



Production of recombinant proteins in *Escherichia coli*

Wolfgang Schumann¹ and Luis Carlos S. Ferreira²

¹University of Bayreuth, Institute of Genetics, Bayreuth, Germany.

²Universidade de São Paulo, Instituto de Ciências Biomédicas, Departamento de Microbiologia, São Paulo, SP, Brazil.

Abstract

Attempts to obtain a recombinant protein using prokaryotic expression systems can go from a rewarding and rather fast procedure to a frustrating time-consuming experience. In most cases production of heterologous proteins in *Escherichia coli* K12 strains has remained an empirical exercise in which different systems are tested without a careful insight into the various factors affecting adequate expression of the encoded protein. The present review will deal with *E. coli* as protein factory and will cover some of the aspects related to transcriptional and translational expression signals, factors affecting protein stability and solubility and targeting of proteins to different cell compartments. Based on the knowledge accumulated over the last decade, we believe that the rate of success for those dedicated to expression of recombinant proteins based on the use *E. coli* strains can still be significantly improved.

Key words: expression vectors, secretion, molecular chaperones.

Received: January 16, 2004; Accepted: March 5, 2004.

Introduction

High-level production of recombinant proteins as a prerequisite for subsequent purification has become a standard technique. Important applications of recombinant proteins are: (1) immunization, (2) biochemical studies, (3) three-dimensional analysis of the protein, and (4) biotechnological and therapeutic use. Production of recombinant proteins involves cloning of the appropriate gene into an expression vector under the control of an inducible promoter. But efficient expression of the recombinant gene depends on a variety of factors such as optimal expression signals (both at the level of transcription and translation), correct protein folding and cell growth characteristics. Display of recombinant proteins on the bacterial surface has many potential biotechnological applications and requires further knowledge on targeting motifs present on carrier proteins usually used as fusion partners. In addition, the selection of a particular expression system requires a cost breakdown in terms of design, process and other economic considerations. The relative merits of bacterial, yeast, insect and mammalian expression systems have been reviewed in Marino (1989).

This review article deals exclusively with *Escherichia coli* cells as a protein factory. Despite its extensive

knowledge on genetics and molecular biology, there is no a priori guarantee that every gene can be expressed efficiently in this Gram-negative bacterium. Factors influencing the expression level include unique and subtle structural features of the gene sequence, the stability and efficiency of mRNA, correct and efficient protein folding, codon usage, degradation of the recombinant protein by ATP-dependent proteases and toxicity of the protein. The objectives of this article are to review the potential influence of these different parameters on the yield of recombinant proteins and to provide the reader with practical suggestions allowing optimization of recombinant protein production and targeting to different compartments of the bacterial cell. For earlier reviews on high-level of gene expression in *E. coli* see Makrides (1996) and Swartz (2001).

DNA sequences involved in transcription

Three different DNA sequences and one multi-component protein are involved in transcription of genes: (1) the promoter, (2) the transcriptional terminator, (3) the regulatory sequence, and (4) the RNA polymerase. The RNA polymerase consists of five different components termed α , β , β' , ω and σ . While $\alpha_2\beta\beta'\omega$ constitute the core enzyme, addition of σ conferring promoter specificity makes up the holoenzyme. The N-terminal part of α is involved in dimer formation and binding to β and β' , and its C terminus, tethered through a flexible linker to its N termi-

nus, is responsible for interaction with the UP element present upstream of some promoters (see below) or with some transcriptional activators. The β subunit binds the rNTPs, contains the catalytic domain and is the target for the antibiotic rifampicin while β' allows unspecific binding to DNA. The role of ω is largely unknown but it is assumed to play a role in RNA polymerase assembly. While all bacterial species analyzed so far contain only one gene each coding for the components of the core enzyme, most species possess genes encoding multiple σ factors. One of these factors functions as the primary or housekeeping σ factor and is involved in the transcription of all those genes needed for growth during the vegetative phase. The additional σ factors are called secondary or alternative σ factors and are needed only under specific growth conditions (Gruber and Gross, 2003). *E. coli* codes for six alternative factors where σ^{32} is needed after a sudden temperature upshift and σ^S replaces the housekeeping σ factor σ^{70} during the stationary phase. So far, only σ^{70} is used in the production of recombinant proteins.

As mentioned above, the σ factor is responsible for the recognition of the promoter, and it follows that each σ factor recognizes a different promoter. Promoters normally consist of three regions called the -35 and the -10 box and the spacer region separating both boxes. Alignment of many promoters allows the deduction of a so-called consensus sequence, and the consensus sequence for σ^{70} is TTGACA - N₁₇ - TATAAT. This sequence represents the optimal promoter sequence with a spacer region of 17 nucleotides. It should be mentioned that there is not a single promoter present on the *E. coli* chromosome identical to the consensus sequence. In most cases, there are one or two deviations in both the -35 and the -10 box. In addition, some promoters contain a fourth region, the UP element located upstream of the -35 box. The UP element consists of an AT-rich sequence allowing interaction with the C-terminal domain of the α subunit thereby increasing the promoter strength. It functions as an independent promoter module, and when fused to other promoters such as *lacUV5*, it stimulates transcription (Rao *et al.*, 1994). None of the promoters directing the production of recombinant proteins makes use of the UP element.

Besides the promoter, a transcriptional terminator is required to allow termination of transcription. Two classes of terminators have been described, factor-independent and -dependent terminators. The first class consists of an inverted repeat followed by several A residues on the template DNA strand. When the RNA polymerase has transcribed the inverted repeat, it folds immediately into a stem-loop structure at the level of mRNA to cause pausing of the enzyme. Since the stem-loop structure is followed by several U residues which make a weak interaction with the A residues on the template DNA, dissociation of the enzyme results. But no terminator will result in the dissociation

of each RNA polymerase molecule resulting in readthrough-transcription into the neighboring gene(s). To reduce this read-through, often two different transcriptional terminators are placed in tandem on the expression vectors. Particularly effective are the two tandem transcription terminators T1 and T2, derived from the *rrnB* rRNA operon of *E. coli* (Brosius *et al.*, 1981). Protein-dependent terminators have a more complex organization and some mechanistic aspects are still not fully understood. So far, Rho factor-dependent terminators have not been used in any expression system aimed at producing of recombinant proteins in *E. coli* strains and will not be discussed here.

Genes are either expressed constitutively or regulated. Two different classes of regulators have been described, transcriptional repressors and transcriptional activators. Repressors bind to operators located either within the promoter region or immediately downstream from it and, in most cases, prevent RNA polymerase-promoter binding or act as a road-block. To relieve repression, the repressor has to dissociate from its operator. In some cases, an inducer will be either synthesized by the cell or taken up from the environment which binds to the repressor causing dissociation from its operator. The LacI repressor is the best studied example and will be discussed below. Another class of repressors need a corepressor to bind to the cognate operator. As long as high amounts of corepressor are present in the cell, repression is exerted. If the corepressor is being used up by the cells, the repressor fails to bind to its operator. The TrpR repressor and its corepressor tryptophan are the most prominent examples. A third, though artificial possibility, are temperature-sensitive repressors. These repressor alleles are isolated after mutagenesis of the repressor gene and cause an amino acid replacement leading to the synthesis of a temperature-sensitive protein. At low temperatures (30-32 °C), the repressor is active and binds to its operator. When cells are shifted to high temperatures (40-42 °C), the repressor alters its conformation and dissociates from its operator. This principle is used with the cI repressor of bacteriophage λ .

Transcriptional activators in general bind upstream from the promoter to a sequence designated upstream activating sequence (UAS). By binding to the UAS, the activator increases the possibility of the RNA polymerase to bind to its promoter and further activates transcription initiation by interaction with one of the subunits, in most cases the α or the σ subunit. No expression system has been described using a transcriptional activator.

DNA sequences involved in translation

Due to the complexity of the process the determinants of protein synthesis initiation have been difficult to decipher. It became clear that the wide range of efficiencies in translation of different mRNAs is predominantly due to the structure at the 5' end of each mRNA species. Therefore, no universal sequence for the efficient initiation of translation

has been devised. The translation initiation region comprises four different sequences: (1) the Shine-Dalgarno sequence, (2) the start codon, (3) the spacer region between the Shine-Dalgarno sequence and the start codon, and (4) sometimes translational enhancers.

Shine and Dalgarno identified a sequence in the ribosome-binding sites (RBS) of bacteriophage mRNAs and suggested that this region interacts with the complementary 3' end of the 16S rRNA during translation initiation (Shine and Dalgarno, 1974). In *E. coli*, the initiation codon AUG is used predominantly (91%) followed by GUG (8%) and UUG (1%) (Gualerzi and Pon, 1990). This preference coincides with the translational efficiency where AUG dominates (Vellanoweth and Rabinowitz, 1992). The spacing between the Shine-Dalgarno sequence and the initiation codon varies from 5 to 13 nucleotides and influences the efficiency of translation, too (Gold, 1988). Extensive studies have been carried out to determine the optimal nucleotide sequence of the translation initiation region and led to the following results: (1) The Shine-Dalgarno sequence UAAGGAGG enables 3- to 6-fold higher protein production than AAGGA for every spacing; (2) the optimal spacing for UAAGGAGG has been determined to be 4 to 8 nucleotides and 5 to 7 for AAGGA (Rinquist *et al.*, 1992).

Furthermore, the secondary structure at the translation initiation region of the mRNA plays an important role in the efficiency of gene expression. It has been shown that occlusion of the Shine-Dalgarno sequence and/or the start codon by a stem-loop structure prevents accessibility to the 30S ribosomal subunit and inhibits translation (Ramesh *et al.*, 1994). There are two reported cases where this principle is used to significantly reduce translation of the downstream reading frame namely the *rpoH* mRNA coding for the heat shock sigma factor σ^{32} in *E. coli* and mRNAs coding for small heat shock proteins in rhizobiae (Morita *et al.*, 1999; Nocker *et al.*, 2001). In both cases, translation of these mRNAs is achieved under heat shock conditions leading to the melting of the secondary structure. There are possibilities to minimize mRNA secondary structure in the region of translation initiation. While the enrichment of the RBS with adenine and thymine residues enhanced expression of certain genes (Chen *et al.*, 1994), the mutation of specific nucleotides up- or downstream from the Shine-Dalgarno sequence suppressed the formation of mRNA secondary structures and enhanced the translation efficiency (Coleman *et al.*, 1985; Gross *et al.*, 1990). Sequences have been identified that markedly enhance the expression of recombinant genes, and these modules have been called translational enhancers. One example is an U-rich region immediately upstream of the Shine-Dalgarno sequence in the *E. coli atpE* gene (McCarthy *et al.*, 1985). This 30-base sequence has been successfully used to overexpress the human interleukin-2 and interferon beta genes (McCarthy *et al.*, 1986).

Protein quality control: molecular chaperones and ATP-dependent proteases

Proteins contain within their complete amino acid sequence all of the information necessary for attaining their functional three-dimensional structure. But all newly synthesized proteins face challenges in reaching their native state within the crowded environment of the cell. While some domains of a nascent chain might be capable of folding spontaneously, the folded structure cannot be obtained until the entire domain is synthesized. This time lag increases the chance that hydrophobic sequences normally buried in the interior of the protein will become exposed, resulting in protein aggregation. About 40 amino acids of the nascent chain are protected from the cytosol by the ribosome exit tunnel. When the chain leaves the tunnel, molecular chaperones bind preventing aggregation. Molecular chaperones are ubiquitous and highly conserved proteins that help other polypeptides to reach their native conformation without becoming part of the final structure. They are not true folding catalysts, since they do not accelerate folding rates. Instead, they prevent off-pathway aggregation reactions by transiently binding hydrophobic domains in partially folded or unfolded polypeptides collectively designated as non-native proteins.

For the vast majority of polypeptides, folding is a spontaneous process directed by the amino acid sequence and the solvent conditions. Yet, even though the native state is thermodynamically favored, the time-scale for folding can vary from milliseconds to days. While protein folding in the absence of kinetic barriers is extremely fast, such barriers which include disulfide bond formation, *cis/trans* isomerization of the polypeptide chain around proline peptide bonds, preprotein processing, and the ligation of prosthetic groups can significantly delay correct folding of proteins. The presence of kinetic barriers results in the accumulation of partially folded species, or folded intermediates, that contain exposed hydrophobic 'sticky' surfaces which promote self-association (Wetzel, 1994; Georgiou *et al.*, 1994). The self-association of folding intermediates is the basis for protein aggregation *in vitro* and for the formation of inclusion bodies. Aggregation can occur during *de novo* folding or as a consequence of unfolding of native proteins induced by heat shock and other types of stress. Cells have evolved an elaborate protein quality control system which consists of molecular chaperones and ATP-dependent proteases acting together to prevent aggregation, assist refolding and degrade misfolded polypeptides (Gottesman *et al.*, 1997).

Molecular chaperones are divided into two distinct classes, folder and holder chaperones. Both classes of chaperones interact with non-native polypeptide chains through exposed hydrophobic surfaces, and while folder chaperones mediate their refolding in an ATP-dependent process, holder chaperones bind non-native proteins and prevent their aggregation. Protein aggregation is frequently

observed upon synthesis of recombinant proteins in *E. coli* which can lead to the formation of insoluble inclusion bodies.

In the cytoplasm of *E. coli* cells (and other bacterial species), there are two multi-component chaperone complexes with broad specificity. The first comprises the 60 kDa heat shock protein GroEL (60 kDa) and the smaller accessory protein GroES (10 kDa). GroEL forms a characteristic doublet of heptameric rings which, during the catalytic cycle, associate one or two heptameric rings of GroES. The GroEL chaperone has a very broad specificity and is essential for viability. The second complex comprises DnaK and the two cochaperones DnaJ and GrpE (the KJE complex). Nascent polypeptide chains are most probably recognized and bound by DnaK. Details of the reaction pathways of these two chaperone systems can be found in an excellent review article (Bukau and Horwich, 1998).

There are many examples that overexpression of molecular chaperones in *E. coli* can facilitate the assembly of heterologous proteins. A systematic investigation of the effects of growth conditions and chaperone co-expression on recombinant protein solubility using a β -galactosidase fusion as a model has recently been completed (Thomas and Baneyx, 1996). GroESL co-expression was found to increase protein expression at 30 °C, but not at 37 or 42 °C; the KJE complex conferred a more substantial increase in the expression of soluble proteins at all temperatures tested. Addition of 3% ethanol was shown to have a synergistic effect with chaperone co-expression and led to the production of protein that was nearly all soluble. For any given recombinant protein, only the chaperone that interacts productively with an aggregation-prone folding intermediate will have a beneficial effect on the production of native protein. Unfortunately, the current substrate-chaperone match has to be found by trial and error.

Two important holder chaperones are the trigger factor and the small heat shock proteins IbpA and IbpB (inclusion body binding proteins A and B). The trigger factor occurs at about 20,000 copies per exponentially growing cell, and is found at the exit tunnel of the ribosomes where it binds to virtually all nascent polypeptide chains to prevent their premature folding. In addition to its holder chaperone activity, it acts as a peptidyl-prolyl *cis/trans* isomerase (PPIase). These enzymes catalyse the interconversion between *cis* and *trans* forms of the peptide bond preceding proline residues. While polypeptide chains are synthesized with the peptide bonds in the *cis* form, about 5% of these are converted into the *trans* form by PPIases. Besides the trigger factor additional PPIases are present within the cytoplasm and the periplasm (Missiakos and Raina, 1997).

ATP-dependent proteases recognize non-native proteins in the cytoplasm and degrade them to peptides of a length of about ten amino acid residues. The current model for proteolytic degradation involves three steps: (1) Recog-

niton. The protease selects a protein for degradation, either because it has an accessible tag located at the N- or C-terminus or because an internal degradation signal has become exposed. (2) Translocation. ATP-hydrolysis promotes both unfolding and translocation into the proteolytic chamber (dual role of ATP). (3) Proteolysis. Proteins are hydrolysed to small peptides which are released from the chamber into the cytoplasm. Five different ATP-dependent proteases have been identified in *E. coli* (Lon, ClpAP, ClpXP, ClpYQ and FtsH where only FtsH is essential) which all form ring-like structures.

DNA sequences involved in translocation of proteins into the periplasm

Proteins present in the cytoplasm are present in the reduced form and do not contain disulfide bonds. There are three reasons to keep proteins in the reduced form: (1) a number of enzymes rely on a reduced cysteine residue in their active site (*e.g.*, ribonuclease reductase, methionine sulfoxide reductase), (2) most proteins present in the periplasm are translocated in an unfolded conformation, and (3) a number of virulence factors and toxins contain multiple disulfide bonds.

How is the formation of disulfide bonds prevented in the cytoplasm? An extreme reducing environment of the cytoplasm, maintained by one or more systems (thioredoxin/thioredoxin reductase, glutathione/glutathione reductase, glutaredoxin/glutaredoxin reductase) and enzymes catalyzing disulfide bonds are absent in the cytoplasm. The periplasm contains several enzymes involved in the formation of disulfide bonds which are grouped into two pathways, the oxidation and the isomerization pathway. In the oxidation pathway, DsbA with two oxidized thiol groups transfers its disulfide to pairs of cysteines in substrate proteins by a thiol-disulfide exchange reaction and becomes reduced. To get oxidized again, it interacts with DsbB, an integral membrane protein which contains two disulfide bonds. The electrons are then transferred during aerobic growth conditions via ubiquinone and cytochrome oxidases to O₂ and during anaerobic growth via menaquinone to anaerobic electron acceptors such as fumarate or nitrate. If the target protein contains more than two thiol groups, DsbA may form a wrong disulfide bond. This is recognized by the isomerization system which consists of three proteins. The reduced forms of DsbC and DsbG can recognize wrongly formed disulfide bonds on target proteins and catalyze the formation of the correct bonds thereby becoming oxidized. Reduction of the disulfide bonds occurs through interacting with the integral membrane protein DsbD which in turn becomes reduced again through interaction with thioredoxin (Hiniker and Bardwell, 2003). Release of the recombinant proteins from the periplasm occurs by osmotic shock.

There are two different systems involved in the translocation of proteins through the inner membrane, the Sec

and the Tat pathway. Both systems differ in both the components facilitating the translocation step and the conformation of the substrate protein. With both systems, proteins to be translocated contain a signal-sequence at their N-terminal end. This signal-sequence has a length of 15-30 amino acid residues and is composed of three different regions termed N, H and C domain. The N domain contains three or four positively charged amino acid residues, the H domain a hydrophobic core and the C domain the type I signal peptidase cleavage site A-X-A, where cleavage occurs after the second A residue (see below). The Tat-type signal sequences are identical in their composition, but contain two consecutive arginine residues (RR) within the N domain which led to the designation of this pathway (Tat stands for *t*wi*n*-*a*rgi*n*i*n*e *t*ransport). Besides the signal-peptide present on the protein to be translocated several other proteins are involved in the translocation process. In the case of the Sec pathway, these are SecA and SecYEG.

To become secreted by the Sec pathway, proteins have to be maintained in an export-competent state. There are several possibilities to reach this goal: (1) The protein may be translocated across the cytoplasmic membrane simultaneously with translation. This process is called cotranslational secretion and is aided by the signal recognition particle (SRP). The prokaryotic SRP is composed of one protein (Ffh) and a 4.5S RNA and seems to recognize signal sequences with an apparent hydrophobicity that is greater than the hydrophobicity of the average signal sequence (see below). (2) Proteins which are exported posttranslationally are prevented from folding in the cytoplasm by molecular chaperones. Here, SecB, active as a homotetramer binding to nascent polypeptide chains when they emerge from the ribosomes, has been identified as the most prominent antifolding factor. (3) In some cases, the signal sequence can act as an intrapeptide chaperone to prevent rapid folding (Liu *et al.*, 1989). In all these cases, the polypeptide interacts with SecA, a homodimer, binding first to the signal peptide. Next, the SecA-polypeptide complex interacts with SecYEG which forms a pore within the inner membrane called translocon. SecA catalyzes translocation of the polypeptide chain through the translocon in a step-wise manner, and this process is driven by the hydrolysis of ATP. About 2.5 kDa of the preprotein is translocated per step. In contrast, the Tat pathway accepts only folded proteins and details of the secretion process are elusive.

DNA sequences involved in surface display of proteins

Surface display of heterologous peptides on Gram-negative bacteria may be advantageous for specific situations such as the development of live-bacterial vaccine delivery systems (Georgiou *et al.*, 1997; Lee *et al.*, 2000), generation of whole-cell biocatalysts by immobilization of enzymes for environmental or biotechnological purposes

(Dhillon *et al.*, 1999; Kim *et al.*, 2000), and expression of ligand binding peptides as an approach for generating new diagnostic tools or as biosensors (Daugherty *et al.*, 1998; Westerlund-Wikstrom *et al.*, 1997). Expression of peptides on the surface of Gram-negative bacterial species, such as *E. coli*, has been achieved mainly by the genetic fusion of the heterologous protein with anchoring motifs present on carrier proteins found in high numbers at the outer surface of the bacterial cell envelope, as outer membrane proteins and subunit components of fimbriae and flagella. The carrier protein should supply all information for the efficient translocation and membrane anchoring of the fusion peptide. Moreover, choosing of the appropriate carrier and fusion strategy are of particular relevance for maintenance of native conformation and biological function of the recombinant peptide.

Outer membrane proteins usually consist of a series of membrane-spanning β -sheets connected by amino acid loops facing either the periplasmic space or the outer environment. Targeting sequences of outer membrane proteins are usually located at the N-terminal end, and expression of recombinant peptides may be attained either by sandwich fusion at internal surface-exposed loops or by terminal fusion at the C-terminal end of the carrier protein (Hofnung, 1991). The expression system based on the fusion of the signal sequence and the first nine N-terminal amino acids of Braun's lipoprotein (Lpp), and five transmembrane segments of the outer membrane protein A (OmpA), supplying the adequate targeting and anchoring signals, have been successfully used to expose heterologous proteins on the surface of *E. coli* cells (Stathopoulos *et al.*, 1996). Diverse proteins, including β -lactamase, bacterial endoglucanases, organophosphorous hydrolase, green fluorescent protein and scFv antibodies, have been successfully expressed in active forms on the surface of bacterial cells using the Lpp-OmpA expression system (Stathopoulos *et al.*, 1996; Francisco *et al.*, 1993; Georgious *et al.*, 1996). Peptides can also be inserted within permissive sites of outer membrane proteins such as LamB, PhoE and OmpC, and displayed on the cell surface (Hofnung, 1991; Agterberg *et al.*, 1990; Xu and Lee, 1999). Nonetheless, conformational constraints affecting correct localization and stability of the chimeric protein restricts the size of inserted peptides to a maximum of approximately 100 residues.

Bacterial flagella are composed of a single structural subunit, flagellin, with a surface-exposed hypervariable domain located at the central region of the protein where heterologous peptides can be inserted without affecting flagellar structure and motility (He *et al.*, 1994). The remarkable immunological properties of flagellin and the possibility of expressing heterologous peptides in a polymeric form render the flagellin expression fusion system especially suited for the development of vaccines against pathogenic microorganisms (Newton *et al.*, 1991; Gewirtz *et al.*, 2001; McSorley *et al.*, 2002). Export of flagellin sub-

units is mediated by the type III export pathway, and each subunit diffuses along a narrow channel of the growing flagellum to assemble at the distal end (Macnab, 2003). Display of peptides genetically fused to flagellin can be attained after introduction of heterologous sequences into a cloned flagellin gene expressed in bacterial strains devoid of a chromosomally-encoded structural subunit but proficient in all other genes required for flagellar expression, processing and assembly. One particularly interesting expression system based on *E. coli* flagellin relies on the insertion of thioredoxin into a central hypervariable surface-exposed flagellin domain (Lu *et al.*, 1995). Thioredoxin represents by itself a versatile scaffold for display of fused peptides at conformations compatible with binding to other peptides and fusion with flagellin targets the hybrid protein to the cell surface. Based on this approach, peptides bound by monoclonal antibodies have been precisely identified from expressed random peptide libraries (Tripp *et al.*, 2001).

Expression systems for *E. coli*

Tight expression of transcription of recombinant genes is often desirable or necessary since leaky expression can be detrimental or even lethal to cell growth. Regulated gene expression requires an inducible or repressible system, and therefore, all expression systems are based on controllable promoters. Promoters allowing constitutive expression turned out not to be adequate for the production of recombinant proteins due to two main reasons: First, they do not allow the production of toxic proteins and second, even non-toxic proteins produced at physiological concentrations can be deleterious to the cells when produced at higher levels. One prominent example are integral membrane proteins which, when overproduced, cause jamming of the inner membrane leading to cell death. Four regulatable promoter systems are widely used, where three are based on the repressors already mentioned (LacI, TrpR and phage λ cI) and the fourth on a phage RNA polymerase.

The *lac* system consists of the promoter/operator region preceding the *lac* operon and the LacI repressor encoded by the *lacI* gene. In the absence of an inducer, the Lac repressor binds to its operator situated immediately downstream from the promoter as a homotetramer. The wild-type *lac* promoter sequence is presented in Table 1 and contains one deviation in the -35 and two in the -10 box, and the spacer region encompasses 18 nucleotides if compared to the consensus sequence. One of the many promoter mutations isolated has been termed *lacUV5*. If its DNA sequence is compared to that of the wild-type promoter, it becomes apparent that two nucleotides have been exchanged resulting in the consensus -10 box (Table 1). The promoter strength of *lacUV5* has increased 2.5-fold, and mutations increasing the promoter strength are called promoter-up mutations in general. The promoter of the *trp* operon exhibits the consensus -35 box and the optimal spacer length, but

three deviations within the -10 box (Table 1). Based on the *lacUV5* and the *trp* promoters, an artificial promoter was constructed exhibiting the consensus sequence of σ^{70} -dependent promoters and termed P_{tac} (from *trp* and *lac*; Table 1) (de Boer *et al.*, 1983).

How are the LacI and TrpR repressors inactivated to initiate expression of the recombinant genes? In the case of the P_{lac} , the P_{lacUV5} and the P_{tac} promoters, the repressor is inactivated by addition of isopropyl- β -D-thiogalactopyranoside (IPTG). This compound binds to the active LacI repressor and causes dissociation from its operator. IPTG has two advantages over lactose: First, its uptake is not dependent on the Lac permease (it diffuses through the inner membrane) and second, it cannot be cleaved by β -galactosidase preventing turn-off of transcription. The *lacI* gene is either part of the expression plasmid or it is present within the chromosome. Since the wild-type level of the LacI repressor is not sufficient to repress expression of the recombinant gene in the absence of IPTG, two derivatives have been isolated resulting in an increase in the amount of repressor based on promoter-up mutations called *lacI^q* and *lacI^{q1}* (Müller-Hill *et al.*, 1968; Glascock and Weickert, 1998). The sequence of the three promoters is given in Table 2 for comparisons. Expression systems based on the *trp* system make use of synthetic media with a defined tryptophan concentration. The concentration is chosen in such a way that the system becomes self-inducible when

Table 1 - DNA sequences of promoters used in expression vectors recognized by the housekeeping sigma factor σ^{70} .

Promoter	-35 region	spacer	-10 region
P_{lac}	TTtACA	18 bp	TATgtT
P_{lacUV5}	TTtACA	18 bp	TATAAT
P_{trp}	TTGACA	17 bp	TtaAcT
P_{tac}	TTGACA	17 bp	TATAAT
λP_L	TTGACA	17 bp	gATAcT
λP_R	TTGACt	17 bp	gATAAT
Consensus	TTGACA	17 bp	TATAAT

Nucleotides present in the consensus sequence are shown in capital letters, those not present in the consensus sequence in small letters.

Table 2 - DNA sequences of the wild-type *lacI* promoter and of two different promoter-up mutations

Promoter	-35 region	spacer	-10 region
P_{lacI}	gcGcaA	17 bp	cATgAT
P_{lacI^q}	gTGcaA	17 bp	cATgAT
$P_{lacI^{q1}}$	TTGACA	18 bp	cATgAT
Consensus	TTGACA	17 bp	TATAAT

Nucleotides present in the consensus sequence are shown in capital letters, those not present in the consensus sequence in small letters.

the tryptophan concentration within the cells falls below a threshold level (Masuda *et al.*, 1996). Additionally, 3- β -indole-acrylic acid can be added which inactivates the TrpR repressor (Rose and Yanofsky, 1974) and inhibits charging of tRNA^{trp} by tryptophanyl-tRNA synthetase (Doolittle and Yanofsky, 1968).

The third system makes use of the bacteriophage λ repressor cI. This repressor is synthesized from the λ prophage and prevents expression of all the lytic genes by interacting with two operators termed O_L and O_R. These two operators overlap with two strong promoters, P_L and P_R, respectively (see Table 1), and as long as the cI repressor is bound to its two operators, binding of RNA polymerase is prevented. Expression vectors carry the cI repressor gene and either P_LO_L or P_RO_R. How can the λ expression system be induced? The wild-type cI repressor protein can be inactivated by UV-irradiation or treatment of the cells by mitomycin C. A more convenient way is the application of a temperature-sensitive version of the cI repressor called *cI857*. Therefore, *E. coli* cells carrying a λ -based expression system are grown to mid-exponential phase at low temperature and then transferred to high temperature to induce expression of the recombinant gene (Elvin *et al.*, 1990).

The most widely applied expression system makes use of the phage T7 RNA polymerase which recognizes only promoters found on the T7 DNA, and not promoters present on the *E. coli* chromosome. Therefore, the expression vectors contain one of the T7 promoters (normally the promoter present in front of gene 10) to which the recombinant gene will be fused. The gene coding for the T7 RNA polymerase is either present on the expression vector itself or on a second compatible plasmid or integrated into the *E. coli* chromosome. In all three cases, the gene is fused to an inducible promoter allowing its transcription and translation during the expression phase. The T7 RNA polymerase offers three advantages over the *E. coli* enzyme: First, it consists of only one subunit, second it exerts a higher processivity, and third it is insensitive towards rifampicin. The latter characteristic can be used especially to enhance the amount of recombinant protein by adding this antibiotic about 10 min after induction of the gene coding for the T7 RNA polymerase. During that time, enough polymerase has been synthesized to allow high-level expression of the recombinant gene, and inhibition of the *E. coli* enzyme prevents further expression of all the other genes present on both the plasmid and the chromosome. Since all promoter systems are leaky, low-level expression of the gene coding for T7 RNA polymerase may be deleterious to the cell in those cases where the recombinant gene codes for a toxic protein. These polymerase molecules present during the growth phase can be inhibited by expressing the T7-encoded gene for lysozyme. This enzyme is a bifunctional protein that cuts a bond in the cell wall of *E. coli* and selectively inhibits the T7 RNA polymerase by binding to it, a

feed-back mechanism that ensures a controlled burst of transcription during T7 infection (Studier, 1991).

Another expression system not widely used so far is induced upon a cold shock. When a mid-exponential phase culture of *E. coli* is rapidly transferred from 37 °C to the 10-15 °C temperature range, the synthesis of most cellular proteins significantly decreases, while that of about 15 cold-shock proteins is transiently upregulated (Jones *et al.*, 1987). CspA, the major cold-shock protein, is virtually undetectable at 37 °C, but more than 10% of the total protein synthesis is devoted to its production 1 h following the temperature downshift (Goldstein *et al.*, 1990). The *cspA* mRNA is transcribed with a 150 nucleotide-long 5' untranslated region that confers high instability to the transcript at 37 °C ($t_{1/2} \sim 10$ s) (Brandi *et al.*, 1996; Goldenberg *et al.*, 1996), but the transcript stability increases by two orders of magnitude upon transfer of the cells to 15-10 °C (Jiang *et al.*, 1993; Brandi *et al.*, 1996). A vector has been constructed based on the *cspA* promoter followed by its untranslated region to express recombinant proteins at low temperatures (Mujacic *et al.*, 1999). Very recently, it could be shown that while the growth rate of an *E. coli* strain dropped rapidly as incubation temperatures decreased to 20 °C, addition of the *groESL* operon of *Oleispira antarctica*, isolated from Antarctic seawater, allowed 3-fold faster growth at 15 °C and an even 36-fold faster at 10 °C (Ferrer *et al.*, 2003). These authors could also show that both molecular chaperones exhibited high protein folding activities *in vitro* at temperatures of 4-12 °C. This result suggests that such an engineered *E. coli* strain could produce high amounts of correctly folded recombinant protein at low temperatures.

Cytoplasmic or periplasmic localization of the recombinant protein?

There are four reasons to translocate recombinant proteins into the periplasm: (1) the oxidizing environment facilitates the formation of disulfide bonds, (2) it contains only 4% of the total cell protein (~100 different proteins), (3) there is less protein degradation, and (4) easy purification by osmotic shock. Formation of disulfide bonds also occurs spontaneously after purification of the protein. There is now an *E. coli* strain available where disulfide bonds are formed within the cytoplasm. This strain called Origami contains four mutations: knock-outs of the genes coding for thioredoxin and glutathione reductase, a third allows cytoplasmic expression of the DsbC isomerase and the fourth is within a so far uncharacterized suppressor gene allowing improved growth of this strain (Bessette *et al.*, 1999).

To translocate recombinant proteins through the inner membrane, any signal sequence can be fused to the protein of interest. But two classes of proteins may pose severe problems to be secreted. These are proteins with extended hydrophobic regions which will be captured within the

membrane. A solution to this problem may be to secrete them using the Tat pathway. The other class of proteins are those which fold too rapidly within the cytoplasm. These proteins may be also secreted in their folded form using a Tat signal sequence, or, alternatively, fused to the signal sequence of the DsbA oxidoreductase. This signal sequence directs the nascent polypeptide chain to the SRP export pathway which is largely cotranslational (Schierle *et al.*, 2003). This ensures that the recombinant protein is translocated across the membrane simultaneously with translation of the protein, thereby preventing the formation of secondary structures in the cytoplasm.

Enhancing post-transcriptional expression (Troubleshooting)

If expression of the recombinant gene is low, several factors may be responsible for the reduced expression: (1) stability of the mRNA, (2) occurrence of secondary structure(s) near the 5' end of the mRNA, (3) rare codons and (4) weak Shine Dalgarno sequence. mRNA molecules are relatively short-lived with a half-life of around 2 min. The following factors are involved in and influence the degradation of transcripts: exonucleases, endonucleases, secondary structures and ribosome-binding sites. In *E. coli*, two exonucleases have been identified, RNase II (*rmb*) and polynucleotid phosphorylase (*pnp*); both attack mRNA molecules at their 3' end. No exonuclease has been identified attacking from the 5' end. 3' → 5' degradation of transcripts by one of the two exonucleases (which are functionally redundant) can be delayed by secondary structure(s) present at or near the 3' ends. Some of these stem-loop structures may act as stabilizers when fused to heterologous mRNAs. This has been shown for the element present within the transcription terminator of the crystal protein gene of *Bacillus thuringiensis*, which had increased the half-life of the human interleukin-2 and of a penicillinase and thereby the final protein yields (Wong and Chang, 1986). Major endonucleases involved in cleavage of transcripts are RNase E, RNase II and RNase P. All three recognize elements, mainly stem-loop structures within the transcripts, and cleave at or near these secondary structures with two different consequences: in most cases, the endonucleolytic cut will lead to the inactivation of the transcript, while in rare cases this cut is part of a processing reaction involving polycistronic mRNAs. RNase E seems to be the most powerful endonuclease which, together with other proteins (exonuclease, RNA helicase, enolase), constitutes the RNA degradosome (Liou *et al.*, 2001). A stabilizing element for the 5' end of transcripts is the 5' untranslated region of the *E. coli ompA* mRNA which prolongs the half-life of a number of heterologous mRNAs in *E. coli* (Emory *et al.*, 1992).

Secondary structures at the 5' end sequestering the Shine-Dalgarno and/or the start codon within a double-stranded stem significantly reduce translation of that tran-

script since it will be barely recognized by the 30S ribosomal subunit. mRNA secondary structures can be detected by appropriate computer programs. There are two experimental solutions to this problem, exchange of nucleotides to prevent formation of inhibitory secondary structures or using a construct allowing translational coupling. Translational coupling requires at least a one-nucleotide overlap between the stop and the start codon, *e.g.* UGAUG, of the upstream and the downstream gene. If translating ribosomes arrive at the stop codon they slide back a few nucleotides on the transcript till they reach the Shine-Dalgarno sequence of the downstream gene. Translation of the downstream gene is normally prevented by a secondary structure near the end of the upstream gene sequestering the Shine-Dalgarno sequence of the downstream gene. This mechanism can be explored to ensure efficient translation of recombinant genes avoiding impairment of translation by secondary structures reducing binding of the 30S subunit. Vectors have been developed ensuring translational coupling of recombinant genes (Tarragona *et al.*, 1992; Birikh *et al.*, 1995).

More than one codon encodes most amino acids and the relative abundance of cognate tRNAs determines codon usage. The codon usage by the different species can be quite different. As an example, codon usage for arginine of four different species is presented in Table 3. While the codons AGA and AGG are rare codons in *E. coli*, they represent frequently used codons in *Saccharomyces cerevisiae* and *Homo sapiens*. Overexpression of genes with high contents of rare arginine codons may result in defective synthesis of the corresponding protein. Besides the amount, the location of rare codons within the coding region can significantly influence the translation level. Chen and Inouye (1990) demonstrated that the closer AGG codons were to the initiation codon, the stronger the effect on protein synthesis. They showed that single and, particularly, tandems of two to five AGG have stronger effects when placed closer to the translation start. Why? Rare codons close to the initiator may stall the ribosome and prevent the entry of new incoming ribosomes (Chen and Inouye, 1994). There are two experimental solutions to this problem: increase in the amount of the appropriate cognate tRNA or alteration of

Table 3 - Frequency of arginine codon usage for four different species

Codon	<i>E. coli</i>	<i>B. subtilis</i>	<i>S. cerevisiae</i>	<i>H. sapiens</i>
CGU	38	18	14	8
CGC	40	21	6	19
CGA	6	10	7	11
CGG	10	16	4	22
AGA	4	26	48	20
AGG	2	9	21	20

Codon usage tables for all major species can be found under <http://www.kazusa.or.jp/codon/>.

these codons to frequently used ones by sequence-specific mutagenesis.

Inclusion bodies and how to prevent their formation

Rapid production of recombinant proteins can lead to the formation of insoluble aggregates designated as inclusion bodies (Betts and King, 1999). These are large, spherical particles which are clearly separated from the cytoplasm and result from the failure of the quality control system to repair or remove misfolded or unfolded protein. The formation of inclusion bodies does not correlate with (1) the size of the synthesized polypeptide, (2) the use of the fusion construct, (3) the subunit structure and (4) the relative hydrophobicity of the recombinant protein. Overproduction by itself (the increase in the concentration of the nascent polypeptide chains) can be sufficient to induce the formation of inclusion bodies. These aggregates do not consist of pure recombinant polypeptide chains, but contain several impurities such as host proteins (RNA polymerase, outer membrane proteins), ribosomal components and circular and nicked forms of plasmid DNA. In addition, they might contain the small heat shock proteins IbpA and IbpB. Strategies to prevent the formation of inclusion bodies are aimed to slow down the production of recombinant proteins and include (1) low-copy number vectors, (2) weak promoters, (3) low temperature, (4) coexpression of molecular chaperones, (5) use of a solubilizing partner, and (6) fermentation at extreme pH values.

A lower level of protein synthesis from a weaker promoter or from a strong promoter under conditions of partial induction is found to result in a higher amount of soluble protein and greater specific activity (Hockney, 1994). Growth at lower temperatures is a well known technique for facilitating correct folding. The reason why a lower temperature favors the native state is related to a number of factors, including a decrease in the driving force for protein self-association, a slower rate of protein synthesis, changes in the folding kinetics of the polypeptide chain, etc. We have mentioned an expression system which is specifically induced at low temperature, and together with the molecular chaperones derived from the Antarctic seawater bacterium, it may create a new and powerful system to obtain correctly folded proteins.

The aggregation of proteins secreted into the periplasmic space can be suppressed by growing cells in the presence of relatively high concentrations of polyols or sucrose, a non-metabolizable sugar for *E. coli*. In the optimal concentration range, these additives do not affect cell growth, protein synthesis or export and, therefore, they influence directly the physicochemical processes that result in protein-protein association. Polyols and sucrose do not permeate through the cell membrane and consequently cannot exert a direct effect on the folding of cytoplasmic proteins. An increase in the osmotic pressure, however, leads to the accumulation of osmoprotectants, such as glycine betaine,

which have an effect similar to sucrose in stabilizing the native protein structures. It has been shown that cells grown in the presence of sorbitol at 25 °C produce 400-fold higher levels of recombinant protein than control cultures (Blackwell and Horgan, 1991).

Vector plasmids are tentatively divided into four classes based on their copy number (the copy number is defined as the number of plasmid copies per chromosome): very high-copy-number vectors are present in more than 100 copies per chromosome (pUC vectors), high-copy-number vectors (15-60 copies; pBR322), medium-copy-number vectors (about 10 copies; pACYC177, pACYC184 and pSC101) and low-copy-number vectors (1-2 copies; mini-F). Here, medium-copy-number vectors might reduce the amount of recombinant protein sufficiently to prevent their aggregation. Alternatively, high-copy-number vectors can be used in combination with a weak promoter such as the wild-type *lac* promoter. Reducing the growth temperature down to 25 or 20 °C also lowers the productivity of the cells. Coexpression of folder chaperones such as the DnaK or the GroE system might help in some cases to keep the recombinant proteins soluble (Nishihara *et al.*, 1998). Solubilizing partners are other proteins which are fused to the recombinant proteins and keep the hybrid proteins soluble. When three different proteins known to increase the solubility (maltose-binding protein [MBP], glutathione-S-transferase [GST] and thioredoxin [TRX]) were fused to six different recombinant proteins, MBP turned out to be superior (Kapust and Waugh, 1999).

Sometimes, it might be desirable to produce recombinant proteins as inclusion bodies. How can active proteins become recovered from aggregates? This involves a four-step procedure. During the first step, the inclusion bodies are harvested by cell lysis and centrifugation of the cell lysate at 5,000 to 12,000 x g. Under these conditions, the protein aggregates will be present in the pellet. The second step involves solubilization of the inclusion bodies by resuspension of the pellet in a buffer with a denaturant agent such as 6 M guanidinium chloride or 6-8 M urea. During the next step, the solubilized polypeptide chains are purified by ion exchange chromatography in the presence of nonionic denaturants such as urea. The fourth and last step results in *in vitro* protein folding. Folding can be aided by the addition of low-molecular weight folding enhancers such as 1.0-1.3 M guanidiniumchloride, 2 M urea or polyethyleneglycol. If the recombinant protein contains one or more disulfide bonds, generation of native bonds can be sustained by addition of reduced and oxidized glutathione.

Design of an optimal expression system for *E. coli*

Based on our present knowledge, we can propose the design of an optimal expression system for *E. coli*. It should be composed of DNA elements directing efficient transcription, stabilizing the transcript, powerful translation, resulting in authentic recombinant protein without any con-

tamination by truncated or extended versions, and it should stay soluble and accumulate to about 20% of the total cellular protein. Such an expression system contains the consensus promoter recognized by the housekeeping promoter σ^{70} and can be further enhanced by addition of an UP element. Readthrough transcription into neighbouring genes is prevented by two strong factor-independent transcriptional terminators arranged in tandem. The transcript itself is stabilized by inverted repeats present at both ends able to form stem-loop structures impairing endonuclease attack at the 5' end and exonucleolytic degradation from the 3' end but not translation. Last but not least, efficient translation is assured by a strong Shine-Dalgarno sequence, an AUG start codon located about 8 bp downstream and the extended UAAU stop codon. Folding of the nascent polypeptide chains is aided by coexpression of folder chaperones. But it has to be mentioned at the end that there is no optimal expression system working with all recombinant proteins. Each protein poses a new problem, and a high level of synthesis has to be optimized in each single case by empirical variation of the different parameters.

Acknowledgments

This work is a result of an international cooperation program (PROBRAL) performed between the German and Brazilian groups and supported by DAAD and CAPES.

References

- Agterberg M, Adriaanse H, van Bruggen A, Karperien M and Tommassen J (1990) Outer-membrane PhoE protein of *Escherichia coli* K-12 as an exposure vector: Possibilities and limitations. *Gene* 88:37-45.
- Bessette PH, Åslund F, Beckwith J and Georgiou G (1999) Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc Natl Acad Sci USA* 96:13703-13708.
- Betts S and King J (1999) There's a right way and a wrong way: *In vivo* and *in vitro* folding, misfolding and subunit assembly of the P22 tailspike. *Structure* 7:R131-R139.
- Birikh KR, Lebedenko EN, Boni IV and Berlin YA (1995) A high-level expression system: Synthesis of human interleukin 1 α and its receptor antagonist. *Gene* 164:341-345.
- Blackwell JR and Horgan R (1991) A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Lett* 295:10-12.
- Brandi A, Pietroni P, Gualerzi CO and Pon CL (1997) Post-transcriptional regulation of CspA expression in *Escherichia coli*. *Mol Microbiol* 19:231-240.
- Brosius J, Ullrich A, Raker MA, Gray A, Dull TJ, Gutell RG and Noller HF (2003) Construction and fine mapping of recombinant plasmids containing the *rrnB* ribosomal RNA operon of *E. coli*. *Plasmid* 6:112-118.
- Chen G-FT and Inouye M (1990) Suppression of the negative effect of minor arginine codons on gene expression: Preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. *Nucleic Acids Res* 18:1465-1473.
- Chen G-FT and Inouye M (1994) Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. *Genes Dev* 8:2641-2652.
- Chen HY, Pomeroy LR, Bjerknes M, Tam J and Jay E (1994) The influence of adenine-rich motifs in the 3' portion of the ribosome binding site on human IFN- γ gene expression in *Escherichia coli*. *J Mol Biol* 240:20-27.
- Coleman J, Inouye M and Nakamura K (1985) Mutations upstream of the ribosome-binding site affect translation efficiency. *J Mol Biol* 181:139-143.
- Daugherty PS, Olsen MJ, Iverson BL, and Georgiou G (1999) Development of an optimised expression system for the screening of antibody libraries displayed on the *Escherichia coli* surface. *Protein Engin* 12:613-621.
- De Boer P, Comstock LJ and Vasser M (1983) The *tac* promoter: A functional hybrid derived from the *trp* and *lac* promoters. *Proc Natl Acad Sci USA* 80:21-25.
- Dhillon JK, Drew PD and Porter AJR (1999) Bacterial surface display of an anti-pollutant antibody fragment. *Lett Appl Microbiol* 28:350-354.
- Doolittle WF and Yanofsky C (1968) Mutants of *Escherichia coli* with an altered tryptophanyl-transfer ribonucleic acid synthetase. *J Bacteriol* 95:1283-1294.
- Elvin CM, Thompson PR, Argall ME, Hendry P, Stamford NPJ, Lilley E and Dixon NE (1990) Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. *Gene* 87:123-126.
- Emory SA, Bouvet P and Belasco JG (1992) A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. *Genes Dev* 6:135-148.
- Ferrer M, Chernikova TN, Yakimov MM, Golyshin PN and Timmis KN (2003) Chaperonins govern growth of *Escherichia coli* at low temperatures. *Nature Biotechnol* 21:1266-1267.
- Francisco JA, Stathopoulos C, Warren RAJ, Kilburn DG and Georgiou G (1993) Specific adhesion and hydrolysis of cellulose by intact *Escherichia coli* expressing surface anchored cellulase or cellulose binding domains. *Biotechnol* 11:491-495.
- Georgiou G, Stephens DL, Stathopoulos C, Poestshie HL, Mendenhall J and Earhart CF (1996) Display of β -lactamase on the *Escherichia coli* surface: Outer membrane phenotypes conferred by Lpp'-OmpA'- β -lactamase fusions. *Protein Engin* 9:239-247.
- Georgiou G, Staphopoulos C, Daugherty PS, Nayak AR, Iverson BL and Curtiss III R (1997) Display of heterologous proteins on the surface of microorganisms: From the screening of combinatorial libraries to live recombinant vaccines. *Nature Biotechnol* 15:29-34.
- Georgiou G, Valax P, Ostermeier M and Horowitz PM (1994) Folding and aggregation of TEM beta-lactamase: Analogies with the formation of inclusion bodies in *Escherichia coli*. *Protein Sci* 3:1953-1960.
- Gewirtz AT, Navas TA, Lyons S, Godowski PJ and Madara JL (2001) Bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. *J Immunol* 167:1882-1885.
- Gluscock CB and Weickert MJ (1998) Using chromosomal *lacP^{ol}* to control expression of genes on high-copy-number plasmids in *Escherichia coli*. *Gene* 223:221-231.

- Gold L (1988) Posttranscriptional regulatory mechanisms in *Escherichia coli*. *Annu Rev Biochem* 57:199-233.
- Goldenberg D, Azar I and Oppenheim AB (1996) Differential mRNA stability of the *cspA* gene in the cold-shock response of *Escherichia coli*. *Mol Microbiol* 19:241-248.
- Goldstein J, Pollitt NS and Inouye M (1990) Major cold shock protein of *Escherichia coli*. *Proc Natl Acad Sci USA* 87:283-287.
- Gottesman S, Wickner S and Maurizi MR (1997) Protein quality control: Triage by chaperones and proteases. *Genes Dev* 11:815-823.
- Gross G, Mielke C, Hollatz I, Blöcker H and Frank R (1990) RNA primary sequence or secondary structure in the translational initiation region controls expression of two variant interferon- β genes in *Escherichia coli*. *J Biol Chem* 265:17627-17636.
- Gruber TM and Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* 57:441-466.
- Gualerzi C and Pon CL (1990) Initiation of mRNA translation in prokaryotes. *Biochemistry* 29:5881-5889.
- He XS, Rivkina M, Stocker BAD and Robinson WS (1994) Hypervariable region IV of *Salmonella* gene *fljC* encodes a dominant surface epitope and a stabilizing factor for functional flagella. *J Bacteriol* 176:2406-2414.
- Hiniker A and Bardwell JCA (2003) Disulfide bond isomerization in prokaryotes. *Biochemistry* 42:1179-1185.
- Hockney RC (1994) Recent developments in heterologous protein production in *Escherichia coli*. *Trends Biotechnol* 12:456-463.
- Hofnung M (1991) Expression of foreign polypeptides at the *Escherichia coli* cell surface. *Methods Cell Biol* 34:77-105.
- Jiang W, Jones P and Inouye M (1993) Chloramphenicol induces the transcription of the major cold shock gene of *Escherichia coli*, *cspA*. *J Bacteriol* 175:5824-5828.
- Jones PG, VanBogelen RA and Neidhardt FC (1987) Induction of proteins in response to low temperature in *Escherichia coli*. *J Bacteriol* 169:2092-2095.
- Kapust RB and Waugh DS (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* 8:1668-1674.
- Kim YS, Jung HC and Pan, JG (2000) Bacterial cell surface display of an enzyme library for selective screening of improved cellulase variants. *Appl Environ Microbiol* 66:788-793.
- Lee JS, Shin KS, Pan JG and Kim CJ (2000) Surface-displayed viral antigens on *Salmonella* carrier vaccine. *Nat Biotechnol* 18:645-648.
- Liou GG, Jane WN, Cohen SN, Lin NS and Lin-Chao S (2001) RNA degradosomes exist *in vivo* in *Escherichia coli* as multicomponent complexes associated with the cytoplasmic membrane via the N-terminal region of ribonuclease E. *Proc Natl Acad Sci USA* 98:63-68.
- Liu G, Topping TB and Randall LL (1989) Physiological role during export for the retardation of folding by the leader peptide of maltose-binding protein. *Proc Natl Acad Sci USA* 86:9213-9217.
- Liu Z, Murray KS, van Celave V, LaVallie ER, Stahl ML and McCoy JM (1995) Expression of thioredoxin random peptide libraries on the *Escherichia coli* cell surface as functional fusions to flagellin: A system designed for exploring protein-protein interactions. *Bio/Technology* 13:366-372.
- Macnab RM (2003) How bacteria assemble flagella. *Ann Rev Microbiol* 57:77-100.
- Makrides SC (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev* 60:512-538.
- Marino MH (1989) Expression systems for heterologous protein production. *BioPharm* 2:18-33.
- Masuda K, Kamimura T, Kanesaki M, Ishii K, Imaizumi A, Sugiyama T, Suzuki T and Ohtsuka E (1996) Efficient production of the C-terminal domain of secretory leukoprotease inhibitor as a thrombin-cleavable fusion protein in *Escherichia coli*. *Protein Engin* 9:101-106.
- McCarthy JEG, Schairer HU and Sebald W (1985) Translational initiation frequency of *atp* genes from *Escherichia coli*: Identification of an intercistronic sequence that enhances translation. *EMBO J* 4:519-526.
- McCarthy JEG, Sebald W, Gross G and Lammers R (1986) Enhancement of translation efficiency by the *Escherichia coli atpE* translational initiation region: Its fusion with two human genes. *Gene* 41:201-206.
- McSorley SJ, Ehst BD, Yu Y and Gewirtz AT (2002). Bacterial flagellin is an effective adjuvant for CD4+ T cells *in vivo*. *J Immunol* 169:3914-3919.
- Missiakas D and Raina S (1997) Protein folding in the bacterial periplasm. *J Bacteriol* 179:2465-2471.
- Morita MT, Tanaka Y, Kodama TS, Kyogoku Y, Yanagi H and Yura T (1999) Translational induction of heat shock transcription factor σ^{32} : Evidence for a built-in RNA thermosensor. *Genes Dev* 13:655-665.
- Mujacic M, Cooper KW and Baneyx F (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* 238:325-332.
- Müller-Hill B, Crapo L and Gilbert W (1998) Mutants that make more *lac* repressor. *Proc Natl Acad Sci USA* 95:1259-1264.
- Newton SMC, Kotb M, Poirer TP, Stocker BAD and Beachey EH (1991) Expression and immunogenicity of a streptococcal M protein epitope inserted in *Salmonella* flagellin. *Infect Immun* 59:2158-2165.
- Nishihara K, Kanemori M, Kitagawa M, Yanagi H and Yura T (1998) Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GroE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2 in *Escherichia coli*. *Appl Environ Microbiol* 64:1694-1699.
- Nocker A, Hausherr T, Balsiger S, Krstulovic NP, Hennecke H and Narberhaus F (2001) A mRNA-based thermosensor controls expression of rhizobial heat shock genes. *Nucleic Acids Res* 29:4800-4807.
- Ramesh V, De A and Nagaraja V (1994) Engineering hyper-expression of bacteriophage Mu C protein by removal of secondary structure at the translation initiation region. *Protein Engin* 7:1053-1057.
- Rao L, Ross W, Appleman JA, Gaal T, Leirmo S, Schlax PJ, Record MT and Gourse RL (1994) Factor independent activation of *rnnB* P1-an "extended" promoter with an upstream element that dramatically increases promoter strength. *J Mol Biol* 235:1421-1435.

- Ringquist S, Shinedling S, Barrick D, Green L, Binkley J, Stormo GD and Gold L (1992) Translation initiation in *Escherichia coli*: Sequences within the ribosome-binding site. *Mol Microbiol* 6:1219-1229.
- Rose JK and Yanofsky C (1974) Interaction of the operator of the tryptophan operon with repressor. *Proc Natl Acad Sci USA* 71:3134-3138.
- Schierle CF, Berkmen M, Huber D, Kumamoto C, Boyd D and Beckwith J (2003) The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the *Escherichia coli* periplasm via the signal recognition particle pathway. *J Bacteriol* 185:5706-5713.
- Shine J and Dalgarno L (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to non-sense triplets and ribosome binding sites. *Proc Natl Acad Sci USA* 71:1342-1346.
- Staphopoulos C, Georgiou G and Earhart CF (1996) Characterization of *Escherichia coli* expressing an Lpp-OmpA (46-159)-PhoA fusion protein localized in the outer membrane. *Appl Microbiol Biotechnol* 45:112-119.
- Studier FW (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol* 219:37-44.
- Studier FW and Moffat BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189:113-130.
- Swartz JR (2001) Advances in *Escherichia coli* production of therapeutic proteins. *Curr Opin Biotechnol* 12:195-201.
- Tarragona-Fiol A, Taylorson CJ, Ward JM and Rabin BR (1992) Production of mature bovine pancreatic ribonuclease in *Escherichia coli*. *Gene* 118:239-245.
- Thomas JG and Baneyx F (1996) Protein folding in the cytoplasm of *Escherichia coli*: Requirements for the DnaK-DnaJ-GrpE and GroEL-GroES molecular chaperone machines. *Mol Microbiol* 21:1185-1196.
- Tripp BC, Lu ZJ, Bourque K, Sookdeo H and McCoy JM (2001) Investigation of the 'switch-epitope' concept with random peptide libraries displayed as thioredoxin loop fusions. *Protein Engineering* 14:367-377.
- Vellanoweth RI and Rabinowitz JC (1992) The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli*. *Mol Microbiol* 6:1105-1114.
- Xu Z and Lee SY (1999) Display of polyhistidine peptides on the *Escherichia coli* cell surface by using outer membrane protein C as an anchoring motif. *Appl Environ Microbiol* 65:5142-5147.
- Westelund-Wikstrom B, Tanskanen J, Virkola R, Hacker J, Lindberg M, Skurnik M and Korhonen TK (1997) Functional expression of adhesive peptides as fusions to *Escherichia coli* flagellin. *Protein Engin* 10:1319-1326.
- Wetzel R (1994) Mutations and off-pathway aggregation of proteins. *Trends Biotechnol* 12:193-198.
- Wong HC and Chang S (1986) Identification of a positive retro-regulator that stabilizes mRNAs in bacteria. *Proc Natl Acad Sci USA* 83:3233-3237.

Associate Editor: Sergio Olavo Pinto da Costa