris-HCl buffer, pH 8.0), 1 mg of lysozyme per ml, and 125 U of benzonase (Merck) per ml. After incubation for 1 h at room temperature with shaking, the suspensions were centrifuged (16,000 × g, 8 min) and the supernatant (periplasm) and pellet fractions were collected.

Analysis of soluble scFv-phOx was performed by enzyme-linked immunosorbent assay (ELISA). Maxisorb plates (Nunc) were coated with phOx-bovine serum albumin (10 μg/ml) in Tris-buffered saline (TBS; 20 mM Tris-HCl, pH 7.5, 150 mM NaCl) overnight at 4°C. All further incubations were performed for 1 h at room temperature. The wells were washed three times with TBS and blocked with Blotto (1% skim milk powder and 0.02% antifoam A in TBS). Serial dilutions of samples in Blotto were then added to the wells, and after washing three times with TBST (TBS containing 0.05% Tween), anti-c-Myc (Invitrogen) (1:5,000 in Blotto) was added. The wells were then washed three times with TBST, and rabbit anti-mouse immunoglobulin-peroxidase (Dako) (1:1,000 in Blotto) was added. Finally, the wells were washed three times with TBST and

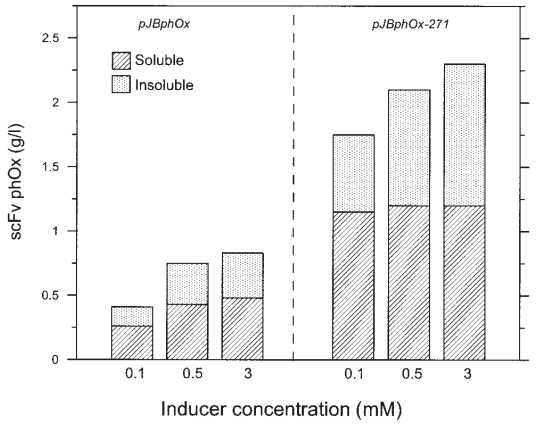

FIG. 3. Total volumetric yields of scFv-phOx (soluble and insoluble) at different induction levels in fermentations obtained with RV308 strains harboring pJBphOx (low copy number) and pJBphOx-271 (medium copy number). Strains harboring pJBphOx-251 (high copy number) were not included in this experiment due to massive loss of this plasmid prior to induction (see the text).

once with TBS, the substrate 2,2′-ethylbenzthiasolinesulfonic acid (ABTS) (Bio-Rad) was added, and the absorbance was read at 405 nm after 10 min.

Analysis of insoluble scFv-phOx in the pellet fractions was performed by Western blot. The pellets were added to 1 ml of sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) running buffer (Bio-Rad) and 1-mm glass beads (KeboLab) and resuspended with vortexing for 1 to 2 min, and dilution series were made and prepared by boiling in denaturing sample buffer. The samples were separated by SDS–12.5% PAGE and transferred to a nitrocellulose membrane, and the membrane was blocked with Blotto. Recombinant scFv-phOx was detected with the same antibodies and washing buffers as described for the ELISA (above) except that tetramethylbenzidine (Pharmingen) was used as the substrate. Comparison of band intensities with standard scFv-phOx of known concentrations was performed and used to estimate the yields of insoluble scFv-phOx. The scFv-phOx standard used in both the ELISA and Western blot experiments had been purified by Ni-Sepharose (Amersham Bioscience) according to the supplier's recommendations, and the concentrations of scFv-phOx were calculated from the absorbance at 280 nm.

RESULTS AND DISCUSSION

Construction and testing of vector pJBphOx for the production of soluble scFv-phOx in high-cell-density cultivation of *E. coli*. As a model protein in our study we chose scFv-phOx, which recognizes the hapten 2-phenyloxazol-5-one (phOx) (17). This important hapten can be easily conjugated to proteins and peptides and the clinical application of scFv-phOx for selective targeting of imaging and therapeutic agents in vivo has been demonstrated (8). In a previous report (15) the expression of 16.5 mg of soluble scFv-phOx/liter was obtained in *E. coli* shake-flask cultures with the *pelB* signal sequence. To our knowledge, no reports describe scFv-phOx expression under high-cell-density cultivation. The DNA fragment of plasmid pHOG21-phOx including the scFv-phOx gene fused to the c-*myc* coding region was cloned in frame between the *pelB* and the His6 coding sequences of plasmid pHKK (Table 1). From this construct the region encoding the *pelB-scFv-phOx-c-myc*-his6 fusion gene, as well as *hok-sok*, was excised and cloned downstream of the *Pm* promoter of pJB658, yielding plasmid pJBphOx (Fig. 1). In this way, scFv-phOx is expressed as a fusion protein targeted for translocation (governed by PelB) and with C-terminal tags (c-*myc*-his6) for facilitated detection and purification.

Three parallel fermentors with strain RV308 (pJBphOx) were run and induced at an OD_{600} of 90 (about 30 g dry weight per liter) by adding 0.1, 0.5, and 3.0 mM *m*-toluate. The production rate of soluble scFv-phOx increased as more inducer was added, and the highest yield was reached 13 h after induction in all cases (Fig. 2). At 3 mM inducer, 0.48 g/liter of soluble scFv-phOx was produced, which is comparable to the results obtained with similar proteins (13). Note that at 13 h most of the soluble product was localized in the growth medium. This may be due to toxic accumulation of scFv-phOx in the periplasm, leading to lysis of the producing cells (see OD_{600} values in Fig. 2). Representative samples were analyzed for insoluble product, and the results of these experiments (Fig. 3) showed that insoluble scFv-phOx yield also increased as more inducer was added, and it constitutes about 35 to 40% of the total scFv-phOx (soluble and insoluble) expressed. It seems probable that these insoluble fractions largely represent intracellular (not translocated) scFv-phOx proteins, while the formation of insoluble inclusion bodies in the periplasm could not be ruled out.

Manipulations with pJBphOx copy number and effects on recombinant scFv-phOx production levels. Plasmid pJB658 has a copy number of five to seven per genome (6) and from the results described above, we predicted that an elevated scFv-phOx gene dosage may accelerate its expression rate and thereby possibly lead to even higher levels of soluble product before cell lysis. To test this hypothesis, the 1,853-bp ClaI/XmaI fragment (including the *trfA* gene; see Fig. 1) was substituted by the corresponding fragment of plasmids pJB655cop271Cluc and pJB655cop251Mluc (Table 1). The expected copy numbers of the resulting derivatives, pJBphOx-271 and pJBphOx-251, should be about three- to fourfold and eightfold higher than that of the parental plasmid pJB658, respectively (6, 10, 12).

To experimentally verify these copy numbers, we took advantage of the ability to cotransform our recombinant strains with an additional replicon, pBR322, with a known copy number (about 30 per genome in *E. coli*). The 752-bp EcoRI-PstI region of pBR322 including the 5′ end of the β-lactamase gene was deleted to yield pBR322Δbla (Table 1), and this plasmid was introduced into *E. coli* strains harboring pJB658, pJBphOx, pJBphOx-271, and pJBphOx-251. The resulting recombinant strains were grown overnight, and dilution series of the plasmids isolated from these cultures were analyzed by gel electrophoresis. Comparison of the band intensities was used to estimate the copy number of each derivative, and the data obtained (not shown) were in good agreement with the corresponding expected values as indicated above.

Strain RV308(pJBphOx-271) was tested for production of scFv-phOx as described above, and the volumetric yield of soluble scFv-phOx obtained was more than twofold higher (1.2 g/liter) than with pJBphOx (Fig. 2), which is in agreement with the hypothesis above. Inducer concentrations above 0.1 mM had no positive effects on the final volumetric yield of soluble scFv-phOx, whereas the production rate in the first 3 h was higher with 0.5 mM. Presumably, a higher production rate leads to an accelerated release of soluble product into the

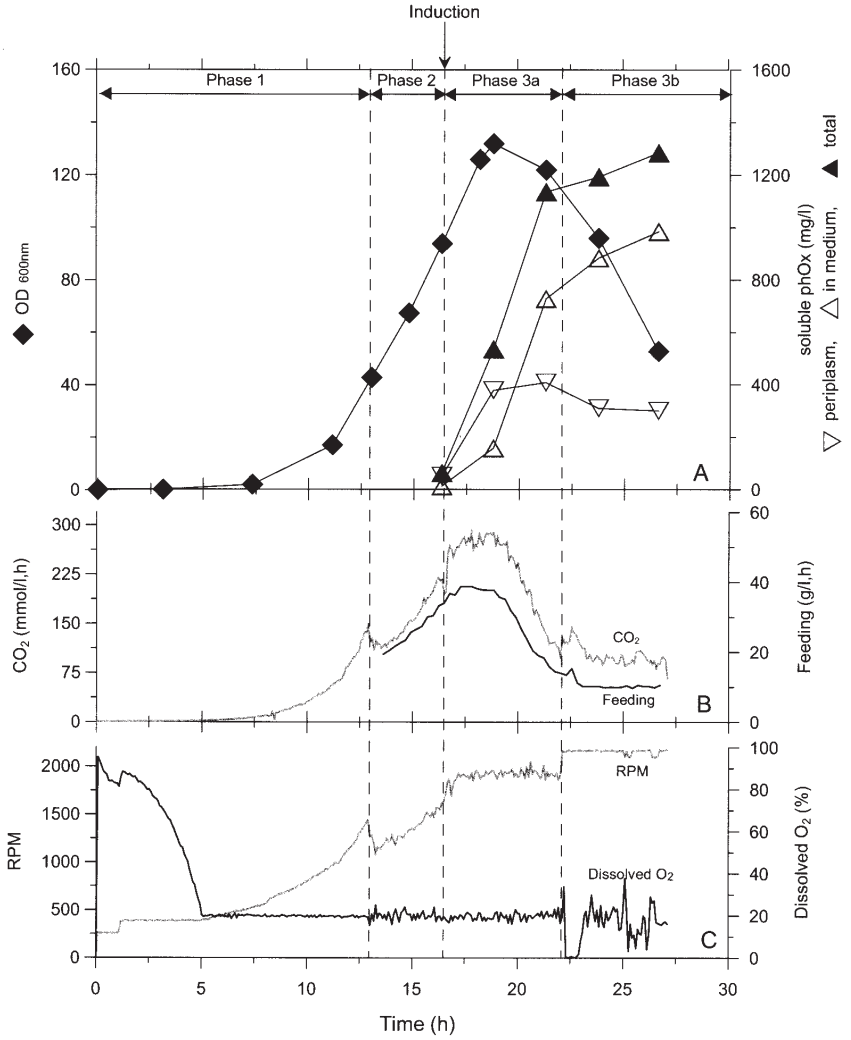

FIG. 4. Fermentation course of recombinant RV308 (pJBphOx-271) under high-cell-density cultivation. Time courses of bacterial growth (OD_{600}) and scFv-phOx production (A), exhausted CO_2 and feeding rate (B), and stirrer speed and dissolved oxygen (C). Phases 1 and 2 are the batch and exponential feeding phase of the fermentation, respectively (see Materials and Methods). The time of induction (addition of 0.5 mM *m*-toluate) is marked with an arrow. Phases 3a and b represent the initial and prolonged induction phases, respectively. In phase 3a, the flexible feeding solution 1 control was applied. In phase 3b, constant feeding and stirrer rates were used.

culture medium (Fig. 2), and the relative fraction of released scFv-phOx increased with incubation time after induction. The maximum cell densities obtained with this strain (OD_{600} between 115 and 144) were lower than those obtained with cells harboring pJBphOx (OD_{600} of about 200) and decreased throughout the induction phase (Fig. 2). These observations support the assumption that high recombinant expression levels lead to rapid toxic accumulation of soluble scFv-phOx in the periplasm, resulting in cell lysis. In spite of this, a faster production rate allowed more total product to be made.

Analysis of RV308(pJBphOx-271) cells for insoluble product demonstrated that about twofold more insoluble scFv-phOx was produced with 3 mM compared to 0.1 mM inducer, even though the amount of soluble product did not increase (Fig. 3). This probably means that translocation is a limiting parameter for further improved production of soluble scFv-phOx under the present conditions, leading to a total production of about 2.3 g/liter of scFv-phOx.

In order to analyze the production course of RV308 (pJBphOx-271), the fermentation with this strain was repeated (five parallel fermentors) with an induction level of 0.5 mM, and a typical fermentation course is shown in Fig. 4. The

volumetric yield of soluble scFv-phOx obtained in these experiments ranged from 1.1 to 1.3 g/liter, demonstrating good reproducibility. The production of scFv-phOx was initiated immediately after the induction, and 90% of the soluble scFv-phOx was produced within 6 h (phase 3a in Fig. 4A). In this period the cell density reached its maximal value and then gradually dropped. The drop in optical density was accompanied by a substantial leakage of scFv-phOx to the extracellular environment, consistent with cell lysis.

In the induction phase (phase 3a, Fig. 4) the feeding rate and dissolved oxygen are maintained constant at the setpoint values by automatic adjustments of the feeding rate. The feeding rate is at its highest level at induction start and is thereafter automatically gradually reduced to one-third of the maximum rate during the next 6 h (phase 3a in Fig. 4B). The reduction of the feeding rate is a consequence of reduced oxygen transfer capacity in the fermentation broth caused by cell lysis. In order to push the system to a maximum level of production, the induction phase was prolonged beyond 6 h (phase 3b in Fig. 4). To control the foaming caused by cell lysis, substantial amounts of antifoam were added, leading to problems with the automatic control of the fermentor. The prolonged induction phase had little impact on the volumetric yield (additional 10%), and the optical density continued to drop throughout this period.

It could not be excluded that production yield of soluble scFv-phOx could be further improved by increasing the plasmid copy number beyond that of pJBphOx-271. Surprisingly, irrespective of the inducer concentration applied (data not shown), the maximum level of soluble scFv-phOx produced by strain RV308(pJBphOx-251) was only 0.15 g/liter (Fig. 2), which is even lower than that of cells harboring the parental vector. The reasons for this unexpected result could later be shown to correlate with extensive plasmid loss before induction (see below).

Effects of high plasmid copy numbers on plasmid stability and viability of recombinant cells. It is generally known that the metabolic burden associated with elevated plasmid copy numbers and/or high recombinant gene expression levels may lead to impaired cell growth and concomitant plasmid loss from recombinant cells (1, 2). We analyzed the various fermentation cultures for plasmid-free cells, and the results of these experiments demonstrated that the low (pJBphOx) and medium (pJBphOx-271) copy number vectors were maintained throughout the fermentation. In contrast, vector pJBphOx-251 was found to be highly unstable, and about 90% of the cells originally harboring this vector were already plasmid-free prior to induction.

We speculated that this massive plasmid loss could be due to high leakage of scFv-phOx expression from *Pm*, and to test this we measured soluble scFv-phOx levels in the fermentor cultures prior to induction. Only traces (below 5 mg/liter) of recombinant protein were detected with the low-copy number plasmid pJBphOx, while significant levels of soluble scFv-phOx were detected with pJBphOx-271 (50 mg/ml) and pJBphOx-251 (120 mg/liter). Considering that the majority of the cells in the latter culture were plasmid-free under these conditions (see above), this result suggested that the leakage from *Pm* in strain RV308 (pJBphOx-251) is high.

To rule out the possibility that the observed plasmid loss was somehow caused by the vector itself, the 830-bp region including the scFv-phOx-encoding gene was deleted from pJBphOx-251, yielding construct pJB-251E (Table 1). Recombinant strains harboring this plasmid were analyzed as described above, and the results showed that this derivative was 100% stable during the entire fermentation period, supporting the assumption that the instability of pJBphOx-251 is related solely to leaky expression of scFv-phOx.

Both data previously reported by us (6) and the results presented here suggest that uninduced expression from *Pm* increases much more than the increase in copy number, and it has been reported that high XylS levels may accelerate constitutive transcription from *Pm* (24). To investigate this, we deleted the 776-bp region including the 5′-terminal region of *xylS* and upstream sequences of the copy-up derivatives pJBphOx-271 and pJBphOx-251, yielding plasmids pJBphOx-271ΔxylS and pJBphOx-251ΔxylS, respectively (Table 1). Recombinant RV308 strains harboring these constructs produce only 5 and 30 mg/liter of soluble scFv-phOx, respectively, under noninduced conditions, which is far below the levels detected with the parental vectors (see above). Interestingly, no plasmid-free cells were detected in these cultures. Together, these results indicate that high basal expression from *Pm* present in a high-copy-number vector is mediated solely by the simultaneous presence of a correspondingly increased copy number of the *xylS* gene. By controlling the *xylS* expression level from a regulated promoter, the problems with plasmid loss at high copy numbers might be avoided, and one might possibly also envision improved yields of scFv-phOx or other protein products of interest.

Plasmid pJBphOx-271 is stably maintained under high-cell-density cultivation without antibiotic selection. The numbers of generations obtained in the preculture medium and the main culture medium under the conditions used in the present study are 9 and 11 to 12, respectively. To further test the stability of our expression vectors, high-cell-density cultivation fermentation of RV308(pJBphOx-271) was performed without antibiotic selection. Moreover, to obtain conditions comparable to a large-scale fermentation, preculture medium cultivation was repeated twice before transfer to the main culture medium, resulting in approximately 30 generations for the three successive cultivations. As a control, a parallel experiment was run with the standard amount of ampicillin (0.1 g/ml) added to the media. The results of these experiments demonstrated that pJBphOx-271 was stably maintained during the entire fermentations and the production levels of soluble scFV-phOx were also similar (data not shown). The *hok-sok* suicide elements have been shown to significantly enhance the stability when present on other replicons (20), but their direct contribution to the stability observed in our constructs remains unknown. In any case, the option of running industrial processes without antibiotics in the production medium confers both economic and regulatory advantages.

Different signal sequences display considerable effects on the expression levels of recombinant scFv-phOx. In *E. coli* it has been shown that the first codons at the 5′ end of a structural gene may have a severe impact on translation initiation (27). In our constructs it is likely that the 5′-terminally fused *pelB* sequence, which is of bacterial origin, contributes to effective expression of scFv-phOx in addition to governing trans-

location. Although the majority of the scFv-phOx expressed in our strains is presumably exported, we could not rule out that translocation is limiting for higher yields of soluble scFv-phOx under the conditions tested. In order to test this, we applied the two alternative signal sequences *ompA* and consensus signal peptide. The former has been used for effective translocation of various recombinant proteins in *E. coli* (22), whereas consensus signal peptide was designed here based on published sequence alignments of several bacterial signal peptides (18). Both signal sequences were made synthetically (see Materials and Methods) and used to substitute *pelB* in pJBphOx-271, yielding derivatives pJBphOx-271ompA and pJBphOx-271CSP (Table 1).

Recombinant strains harboring these plasmids produced soluble scFv-phOx at significantly lower yields (0.15 and 0.43 g/liter, respectively) than the strain harboring the parental vector. We proceeded to measure insoluble scFv-phOx in these samples, and the data showed that these levels were also low (0.10 g/liter for OmpA and 0.17 g/liter for consensus signal peptide). Together our results indicate that the overall scFv-phOx expression levels obtained with OmpA and consensus signal peptide are low compared to those obtained with PelB. Thus, our results show that the choice of signal sequence has a strong impact on both the translocation and expression level of recombinant scFv-phOx under the conditions tested.

ACKNOWLEDGMENTS

We are grateful to Uwe Horn for kindly providing us with plasmid pHKK. We thank Inger Løberslie, Anne Tøndervik, and Marianne Korsnes for helping with scFv-phOx analyses and certain cloning experiments, Mona Senneset for help with fermentations, and Geir Klinkenberg for help with programming the control systems needed for the high-cell-density cultivation fermentations.

The work was financed by Affitech AS and the Research Council of Norway.

REFERENCES

1. **Bentley, W. E., N. Mirjalili, D. C. Andersen, R. H. Davis, and D. S. Kompala.** 1990. Plasmid encoded protein: The principal factor in the "metabolic nurden" associated with recombinant bacteria. Biotechnol. Bioeng. **35:**668–681.
2. **Betenbaugh, M. J., C. Beaty, and P. Dhurjati.** 1989. Effects of plasmid amplification and recombinant gene expression on the growth kinetics of recombinant *E. coli*. Biotechnol. Bioeng. **33:**1425–1436.
3. **Better, M., C. P. Chang, R. R. Robinson, and A. H. Horwitz.** 1988. *Escherichia coli* secretion of an active chimeric antibody fragment. Science **242:** 1041–1043.
4. **Bird, R. E., K. D. Hardman, J. W. Jacobsen, S. Johnson, B. M. Kaufman, S-M. Lee, T. Lee, S. H. Pope, G. S. Riordan, and M. Whitlow.** 1988. Single-chain antigen-binding proteins. Science **242:**423–426.
5. **Blatny, J. B., T. Brautaset, H. C. Winther-Larsen, K. Haugan, and S. Valla.** 1997. Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. Appl. Environ. Microbiol. **63:** 370–379.
6. **Blatny, J. B., T. Brautaset, H. C. Winther-Larsen, P. Karunakaran, and S. Valla.** 1997. Improved broad-host-range RK2 vectors useful for high and low regulated gene expression levels in gram-negative bacteria. Plasmid **38:**35–51.
7. **Brautaset, T., S. B. Petersen, and S. Valla.** 2000. *In vitro* determined kinetic properties of mutant phosphoglucomutases and their effects on sugar catabolism in *Escherichia coli*. Metab. Eng. **2:**104–114.
8. **Cheng, T. L., Liao, K. W., Tzou, S. C., Cheng, C. M., Chen, B. M., and Roffler, S. R.** 2004. Hapten-directed targeting to single-chain antibody receptors. Cancer Gene Ther. **11:**380–388.
9. **Donini, M., V. Morea, A. Desidero, D. Pashkoulov, M. E. Villani, A. Tramontano, and E. Benvenuto.** 2003. Engineering stable cytoplasmic intrabodies with designed specificity. J. Mol. Biol. **330:**323–332.
10. **Durland, R. H., A. Toukdarian, F. Fang, and D. R. Helinski.** 1990. Mutations in the *trfA* replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy number. J. Bacteriol. **172:**3859–3867.
11. **Guo, J.-Q., S-Y. You, L. Li, Y-Z. Zhang, J.-N. Huang, and C-Y. Zhang.** 2003. Construction and high-level expression of a single-chain Fv antibody fragment specific for acidic isoferritin in *Escherichia coli*. J. Biotechnol. **102:** 177–189.
12. **Haugan, K., P. Karunakaran, A. Tøndervik, and S. Valla.** 1995. The host-range of RK2 minimal replicon copy up mutants is limited by species-specific differences in the maximum tolerable copy number. Plasmid **33:**27–39.
13. **Horn, U., W. Strittmatter, A. Krebber, U. Knüpfer, M. Kujau, R. Wenderoth, K. Müller, S. Matzku, A. Plückthun, and D. Riesenberg.** 1996. High volumetric yields of functional dimeric miniantibodies in *Escherichia coli*, using an optimized expression vector and high-cell-density fermentation under non-limited growth conditions. Appl. Microbiol. Biotechnol. **46:**524–532.
14. **Humphreys, D. P.** 2003. Production of antibodies and antibody fragments in *Escherichia coli* and a comparison of their functions, uses and modification. Curr. Opin. Drug Discov. Dev. **6:**188–196.
15. **Kipriyanov, S. M., G. Moldenhauer, and M. Little.** 1997. High level production of soluble single chain antibodies in small-scale *Escherichia coli* cultures. J. Immunol. Methods **200:**69–77.
16. **Kumar, P. K. R., H-E. Maschke, K. Friehs, and K. Schügerl.** 1991. Strategies for improving plasmid stability in genetically modified bacteria in bioreactors. Trends Biotechnol. **9:**279–284.
17. **Marks, J. D., A. D. Griffiths, D. Andrew, M. Malmquist, T. Clackson, J. Bye, and G. Winter.** 1992. By-passing immunization: building high affinity human antibodies by chain shuffling. Bio/Technology **10:**779–783.
18. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne.** 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:**1–6.
19. **Oelschlaeger, P., S. Lange, J. Schmitt, M. Siemann, M. Reuss, and R. D. Schmid.** 2003. Identification of factors impeding the production of a single-chain antibody fragment in *Escherichia coli* by comparing in vivo and in vitro expression. Appl. Microbiol. Biotechnol. **61:**123–132.
20. **Pecota, D. C., C. S. Kim, K. Wu, K. Gerdes, and T. K. Wood.** 1997. Combining the *hok/sok, parDE*, and pnd postsegregational killer loci to enhance plasmid stability. Appl. Environ. Microbiol. **63:**1917–1924.
21. **Pierce, J. J., C. Turner, E. Keshacarz-Moore, and P. Dunnhill.** 1997. Factors determining more efficient large-scale release of a periplasmic enzyme from *E. coli* using lysozyme. J. Biotechnol. **58:**1–11.
22. **Pines, O., and M. Inouye.** 1999. Expression and secretion of proteins in *E. coli*. Mol. Biotechnol. **12:**25–34.
23. **Pini, A., and L. Bracci.** 2000. Phage display of antibody fragments. Curr. Protein Peptide Sci. **1:**155–169.
24. **Ramos, J. L., C. Michan, F. Rojo, D. Dwyer, and T. Timmis.** 1990. Signal-regulator interactions: Genetic analysis of the effector binding site of XylS, the benzoate-activated positive regulator of Pseudomonas TOL plasmid meta-cleavage pathway operon. J. Mol. Biol. **211:**373–382.
25. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
26. **Skerra, A., and A. Plückthun.** 1988. Assembly of functional immunoglobulin Fv fragment in *Escherichia coli*. Science **240:**1038–1041.
27. **Stenström, C. M., E. Holmgren, and L. A. Isakson.** 2001. Cooperative effects by the initiation codon and its flanking regions on translation initiation. Gene **273:**259–265.
28. **Summers, D. K.** 1999. The kinetics of plasmid loss. Trends Biotechnol. **9:** 273–278.
29. **Winther-Larsen, H. C., K. D. Josefsen, T. Brautaset, and S. Valla.** 2000. Parameters affecting gene expression from the *Pm* promoter in gram-negative bacteria. Metab. Eng. **2:**79–91.
30. **Winther-Larsen, H. C., J. B. Blatny, B. Valand, T. Brautaset, and S. Valla.** 2000. *Pm* promoter expression mutants and their use in broad-host-range RK2 plasmid vectors. Metab. Eng. **2:**92–103.

Paper II

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Feb. 2007, p. 906–912 Vol. 73, No. 3
0099-2240/07/$08.00+0 doi:10.1128/AEM.01804-06

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*▽

H. Sletta,[1] A. Tøndervik,[2] S. Hakvåg,[2] T. E. Vee Aune,[2] A. Nedal,[2] R. Aune,[1] G. Evensen,[3] S. Valla,[2] T. E. Ellingsen,[1] and T. Brautaset[1]*

SINTEF Materials and Chemistry, Department of Biotechnology, SINTEF, Trondheim, Norway[1]; Department of Biotechnology, Norwegian University of Science and Technology, Trondheim, Norway[2]; and Alpharma AS, Oslo, Norway[3]

Received 31 July 2006/Accepted 18 November 2006

Genetic optimizations to achieve high-level production of three different proteins of medical importance for humans, granulocyte-macrophage colony-stimulating factor (GM-CSF), interferon alpha 2b (IFN-α2b), and single-chain antibody variable fragment (scFv-phOx), were investigated during high-cell-density cultivations of *Escherichia coli*. All three proteins were poorly expressed when put under control of the strong *Pm*/*xylS* promoter/regulator system, but high volumetric yields of GM-CSF and scFv-phOx (up to 1.7 and 2.3 g/liter, respectively) were achieved when the respective genes were fused to a translocation signal sequence. The choice of signal sequence, *pelB*, *ompA*, or synthetic signal sequence CSP, displayed a high and specific impact on the total expression levels for these two proteins. Data obtained by quantitative PCR confirmed relatively high in vivo transcript levels without using a fused signal sequence, suggesting that the signal sequences mainly stimulate translation. IFN-α2b expression remained poor even when fused to a signal sequence, and an alternative IFN-α2b coding sequence that was optimized for effective expression in *Escherichia coli* was therefore synthesized. The total expression level of this optimized gene remained low, while high-level production (0.6 g/liter) was achieved when the gene was fused to a signal sequence. Together, our results demonstrate a critical role of signal sequences for achieving industrial level expression of three human proteins in *E. coli* under the conditions tested, and this effect has to our knowledge not previously been systematically investigated.

Human cytokines are proteins that promote immune responses, and they have a broad range of medical uses, such as treatment of microbial and viral infections and vaccination against cancer. In 2004, the global cytokine market was 6.5 billion U.S. dollars, and according to a Research and Markets report, the demand for existing cytokines is expected to grow significantly in the next years (http://www.researchandmarkets.com/reports/314808/). Thus, cytokines are needed in large quantities, and the development of high-cell-density cultivations (HCDC) has led to production at high volumetric yields of such pharmaceutically important proteins in heterologous hosts like *Escherichia coli*. Despite the many advantages of this organism, high-level heterologous expression is not routinely achieved, typically due to biased codon usage, gene product toxicity, low gene product solubility, mRNA secondary structure formation, and low mRNA stability (15, 26). The broad-host-range plasmid pJB658 harbors the inducible *Pm*/*xylS* promoter/regulator elements for recombinant expression of cloned genes in a wide range of gram-negative bacteria (3, 4). We recently modified this plasmid to express high volumetric yields of secreted recombinant single-chain antibody variable fragment (scFv-phOx) during HCDC of *E. coli* (26). To obtain secretion, we added (at the DNA level) a signal sequence at the N terminus of scFv-phOx, a method that is commonly used if intracellular production is undesired.

The important cytokine granulocyte-macrophage colony-stimulating factor (GM-CSF) is one of four specific glycoproteins that stimulate generation of the white blood cells granulocytes and macrophages (18). Recombinant GM-CSF has been expressed in bacterial, yeast, and mammalian cells and is now produced for clinical uses. This protein has greatly reduced the infection risk associated with bone marrow transplantation (19). GM-CSF produced recombinantly in *E. coli* ends up in inclusion bodies (IBs) and has certain drawbacks, including complex processing, low specific activity, and poor in vitro renaturation (2). The GM-CSF product obtained as an IB has an added methionine at the N terminus that leads to stimulation of antibody production in the human body, hence influencing the therapeutic value (8, 33). Interferon alpha 2b (IFN-α2b) belongs to the IFN family of cytokines, which can induce antiproliferative, immunomodulatory, and potent antiviral activities against a wide range of mammalian viruses (7). IFN-α2b is used to treat several diseases, including some types of cancer and hepatitis, in particular hepatitis C. Since it can increase the intensity of antigen expression on certain tumors, IFN-α2b has a potential for use in diagnostics and therapeutics (20). As for GM-CSF, IFN-α2b has also been expressed recombinantly into active form in *E. coli* (28).

* Corresponding author. Mailing address: SINTEF Materials and Chemistry, Department of Biotechnology, SINTEF, Sem Selands vei 2, 7465 Trondheim, Norway. Phone: 47 98 28 39 77. Fax: 47 73 59 69 95. E-mail: trygve.brautaset@sintef.no.

▽ Published ahead of print on 1 December 2006.

In this report, we describe the use of the pJB658-based expression system (26) to produce GM-CSF, IFN-α2b, and scFv-phOx both intracellularly (lacking signal sequences) and through secretion (with signal sequences) during HCDC. Surprisingly, we found that the presence of signal sequences very strongly stimulated not only secretion but also the total production levels of all three of these proteins. Such effects are to our knowledge not commonly referred to in the scientific literature but should be of significant importance in this field of biotechnology.

MATERIALS AND METHODS

Strains, DNA manipulations, and growth conditions. *E. coli* strains used in this study were as follows: DH5α (BRL) was used as a cloning host, while RV308 (ATCC 31608) was the standard recombinant production strain used for HCDC of *E. coli*. Strain BL21(DE3) (Stratagene) is deficient in the OmpT and Lon proteases and was used as an alternative test strain for recombinant protein production. BL21-CodonPlus (DE3)-RIPL (Stratagene) carries *argU*, *ileY*, and *leuW* genes encoding tRNA molecules that recognize rare codon triplets. ATCC 67979 and ATCC 53157 carry plasmids with the IFN-α2b and GM-CSF coding regions, respectively (ATCC). Standard cloning experiments were performed as described elsewhere (23), and recombinant *E. coli* strains were grown at 37°C in liquid Luria-Bertani (LB) medium or on solid LB agar plates. For small-scale production analyses, cells were grown at 30°C with shaking (225 rpm; orbital moment, 2.5-cm amplitude) in 250-ml baffled shake flasks containing 40 ml HiYe medium, which is composed as follows: $Na_2HPO_4 \cdot 2H_2O$, 8.6 g/liter; KH_2PO_4, 3 g/liter; NH_4Cl, 1 g/liter; NaCl, 0.5 g/liter; glucose, 2 g/liter; glycerol, 10 g/liter; yeast extract, 10 g/liter; $MgSO_4$, 2.5 mM; Fe(III)-citrate, 250 μM; H_3BO_3, 49 μM; $MnCl_2$, 79 μM; EDTA, 23 μM; $CuCl_2$, 9 μM; Na_2MoO_4, 10 μM; $CoCl_2$, 11 μM; and Zn-acetate, 36 μM. Site-specific mutagenesis was performed by using a QuickChange site-directed mutagenesis kit from Stratagene according to the manufacturer's instructions. Oligonucleotides used in this study are listed in Table 1. The defined preculture and the main culture media were prepared as described previously (26). When appropriate, the media were supplemented with ampicillin (100 μg/ml) and chloramphenicol (10 μg/ml). Design of the optimized IFN-α2bS gene was performed by GenScript, a 501-bp double-stranded DNA fragment encoding this sequence was synthesized (GeneScript), and unique NdeI and NotI restriction sites were introduced at the 5′ and 3′ ends, respectively, and cloned into the corresponding sites of pUC57, yielding pUC57IFN-S.

Vector constructions. All oligonucleotide primers used hereunder are listed in Table 1, and all expression vectors constructed were verified by DNA sequencing.

pJBphOx-271d. Two unique restriction sites, SacI and AgeI, were introduced, flanking the *xylS* coding region of plasmid pJT19bla (31), by site-directed mutagenesis with the mutagenic oligonucleotide pairs sacI-F and sacI-R and ageI-F and ageI-R, respectively. The 1,025-bp SacI-AgeI fragment was subcloned into the corresponding sites of pLITMUS28 (New England Biolabs). The NcoI restriction site in the *xylS* gene then was specifically mutated in the latter construct with the mutagenic oligonucleotides ncoI-F and ncoI-R, yielding pTA4-NcoI. In parallel, plasmid pJB655cop271 (3) was digested with XmaI-BseRI, and the cohesive ends of the vector fragment were blunted and religated. The unique NdeI site of the *trfA* gene was then specifically mutated with the mutagenic oligonucleotides ndeI-F and ndeI-R, yielding plasmid pTA20. The 1,171-bp ClaI-SexA1 fragment of pTA20, including the mutated *trfA* gene, and the 339-bp HpaI-MunI fragment of pTA4-NcoI, including the *xylS* gene, were isolated and used to substitute for the corresponding fragments of vector pJBphOx-271 (26), yielding the cassette cloning and expression vector pJBphOx-271d.

pGM29. The coding region of the mature GM-CSF gene was PCR cloned from total DNA isolated from *E. coli* ATCC 53157 by using primers gm-1F and gm-1R. The resulting DNA fragment was end digested with NdeI and NotI, and the 389-bp fragment was used to substitute with the corresponding scFv-phOx fragment of pJBphOx-271d, yielding pGM29 (expressing GM-CSF without signal sequence).

pGM29pelB. The coding region of the mature GM-CSF gene was PCR cloned from *E. coli* ATCC 53157 total DNA with the primers gm-2F and gm2R. The resulting DNA fragment was end digested with NcoI and NotI, and the 390-bp fragment was used to substitute for the corresponding scFv-phOx fragment of plasmid pJBphOx-271d, yielding pGM29pelB (expressing GM-CSF with *pelB* signal sequence).

TABLE 1. Oligonucleotides used in this study

Name	Sequence (5′–3′)[a]
sacI-F	CTCTTATTTTAATGTGAGCTCCTTGGTGTGATGTAG
sacI-R	CTACATCACACCAAGGAGCTCACATTAAAATAAGAG
ageI-F	GCGATGCCAACCGGTCCCTTCTTCGGC
ageI-R	GCCGAAGAAGGGACCGGTTGGCATCGC
ncoI-F	GGCGTGGCCATGGCCAGGAGCAC
ncoI-R	GTGCTCCTGGCCATGGCCACGCC
ndeI-F	CTACACGAAATTCATCTGGGAGAAGTACCGC
ndeI-R	GCGGTACTTCTCCCAGATGAATTTCGTGTAG
gm-1F	TTTTCATATGGCACCCGCCCGCTCGCC
gm-1R	TTCGGATCCAGCGGCCGC
gm-2F	CACCATGGCTGCACCCGCCCGCTCGCC
gm-2R	GCGCGGCCGCCTCCTGGACTGGCTCCC
phox-1F	TTTTCATATGGCCCAGGTGCAGCTGGTGCAGTCTGGAG
phox-1R	ATCAGCTTTTGTTCGGATCCAGCGGCCGCACC
IFN-1R	TTTTCCATGGCTTGTGATCTGCCTCAAACCC
IFN-1R	TTTTGCGGCCGCTTCCTTACTTCTTAAACTTTCTTGC
IFN-2F	TTTTCATATGTGTGATCTGCCTCAAACCC
IFN-3F	TTTTCCATGGCTTGCGATCTGCCGCAGACCC
IFN-2R	CAGCTTTTGTTCGGATCCAGCGGCCGCTTCTTTGC

[a] Relevant restriction sites used for cloning of PCR products are underlined.

pGM29ompA and pGM29CSP. The DNA fragments with *ompA* and CSP coding sequences were prepared by annealing of synthetic oligonucleotides as described previously (26) and used to substitute for the corresponding NdeI-NcoI *pelB* fragment of pGM29pelB, to yield pGM29ompA and pGM29CSP, respectively (expressing GM-CSF with *ompA* and CSP signal sequences, respectively).

pAT65. The scFv-phOx-cmyc-his6 coding region of pJBphOx was PCR amplified by using the primers phox-1F and phox-1R and end digested with NdeI-NotI, and the resulting 770-bp fragment was used to substitute for the corresponding fragment of plasmid pGM29pelB, yielding pAT65 (expressing scFv-phOx without signal sequence).

pIFN30pelB. The IFN-α2b coding region was PCR amplified from DNA isolated from *E. coli* strain ATCC 67979 by using the primers IFN-1F and IFN-1R and was end digested with NcoI-NotI, and the 504-bp fragment was ligated into the corresponding sites of pGEM-5zf (Promega). From the resulting plasmid, the 504-bp NcoI-NotI insert was isolated and used to substitute for the corresponding region of pGM29pelB, yielding pIFN30pelB (expressing IFN-α2b with *pelB* signal sequence).

pIFN30. The IFN-α2b coding region was PCR amplified from DNA isolated from strain ATCC 67979 by using the primers IFN-2F and IFN-1R and was end digested with NdeI-NotI, and the 501-bp fragment was ligated into the corresponding sites of pGEM-5zf. From the resulting plasmid, the 501-bp NdeI-NotI insert was isolated and used to substitute for the corresponding fragment of pGM29pelB, yielding pIFN30 (expressing IFN-α2b without signal sequence).

pIFN30S and pIFN30SpelB. The 501-bp NdeI-NotI fragment of plasmid pUC57IFN-S (see above) was used to substitute for the corresponding fragment of plasmid pGM29pelB, yielding pIFN30S (expressing IFN-α2bS without signal sequence). The IFN-α2bS insert was PCR amplified from pIFN30S by using primers IFN-3F and IFN-2R and was end digested with NcoI-NotI, and the 504-kb fragment was used to substitute for the corresponding fragment of pGM20pelB, yielding pIFN30SpelB (expressing IFN-α2bS with *pelB* signal sequence).

Production analyses in shake flask cultures. Overnight cultures of recombinant cells were diluted into fresh prewarmed HiYe medium (100 ml) to an optical density at 600 nm (OD_{600}) of 1, and cell growth was continued to an OD_{600} of 3. At this point, the cultures were induced by adding *m*-toluic acid (0.5 mM) and cell growth was continued for 4 h (OD_{600} = 10 to 15) before cell harvesting. Preparation of cell samples and product detections were performed as described below.

Production analyses in HCDC. Preparation of fermentation inoculum and HCDC of recombinant *E. coli* strains was performed as described previously (26). The cell cultures were induced at an OD_{600} of 100, and cell growth was

continued for 4 to 6 h, typically corresponding to an OD_{600} of 180 to 200. All quantitative protein data given from the high-yielding HCD fermentations are based on values from at least two independent cultivations.

Preparation of cell samples for protein product analyses by ELISA and Western blotting. Cell sample preparations of medium (designated as S1 phase), soluble (designated as S2 phase), and pellet (designated as P2 phase) fractions for production analyses, and Western analyses of recombinant IFN-α2b, scFv-phOx, and GM-CSF proteins, were performed as described elsewhere (26). Recombinant soluble GM-CSF was quantified by using a human GM-CSF enzyme-linked immunosorbent assay (ELISA) set kit from BD Biosciences, in accordance with the manufacturer's instructions. Purified scFv-phOx protein served as a standard in all cases.

N-terminal sequencing of recombinant proteins. Samples (0.2 ml) collected from the soluble fractions (see above) of the cell extracts were purified by using the Ni-nitrilotriacetic acid spin protocol from QIAGEN under denaturating conditions, in accordance with manufacturer's instructions. About 15 mg of purified protein (judged by using the Bio-Rad method) was then subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis and concomitant Western blotting, as described above. The membrane was stained with Coomassie brilliant blue R250, and the band corresponding to the desired mass was excised and used as a template for the sequencing which was performed by the Biotechnology Centre of Oslo, Norway (K. Sletten and S. Kjæraas).

Isolation of total RNA, reverse transcription, and quantitative PCR. The experiments hereunder were performed essentially as described previously (9). Cell samples (0.1 ml) were harvested from the HCDC fermentations at an OD_{600} of 130 to 160, diluted in fresh medium to an OD_{600} of 10, and immediately treated with RNA Protect (QIAGEN). Total RNA was isolated by using an RNAqueous kit (Ambion) according to the manufacturer's instructions, and 3 μg of the obtained material was treated with a DNA-free kit (Ambion) and used as template for cDNA synthesis by using a first-strand cDNA synthesis kit (Amersham). The quantitative PCR profile was as follows: segment 1 (1 cycle), 95°C at 10 min; segment 2 (40 cycles), 95°C for 30 s, 55°C for 60 s, and 68°C for 30 s; and segment 3 (1 cycle), 95°C for 60 s, 55°C for 30 s, and slowly up to 95°C (dissociation curve). The primers used are listed in Table 1. We used the iTaq SYBR Green Supermix with Rox (Bio-Rad), and amplification reaction mixtures contained 5 μl of diluted cDNA templates and 3 pmol of each primer in a final volume of 20 μl. The PCR products were detected by monitoring the increase in fluorescence by using an Mx3000P cycler system (Stratagene). Thresholds (C_T) were set which intersected the amplification curves in the linear region of the semilogarithmic plots, and inspection of the dissociation curves confirmed negligible levels of primer self-hybridizations. Amplification of the β-lactamase gene present on the expression vector was used for sample normalization. All experiments were performed in three parallels.

In silico analysis of mRNA secondary structures. The deduced transcripts of the gene variants were analyzed for secondary structures by using the program MFOLD (34) (http://www.bioinfo.rpi.edu/applications/mfold/old/rna/). Regions covering about 50 to 60 bp from the *Pm* transcription start point (22), including the translation initiation codon and its flanking regions, were used for these analyses.

RESULTS

Construction of the cloning and cassette expression vector pJBphOx-271d. Plasmid pJBphOx-271, used previously to produce scFv-phOx, was modified (generating pJBphOx-271d; Fig. 1) to facilitate the construction of the expression cassettes required for this study. In this new vector, the gene encoding scFv-phOx can easily be substituted by any gene of interest by a one-step cloning procedure, using either the NdeI/NotI sites (no signal sequence) or the NcoI/NotI sites (DNA sequence encoding the *pelB* signal sequence fused in frame to the 5′-terminal end). The coupled affinity detection sequence c-*myc*-his6 is fused to the 3′-terminal end of the cloned gene. The complete fusion genes are at the 5′ end, with or without a secretion signal sequence, placed in frame under translational control of the ribosome binding sequence (*rbs*) region and under transcriptional control of the tightly controlled and inducible *Pm*/*xylS* promoter/regulator system (3, 4) of the plasmid. Recombinant genes to be expressed, the type of signal sequence (*pelB*, *ompA*, or CSP) used, or the vector copy number (see reference 26) can thus easily be changed by one-step cloning procedures.

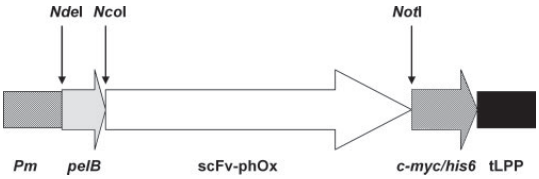

FIG. 1. Physical map of part of the cassette cloning and expression vector pJBphOx-271d. Relevant restriction sites useful for gene cloning with and without a fused signal sequence, as well as for shifting out *pelB* with *ompA* and CSP, are indicated (see also Materials and Methods). tLPP, transcriptional terminator.

GM-CSF and IFN-α2b are not expressed at detectable levels in pJBphOx-271d in the absence of a signal sequence. In contrast to protein scFv-phOx, GM-CSF and IFN-α2b do not contain disulfide bridges. Consequently, the translocation of these two proteins to the periplasm to obtain correct product folding was not considered necessary, and we therefore chose to express them without the use of a secretion signal sequence. The gene encoding scFv-phOx was therefore substituted with those of GM-CSF and IFN-α2b as NdeI/NotI fragments (Fig. 1), generating plasmids pGM29 and pIFN30, respectively. These plasmids were transformed into the *E. coli* production host strain RV308, and the resulting transformants were initially subjected to production analyses in shake flask cultures. Western blot analyses of both the soluble (S1 and S2) and the pellet (P2) fractions (see Materials and Methods) of the cells showed that no recombinant product could be detected (below 1 mg/liter) from any of the two strains. To rule out the possibility that the poor result is linked to the RV308 host strain used, we also analyzed the production in the protease-modified *E. coli* strain BL21(DE3) (see Materials and Methods). No detectable production was observed in this host either, indicating that the low production levels observed are not a strain-specific property.

The production levels of GM-CSF and scFv-phOx, but not of IFN-α2b, are dramatically stimulated by 5′-terminal fusion of the *pelB* signal sequence. We previously demonstrated that the choice of signal sequence used, *pelB*, *ompA*, or CSP, is important to achieve both effective translocation and high total expression level of scFv-phOx, and the best overall result was obtained with *pelB* (26). Based on the somewhat surprising results obtained with proteins GM-CSF and IFN-α2b (see above), it was of interest to also test the impact of expressing scFv-phOx protein without the use of a secretion signal sequence. We therefore constructed vector pAT65 which has no signal sequence and with the scFv-phOx coding sequence positioned in frame with the relevant translational start codon of the vector (see Materials and Methods). Production analysis of the S1, S2, and P2 fractions of strain RV308(pAT65) showed that no recombinant product (below 1 mg/ml) could be detected. This result showed that scFv-phOx is dependent on a signal sequence to be effectively expressed under these conditions, suggesting that efficient expression of GM-CSF and IFN-α2b might also potentially be obtained by fusing the corre-

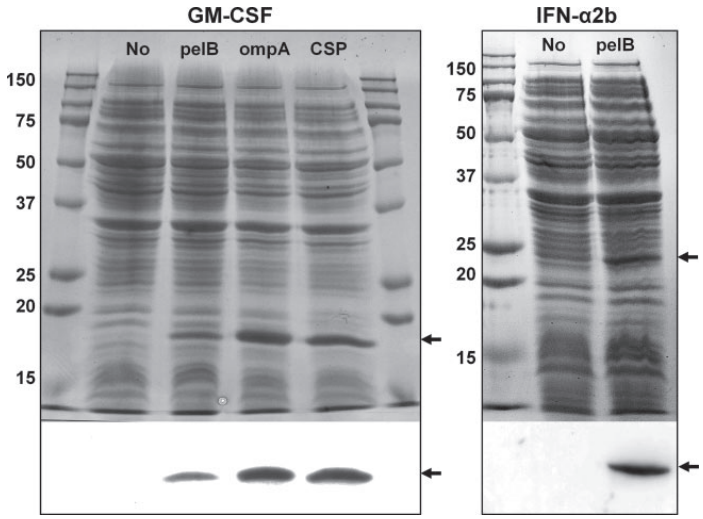

FIG. 2. Images of sodium dodecyl sulfate-polyacrylamide gel electrophoresis (top panels) and the corresponding Western blots (bottom panels) of insoluble fractions of recombinant RV308 strains producing GM-CSF and IFN-α2b proteins, during HCDC. The type of signal sequence used, *pelB*, *ompA*, CSP, or no signal sequence, is indicated in each lane, and the recombinant proteins are indicated with arrows. Numbers on the left side of each image represent molecule sizes in kilodaltons.

sponding gene sequences to the *pelB* sequence. Strains (RV308) containing the resulting plasmids pGM29pelB (expressing GM-CSF with *pelB*) and pIFN30pelB (expressing IFN-α2b with *pelB*) were therefore subjected to production analysis in shake flask cultures. Interestingly, a substantial level (about 50 mg/liter) of recombinant GM-CSF protein was now produced, while no detectable IFN-α2b production could be detected. Here we show that the poor IFN-α2b production observed is partly due to unfavorable codon usage in the IFN-α2b coding region (see below). Under HCDC conditions, high-level production of GM-CSF was achieved (0.8 g/liter) and about 50% of the detected protein was present as soluble product (Fig. 2 and Table 2). This result therefore clearly indicated that efficient GM-CSF expression is (similar to scFv-phOx) dependent on a fused signal sequence to be effectively expressed, under the conditions tested.

TABLE 2. Expression data of GM-CSF-producing strains during HCDC[a]

Plasmid	Signal sequence	OD_{600}[b]	Protein product		mRNA[e]
			Soluble[c]	Insoluble[d]	
pGM29	None	183	BD[f]	BD	1.0× ± 0.1×
pGM29pelB	*pelB*	190	0.39 ± 0.03[g]	0.4 ± 0.05	3.0× ± 0.5×
pGM29CSP	CSP	181	0.80 ± 0.11	0.7 ± 0.15	3.4× ± 0.3×
pGM29ompA	*ompA*	183	0.70 ± 0.09	1.0 ± 0.25	8.0× ± 0.9×

[a] Recombinant cell cultures were grown in the fermentors to an OD_{600} of 100 before inducer was added (see Materials and Methods), and the maximum protein product and corresponding mRNA levels are given.
[b] OD_{600} values obtained in the fermentors at the time of maximum product yields.
[c] Soluble protein was determined by using ELISA as described in Materials and Methods.
[d] Insoluble protein was quantified from dilution series of insoluble fractions subjected to Western blotting as shown in Fig. 2.
[e] The GM-CSF mRNA levels were measured by quantitative PCR as described in Materials and Methods; the mRNA level of cells harboring plasmid pGM29 (no signal sequence) was defined here as 1.0×.
[f] Below detection limit (below 1 mg/ml).
[g] Values for soluble and insoluble proteins are in g/liter.

Exchange of *pelB* with the alternative signal sequences CSP and *ompA* caused up to twofold-improved GM-CSF production levels. To analyze the effect of using different signal sequences on the expression level of GM-CSF, the *pelB* coding region of plasmid pGM29pelB was substituted with the *ompA* and CSP coding regions to yield constructs pGM29ompA and pGM29CSP, respectively. Production analysis during HCDC with cells harboring these two constructs (Fig. 2 and Table 2) demonstrated that the highest total production level (1.7 g/liter) was obtained by using *ompA*. Interestingly, the rankings of signal sequences with respect to the highest total expression level achieved for GM-CSF and scFv-phOx are very different (Fig. 3). As observed with scFv-phOx, the fraction that is soluble relative to the total GM-CSF produced remained relatively constant, irrespective of the total expression level (about 50 to 60%).

Unfavorable codon usage partly contributes to the poor expression of IFN-α2b. The strong stimulatory effects of secretion signal sequences on the expression levels for both scFv-phOx and GM-CSF were not observed with IFN-α2b (see above), and we wanted to identify the reasons for this. First, the respective coding DNA sequences were analyzed and compared (Table 3). All three genes are relatively small (between 378 and 777 nucleotides), and the total numbers of rare codons are also similar (between 10 and 13). We also noticed that the IFN-α2b gene has a low GC content compared to those of both the GM-CSF and scFv-phOx genes, and the possible relevance of this for expression is unknown. To experimentally test the impact of the rare codons on gene expression, we transformed the vectors pIFN30 (no signal sequence) and pIFN30pelB

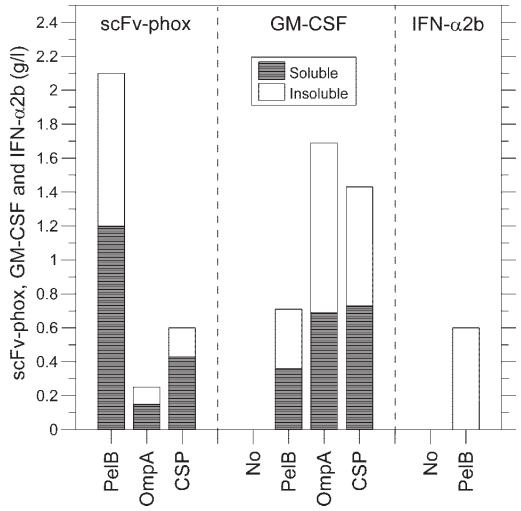

FIG. 3. Production levels of scFv-phOx and GM-CSF as a function of different signal sequences, and no signal sequence, during HCDC. The scFv-phOx production data presented here were imported from our previous study (26). Both soluble and insoluble fractions are given for all recombinant strains tested. Production analysis of scFv-phOx without using a signal sequence resulted in no detectable product (see the text) and was therefore not performed during HCDC. Production of INF-α2b was not tested with signal sequences *ompA* and CSP.

(*pelB* signal sequence) into the *E. coli* strain BL21-CodonPlus (DE3)-RIPL, which can effectively translate the low-usage codons AGG/AGA, AUA, and CUA (see Materials and Methods). The resulting recombinant strains were subjected to shake flask production analyses as described above. Interestingly, substantial IFN-α2b production (about 30 mg/liter) was achieved with cells harboring pIFN30pelB, while low production was still obtained (about 5 mg/liter) with cells harboring pIFN30 (Table 4). These results confirm our assumption that the codon usage of IFN-α2b is one major reason for the poor expression obtained for this protein in strain RV308. They also indicate that expression of this gene, as observed for scFv-phOx and GM-CSF genes, is strongly stimulated by the *pelB* secretion signal sequence under these conditions.

TABLE 3. Comparison of coding sequences of recombinant genes used in this study

Gene product	Gene length (nucleotides)	GC content (%)	No. of rare codons[a]
scFv-phOx	777	55.0	12
GM-CSF	378	57.7	10
IFN-α2b	495	48.4	13
IFN-α2bS	495	53.6	0

[a] Rare codons are as follows. scFv-phOx: arginine, CGA (3 codons), AGA (4 codons), and AGG (1 codon); leucine, CTA (1 codon); isoleucine, ATA (1 codon); and proline, CCC (2 codons). GM-CSF: arginine, CGG (2 codons) and AGA (1 codon); leucine, CTA (1 codon); and proline, CCC (6 codons). IFN-α2b: arginine, AGG (4 codons) and AGA (5 codons); leucine, CTA (1 codon); isoleucine, ATA (1 codon); and proline, CCC (2 codons).

TABLE 4. Production levels of recombinant IFN-α2b protein obtained in shake flasks by using original and synthetic genes and in different genetic backgrounds

Plasmid	Production level (mg/liter)		
	RV308[a]	BL21(DE3)	BL21-CodonPlus (DE3)-RIPL
pIFN30	BD[b]	BD	5 ± 2
pIFN30S	BD		
pIFN30pelB	BD	BD	30 ± 9
pIFN30SpelB	40 ± 12		

[a] The three different *E. coli* host strains used, RV308, BL21(DE3), and BL21-CodonPlus (DE3)-RIPL, are described in Materials and Methods. For IFN-α2b production levels with host RV308 under HCDC, see Fig. 2 and 3. All experiments were run in three parallels, and the standard errors are indicated.
[b] BD, below detection limit (under 1 mg/liter).

Design and construction of a synthetic IFN-α2b gene, designated IFN-α2bS, useful for high-level IFN-α2b production. To optimize the coding sequence of IFN-α2b for high expression in our preferred production host strain RV308 (26), the entire gene sequence was completely redesigned through complete gene synthesis (using computer software; see Materials and Methods). The synthetic gene is 80.4% identical to its parental version and has 99 nucleotide substitutions affecting 77 of its codons (46% of all its codons), still maintaining the original sequence of the protein product. The GC content of the synthetic gene is 53.6% compared to 48.4% of the original gene, and it has no rare codons (Table 3). It was of interest to test if this optimized IFN-α2bS gene could be efficiently expressed without a signal sequence in RV308, and it was therefore used to substitute for IFN-α2b in plasmid pIFN30, yielding plasmid pIFN30S. Interestingly, shake flask experiments with RV308(pIFN30S) showed that the IFN-α2b production was still below the detection level (Table 4). Therefore, we constructed the analogous vector pIFN30SpelB containing *pelB* and a similar analysis of RV308 harboring this plasmid showed that the IFN-α2b protein was then produced at about 40 mg/liter (Table 4). Strains RV308(pIFN30S) and RV308 (pIFN30SpelB) were then analyzed during HCDC, and a total of 0.6 g/liter of IFN-α2b protein was produced with plasmid pIFN30SpelB, while no product was detected with plasmid pIFN30S (Fig. 2 and 3). Interestingly, all recombinant IFN-α2b production was present as IBs. As a control in these HCDC experiments, we also included strain RV308(pIFN30pelB) with the original IFN-α2b gene fused to *pelB*, and the total production level of this strain was found to be low, as expected (about 0.04 g/liter). Thus, the optimized IFN-α2bS DNA coding sequence could be used to achieve a high production level for IFN-α2b protein in *E. coli* RV308, but only provided that a signal sequence is fused to the 5′-terminal end of the coding region.

Quantitative PCR analyses suggest that both the transcript levels and translation efficiencies are higher when using signal sequences. To increase the understanding of how the signal sequences stimulate gene expression, we analyzed a selection of our recombinant strains under relevant production conditions by using quantitative real-time PCR. The rankings of GM-CSF transcript levels in the four different GM-CSF-producing strains tested were similar to the corresponding rankings of total protein production levels (Table 2). Interestingly, the transcript level of RV308(pGM29), which has no signal

sequence and produces no detectable GM-CSF protein, is only about threefold lower than the transcript levels detected when using *pelB* or CSP. This result implies that the low GM-CSF expression level of this strain is largely due to inefficient translation and not poor transcription. In bacteria, transcription and translation are tightly coupled in time and space, and for a given mRNA, high translational activity can protect the mRNA from degradation due to ribosome occupancy, which in turn contributes to more translational activity. We therefore believe that our data indicate that signal sequences can increase both the transcript levels and the translation efficiencies of the genes to which they have been fused.

N-terminal sequencing of recombinant proteins confirmed that the signal sequences are cleaved off in vivo. To make sure that the secretion signal sequences were cleaved off in vivo, the soluble fractions (S1 and S2) of the fermented cultures of recombinant strains RV308(pGM29ompA) and RV308(pGM2 9CSP) were subjected to Western analysis and the desired proteins were excised and subjected to N-terminal sequencing. The results of these experiments confirmed that the recombinant products in both samples are equivalent to the primary sequence of the GM-CSF mature protein (data not shown), implying that the signal sequences and the initial Met residue were correctly cleaved off.

DISCUSSION

Secretion signal sequences are generally used in recombinant gene expression for the purpose of achieving translocation of the protein of interest. It is well-known that there exists no general rule guiding the choice of signal sequence that will maximize the level of translocation of any particular protein, and a trial-and-error type of approach is therefore commonly used (6). Here we demonstrate an unexpected and important role of signal sequences in strongly stimulating the levels of expression for three different proteins of human origin during HCDC in *E. coli*. By selecting the appropriate signal sequences, we achieved total product yields of up to 2.3 g/liter (Fig. 3), which is sufficiently high to be commercially interesting. In gram-positive *Lactobacillus lactis*, it has been showed that protein secretion can be an effective way to increase the overall expression level of several heterologous proteins (14). It has been suggested that this effect is due to lack of proteolysis, but this hypothesis is not yet well confirmed experimentally. We have previously demonstrated that the parental vector pJB658 can be used for high-level expression of different bacterial proteins without the use of any signal sequences (3, 4, 31). Plasmid pJBphOx-271d and its derivatives constructed in this study have retained the region covering *Pm* and *rbs* of pJB658 unmodified, so the observations made here do not seem to be related to the vector system as such. Although more GM-CSF transcript is present when the gene is fused to a signal sequence, the increase is not sufficient to explain the vast difference in the final protein product produced. It therefore seems likely that the proteins are either very inefficiently translated in the absence of a signal sequence or they are immediately degraded. Proteolytic degradation cannot be completely excluded, but tests in a strain commonly used to reduce such problems did not give any indication in support of degradation of the protein product as a major reason for the low production levels.

A

```
      ****** *           *    **         *       *
pelB: AUGAAAUACCUAUUGCCUACGGCAGCCGCUGGAUUGUUAUUACUCGCGGC
ompA: AUGAAAAAAACUGCUAUCGCUAUCGCUGUUGCUCTGGCUGGUUUCGCUAC
CSP:  AUGAAAAAAAAAUUAUUGGCGUUAGCCUUGUUAGCGUUAUUGUUUAACGG

pelB: CCAGCCGGC......
ompA: UGUUGCUCAGGC...
CSP:  CGCGCAGGC......
```

B

```
      **      *
PelB: MKYLLPTAAAGLLLLAAQP.....
OmpA: MKKTAIAIAVALAGFATVAQ....
CSP:  MKKKLLALALLALLFNGAQ.....
```

FIG. 4. Sequence alignments of the signal sequences *pelB*, *ompA*, and CSP. A: mRNA sequences (5′ to 3′). B: Primary sequences. Identical nucleotides and residues in the DNA sequence and primary sequences are indicated with asterisks.

It has been documented that secondary structures in the *rbs* of mRNA can lower translation initiation efficiency and high expression levels can be transferred from an N-terminal fusion partner to a poorly expressing partner as a result of mRNA stabilization (10, 11, 12, 27, 32). Sequence alignments show that *pelB*, *ompA*, and CSP are only 22% and 16% identical at the mRNA and amino acid sequence levels, respectively (Fig. 4). We analyzed the 5′ ends of the different mRNAs, including the initiation codon and its flanking regions, by using the MFOLD software (34). Interestingly, the folding energies of all three genes tested are significantly reduced when fused to a signal sequence (dG values between −1 and −2 kcal) compared to the energy without a signal sequence (dG values between −4 and −10 kcal). As such, these data could possibly partly explain the poor translation of these mRNA molecules when not fused to a signal sequence. However, the calculated binding energy of the optimized INF-α2bS mRNA 5′ end is also low (dG values between +1 and −4 kcal), and the latter result cannot explain why this gene is poorly expressed without being fused to *pelB*. Gene expression is also controlled by the degradation of mRNA, and in bacteria, transcription and translation are coupled processes (27). Possibly, the up to eightfold higher transcript levels accompanying high GM-CSF protein levels were due to a reduced mRNA degradation caused by high translation initiation efficiency and ribosomal protection from mRNA degradation, and not so much due to increased GM-CSF transcription rates.

In contrast with what was observed for scFv-phOx and GM-CSF, the fusion of a signal sequence to the IFN-α2b gene was alone not sufficient to achieve high level expression. It has been previously reported that the expression of this particular gene may be hampered due to many rare codons as well as to secondary structure formation (1, 30, 32). This was confirmed here by the studies with the strain BL21-CodonPlus (DE3)-RIPL, specially designed to deal with such problems, as well as by the use of a redesigned synthetic version of the gene. Interestingly, a fusion signal was still needed in both cases in order to achieve high-level expression of the gene encoding IFN-α2b. It could therefore be concluded that all three tested genes responded similarly to the signal sequence, in spite of their unrelatedness in terms of their primary sequences.

The three signal sequences *ompA* (*E. coli* origin), *pelB* (*Erwinia*

carotovora origin), and CSP (synthetic) target their fusion partners for translocation by the Sec translocase pathway (16). This is a posttranslational pathway where the polypeptide is translocated in an unfolded state that is not dependent on the signal recognition particle (13, 17). CSP was designed by us based on its amino acid sequence (26), and it was not optimized regarding formation of mRNA secondary structures. Alignments showed that these three signal sequences share low sequence identity at both the DNA and the amino acid sequences (Fig. 4). Thus, general sequence similarities in the signal sequences do not seem to explain the observed effects on expression either. Certain point mutations in the *pelB* coding sequence have been shown to affect expression levels of recombinant genes, presumably by affecting the mRNA secondary structures (5). It has been reported that codons immediately downstream of the translation initiation codon can have strong effects on translation initiation efficiency in *E. coli* (21, 25). We noticed that the signal sequences *pelB*, *ompA*, and CSP all possess the AAA triplet in position +2. This particular codon positioned in the +2 codon of the β-galactosidase coding region can cause high-level expression, but this positive effect was highly sensitive to sequence alterations in the upstream *rbs* region (29). Moreover, triplet AAA in position +2 was not optimal for high-level expression of the enterotoxin II protein (25), suggesting that a universal role for this codon (in that particular location) is questionable.

Another very different type of hypothesis would be to assume that the translocation process itself has an effect on the expression rates of the corresponding protein. The Sec translocation apparatus consists of multiple proteins, and the translocation and translation processes are closely coupled (for a review, see reference 17). The targeting of preproteins is governed by the signal sequence, and the preprotein is translocated in an unfolded state. It could therefore be hypothesized that this translocation process somehow contributes to a higher translation rate, but to our knowledge, no experimental support of such a hypothesis has been reported.

ACKNOWLEDGMENTS

This work was financed by Alpharma AS and the Norwegian Research Council.

REFERENCES

1. **Alexandrova, R., M. Eweida, F. Georges, M. B. Dragulev, and I. Ivanov.** 1995. Domains in human interferon alpha-1 gene containing tandems of arginine codons AGG play the role of translational initiators in *E. coli*. Int. J. Biochem. Cell Biol. **27:**469–473.
2. **Belew, M., Y. Zhou, S. Wang, L. E. Nystöm, and J. C. Janson.** 1994. Purification of recombinant human granulocyte-macrophage colony-stimulating factor from the inclusion bodies produced by transformed *Escherichia coli* cells. J. Chromatogr. **679:**67–83.
3. **Blatny, J. M., T. Brautaset, H. C. Winther-Larsen, P. Karunararan, and S. Valla.** 1997. Improved broad-host range RK2 vectors useful for high and low regulated gene expression levels in gram-negative bacteria. Plasmid **38:** 35–51.
4. **Brautaset, T., S. B. Petersen, and S. Valla.** 2000. In vitro determined kinetic properties of mutant phosphoglucomutases and their effect on sugar catabolism in *Escherichia coli*. Metab. Eng. **2:**104–114.
5. **Calvez, H. L., J. M. Green, and D. Baty.** 1996. Increased efficiency of alkaline phosphatase production levels in *Escherichia coli* using a degenerate PelB signal sequence. Gene **170:**51–55.
6. **Choi, J. H., and S. Y. Lee.** 2004. Secretory and extracellular production of recombinant proteins using *Escherichia coli*. Appl. Microbiol. Biotechnol. **64:**625–635.
7. **Emanuel, S. L., and S. Petska.** 1993. Human interferon-αA, -α2, and -α2 (Arg) genes in genomic DNA. J. Biol. Chem. **17:**12565–12569.
8. **Greenberg, R., D. Lundell, and Y. Alroy.** 1988. Expression of biologically active, mature human granulocyte-macrophage colony-stimulating factor with an *E. coli* secretory expression system. Curr. Microbiol. **17:**321–322.
9. **Jakobsen, Ø. M., A. Benichou, M. C. Flickinger, S. Valla, T. E. Ellingsen, and T. Brautaset.** 2006. Upregulated transcription of plasmid and chromosomal RuMP pathway genes is critical for methanol assimilation and methanol tolerance level in *Bacillus methanolicus*. J. Bacteriol. **188:**3063–3072.
10. **Kane, J. F.** 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Curr. Opin. Biotechnol. **6:**494–500.
11. **Kozak, M.** 2005. Regulation of translation via mRNA secondary structure in prokaryotes and eukaryotes. Gene **361:**13–37.
12. **Laursen, B. S., H. P. Sorensen, and K. Mortensen.** 2005. Initiation of protein synthesis in bacteria. Microbiol. Mol. Biol. Rev. **69:**101–123.
13. **Lee, H. C., and H. D. Bernstein.** 2001. The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. Proc. Natl. Acad. Sci. USA **98:**3471–3476.
14. **Le Loir, Y., V. Azevedo, S. C. Oliveira, D. A. Freitas, A. Miyoshi, L. G. Bermúdez-Humarán, S. Nouaille, L. A. Ribeiro, S. Leclercq, J. E. Gabriel, V. D. Guimaraes, M. N. Oliveira, C. Charlier, M. Gautier, and P. Langella.** 2005. Protein secretion in *Lactococcus lactis*: an efficient way to increase the overall heterologous protein production. Microb. Cell Fact. **4:**2.
15. **Makrides, S. C.** 1996. Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol. Rev. **60:**512–538.
16. **Manting, E. H., and A. J. Driessen.** 2000. *Escherichia coli* translocase: the unravelling of a molecular machine. Mol. Microbiol. **37:**226–248.
17. **Mergulhão, F. J. M., D. K. Summers, and G. A. Monteiro.** 2005. Recombinant protein secretion in *Escherichia coli*. Biotechnol. Adv. **23:**177–202.
18. **Metcalf, D.** 1985. The granulocyte-macrophage colony-stimulating factors. Science **229:**16–22.
19. **Metcalf, D.** 1991. Control of granulocytes and macrophages: molecular, cellular, and clinical aspects. Science **254:**529–533.
20. **Neves, F. O., P. L. Ho, I. Raw, C. A. Pereira, C. Moreira, and A. L. T. O. Nascimento.** 2004. Overexpression of a synthetic gene encoding human alpha interferon in *Escherichia coli*. Prot. Expr. Purif. **35:**353–359.
21. **Puri, N., K. B. Appa Rao, S. Menon, A. K. Panda, G. Tiwari, L. C. Garg, and S. M. Totey.** 1999. Effect of the codon following the ATG start site on the expression of bovine growth hormone in *Escherichia coli*. Prot. Expr. Purif. **17:**215–223.
22. **Ramos, J. L., N. Mermod, and K. N. Timmis.** 1987. Regulatory circuits controlling transcription of TOL plasmid operon encoding meta-cleavage pathway for degradation of alkylbenzoates by *Pseudomonas*. Mol. Microbiol. **1:**293–300.
23. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
24. **Schmidt, F. R.** 2004. Recombinant expression systems in the pharmaceutical industry. Appl. Microbiol. Biotechnol. **65:**363–372.
25. **Simmons, L. C., and D. G. Yansura.** 1996. Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. Nat. Biotechnol. **14:**629–634.
26. **Sletta, H., A. Nedal, T. E. V. Aune, H. Hellebust, S. Hakvåg, R. Aune, T. E. Ellingsen, S. Valla, and T. Brautaset.** 2004. Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host-toxic single-chain antibody fragment in *Escherichia coli*. Appl. Environ. Microbiol. **70:**7033–7039.
27. **Sørensen, H. P., and K. K. Mortensen.** 2005. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J. Biotechnol. **115:**113–128.
28. **Srivastava, P., P. Bhattacharaya, G. Pandey, and K. J. Mukherjee.** 2005. Overexpression and purification of recombinant human interferon alpha2b in Escherichia coli. Prot. Expr. Purif. **41:**313–322.
29. **Stenström, C. M., E. Holmgren, and L. A. Isaksson,.** 2001. Cooperative effects by the initiation codon and its flanking regions on translation initiation. Gene **273:**259–265.
30. **Valente, C. A., D. M. F. Prazeres, J. M. S. Cabral, and G. A. Monteiro.** 2004. Translational features of human alpha 2b interferon production in *Escherichia coli*. Appl. Environ. Microbiol. **70:**5033–5036.
31. **Winther-Larsen, H. C., K. D. Josefson, T. Brautaset, and S. Valla.** 2000. Parameters affecting gene expression from the *Pm* promoter in gram-negative bacteria. Metab. Eng. **2:**79–91.
32. **Wu, X., H. Jörnvall, K. D. Berndt, and U. Oppermann.** 2004. Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance. Biochem. Biophys. Res. Commun. **313:**89–96.
33. **Zhang, X. W., T. Sun, D. X. Gu, and Z. Q. Tang.** 1999. Production of granulocyte-macrophage colony-stimulating factor (GM-CSF) by high cell density fermentation of secretory recombination *Escherichia coli*. Process Biochem. **34:**55–58.
34. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:**3406–3415.

Paper III

Paper III is not included due to copyright.

Paper IV

Paper IV is not included due to copyright.